UNIVERSIDAD YACHAY TECH

UNIVERSIDAD DE INVESTIGACIÓN DE TECNOLOGÍA EXPERIMENTAL YACHAY

Escuela de Ciencias de la Tierra, Energía y Ambiente

ANALYSIS OF LANDSLIDE TRIGGERING MECHANISMS IN THE IMBABURA PROVINCE

Trabajo de integración curricular presentado como requisito para la obtención del título de Geóloga

Autor:

Rivera Añazco Ariana Guiscelle

Tutor:

Ph.D. Pineda Ordóñez Luis Eduardo

Urcuquí, julio 2021

SECRETARÍA GENERAL
(Vicerrectorado Académico/Cancillería)
ESCUELA DE CIENCIAS DE LA TIERRA, ENERGÍA Y AMBIENTE
CARRERA DE GEOLOGÍA
ACTA DE DEFENSA No. UITEY-GEO-2021-00007-AD

A los 7 días del mes de julio de 2021, a las 10:00 horas, de manera virtual mediante videoconferencia, y ante el Tribunal Calificador, integrado por los docentes:

Presidente Tribunal de Defensa	Dr. GOMEZ PAREDES, JORGE ESTEBAN , Ph.D.
Miembro No Tutor	Mgs. FLORES DE LA TORRE, JOSE LUIS
Tutor	Dr. PINEDA ORDOÑEZ, LUIS EDUARDO , Ph.D.

El(la) señor(ita) estudiante **RIVERA AÑAZCO, ARIANA GUISCELLE**, con cédula de identidad No. **1104554199**, de la **ESCUELA DE CIENCIAS DE LA TIERRA, ENERGÍA Y AMBIENTE**, de la Carrera de **GEOLOGÍA**, aprobada por el Consejo de Educación Superior (CES), mediante Resolución **RPC-SE-10-No.031-2016**, realiza a través de videoconferencia, la sustentación de su trabajo de titulación denominado: **ANALYSIS OF LANDSLIDE TRIGGERING MECHANISMS IN THE IMBABURA PROVINCE** , previa a la obtención del título de **GEÓLOGO/A**.

El citado trabajo de titulación, fue debidamente aprobado por el(los) docente(s):

Tutor	Dr. PINEDA ORDOÑEZ, LUIS EDUARDO , Ph.D.

Y recibió las observaciones de los otros miembros del Tribunal Calificador, las mismas que han sido incorporadas por el(la) estudiante.

Previamente cumplidos los requisitos legales y reglamentarios, el trabajo de titulación fue sustentado por el(la) estudiante y examinado por los miembros del Tribunal Calificador. Escuchada la sustentación del trabajo de titulación a través de videoconferencia, que integró la exposición de el(la) estudiante sobre el contenido de la misma y las preguntas formuladas por los miembros del Tribunal, se califica la sustentación del trabajo de titulación con las siguientes calificaciones:

Tipo	Docente	Calificación
Presidente Tribunal De Defensa	Dr. GOMEZ PAREDES, JORGE ESTEBAN , Ph.D.	9,7
Tutor	Dr. PINEDA ORDOÑEZ, LUIS EDUARDO , Ph.D.	9,5
Miembro Tribunal De Defensa	Mgs. FLORES DE LA TORRE, JOSE LUIS	9,0

Lo que da un promedio de: **9.4 (Nueve punto Cuatro)**, sobre 10 (diez), equivalente a: **APROBADO**

Para constancia de lo actuado, firman los miembros del Tribunal Calificador, el/la estudiante y el/la secretario ad-hoc.

RIVERA AÑAZCO, ARIANA GUISCELLE
Estudiante

Dr. GOMEZ PAREDES, JORGE ESTEBAN , Ph.D.
Presidente Tribunal de Defensa

Dr. PINEDA ORDOÑEZ, LUIS EDUARDO , Ph.D.
Tutor

Mgs. FLORES DE LA TORRE, JOSE LUIS
Miembro No Tutor

TERÁN ROSALES, ANDREA YOLANDA
Secretario Ad-hoc

AUTORÍA

Yo, **ARIANA GUISCELLE RIVERA AÑAZCO**, con cédula de identidad 1104554199, declaro que las ideas, juicios, valoraciones, interpretaciones, consultas bibliográficas, definiciones y conceptualizaciones expuestas en el presente trabajo; así cómo, los procedimientos y herramientas utilizadas en la investigación, son de absoluta responsabilidad de el/la autora (a) del trabajo de integración curricular. Así mismo, me acojo a los reglamentos internos de la Universidad de Investigación de Tecnología Experimental Yachay.

Urcuquí, Julio 2021.

_____
Ariana Guiscelle Rivera Añazco

CI: 1104554199

AUTORIZACIÓN DE PUBLICACIÓN

Yo, **ARIANA GUISCELLE RIVERA AÑAZCO,** con cédula de identidad 1104554199, cedo a la Universidad de Investigación de Tecnología Experimental Yachay, los derechos de publicación de la presente obra, sin que deba haber un reconocimiento económico por este concepto. Declaro además que el texto del presente trabajo de titulación no podrá ser cedido a ninguna empresa editorial para su publicación u otros fines, sin contar previamente con la autorización escrita de la Universidad.

Asimismo, autorizo a la Universidad que realice la digitalización y publicación de este trabajo de integración curricular en el repositorio virtual, de conformidad a lo dispuesto en el Art. 144 de la Ley Orgánica de Educación Superior

Urcuquí, julio 2021.

<div align="center">

_____
Ariana Guiscelle Rivera Añazco
CI: 1104554199

</div>

A mi familia.

ACKNOWLEDGEMENTS

RESUMEN

Los deslizamientos de tierra se producen por interacciones mecánicas y complejas entre factores predisponentes y desencadenantes que no se entienden en su totalidad. La mayoría de los deslizamientos de tierra en la provincia de Imbabura, en el norte de Ecuador, se registran como eventos de riesgo geológico cuya ocurrencia se asume que es provocada por fuertes lluvias. Pocos estudios han intentado discriminar la influencia de los factores desencadenantes en la ocurrencia de deslizamientos de tierra para verificar tal afirmación. Este estudio investiga la relación causa-efecto que desencadenó deslizamientos de tierra en la provincia de Imbabura. Para ello, se utilizaron modelos lineales generalizados (GLM) y modelos aditivos generalizados (GAM) en un entorno de modelado explicativo para analizar la relación entre los puntos de inicio de deslizamientos mapeados y los factores predisponentes de terreno, topográficos y climáticos. Ambos modelos estadísticos mostraron un desempeño de alta predictibilidad, donde la precipitación tuvo la mayor influencia en la ocurrencia de deslizamientos de tierra con estimaciones de 6.35 y 6.78, respectivamente, seguidas del factor elevación con una estimación de 5.98 en GLM, y el factor pendiente con un estimado de 1.72 en GAM, bajo un nivel de significancia del 5%. Esto muestra claramente la importancia de los factores predisponentes climáticos en la ocurrencia de deslizamientos de tierra debido a un tipo de precipitación estratiforme que aumenta la saturación de humedad del suelo en el área de estudio. Además, el uso de la tierra agrícola y la ganadería mostraron una influencia negativa, lo que requiere más investigación para determinar el papel que el uso de la tierra puede desempeñar en la contribución a la ocurrencia de deslizamientos de tierra. El presente estudio puede servir como base para evaluar la efectividad de las declaraciones sobre los mecanismos desencadenantes de deslizamientos de tierra en el país.

Palabras clave:

Ocurrencia de deslizamientos, regresión logística, modelo aditivo generalizado, provincia de Imbabura.

ABSTRACT

Landslides are produced by mechanical and complex interactions between predisposing and triggering factors that lack understanding. Most landslides in the Imbabura Province, northern Ecuador, are recorded as geohazard events whose occurrence is assumed to be triggered by heavy rainfall. Few studies have attempted to discriminate the influence of triggering factors on the landslide occurrence to verify such statement. This study investigates the cause-effect relationship that triggered landslides in the province of Imbabura. To do this, generalized linear models (GLM) and generalized additive models (GAM) were used in an explanatory modelling setting to analyse the relationship between mapped landslide initiation points and terrain, topographic, and climatic predisposing factors. Both statistical models showed high predictability performance, where precipitation had the highest influence on the occurrence of landslides with estimates of 6.35 and 6.78, respectively, followed by the factor elevation with estimates of 5.98 in GLM, and the factor slope aspect with an estimate of 1,72 in GAM, under a 5% significance level. This clearly shows the importance of climatic predisposing factors in landslide occurrence which is due to a stratiform type of precipitation that increase the soil moisture saturation in the study area. Furthermore, agricultural land use and livestock showed a negative influence, which calls for further research to determine the role that land use may play in contributing landslide occurrence. The present study may serve as a baseline to assess the effectiveness of statements on landslide triggering mechanisms in the country.

Keywords:

Landslide occurrence, logistic regression, generalized additive model, Imbabura province.

INDEX

LIST OF TABLES

LIST OF FIGURES

I. INTRODUCTION

Historically, Ecuadorian ancestors have settled on places prone to natural hazards. Governmental institutions such as the National Service of Risk Management and Emergencies (SNGRE, 2021) are responsible for the management of natural hazards, protecting the environment, and reducing geohazards' risks to population. To do so, these organizations are obliged to make the population aware about the risks of living in hazard-prone places. For appropriate and cost-effective geohazard decision making, understanding conditions, driving processes as well as providing well-categorized information about these hazards is essential.

Geohazards are phenomena caused by geological processes that can cause significant damage to human life, their possessions and activities. These can be natural, induced or worsened by the intervention of humans (McCall, Laming, & Scott, 1992). Ecuador is subjected to several types of geohazards like earthquakes, volcanic eruptions, floods, fires, and landslides. Although there is a wide range of related studies in Ecuador, few of them have explored the space-time variability, the geophysical drivers, and the mechanistic processes leading to their occurrence. For example, for earthquakes, spatio-temporal distribution of seismicity has been applied (Soto-Cordero, 2019); for intense storms, climatological reconstructions (Changoluisa & Pineda, 2020); for flooding risk assessment, estimations of runoff and river discharges (Campos Cedeño, Salas Guillén, Macias Ramos, Sinichenko, & Gritsuk, 2019; Torres Alves, 2018); and for local landslides, analysis of spatio-temporal distributions (Salazar, 2017).

The study presented in this thesis is focused on landslides, which are processes that result from gravity-based movements down and outwards of slope-forming materials such as rock, soil, artificial fill, or a mixture of these (Lyyn, 2014). They can move by falling, toppling, sliding (rotational or translationally), spreading, or flowing (Lyyn, 2014). Depending on the materials involved and their movement, several types of classifications have been made and updated to improve the terminology and its features (Cruden & Varnes, 1996; Hungr, Leroueil, & Picarelli, 2013; Varnes, 1978).

Worldwide, several studies have been done to evaluate and communicate landslide hazards with the use of direct information and technological tools (Highland & Bobrowsky, 2008). This is done by reporting features (Zhang, Wang, Bao, & Zhao, 2019),

producing susceptibility maps (Ali, Biermanns, Haider, & Reicherter, 2019), using remote sensing (Tsai, Hwang, Chen, & Lin, 2010), and monitoring local areas that are susceptible to landsliding, as well as by knowing thresholds and conditions that increase and trigger landslides (Brunetti et al., 2010). These studies provide the likelihood of landslides' occurrence, expected location, and the estimated severity of their effects, for engineers, planners, and decision-makers.

Landslides happen due to natural or human causes (Highland & Bobrowsky, 2008). The occurrence of landslides can be seen as a result of mechanical and complex interactions between predisposing and triggering factors (Leonarduzzi, 2020). Predisposing factors such as the soil's structure and the slope's stability are determined by the physical properties at the moment of the rock formation. Environmental conditions such as erosion, weathering, and crustal movements generate subsequent triggering factors that lead to a decrease in soil shear strength and slope failure (Varnes, 1978). Wieczorek (1996) concluded that several causes contribute to the occurrence of a landslide, but there is only one trigger or external stimulus such as intense rainfall, earthquakes, volcanic eruptions, storm waves, rapid stream.

Most studies on landslide triggering seek to find statistical data on the relationship and influence of triggering factors. They do this through several geostatistical methods such as spatial analysis (Tsai et al., 2010), numerical models (Havenith, Strom, Calvetti, & Jongmans, 2003), and Bayesian inference methods (Brunetti et al., 2010; Vennari et al., 2014). Rainfall intensity-duration (ID) threshold curves are also very common in the quantification of the joint occurrence of rainfall and landslides (Larsen & Simon, 1993; Leonarduzzi, Molnar, & McArdell, 2017; Ma, Li, Lu, & Bao, 2015; Marra, Morin, Peleg, Mei, & Anagnostou, 2017). Some studies try to obtain the impact of climatic phenomena on precipitation and landslides (Zêzere, Trigo, & Trigo, 2005), others take into account new preconditioning factors such as the soil moisture to derive an empirical rainfall threshold that triggers landslides (Lazzari, Piccarreta, & Manfreda, 2018). Landslide forecasting is the next step, and numerical weather prediction models help predict temporal changes in landslide probabilities (Schmidt, Turek, Clark, Uddstrom, & Dymond, 2008).

Specifically in Ecuador, studies on landslides involve risk assessment using Geographic Information System (GIS) techniques (Buitrón Vinueza, 2014; López Cevallos, 2004; Segovia Puente, 2017), analysis of slope stability (Vázquez, 2005),

impact studies of significant landslides (Domínguez, 2014), local characterizations of geological units subjected to landslides (Burga Cholca, 2019; Dykes & Welford, 2007), susceptibility estimations (Brenning, Schwinn, Ruiz-Páez, & Muenchow, 2015), among others related to these topics.

DesInventar (DesInventar, 2020), the most complete international natural disaster database for Ecuador, has registered 2405 landslides severely affecting communities from 1970 to 2015. Of this account, 1794 landslides have occurred in the northern provinces of the Ecuadorian Sierra. However, landslides are still not recognized as major geohazards due to its high frequency, small-scale and rare menace (McCall et al., 1992, p. 117).

In the last decade, little has been done and studied about the triggers such as antecedent rainfall amount, duration and intensity, slope stability, land cover type, soil type, seismic activity, construction on hillslopes to understand the cause-effect relationship between them and landslides. Several studies on landslide susceptibility have made use of Generalized Linear Models (GLMs), however, applied Generalized Additive Models (GAMs) have shown stronger predictive performance (Goetz, Guthrie, & Brenning, 2011) and are particularly effective at handling complex nonlinearities in datasets with non-normal distributions (Chen et al., 2017; Pourghasemi & Rossi, 2017).

Brenning et al (Brenning et al., 2015) is an example of statistical analysis, in which GLMs and GAMs are used to estimate the effect of several variables such as elevation, slope, geology, rainfall, among others, on the landslide initiation frequency in the Andes of southern Ecuador.

II. PROBLEM STATEMENT

Among the geophysical factors driving landslides, one of the most significant is the increase in pore water pressure from rainfall, and if the times and duration of landslides are well documented, thresholds of landslide triggering by rainfall can be identified (Wieczorek, 1996). To do this, observational and simulation methods that gather regional and on-site information, which also depends on local geologic, geomorphic, and climatologic conditions, are necessary (Borgatti, Vittuari, & Zanutta, 2010; Crosta & Frattini, 2003; Ochoa, Pineda, Willems, & Crespo, 2014; Springman, Kienzler, Casini, & Askarinejad, 2009). However, these types of observatories and local experimental laboratories that contribute to the understanding of factors and to future research on landslide's predictability are lacking in Ecuador; only empirical estimations have been done in Southern Ecuador (Soto et al., 2019).

Most hydrological and geohydrological studies, such as those estimations mentioned above, rely on the quality of rainfall estimates, which can be retrieved from satellite-based data, and rain gauges, or by a combination of both in order to produce meaningful precipitation estimates for impact analysis (Nerini et al., 2015). Both have advantages and disadvantages: remote-sensing satellite data has a broad spatial coverage and scans downward over complex terrains, but since it is an indirect measure, it does not quantify some of the key features of spatiotemporal patterns of precipitation such as the intensity of rainfall, and might underestimate the magnitude of extreme rainfall; on the other hand, rain gauges measure precipitation directly, but only at a given location, and with random errors produced by splash-out during heavy rainfall, lack of sensitivity to light rain rates, under-catching by wind drift, and evaporation (Li et al., 2020). Thus, both by themselves do not reflect a complete estimation of precipitation (Ulloa, Ballari, Campozano, & Samaniego, 2017).

Rain gauges' disadvantages become more severe in tropical regions due to high rainfall variability and scarce data conditions (Nerini et al., 2015), so merging the rainfall estimates from both sources is possible but it is data and computationally expensive. Thus, the use of satellite-based precipitation estimates, such as the Tropical Rainfall Measuring Mission (TRMM), is necessary. Finally, other important limiting factors are the lack of high-quality databases of landslide records, as the data available in DesInventar

4

(DesInventar, 2020) does not accurately show location, dates, or duration of the landslide events.

To overcome these obstacles and make use of the available information for Imbabura, generalized linear models in which the response variable depends on some predictor variables, GLM and GAM can be fit to a dataset of observations when combining linear and nonlinear relationships between geophysical explanatory variables and event occurrences in order to understand the relationships between landslides and its possible drivers or predisposing factors.

III. OBJECTIVES

The objectives of this study are:

1. To analyse attributes related to landslides in the Imbabura province (North of Ecuadorian Sierra) using available landslides' records.

2. To identify the cause-effect relationships between geophysical predisposing factors and the occurrence of landslides using statistical modelling.

IV. STUDY AREA

4.1 Location

The chosen area for this study is located in northern Ecuador, more specifically in the Imbabura province, which comprises an area of approximately 4544 km^2. This area is limited by Esmeraldas, Carchi, Pichincha and Sucumbíos provinces, to the east, north, south and west, respectively (Figure 1). Geographically, it is part of the Interandean Valley, so its elevation ranges from 304 masl in northern and south-western Imbabura, to 4925 masl in central, southern and south-eastern Imbabura.
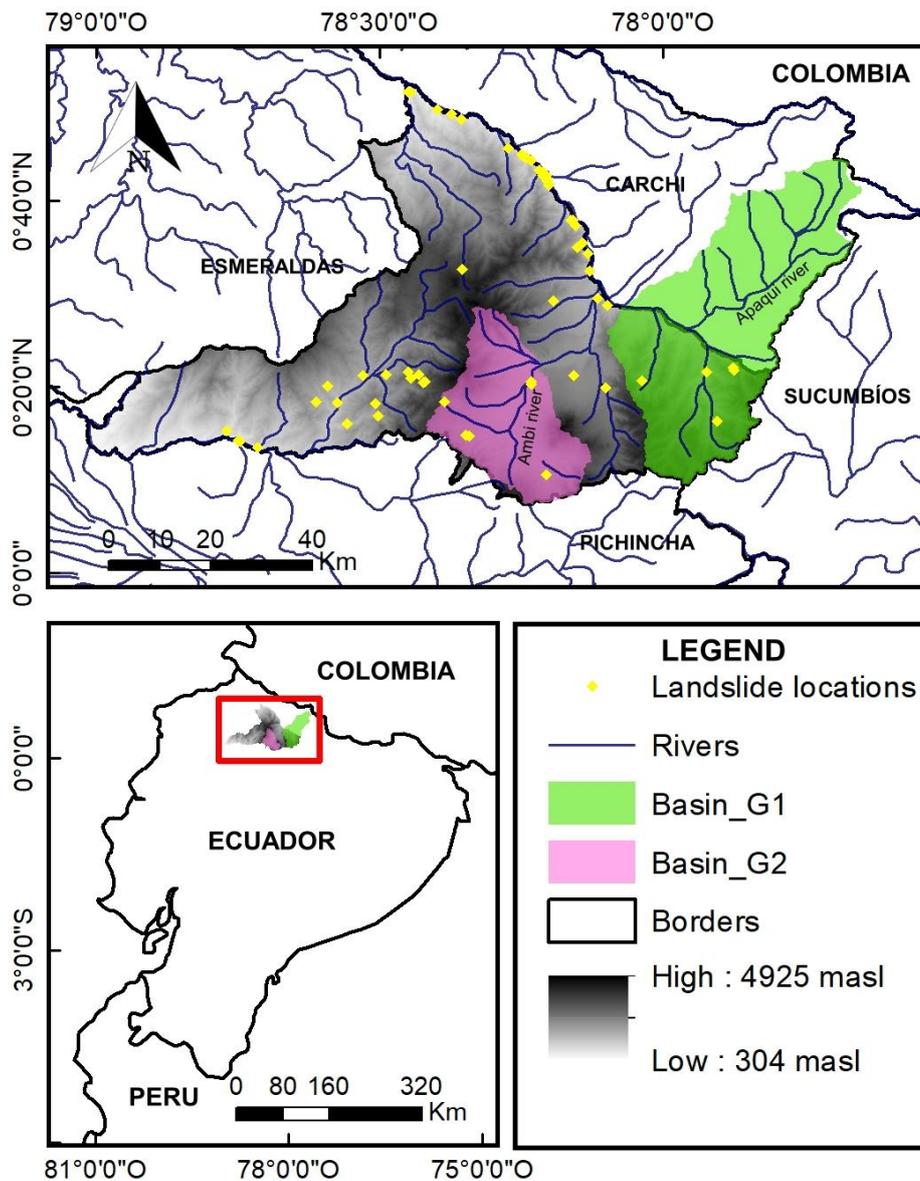
Figure 1. Location of the study area with a topographic map of the study area, showing landslide locations, and the main river network.

The great majority of the landslides registered for the Imbabura province are located near rivers such as the Mira River and Intag River, and along main roads, such as the highway E10, way to San Lorenzo, which borders the Carchi province, and the Cotacachi-Quiroga-Cuicocha road, situated in southern Imbabura.

4.2 Geological overview

The Ecuadorian geology is basically composed of several NNE-oriented parallel morphotectonic zones separated by major fault systems or suture zones, and accretions of island-arc and oceanic plateau terranes that formed an orogenic belt with the same orientation (Ruiz, 2002). These events are responsible for the orogenesis, arc volcanism and surface uplift in the northern Andes, which can be evidenced by igneous rocks of late Mesozoic to Tertiary age, part of which rest on an older basement, and basic to intermediate volcanic and plutonic rocks of Cretaceous to Eocene age cropping out extensively in the coastal area and the Western Cordillera (Lebras, Megard, Dupuy, & Dostal, 1987). For a further understanding on the geology of the Imbabura province, it is necessary to approach the geodynamics and structural geology of the northern Andes.

4.2.1. Geodynamics

Ecuador lies in a geo-dynamically active region that created the Andes mountain range. This region consists of a convergent margin where the subducting oceanic Nazca plate evolves from flat-slab to dipping and oblique subduction eastwardly beneath the western margin of the South American plate at a rate of 5-7 cm/yr (Baby, Rivadeneira, Barragán, & Christophoul, 2013; Roperch et al., 1987). f (E Jaillard et al., 2009).

This geologic process displaces the North Andean block to the northeast at a rate of 6 mm/yr (Baby et al., 2013; Gutscher, Malavieille, Lallemand, & Collot, 1999), and creates a 1000 km long Andean Chain, which is a major morphological feature of the South American continent trending north north-east to south south-west (Etienne Jaillard et al., 2000). An important section of the Nazca subduction that influences specifically the crustal deformation of Ecuador and creates a region of great seismicity is the ENE-oriented Carnegie Ridge (Gutscher et al., 1999).

4.2.2. Structural geology

The Ecuadorian part of the Northern Andes presents six distinct longitudinal morphotectonic regions from west to east (Farinango & Geovanny, 2014; Ruiz, 2002; Toro Álava & Jaillard, 2005; Vallejo Cruz, 2007):

(1) the coastal area (forearc), which received Paleogene to Neogene forearc deposits over the Cretaceous Piñon terrane;

(2) Western Cordillera or Cordillera Occidental, which is mainly composed by mafic magmatic rocks from the Guaranda and San Juan terranes;

(3) Inter-Andean Valley, which is composed by Neogene (Late Miocene) to recent volcanosedimentary and volcanic rocks, and is deformed by the dextral transpressive Dolores Guayaquil Megashear (DGM) zone (Baby et al., 2013);

(4) Eastern Cordillera or Cordillera Real, which is composed of exhumed metamorphic rocks and intrusions;

(5) Sub-Andean Zone, which is part of the Cretaceous Oriente Basin, floored by a thick Jurassic volcanic arc uplifted during the Andean Orogeny;

(6) Oriente Basin, which received marine deposits during part of the Cretaceous and constitutes the present-day retroarc basin.

4.2.3. Imbabura geology

In this context, Imbabura is located in the Inter-Andean Valley, and according to the geologic maps available in the Geological and Energy Research Institute (IIGE) (1979), its geology in this part of the valley is characterized by the Macuchi and Silante Formations, the Chota group, the Angochagua, Yanahurco, Cotacachi, Negro Puño and Imbabura volcanics, glacial, alluvial and colluvial deposits, and intrusives such as the Apuela-Nanegal Batolith.

The Macuchi Formation is outcropping in the western sector of Imbabura and is characterized by Cretaceous altered metapelites and metabasalts intercalated with volcanoclastic sediments and green andesite lavas. The Silante Formation is located in the north, and it is a very faulted, altered and fractured formation from the Upper Cretaceous, composed by volcanic conglomerates intercalated with green and violet greywackes, sandstones and red mudstones. The Chota unit is located along the Mira river and belongs to the Upper Tertiary, Miocene period, and it is composed mainly by sedimentary sequences that range between latest Miocene and Pleistocene epoch

(Spikings et al., 2015), specifically intercalations of volcanic conglomerates and breccias with volcanic sandstones and very fine limestones. The Angochagua, Yanahurco, Cotacachi, Negro Puño and Imbabura volcanics are Miocene to Holocene deposits located in the center of Imbabura. Also, terraces, alluvial and colluvial deposits of sand and cangagua are widely spread in the whole province and at the margins of the main rivers Mira, Lita and San Juan since the beginning of the Holocene. Finally, glacial deposits are located in south-western side of Imbabura belonging to the Cotacachi, Imbabura, Cusin and Cayambe volcanos, while intrusives such as the Apuela-Nanegal Batolith are in south-eastern Imbabura.

4.3 Climate setting

Ecuador presents climatic divides, major climatic regions or dominant meridional zones: Coast, Inter-Andean valleys (Sierra) and Amazonia (Eidt, 1969; Emck, 2007). The coastal area has a single precipitation maximum during summer (December to February), the lowlands of the eastern Andean slope have a single rainfall maximum at heights of 1000 to 3500 masl in winter (July), a well-defined invernal dry season (July to August) and a shorter one (December to February), and in the Amazon Lowlands there is a regime with three peaks and precipitation over all seasons (Bendix & Lauer, 1992).

According to the territorial arrangement planning done to the Imbabura province, the climatology includes a minimum of three specific types of climates: dry climate in the Chota Valley, template in the cantonal capitals, and high mountain cold in Intag and Lita sectors (Gad Provincial de Imbabura, 2018). These climates are modified by the influence of a complex interaction among external factors such as the astronomical location, atmospheric circulation, surface features, and by climatic elements such as temperature, rainfall, atmospheric pressures, relative humidity, and wind speeds and directions (Eidt, 1969). This factors lead to complex phenomena and different degrees of variability, and complicate assessment of possible climate change effects (Morán-Tejeda et al., 2016).

4.3.1. Astronomical location

South America has a general humid tropical climate due to the latitude and longitude ranges, width and extension of the continent, which determines major climatic characteristics. In this sense, in tropical and non-tropical areas, there is a small and

attenuated range between hottest and coldest months, so extreme winters and summers are not significant (Eidt, 1969).

4.3.2. Surface features: landforms and ocean currents

The Andes in Ecuador are a landform of great influence on climatology since, when above 2000 masl, they regulate and prevents winds from the Southern Pacific anticyclone from entering into the continent by forming elongated rain shadows (Eidt, 1969; Morán-Tejeda et al., 2016). Also, the presence of highland basins between the cordilleras in Ecuador vary the wind systems in those areas according to the position of the local ranges, the altitude and the exposure. Furthermore, the north coast of Ecuador has a dry winter climate because the north-east trade winds are blocked.

In terms of ocean currents, the more prominent ones are: a cold current which transports water from cooler to warmer latitudes, such as the Humboldt Current, and a warm current which oscillates from east to west in the Tropical Equatorial Pacific Ocean, bringing warm water to the coast of South America at inter-annual scales, such as El Niño-Southern Oscillation (ENSO) (Emck, 2007). The Humboldt Current upwells deep and cold water, and stabilizes the warmer tropical air above producing a continuous inversion throughout the year and limiting rain. On the other hand, when the Humbolt Current is displaced to the west and water temperatures are above 24°C, the El Niño Current brings warmer than normal sea surface temperature, and consequently, heavy and violent rains to the western coasts of South America (Eidt, 1969).

4.3.3. Precipitation

The Pacific-Andean watershed has a complex spatial rainfall distribution and precipitation ranges from 300 to 6000 mm year^{-1}, explained by the occurrence of inter-annual anomalies due to the large-scale circulation phenomena of ENSO (Ochoa et al., 2014). The highest annual precipitation totals are recorded on the warm and humid Cordillera foothills, while on the highlands, long-duration and low-intensity rainfall occurs, and in the inner valleys, precipitation gradients are a result of the influence of landscape relief, i.e. about 300 mm year^{-1} in valleys, increasing to 1200–1400 mm year^{-1} in the mountains.

In Imbabura, high altitudes are cold and wet with low but lasting precipitations, at intermediate zones, temperature increases and precipitation becomes more intense, and at low valleys like Chota, high temperatures and low precipitation (below 300 mm per day) are prevalent. In this sense, northern regions of Ecuador are characterized to be more affected by the influence of the continental climate divide explained above, the annual movement of the inter-tropical convergence zone (ITCZ), and the occurrence of more stratiform clouds on the west and east flanks of cordilleras resulting in deep convection in the inter-Andean valleys, explained next (Ochoa et al., 2014):

- According to Eidt (1969), the ITCZ is the movement of a transitional area or belt of converging northeastern trade winds between the high pressure cells in both hemispheres. The ITCZ shifts seasonally from 4°S to 7°N and, along with the sea surface temperature fluctuations, determine the precipitation regimes (Morán-Tejeda et al., 2016).
- Moreover, precipitation is generated in advective-orographic systems (light rain) and convective systems (Emck, 2007). The latter can be classified into convective and stratiform based on differences in the nature of precipitation (Huang, Wang, & Cui, 2019). Here, stratiform genesis consists of moderate rain, low intensity, aggregates of ice particles and longer duration; and convective genesis, heavier rain, high intensity, rimmed ice particles and shorter duration (Huang et al., 2019)

Furthermore, some seasons are more influenced by the stratiform rainfall (yellow and green shades in the spatial precipitation maps of Figure 2), which can be seen as widespread wet masses during the months of December to April and, where most landslides occur. This might be explained by the continuous input of atmospheric water leading to high saturation in the soil that act as a preconditioning factor for landsliding. While convective rainfall regions (red areas in Figure 2) take place during the months of June to September and November, and do not present landslides, possibly due to its shorter duration leading to lower effects in saturation of soils.
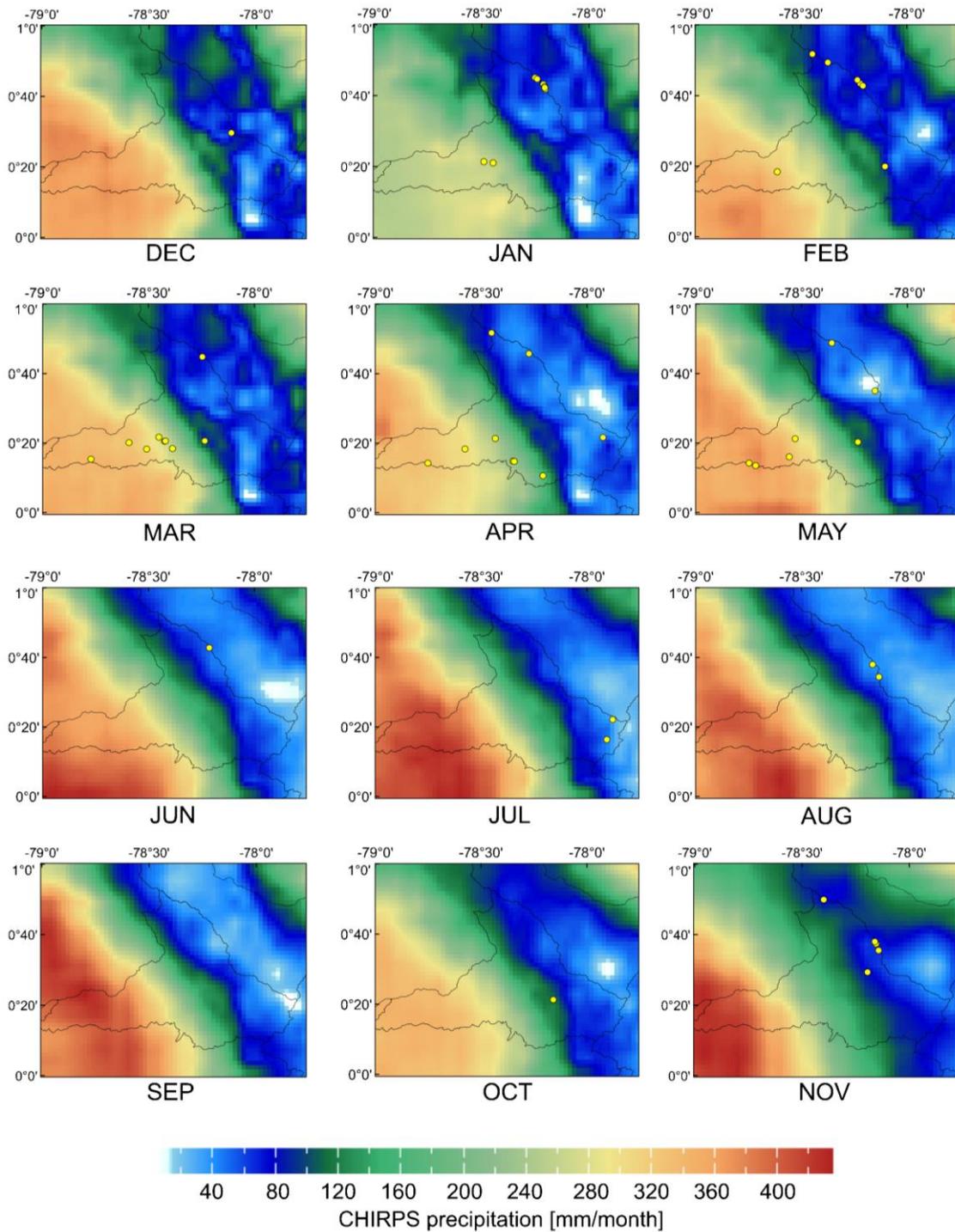
Figure 2. Monthly average precipitation derived from Climate Hazards Group InfraRed Precipitation with Stations data (CHIRPS) data with the location of landslides on the study area (yellow dots).

Moreover, precipitation data from meteorological stations located on the north-east and south-west regions of the study area show a long-term synopsis of the seasonal rainfall patterns and its relationship with landslide occurrence (Figure 3). The monthly

long-term mean rainfall in the inter-Andean valley shows a bimodal rain regime, with two precipitation maxima in March to April and October to December (Villacís, Vimeux, & Denis, 2008).

The seasonality of precipitation dictates the dynamics of the surface and subsurface flows in the wettest season, March to May, and in the following drier period (June to August). Both regions, the north-east and south-west sub-catchments show similar timing but slight variations in magnitude.

The synoptic reason for this seasonal differentiation is the respective position of the ITCZ. During the months of October to December, rainfall mostly of advective origin falls more on the higher levels of the north-east sub-catchments (Apaquí) than on the valley floors; whereas during the months of March to May, high precipitation volumes in April are enhanced by moist-laden rains of oceanic origin. During October, the oceanic ITCZ dwells far north (>= 10°N) and the most intensive rains occur north of the equator. During the months of March to April, the ITCZ is near the equator, and a zone of adynamic winds lies off Ecuador, the offshore waters are at their annual thermal high, and make a rich source of atmospheric water inputs (Emck, 2007), which is depicted in Figure 3 with two maximums during the October to December, and March to May periods.

Regarding landslides, most of them happen during the wetter months in both stations, while in the north-east station some landslides in drier periods can be explained as a result of the water saturation produced form the previous wetter months.
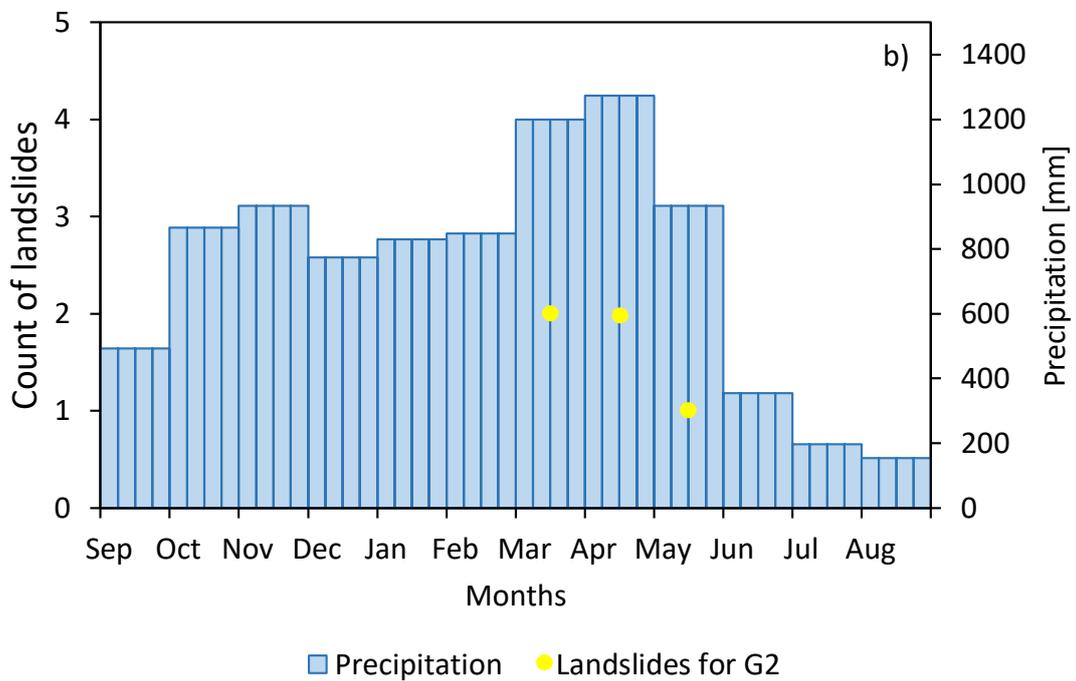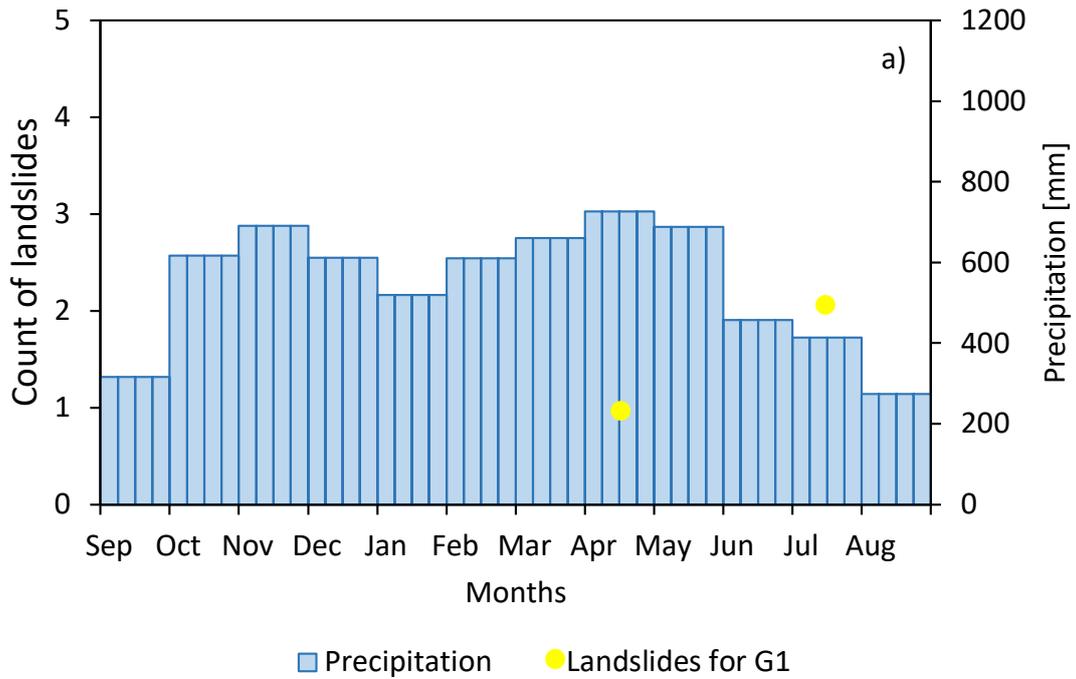
Figure 3. Monthly precipitation averages and count of landslides for the south-west station G1 H0023 Ambi (a), and the north-east station G2 H0016 Apaquí (b).

4.4 Landslides

In the Imbabura province, landslides have been concurrent in the Ibarra canton, specifically in the Lita parish, and the Pimampiro canton, causing road obstructions and damage to infrastructure such as houses, schools and potable water systems, affecting Ambuquí and García Moreno cantons (Gad Provincial de Imbabura, 2018). However, landslides have been relatively under-studied in tropical areas compared to other areas like Europe and USA (Urgilez Vinueza, Robles, Bakker, Guzman, & Bogaard, 2020).

Landslides are a geomorphic process that result from mechanical and complex interactions between predisposing factors and triggering factors, where unique combinations of several factors of these may accelerate or cause slope failure (Leonarduzzi et al., 2017; Prokešová, Alžbeta, Táborík, & Snopková, 2012).

These predisposing or pre-conditional factors should not be mistaken with landslide causes, and triggers (Brönnimann, 2011). Causes refer to long term processes leading to slope instabilities, distinguished between geological, morphological, physical and human causes, triggers are sudden events that releases or reactivates the landslide, and predisposing factors are used to distinguish different characteristics of a slope which build up specific sub-surface conditions that make the area/slope prone to failure, and which may change over a long time span (Brönnimann, 2011; Werner & Friedman, 2010).

The most common predisposing factors that influence landslide occurrence are lithological and hydrogeological conformation defined by permeability and degree of saturation of the slope layer, geomorphology, volcanic activity, land use change, and anthropogenic impacts (Urgilez Vinueza et al., 2020). These are affected by time-dependent factors such as the permanently present rainfall throughout the entire year, tectonic activity, erosion processes and surface conditions (Urgilez Vinueza et al., 2020).

4.4.1. Triggering mechanisms

A triggering mechanism describes the physical, chemical and mechanical function of the triggering process that is connected with the loss of strength of the soil (Brönnimann, 2011). Commonly, rainfall is the most frequently used trigger to describe landslides in reports and the media after strong storms (Prokešová et al., 2012). This is due to several studies in which rainfall has been identified as the threshold parameter that determines the occurrence of landslides. However, there are other factors such as

antecedent water content, time-variant geotechnical parameters within the regolith, vegetation cover and land use, and the tectonic activities around the area that might play an important role triggering landslides (Urgilez Vinueza et al., 2020). This is why it is important to analyse triggering mechanisms before stating a trigger.

Landslides can be initiated naturally or artificially by increasing the effective stress, which is controlled by the increase in pore-water pressures due to intense or long-lasting rainfalls, the change of the slope geometry by undercutting of the slope foot (e.g., by river erosion), reduction of the strength of the slope by weathering or land-use changes (deforestation, reclamation) and/or earthquakes in seismically active areas (Prokešová et al., 2012; Urgilez Vinueza et al., 2020). These triggers are dynamically linked to slope deformation and progressive failure.

The most common natural landslide triggers such as intense rainfall, rapid snowmelt, water-level change, volcanic eruption, earthquake shaking and human activity are described and exemplified according to Wieczorek (1996):

- Intense rainfall as a trigger of landslides has been widely studied in several regions worldwide, and such studies show that thresholds of combined intensity and duration may be necessary to trigger shallow landslides in steep slopes of loose or weak soils and weathered rock during intense storms. This happens when rainfall infiltrates quickly due to soil saturation and a temporary rise in pore-water pressures. Taking into account seasonality, extreme precipitation received in a winter period can act as trigger for deep-seated land-slides, while extreme precipitation received in summer period can acts as important precondition factor (Prokešová et al., 2012).

- Rapid melting of a snowpack happens during warming spells or by rain falling on snow, which recharges shallow fractured bedrock, raises pore-water pressures beneath soils due to a more continuous supply of moisture to soils over a longer time period compared with the usual duration of infiltration from rain.

- A sudden water-level decrease produces high pore pressures in a slope of thick uniform deposits of low permeable clays and silts. This subjects a slope to high shear stresses and potential instability, which usually triggers landslides in earth dams, along coastlines, and on backs of lakes, rivers and reservoirs.

- Volcanic eruptions have historically triggered landslides in form of mud flows, debris flows and slide-debris avalanches. This is due to deposits of volcanic ash that are eroded from intense rainfall or quick snowmelt from pyroclastic flows and surges.

- Earthquake shaking usually produce rock falls, soil slides and rock slides in steep slopes, and earth spreads, earth slumps, earth block slides, and earth avalanches in gentle slopes. This happens because shaking produce liquefaction in the soils where pore-water pressures are temporarily increased and the strength of the soil is reduced. In high seismicity regions with rainy seasons, such as Ecuador, earthquake and climatic triggering mechanisms have complex interactions in which one mechanism could prepare the conditions to trigger a landslide by alternatively fracturing/weakening the medium or increasing the pore pressure, which is primarily driven by precipitation and water infiltration (Zerathe et al., 2016).

- The importance of anthropogenic factors such as human settlements, agriculture, land use change, and mining lies in the fact that in most developing countries found in tropical areas there is not always adequate land planning and controls (Urgilez Vinueza et al., 2020). When these changes of land use are not planned and suddenly modifies the stresses of the ground landslides might be triggered, e.g. removing vegetation in high slopes, road construction, and mining explosions.

V. DATA AND METHODS

5.1 Data

In order to identify factors that trigger areas susceptible to future landsliding, a range of attributes associated with the presence or absence of landslides is typically examined based on the knowledge of past landslide events (Goetz et al., 2011). Pourghasemi (2017) mentions a deep review on conditioning factors mostly used in landslide susceptibility analyses. These factors consider terrain attributes, geology and anthropogenic environmental conditions which are all associated to some degree with the occurrence of a landslide depending on the location (Brenning, 2005). Thus, the data used for the study of landslide triggering mechanisms in the Imbabura province embraces a set of descriptors of geologic and geomorphologic attributes as well as vegetation cover and land use.

Moreover, to apply GLM and GAM methods in the current study, a spatial database of conditioning factors was designed and constructed. Hence, preparing the datasets encompasses five phases as shown in the flow diagram of Figure 4: gathering of initial data such as landslide locations, files of digital elevation maps (DEM) of the study area, processing the gathered data accounting for errors, identifying and choosing the attributes in each landslide location, data quality checking, and analysing the constructed database.
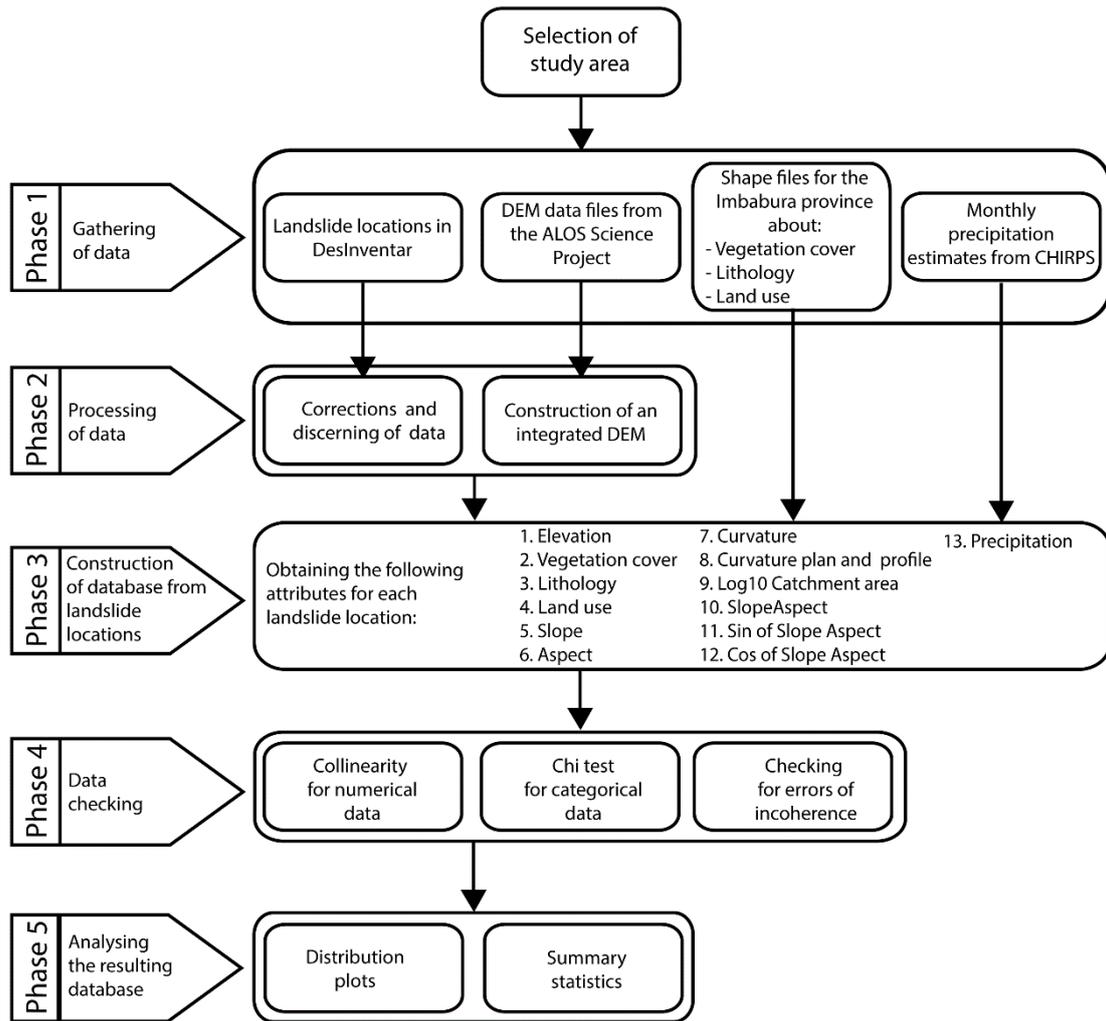
Figure 4. Flow diagram showing the four stages for the construction of the database used in this study. ALOS: Japan Aerospace Exploration Agency.

5.1.1 Attributes

Factors controlling slope stability considered within current physically based models for shallow translational landslides are terrain-based, such as elevation, slope, plan curvature, profile curvature, base 10 logarithm of the contribution or catchment area, the sine and cosine of the slope aspect; topology-based, such as geologic cover, vegetation cover, land use change; and climatic, such as precipitation.

The terrain attributes were derived from a digital terrain map that serve as proxies to these factors for landslide-controlling processes by representing surface processes and geophysical site conditions, and simplify complex geomorphological relationships (Brenning et al., 2015). The terrain attributes considered in this study are:

- Elevation (m). It is one of the most important terrain factors influencing changes in temperature, moisture, and wind at small spatial scales (Morán-Tejeda, López-Moreno, & Beniston, 2013; Pourghasemi & Rossi, 2017).

- Slope angle (°). It is considered to be the most important factor controlling slope stability in physically based models (Vorpahl, Elsenbeer, Märker, & Schröder, 2012). Shear stress in unconsolidated material increases producing a decrease of stability with increasing Slope (Lee & Min, 2001).

- Aspect (°). It is significant in this type of study because it is influenced on the slopes by rainfall direction, the amount of solar radiation, and geological settings (Pourghasemi & Rossi, 2017). That is why the slope's aspect is considered, and specifically their sine and cosine transformations in order to express this circular variable using two independent variables that represent east–west and north–south exposure components, respectively, as followed by Brenning (2015).

- Upslope contributing or catchment area (m^2). It is the area that potentially affects groundwater pressures at a prospective landslide slip surface (Iverson, 2000). It was transformed logarithmically (to the base 10) to reduce skewness (Brenning et al., 2015).

- Profile curvature. It is parallel to the slope and indicates the direction of maximum slope (acceleration and deceleration of flow across the surface), while plan curvature is perpendicular and represent the convergence and divergence of topography (concave, convex) (Goetz et al., 2011; Lee & Min, 2001). This information also can be useful to categorize hillslopes into three regions: 1) hollows, where plan curvature of the contours is concave in the downslope direction and where surface water would converge as it moves downslope, 2) noses, where the plan curvature of the contours is convex in the down-slope direction, and surface water will diverge, and 3) planar regions where the values are 0 (Ohlmacher, 2007). In this sense, this can show subsurface water conditions and substrate properties that have an important influence on landslide occurrence (Goetz et al., 2011).

Also, among the information that depends on topological site characteristics, geological cover, vegetation cover and land use were also selected because they represent relevant environmental conditions (Brenning, 2005; Brenning et al., 2015). Namely:

- Geologic cover. It can determine slide-prone areas due to certain types of rocks and geologic conditions that favour landsliding wherever they occur on slopes. In this sense, fine-grained clastic rocks (mostly formed by clay and silt) are most prone to landsliding, and more if they are poorly consolidated or overlain by more resistant rocks such as limestone, sandstone, or basalts. Moreover, highly sheared or deformed rocks, particularly tectonic melanges, slide extensively. Loose slope accumulations of fine-grained surface debris such as colluvial deposits are highly susceptible to sliding, particularly at times of intense precipitation (Radbruch-Hall & Varnes, 1976).

- Vegetation cover. It is one of the key factors that regulate the surface hydrology in humid climate regimes since it protects the surface from raindrop impact, controls the infiltration rate, and reduces surface runoff (Molina et al., 2008). A lack of vegetation cover can easily show signs of soil cohesion loss in slopes, leading to landslides (Goetz et al., 2011), and this is why planting or encouraging natural growth of vegetation is seen as an effective means of slope stabilization (Highland & Bobrowsky, 2008).

- Land use changes. These changes weaken slope stability. In a study by Zhang (2019), higher occurrence rates were found for both landslide and debris flow on cultivated lands than on forests and grasslands. This is a clear example that the destruction and conversion of natural and dense land cover make these areas more susceptible to natural hazards such as floods, landslides, and rockfalls. Several detailed case studies of landslides in the Andes showed how land cover and land cover conversion, e.g., from forest to pasture, affected slope deformation in south eastern Ecuador (Urgilez Vinueza et al., 2020).

Finally, the climatic attribute precipitation was selected since it is known to be the main driver on the kinematics of landslides (Eidt, 1969; Zerathe et al., 2016).

5.1.2 Gathering of data

The first step to construct a complete database for this study was the gathering of landslide data, including locations, a digital elevation map of the studied area, some available shapefiles such as vegetation cover, lithology, and land use for the Imbabura province, and precipitation estimates. These initial datasets served as a base to derive other attributes such as elevation, slope, catchment area, aspect and curvature, which were used as proxies for landslide-controlling processes altogether.

The landslide locations were obtained from DesInventar (DesInventar, 2020), which is a conceptual and methodological tool for the generation of National Disaster Inventories and the construction of databases of effects on populations, property and infrastructure that are vulnerable to natural and socio-natural phenomena.

About 144 landslide records were registered in the DesInventar database since 1989. These were defined as landslide events, caused by intense rainfall, and producing any type of damage on sectors such as water supply, health sectors, industries, communications, sewerage, education, transportation, power and energy, and agriculture. In total 61 landslide locations since 2014 were mapped and initially used in this study since the rest of the records lacked information of interest such as geographic location.

Files of digital elevation model (DEM) covering the Imbabura province were downloaded from the Advanced Land Observing Satellite (ALOS) project by the Japan Aerospace Exploration Agency (JAXA). The geoTIFF files take in 1° x 1° in latitude and longitude, and their horizontal resolution is approximately 30 meters. This dataset was obtained from the Panchromatic Remote-sensing Instrument for Stereo Mapping (PRISM), which is an optical sensor on board the ALOS (JAXA, 2020).

Precipitation estimates consisted in monthly averages since January 1981 until February 2015 for the Imbabura province, and were obtained from the global database of CHIRPS available in the International Research Institute for Climate and Society online data library (Columbia University, 2021).

Finally, shapefiles of vegetation cover, geologic cover and land use were obtained from the database of the National System of Information (SNI), which unfortunately has not been updated since 2014 (SNI, 2014). The cartographic information of these datasets is specified in Table 1.

Table 1. Cartographic information of the obtained and produced datasets for this study.

Datasets	Format	Year	Scale	Description	Source
Geologic cover	Vector	2005	1:100000	Lithology	MAG[1]
Vegetation cover	Vector	1990	1:250000	Types of vegetation	MAG[1]
Land use	Vector	n.d.	1:250000	Anthropogenic activities	MAG[1]
DEM	Raster	2018	1:80000	Digital elevation model	JAXA[2]
Slope	Raster	2018	1:80000	Slope angles	Author
Aspect	Raster	2018	1:80000	Aspect angles	Author
Curvature	Raster	2018	1:80000	Curvature	Author
Curvature plan and curvature profile	Raster	2018	1:80000	Curvature plane	Author
Catchment area	Vector	2018	1:80000	Catchment area	Author
Slope Aspect	Raster	2018	1:80000	Slope Aspect	Author

5.1.3   Initial data processing

Because of the importance of these first datasets on the construction of more attributes, the dataset had to be processed. In this sense, the obtained landslide points lacked or had switched coordinates' fields, and different coordinate systems, so they were manually discarded or corrected. Then, using ArcGIS software, the resulting landslide locations were projected into the UTM Zone 17N coordinate system, referenced to the WGS1984 datum. Next, the DEM files were merged for an easier management of the raster, also projected into the previously mentioned coordinate system and plotted using ArcGIS.

5.1.4   Construction of the database

Once the initial datasets were ready to be used, the terrain attributes were derived for each landslide location using the available tools of ArcGIS, and recorded in an Excel spreadsheet: elevation, slope, aspect, curvature, curvature plan and profile, catchment area, base 10 logarithm of the catchment area, slope aspect, and the sine and cosine of the slope aspect.

---

[1] Ministerio de Agricultura y Ganadería
[2] Japan Aerospace Exploration Agency

5.1.5 Data checking

In order to avoid inclusion of redundant variables, bias in estimation of coefficients, and model overfitting, the dataset was curated. This was done as follows: 1) a general data check, 2) a multicollinearity analysis for the eight continuous attributes, and 3) a Chi-Square test for independency for the three categorical attributes.

1) The basic check looked for inconsistencies between the attributes, outliers or geographic errors, which lead to the reduction of the amount of landslide locations to 56. Three repeated landslide locations, and two incoherencies between vegetation cover and land use were identified, e.g. where wasteland areas showed also a land use of agriculture.

Moreover, regression models for landslide triggering were needed to code the binary (two level) qualitative response to be incorporated in the model, and this was done by creating an indicator or dummy binary: presence and absence of a landslide (Equation 1) (Casella, Fienberg, & Olkin, 2013).

$$Y = \begin{cases} 0 & \text{if a landslide occurred (presence);} \\ 1 & \text{if a landslide did not toccur (absense).} \end{cases} \quad (1)$$

There are several methods used in this field such as bootstrapping resampling that creates a new dataset based on the actual data (Brenning et al., 2015; Goetz et al., 2011), random sampling (Lombardo & Mai, 2018), using the distance between parameter distribution functions on cross-validated data sets (Motrenko, Strijov, & Weber, 2014), categorization of 10-m cells in all of a gridded area in ArcGIS (Ohlmacher & Davis, 2003), seed cell theory which considers close vicinity areas of landslides as stable slopes (Yesilnacar & Topal, 2005), and stepwise discriminant analyses (Ayalew & Yamagishi, 2005). Therefore, there is no consensus on the criteria nor available guidelines on how to obtain this information. For the sake of practicality in this study, the discrimination of event was conditioned to the occurrence of precipitation anomalies. Monthly averages of precipitation amounts were standardized within 1981-2015 period as to identify an empirical rainfall threshold for landslide occurrence that represents above normal wet conditions associated with a landslide event. In this sense, a monthly precipitation anomaly equal to 0.6 standard score was found to represent a threshold for wetter than normal conditions. The application of the threshold criteria was as follows: if the standardized rainfall for a specific landslide was smaller than 0.6 (value also found by dividing the dataset into two groups with similar size), it was considered a non-landslide event and the value assigned for the landslide occurrence was "0"; and "1", if the value

was larger than 0.6. The monthly precipitation anomaly data was computed and available at the library of IRI.

2) A generalized linear model requires to check for multicollinearity of independent variables, which is a phenomenon where multiple predictor variables have a substantial correlation with each other in a regression model (Kuhn & Johnson, 2013). When this is the case, one variable can be linearly predicted from the others with some degree of accuracy, and the variables are contributing to the total variance because they present the same information (Chen et al., 2017). Some reasons to avoid data with highly correlated predictors are that i) redundant predictors add more complexity to the model than information they provide to the model; and ii) highly unstable models, numerical errors, and degraded predictive performance can be yielded by techniques like linear regression (Kuhn & Johnson, 2013; Mackinnon & Puterman, 1989). However, linear model collinearity diagnostics are not always appropriate in generalized linear models, which means that the assessment of the presence of collinearity should be based on an information matrix and not on a correlation matrix (Mackinnon & Puterman, 1989). This quick check of the collinearity between predictors was done visually using the function "pairs()" in the software "R", and numerically with the function "cor()" which use the Pearson's correlation method.

3) Moreover, there were three categorical variables that needed to be checked for statistical independence or association between them, so that they could be considered statistically significant. In this sense, Davis (2005) highlights the utility of the Chi-Square test for independency, which helps to quantify how well the categorical data fits a distribution like the null distribution . In order to apply this test, certain requirements needed to be met:

- Two categorical variables.
- Two or more categories (groups) for each variable.
- Independence of observations: there is no relationship between the subjects in each group and the categorical variables are not "paired" in any way (e.g. pre-test/post-test observations).
- Relatively large sample size. This means that the expected frequencies for each cell are at least 1, and at least 5 for the majority (80%) of the cells.

If all the requirements are met, the hypotheses are written:

- H_0: the variables are independent, there is no relationship between the two categorical variables.
- H_1: the variables are dependent, there is a relationship between the two categorical variables.
- Level of significance: $\alpha = 0.05$

The formula that describes the Chi-Square test is based on a table of frequencies of the data where rows, R, belong to the levels of one variable, and columns, C, belong to the levels of the other. Basically, it compares the observed values in the dataset to calculated expected values that assume there is no relationship between the variables (Equation 2) (Davis, 2002):

$$\chi^2 = \sum_{i=1}^{R} \sum_{j=1}^{C} \frac{(o_{ij} - e_{ij})^2}{e_{ij}}, \tag{2}$$

where o_{ij} is the observed cell count in the ith row and j[th] column of the table. The term e_{ij} is the expected cell count in the ith row and j[th] column of the table if the null hypothesis is assumed to be true, i.e. the joint probabilities of both variables is equal to individual probabilities multiplied together (Equation 3) (Hand, 2007).

$$e_{ij} = \frac{row\ i\ total * col\ j\ total}{total\ observations} \tag{3}$$

The calculated χ^2 value is then used to obtain a p-value from the right-tailed χ^2 distribution with degrees of freedom $df = (R - 1)(C - 1)$, using the function chisq.test() in the software "R". This value is the probability, under the null hypothesis, that the observed χ^2 value could be obtained by random chance (Davis, 2005), and it has to be equal or larger than the chosen level of significance α, which means that the probability of observing such a difference between the observed and expected frequencies is unlikely, and the alternate hypothesis can be rejected at the chosen level of significance. Or, if using the critical value approach, this value is obtained from the χ^2 distribution table with the same conditions, which creates a rejection zone where if the χ^2 value is large enough to fall into, then the null hypothesis is rejected.

One last detail on logistic regression is the sample size. Studies shown the importance of having sample sizes greater that 100 (Bujang, Sa'at, Tg Abu Bakar Sidik, & Lim, 2018), this is due to the fact that logit models use maximum likelihood estimation techniques, in this sense, a re-categorization of data in the categorical variables vegetation

27

cover and land use was used due to lack of data as shown by Molina et al (2008). The specific changes are detailed as follows and the Table 2 shows a comparison that displays the improvement.

- Dry Scrubland and Wet Scrubland were categorized into Scrubland.
- Crops of tempered zones, Corn orchards, Orchards were categorized into Orchards.
- Paramo vegetation, Shrubby and herbaceous vegetation (C&P) and Wasteland were deleted (reducing the dataset from 56 to 52).
- Livestock (C&P) was added to Livestock.
- Agriculture (C&P) was added to Agriculture.

Table 2. Comparison table of the changes made to improve the dataset respect to landslide occurrence size. Problematic levels in both vegetation cover and land use variables are in bold.

			Landslide occurrence	
			0	1
Vegetation cover (initial dataset)		**Corn orchards**	3	0
		Crops of tempered zones	2	0
		Dry scrubland	9	6
		Humid forest	5	4
		Orchards	0	1
		Paramo vegetation	1	0
		Pasture crop forest	7	16
		Sugar cane crops	1	0
		Wet scrubland	1	0
Vegetation cover (after reorganization and deleting)		Humid forest	5	4
		Orchards	5	1
		Pasture crop forest	6	15
		Scrubland	10	6
Land use (initial dataset)		Agriculture	9	7
		Agriculture (C&P)	3	1
		Agric. and Livestock	5	2
		Forest (C&P)	3	5
		Livestock	6	11
		Livestock (C&P)	1	0
		Shrubby and herbaceous vegetation (C&P)	2	0
		Wasteland	0	1
Land use (after reorganization and deleting)		Agriculture	12	8
		Agric. and Livestock	5	2
		Forest (C&P)	3	5
		Livestock	6	11

5.1.6   Analysing the dataset

Once the dataset was developed, it is necessary to understand more the distribution of variables with the use of summary statistics and distribution plots. The summary statistics found in the Table 3 shows the main statistical characteristics of the sample of landslides in this study, which does not show outliers.

Table 3. Summary statistics of the attributes dataset.

Attributes	Mean	Standard deviation	Min	Max
Elevation	1727.07	804.70	569	3390
Slope	24.71	11.77	5.28	47.66
Curvature profile	0.09	0.96	-1.82	1.936
Curvature plan	-0.08	0.82	-1.60	3.21
log10 of Catchment area	3.62	0.89	2.81	7.14
Sine of the slope aspect	0.04	0.73	-0.99	0.98
Cosine of the slope aspect	0.07	0.66	-0.99	0.99
Precipitation	142.85	84.98	20.50	467.53

In Figure 5, the histograms of the distributions of all continuous attributes show that the data has varying scales or ranges and only a few variables have Gaussian distribution, such as the curvature profile and curvature plan. The curvature profile distribution looks platykurtic, while the curvature plan is more leptokurtic. The precipitation and log10 of the catchment area are mesokurtic and positively skewed. In the case of the elevation, slope and sine and cosine of the slope aspect, the distribution seems to be bimodal, which for the case of elevation and slope, it means that landslides may occur in both very high and low elevations, and high and low slopes.

Due to the characteristics of the continuous data, it is necessary to apply a normalization of the data which will change the values of numeric columns in the dataset to a common scale, without distorting differences in the ranges of values.

In the categorical data, the principal vegetation cover types correspond to forests of pasture crops, dry scrublands and humid forests. The principal land use types correspond to livestock and agriculture (Figure 6). Specifically, the livestock category correspond to mainly pasture crops and natural pasture; forests are protected lands of natural forests or a mixture of 70-30% forest and pasture crops; agriculture corresponds to corn, sugar cane and short cycle crops for agriculture; and the livestock and agriculture category corresponds to shared land use of both 70% pasture crops and 30% crops.
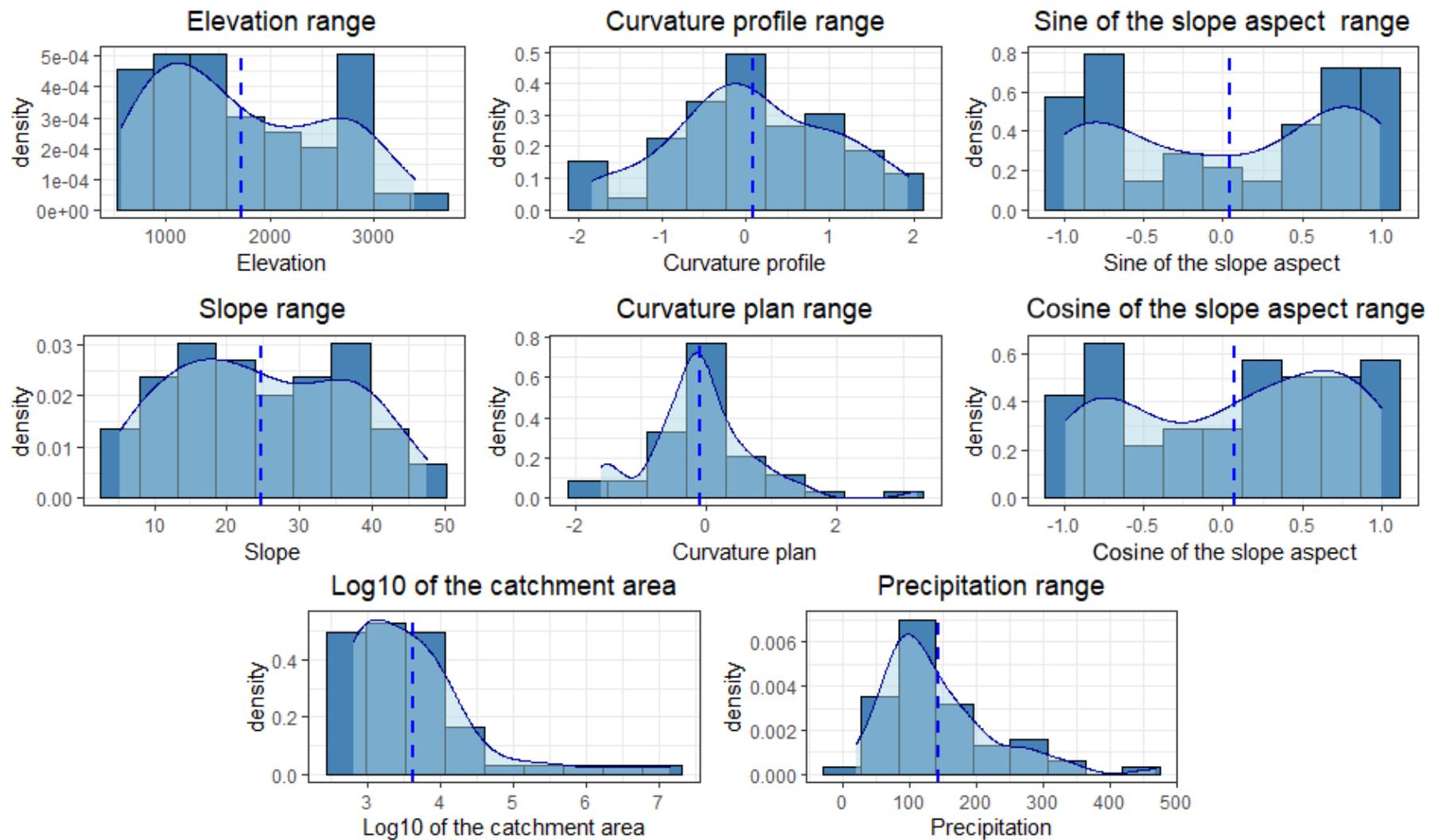
Figure 5. Density plots of the eight continuous attributes. The blue dashed line represents the mean of the attribute.
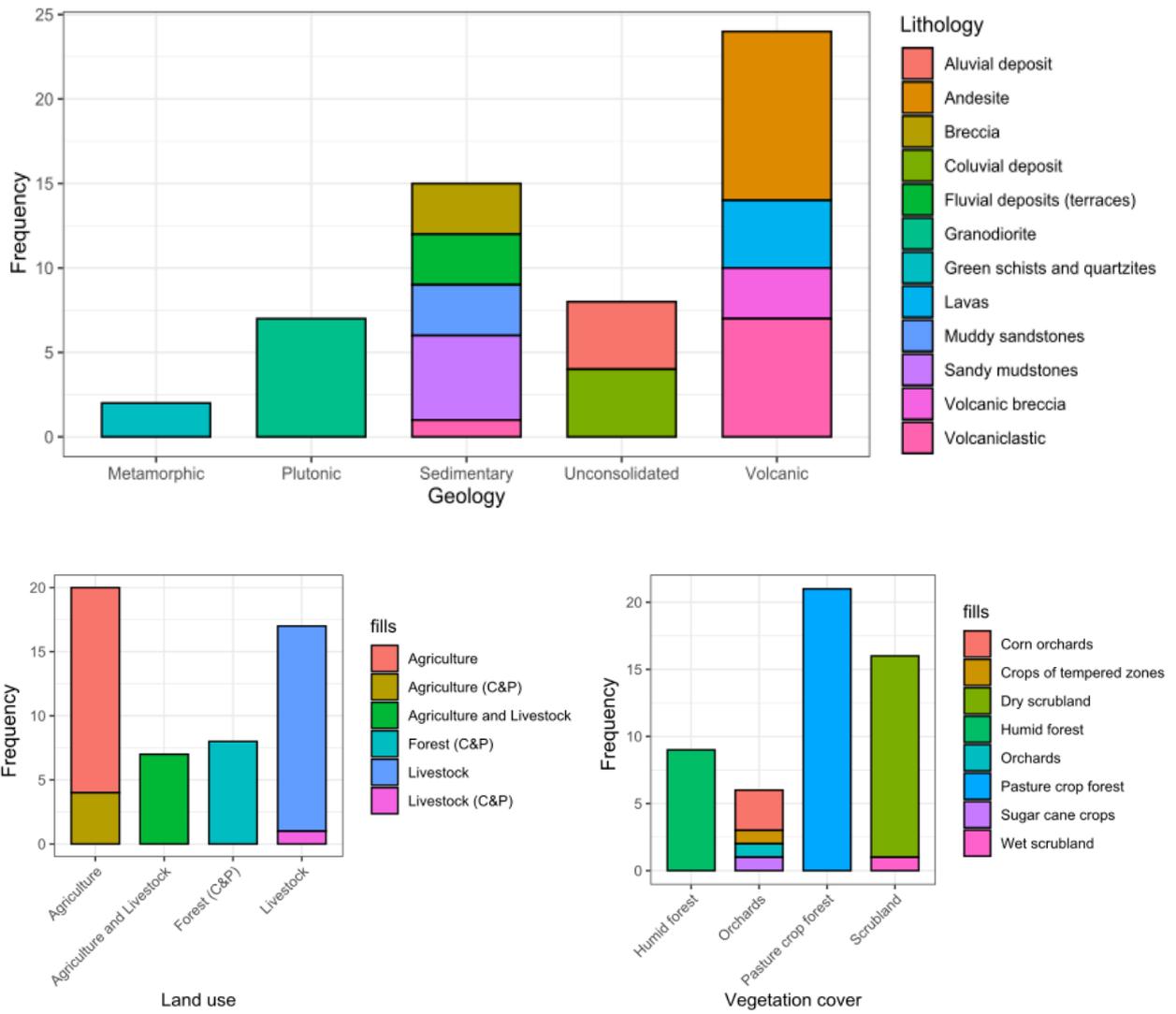
Figure 6. Frequency plots of the categorical attributes: vegetation cover and land use
(C&P: Conservation and protection).

5.2 Methods

The methodology of this study covers approaches for predicting a qualitative response, a process that is known as classification because it involves assigning the observation to a category or class (Casella et al., 2013). GLM and GAM are two classification techniques or classifiers used to predict a qualitative response, or more, which is the case for this study. Namely, to generate and analyse the relationship between the response variable (landslide occurrence) which depends on multiple predictor variables (drivers of landslide occurrence). Moreover, the performance of these models is evaluated using the Area Under Receiver Operating Characteristic Curve (AUC-ROC) approach.

5.2.1 General linear models

There are cases where the relationship between a response variable and a predictor variable is linear (Equation 4). Here, x_i is the ith observation on a predictor variable X_i, and where y_i is the ith observation on response variable Y_i, who has a normal distribution and whose expectation, $\mu_i \equiv \mathbb{E}(Y_i)$.

$$Y_i = \mu_i + \epsilon_i, \qquad \text{where } \mu_i = X_i b \qquad (4)$$

In other words, Y, a continuous response variable, is given by both categorical or continuous predictor variable X multiplied by a constant b or regression slope, which can be estimated by minimizing the residual sum of squares with respect to b, plus a random term ϵ, which is the part of Y that could not be explained by the predictors in the model.

5.2.2 Generalized linear models

Linear models can be "generalized" by allowing the response variable Y to be both continuous or categorical, and dependent on multiple predictor variables, which can themselves be transformations of the original predictors and handle several types of distributions, plus an additive constant.

The structure of a generalized linear model has three components: 1) systematic components such as constants, coefficients and predictor variables, 2) random components such as the response variable and the error term, and 3) link functions that, as its name says, links or relates both random to systematic components in the model. This link function transforms the probabilities of the levels of a categorical response

33

variable to a continuous scale that is unbounded, so that the relationship between the predictors and the response can be modelled with linear regression.

GLMs assume a linear relationship between a link function of the dependent variable and several independent variables (Pourghasemi & Rossi, 2017). Also, these models allow for response distributions different than normal, and for a degree of non-linearity in the model structure (Wood, 2017). They usually make distributional assumptions that the random variable Y_i is independent and belongs to the exponential family, which includes distributions such as binomial, gamma, Poisson and normal (Wood, 2017). Its structure is as in Equation 5.

$$g(\mu_i) = X_i b \qquad (5)$$

where $\mu_i \equiv \mathbb{E}(Y_i)$, g is a smooth monotonic link function (i.e. logistic for a logistic regression) of the dependent or response variable. X_i is the ith row of a model matrix, X, and b is a vector of unknown factors or predictors.

5.2.2.1 Multiple logistic regression

Logistic regression is one of the most popular forms of generalized linear models, characterized by their response distribution (in this study the binomial distribution) and a link function (logit) that transfers the mean value to a scale in which the relation to background variables is described as linear and additive (Dalgaard, 2008). Here, the response variable is sampled as a binary variable, such as presence or absence of a landslide, but rather than modelling it directly, the probability of the presence and absence of landslides (present conditions) is modelled given the observed values of the predictor variables (pre-failure conditions) (Gorsevski, Gessler, & Elliot, 2006). This relationship is written in Equation 6.

$$Y = logit(p) = \log\left(\frac{p}{(1-p)}\right), \text{where } p = Pr(Y = 1 \mid X). \qquad (6)$$

Here, p is the conditional probability (given the predictors X) that the response variable Y equals one, which will range between 0 and 1, and $\frac{p}{(1-p)}$ is the so-called odds or likelihood ratio (Casella et al., 2013; Lee & Predhan, 2007). The logistic regression involves fitting an equation whose log odds or logit is linear in X, in the form of Equation 7.

$$Y = \log\left(\frac{p}{(1-p)}\right) = b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_n x_n, \qquad (7)$$

where b_0 is the intercept of the model, b_i $(i = 0, 1, 2, \ldots, n)$ are the slope coefficients of the logistic regression model, and x_i $(i = 0, 1, 2, \ldots, n)$ are the predictor variables. The coefficients that result from logit regression represent the change in log odds due to incremental-unit changes in the values of the predictors (Ohlmacher & Davis, 2003), so that increasing X by one unit changes the log odds by b_1 (Casella et al., 2013).

Then, the coefficients that fit the logistic regression model are estimated using the available training data and a method called maximum likelihood, where it is necessary to find the coefficients such that plugging them into the model for $p(X)$ yields a number close to one for all predictors that influenced the production of the landslide, and a number close to zero for all predictors that did not (Casella et al., 2013).

In the software "R", the "glm()" function fits generalized models, passing the argument family=binomial to run specifically a logistic regression. The method of variable selection for the model was a forward stepwise selection, which consisted in initially withholding all the predictor variables from the model, then, those variables determined to be significant (i.e. the model has the smallest p-value) are added up to the model while all others are withheld, until all the variables under consideration are included in the model.

5.2.3 Generalized additive model for classification problems

GLM is a well-established tool for landslide susceptibility modelling, however linearity is unrealistic in many environmental modelling situations, and limits the predictive performance (Brenning et al., 2015; Goetz et al., 2011). Hence, in order to relax the linearity assumption while still attempting to maintain as much interpretability as possible while using several predictors, Hastie & Tibsgirani (1986) proposed GAMs. GAMs are a semi-parametric extension of the GLM that combines linear and non-linear relationships between predictor and response variables (Goetz et al., 2011). Nonlinear terms utilize non-linear functions or "smooths", which are calculated separately for each X, and then, all their contributions are added together, hence the "additive" term in this type of models (Casella et al., 2013).

Furthermore, GAMs can be used in situations where Y is qualitative, and in the case where it takes on values zero or one, the multiple logistic regression model explained previously can be extended to allow for non-linear relationships, and such a model results in Equation 8.

$$\log\left(\frac{p}{(1-p)}\right) = b_0 + f_1X_1 + f_2X_2 + \cdots + f_nX_n, \tag{8}$$

where, f_j is a smooth function of the variable X_j, whose smoothness can be summarized via degrees of freedom. The model can be fit using methods such as local regression, polynomial regression, step functions, basis functions, regression splines, smoothing splines, local regression, or any combination of them as building blocks (Casella et al., 2013).

Advantages of the GAMs are that 1) they allow certain non-linear relationships, which make them more flexible and can potentially make more accurate predictions of predictors, and 2) predictors can be examined individually while holding all of the other variables fixed. However, some limitations of these models are, 1) GAMs are less interpretable than linear regression since the relationship between each predictor and the response is now modelled using a curve, and 2) since the model is additive, some interactions between several predictors have to be manually added in the form $X_j \times X_k$ (Casella et al., 2013).

This model was carried out with the software "R" using the "gam ()" function from the mgcv package. Similarly, the method of variable selection for this model was forward stepwise selection too; applying splines for non-parametric smoothing of the variables, s(variable), indicates a nonlinear transformation.

5.2.4 Receiver operating characteristic curves

ROC curves are plots used to assess the accuracy of a diagnostic test that may be used to distinguish between two classes of events, and to visualize classifier performance (Gorsevski et al., 2006). In this study, the area under the ROC curve (AUC) was used as a measure of overall fit, to evaluate the model performance and to compare modelled predictions of landslide occurrence for several GLMs and GAMs. In dichotomous statistical modelling such as logistic regression, ROC curves are very useful for evaluating the predictive accuracy of a chosen model considering the quality and amount

of input predictors (Fustos, Abarca-del-Rio, Moreno-Yaeger, & Somos-Valenzuela, 2020).

What these plots specifically show is the probability of having a true positive rate (correctly predicted response such as the prediction of a landslide where a landslide occurred) versus the probability of a false positive rate (falsely predicted response such as a prediction of a landslide at a location where a landslide did not occur) as the cut-off probability (predicted value) varies (Gorsevski et al., 2006). The true positive rate is often called the sensitivity test, and constitutes the Y axis on the ROC curve, while the false positive rate is often called the specificity test, and one minus this value constitutes the X axis on the ROC curve. If the value of the area under the curve is close to 1, the probability of a true positive is close to 1 and the model is ideal, however it does not provide the optimal cut-off value (optimal decision threshold) nor does it illustrate how landslide occurrence affects cut-off selection.

VI. RESULTS AND DISCUSSION

6.1 Chi-Square for independence between categorical attributes

As observed in the Table 4, the test statistics which contain the attribute Geology are larger than the critical value or the p-value is smaller than the significance level, hence the null hypothesis is rejected. However, between the attributes veg cover and land use, the test statistic is smaller than the critical value or the p-value is larger than the significance level, meaning that there is not enough evidence in this dataset that suggests these variables are dependent.

In this way, the attribute geology cannot be taken into account for this study since this variable has substantial correlation with the other categorical data, and it can affect severely the regression.

6.2 Collinearity test for numerical variables

In this sense, it was regarded that a Pearson >0.7 means high collinearity (Bui, Tuan, Klempe, Pradhan, & Revhaug, 2016), hence, none of the continuous attributes taken into account for this study were considered collinear and there was no need to omit predictors or be aware of multicollinearity in the regression (Table 5). Although, it is important to note that there are some slightly collinear attributes, e.g. precipitation with elevation, and curvature plan with curvature profile (Figure 7, yellow cells in Table 5).

Table 4. Results of the test statistic Chi-Square for independence between the three categorical attributes, with both p-value and critic value approaches, for a significance level 0.05.

attributes\values	χ^2	df	critical value	Rejection zone	p-value	Reject
Geology vs Land use	118.19	77	98.48	[98.48, ∞)	0.00178	YES
Geology vs Veg cover	137.58	88	110.89	[110.89, ∞)	0.00057	YES
Veg cover vs Land use	71.17	56	74.46	[74.46, ∞)	0.08322	NO

Table 5. Numerical correlation of the continuous attributes. Yellow cells represent the most collinear attributes.

|	Elevation	Slope	Curv Profile	Curv Plan	log10 CatchA	Sin SlopeAspect	Cos SlopeAspect	Precipitation
Elevation	1.0000	-0.0377	0.0056	-0.0361	0.0893	-0.1416	-0.1289	-0.6196
Slope	-0.0377	1.0000	0.1640	-0.0635	-0.1880	0.0992	-0.0049	-0.0435
CurvProfile	0.0056	0.1640	1.0000	-0.5588	0.2720	0.0293	-0.0368	0.0552
CurvPlan	-0.0361	-0.0635	-0.5588	1.0000	-0.4159	-0.0776	0.0762	-0.0408
log10CatchA.	0.0893	-0.1880	0.2720	-0.4159	1.0000	0.0667	0.0893	-0.0628
Sin SlopeAspect	-0.1416	0.0992	0.0293	-0.0776	0.0667	1.0000	0.2537	0.1333
Cos SlopeAspect	-0.1289	-0.0049	-0.0368	0.0762	0.0893	0.2537	1.0000	0.2013
Precipitation	-0.6196	-0.0435	0.0552	-0.0408	-0.0628	0.1333	0.2013	1.0000
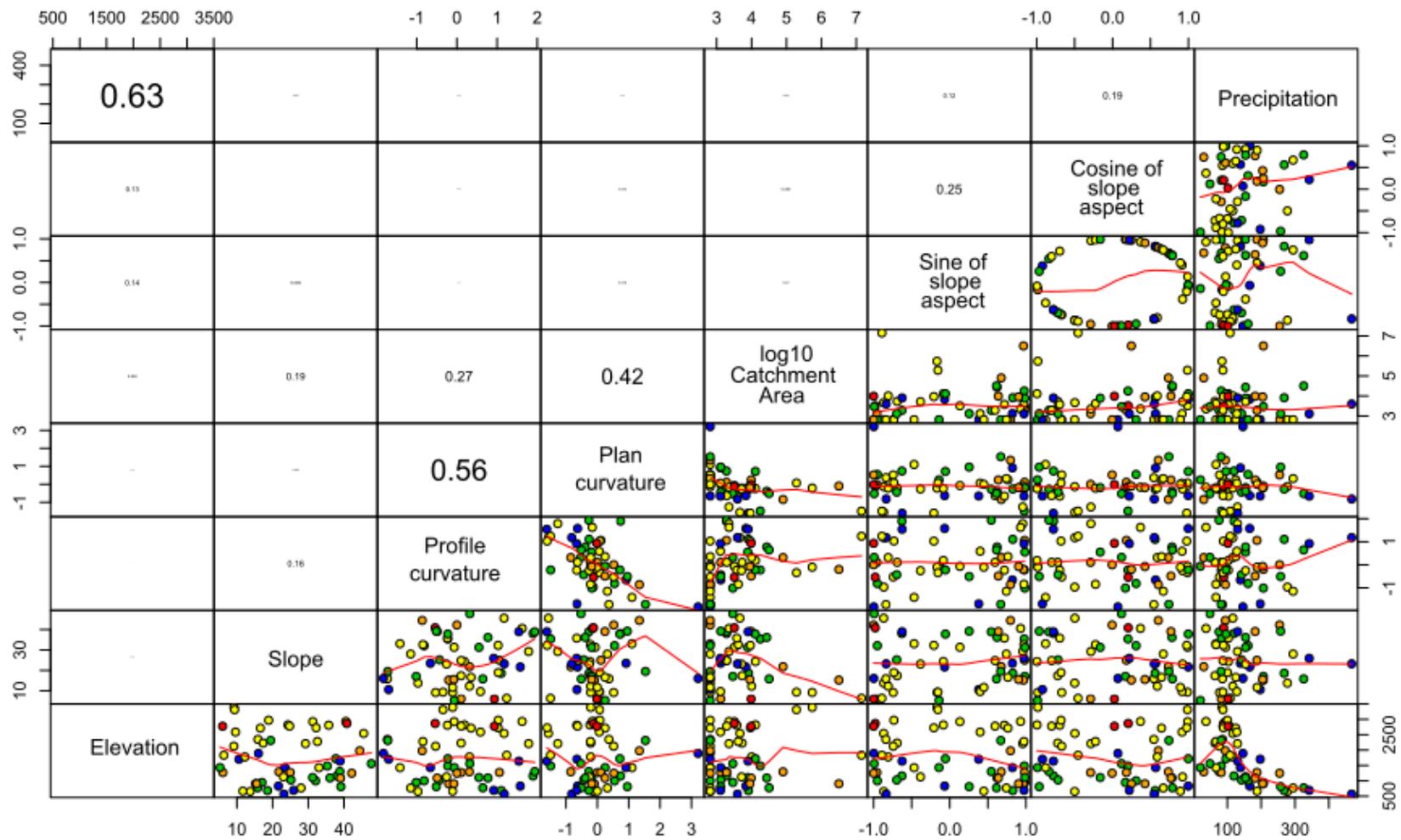
Figure 7. Collinearity test among attributes. The colored dots represent landslides categorized by type of rock (red: metamorphic, blue: plutonic, green: sedimentary, orange: unconsolidated, yellow: volcanic)

6.3 Logistic modelling approach

The selected model for the logistic regression accounted for every predictor and used polynomials in two of them: log10 catchment area and profile of the curvature. The summary of the deviance residuals in Table 6 shows how well the model fit the data measuring the distribution of the deviance residuals for individual cases used in the model. They seem acceptable since they are close to being centered on 0 and are roughly symmetrical.

Based on the coefficients, which give the change in the log odds of the outcome for every one-unit increase in the predictor variable, there are only three variables whose effect size is large according to Table 7: the land use for agriculture and livestock, elevation and precipitation. In other words, for every one unit change of standard deviation in the variables precipitation and elevation (84.13 mm and 804.7 m respectively), the odds of landslide occurrence (versus non-landslide occurrence) increases by 6.35173 and 5.98652, respectively. In terms of odds of landslide occurrence, there is an increase of 395.44 and 572.49 in the odds with every unit change of standard deviation increase in the precipitation and elevation variables. According to the continuous variables like land use agriculture and livestock versus the rest of the levels in land use, there is a decrease in the log odds of landslide occurrence by -7.6835, meaning that if a landslide was about to occur, then it is less likely to happen due to this predictor, i.e., areas with a shared land use of agriculture and livestock are 0.0004 times prone to landslide occurrence than those presenting pure forests, agriculture, and livestock land uses. This result is not coherent with several studies that show that agricultural lands lead to slope instability due to its effects on percolation and aquifer recharge, which decreases the effective soil cohesion because of saturation (Garcia-chevesich et al., 2021). This should be further studied taking into account irrigation systems and the role of short cycle crops versus pasture crops' roots in areas of shared land use.

This is better shown in the partial effects of the linear terms that add up to the overall prediction in Figure 8. Here, the plots express the fitted probabilities of each predictor in landslide occurrence. For the case of land use, agriculture and livestock separately have higher expected probabilities of increasing the odds of landslide occurrence. These categorical plots also include 95% error bars to show the uncertainty

of the estimate. For the categorical case of vegetation cover, there are negative odds of landslide occurrence.

A large absolute "z" value indicates evidence against the null hypothesis $H_0: b_1 = 0$, which implies that the probability of landslide occurrence does not depend on the respective predictor. Furthermore, their p-values are below 0.05, and thus, the log odds and the log odds ratios are both statistically significant at the 0.05 level and the H_0 can be rejected. There is indeed an association between landslide occurrence and land use for agriculture and livestock, elevation and precipitation. For the rest of predictors, the p-values are relatively large, so there is no clear evidence of a real association between landslide occurrence and them.

Furthermore, the predicted probability of landslide occurrence increases as all of the continuous predictors increase, although the confidence bands widen, which indicates there are some probabilities with high predictor scores, and hence less confidence in the predictions.

Finally, for the validation among all the logistic regression models that were tried, the best AUC-ROC test yielded 0.947 (AUC), which shows a high performance (AUC>0.9) considering the quality and the amount of predictors (Figure 9).

Table 6. Summary of deviance residuals.

Deviance Residuals				
Min	1Q	Median	3Q	Max
-1.49	-0.39	-0.0004	0.46	1.76

Table 7. Parameter estimates of the logistic regression model for the prediction of the probability of landslide occurrence using all eleven predictors. Estimated odds ratios and their 97.5% confidence intervals for landslide occurrence while accounting for all the predictors in GLM. Boldface indicates significant tests at the 5% significance level of the null hypothesis that the probability of a landslide does not depend on the predictor variable.

| Predictor variables	Estimate	Std. Error	z value	Pr(>|z|)	Exp(estimates)	2.50%	97.50%
(Intercept)	1.36	1.61	0.84	0.39	3.90E+00	1.90E-01	6.47E+02
Land use: Agriculture and Livestock	**-7.68**	**3.98**	**-1.92**	**0.05**	**4.60E-04**	**1.62E-08**	**1.80E-01**
Land use: Forest (C&P)	-4.68	2.93	-1.59	0.11	9.23E-03	1.09E-05	1.58E+00
Land use: Livestock	1.59	2.12	0.75	0.45	4.94E+00	1.04E-01	7.86E+02
Veg cover: Orchards	-3.35	3.08	-1.08	0.27	3.50E-02	2.32E-05	7.54E+00
Veg cover: Pasture crop forest	1.90	2.007	0.94	0.34	6.71E+00	2.28E-01	1.00E+03
Vege cover: Scrubland	-0.05	2.38	-0.02	0.98	9.48E-01	5.92E-03	1.07E+02
Elevation	**5.98**	**2.61**	**2.28**	**0.02**	**3.98E+02**	**7.29E+00**	**2.88E+05**
Slope	0.65	0.68	0.95	0.34	1.92E+00	5.41E-01	8.84E+00
Curvature profile, poly3,1	0.03	6.99	0.005	0.99	1.03E+00	3.52E-07	9.78E+05
Curvature profile, poly3,1	-6.51	6.24	-1.04	0.29	1.48E-03	1.95E-10	5.01E+01
Curvature profile, poly3,1	10.61	6.64	1.59	0.11	4.06E+04	4.02E-01	4.17E+11
Curvature plan	0.15	0.70	0.21	0.82	1.17E+00	2.61E-01	5.23E+00
Log 10 catchment area, poly3,1	8.46	15.08	0.56	0.57	4.73E+03	1.46E+03	NA
Log 10 catchment area, poly3,1	-2.10	16.25	-0.13	0.89	1.22E-01	3.61E-02	NA
Log 10 catchment area, poly3,1	-6.35	10.11	-0.62	0.52	1.73E-03	3.42E-13	3.27E+13
Sine of the slope aspect	0.41	0.57	0.72	0.46	1.52E+00	5.10E-01	5.45E+00
Cosine of slope aspect	0.71	0.82	0.86	0.38	2.04E+00	4.07E-01	1.27E+01
Precipitation	**6.35**	**2.54**	**2.49**	**0.01**	**5.73E+02**	**1.15E+01**	**3.30E+05**
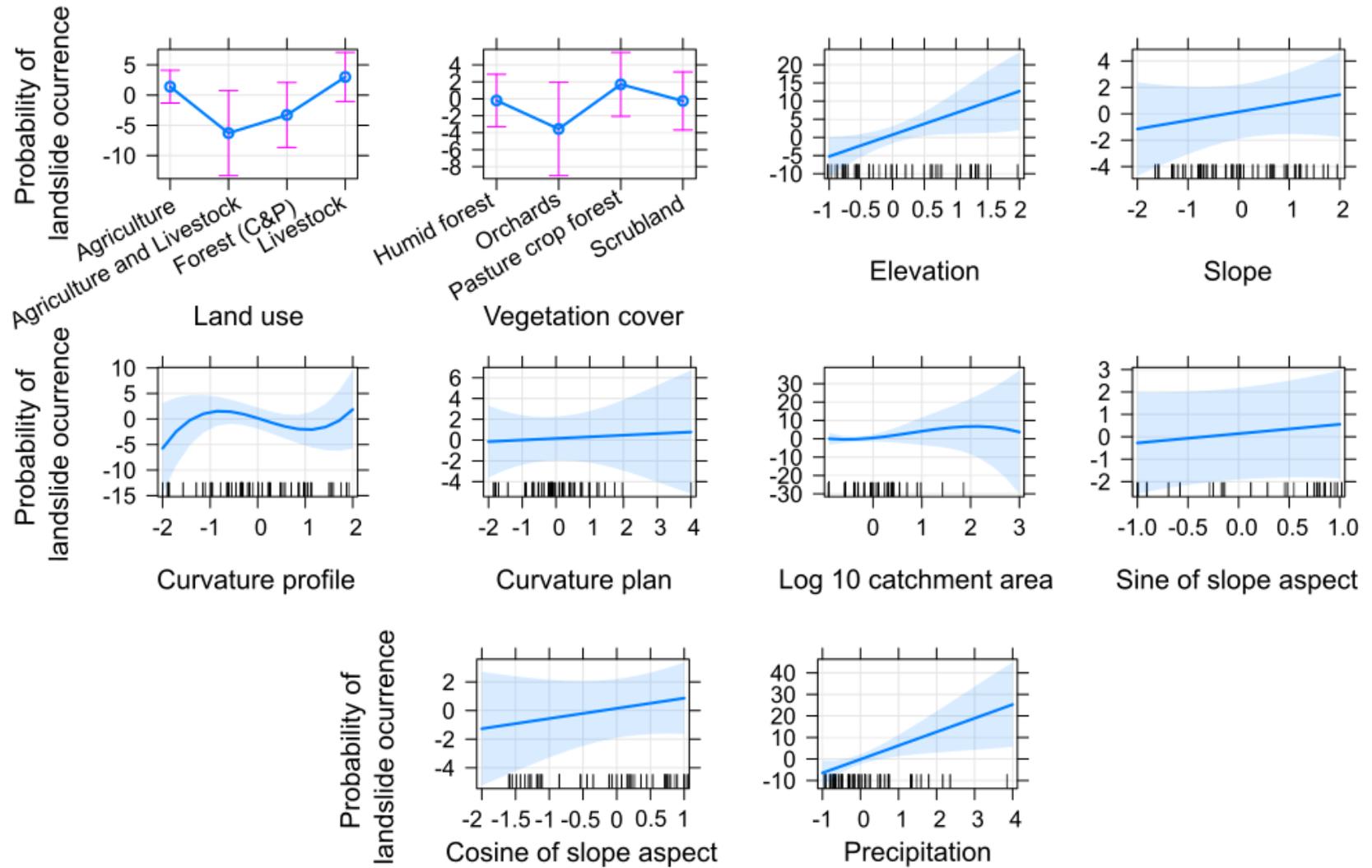
Figure 8. Plot of the log(odds) of each descriptor to landslide occurrence. Dashes on the x-axis are observations.
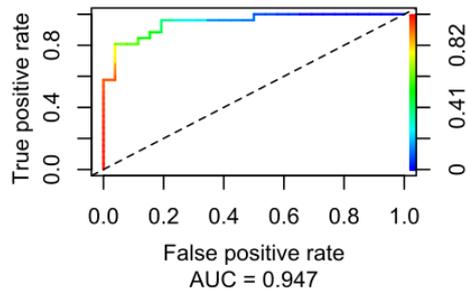
44

Figure 9. Predictive performance of the logistic model from a ROC curve that plots the false positive rate (FPR) on the x-axis and the true positive rate (TPR) on the y-axis.

6.4 Generalized additive model approach

Since the dataset is the same, the model assumes again a Binomial distribution of the errors in the model, and the link function is also a logit, which transforms the predictions into log of the odds.

Among all of the GAMs tried, an AUC-ROC was performed and four of them yielded 0.954 (AUC), which shows even a higher performance (AUC>0.9) than the GLM as shown in Figure 10.

In order to find the best model, an analysis of deviance table (goodness of fit) was constructed (Table 8), and the model with spline functions applied to elevation, log10 of the catchment area, and the sine of the slope aspect showed to be the more significant using a Chi Square test. Although in this model the default basis dimension k=10 was used, the effective degrees of freedom (edf) show little amount of turning points or knots, which show little complexity in the smoothing process. Table 9 shows the significance of these smoothing terms (along with parametric terms) by telling which smooth curves are explaining the landslide occurrence response. However, the real nature of this explanation or significance of the smooth term can only be visualized in the elevation response curve (edf=2.32) of Figure 11, which shows a change after the spline, so that a horizontal line cannot be drawn through the 95% confidence interval, and the original response value is above the average intercept.

Furthermore, other linear parameters showed significance in the model: agriculture and livestock with an estimate of -5.62 under a 10% significance level, and cosine of slope aspect with an estimate of 1.72 under a 5% significance level. This

45

means, as in the GLM, that if a landslide was about to occur, then it is less likely to happen in an area where agriculture and livestock is the main land use, and more likely to depend on the cosine of the slope, since for every one unit change of deviation standard (0.66) in cosine of the slope, the log odds of landslide occurrence (versus non-landslide occurrence) increases by 1.72. In terms of odds of landslide occurrence, there is an increase of 5.78 times in the odds with every unit change of standard deviation increase of the cosine of the slope.

Table 8. Table of analysis of deviance among the best four models with highest AUROC values.

Analysis of deviance table
Model 1: Y ~ VegetationCover + LandUse + s(Elevation) + Slope + Curv_prof + Curv_plan + log10_CatchmntArea + Sin_SlopeAspect + Cos_SlopeAspect + Precipitation
Model 2: Y ~ VegetationCover + LandUse + s(Elevation) + Slope + Curv_prof + Curv_plan + s(log10_CatchmntArea) + Sin_SlopeAspect + Cos_SlopeAspect + Precipitation
Model 3: Y ~ VegetationCover + LandUse + s(Elevation) + Slope + Curv_prof + Curv_plan + s(log10_CatchmntArea) + s(Sin_SlopeAspect) + Cos_SlopeAspect + Precipitation
Model 4: Y ~ VegetationCover + LandUse + s(Elevation) + Slope + Curv_prof + Curv_plan + s(log10_CatchmntArea) + s(Sin_SlopeAspect) + poly(Cos_SlopeAspect, 2, raw = T) + Precipitation

Model	Resid. Df	Resid. Dev	Df	Deviance	Pr(>Chi)
1	35.050	27.893			
2	35.050	27.893	7.9079e-05	0.0000698	0.0003829
3	**35.050**	**27.893**	**1.7778e-05**	**0.0000146**	**9.999e-05**
4	34.053	27.865	9.9678e-01	0.0283032	0.8654708
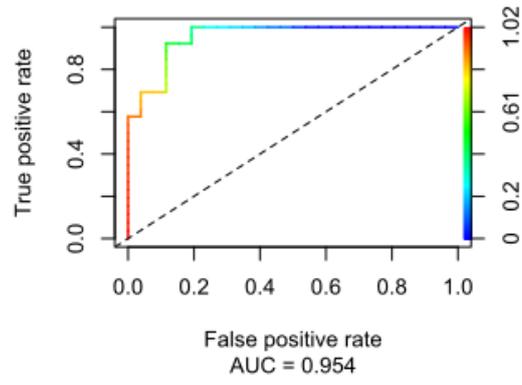
Figure 10. Predictive performance of the selected GAM model from a ROC curve that plots the false positive rate (FPR) on the x-axis and the true positive rate (TPR) on the y-axis.

Table 9. Parameter estimates of the selected generalized additive model for the prediction of the probability of landslide occurrence using all eleven predictors.

Predictor variables	Estimate	Std. Error	z value	Pr(>|z|)	Exp(estimates)
(Intercept)	-2.34	1.58	-1.48	0.14	0.096327638
Vegetation cover: Orchards	2.03	2.68	0.76	0.45	7.614086359
Vegetation cover: Pasture crop forest	2.83	1.8	1.57	0.12	16.94546082
Vegetation cover: Scrubland	3.00	2.09	1.44	0.15	20.08553692
Land use: Agriculture and Livestock	**-5.62**	**3.05**	**-1.84**	**0.07**	**0.003624641**
Land use: Forest (C&P)	-3.02	2.56	-1.18	0.24	0.048801218
Land use: Livestock	1.57	1.67	0.94	0.35	4.806648194
Slope	0.28	0.62	0.45	0.65	1.323129812
Curvature profile	-0.11	0.79	-0.14	0.89	0.895834135
Curvature plan	-0.77	0.73	-1.05	0.30	0.463013068
Cosine of Slope Aspect	**1.72**	**0.88**	**1.94**	**0.05**	**5.584528464**
Precipitation	**6.78**	**2.34**	**2.90**	**0.004**	**880.0687241**

	k	edf	Ref. edf	Chi.sq	p-value	
s(Elevation)	**9**	**2.322**	**2.95**	**8.862**	**0.032**	**10.19604602**
s(Log10 of the catchment area)	9	1	1	0.174	0.6768	2.718281828
s(Sine of the slope aspect)	9	1	1	0.534	0.4649	2.718281828
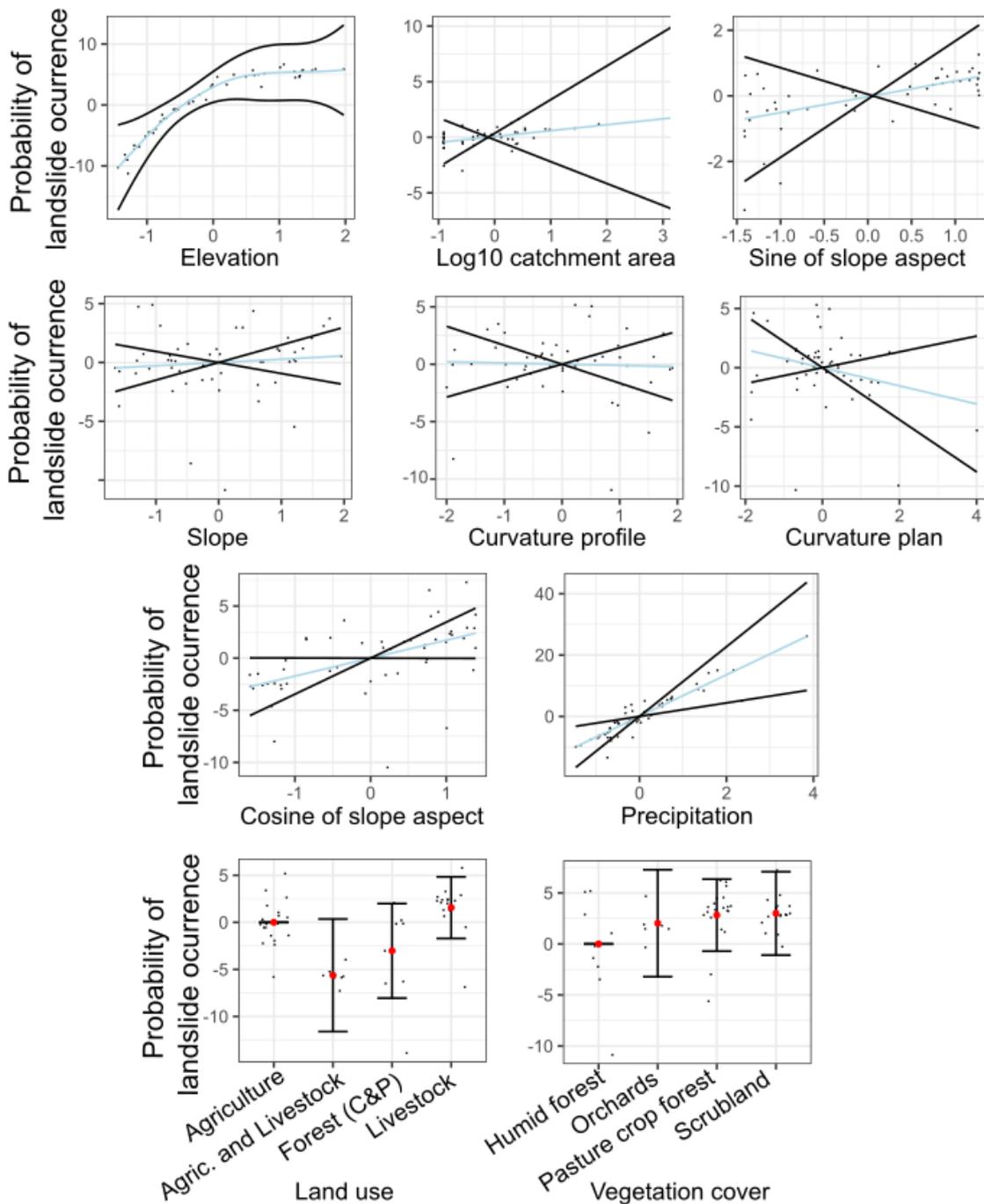
Figure 11. Partial effects of the transformation functions of the selected generalized additive model (GAM) without interaction term. The black lines represent standard errors with 95% confidence interval for the mean shape of the effect.

Visually, it's not always obvious to determine whether enough basis functions were used. We can test for this, though, via the "gam.check()" function. This function plots four standard diagnostic plots of the residuals of the model, smoothing parameter estimation convergence information and the results of the tests that may indicate if the

smoothing dimension is too low for a smooth. In this sense, the model converged after 10 iterations, meaning that there were not too many parameters for the used dataset. Furthermore, small p-values yield by this function shows that there might not be enough basis functions and that the residuals are randomly distributed, which is not the case in this fitting process as reported in Table 10 since the p-values are not significantly small, all of the smooths have enough basis functions and none of them have significant patterns in their residuals.

Plots of the residuals are necessary since they are part of the data that the model does not explain. The normal QQplot shows the deviance residuals against approximate theoretical quantiles of the deviance residual distribution (Figure 12 (a)). Deviance residuals represent the square root of the contribution that each data point has to the overall residual deviance. This plot compares the model deviance residuals to a normal distribution, which is close to a straight line and is interpreted as a good fitting model. Also, a histogram of residuals is shown in Figure 12 (b), where it is expected to have a symmetrical bell shape. Figure 12 (c) shows test for nonlinearities and heteroscedasticity. A band of evenly spread residuals around 0 is expected, however, the shape is caused by the fact that the response has only two possible values, but the predicted values are continuous and are close to 0. Finally, the plot of response against fitted values, Figure 12 (d), should show a straight line, however, since the nature of this problem is categorical and binary, the regression fits an "S" shaped logistic function.

Table 10. Smoothing parameters of the model checking with gam.check()

Predictor variables	k	edf	k-index	p-value
s(Elevation)	9	2.322	0.95	0.23
s(Log10 of the catchment area)	9	1	1.16	0.86
s(Sine of the slope aspect)	9	1	0.95	0.34
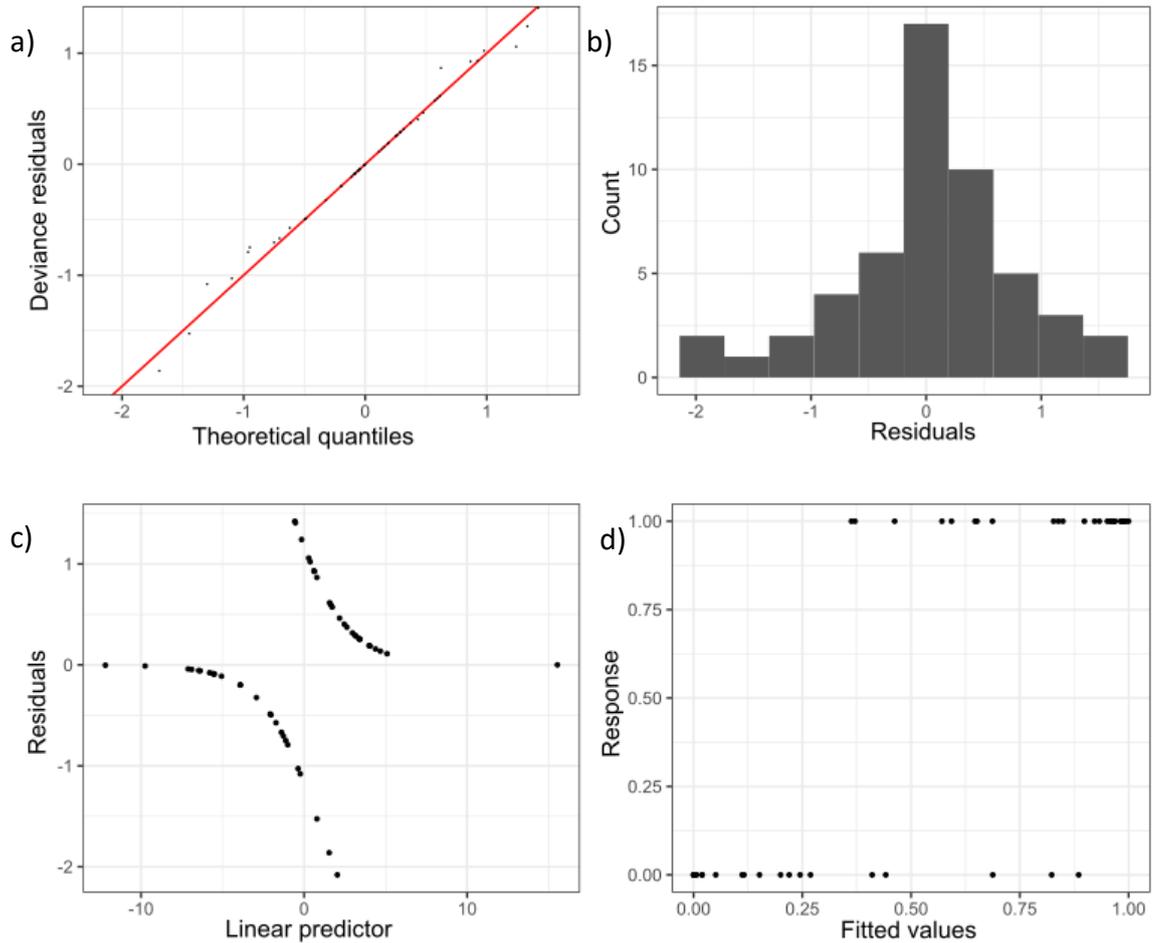
Figure 12. Plots for control check of the selected GAM model a) Q-Q Plot, using the method "Simul1", b) histogram of residuals, c) residuals against linear predictors, and d) response against fitted values.

Finally, both of these results can be expressed as an equation that includes the constant term and the regression coefficient for each variable, which has been found to be significant (usually using $P < 0.05$). The equation provides a model which can be used to predict the probability of an event happening for a particular individual, given his/her profile of predictor factors.

VII. CONCLUSIONS

Landslides are geohazards that can be caused by natural processes or by humans, however, the understanding of landslides in the Andes of Ecuador is still limited. Thus, there is a need to develop knowledge of cause-effect relationships through statistical and empirical studies, taking into account that they are produced by mechanical and complex interactions between predisposing and triggering factors. Predisposing factors correspond to those that make a slope prone to failure, such as long-term human effects or the lithological and hydrogeological background of the slope. On the other hand, triggering factors are sudden events that set off a landslide, such as sudden snowmelt, intense rainfall, an earthquake, and/or human activities such as the construction of roads.

Although the aforementioned factors are to some extent well known, and landslides have been evaluated using several methods of remote sensing and communicated with the use of susceptibility maps in Ecuador (specifically in the north of Ecuador), most studies do not take into account the triggering mechanisms that produce landslides. This has led to an inaccurate explanation of the cause of landslides and a common view of rainfall as the main trigger of most reported landslides.

To analyse the cause-effect relationship between landslides and predisposing factors from a descriptive analysis, a dataset was constructed for the province of Imbabura with terrain attributes, such as elevation, slope, plan curvature, profile curvature, base 10 logarithm of the contribution or catchment area, and the sine and cosine of the slope aspect; topological attributes, such as vegetation cover and land use; and the climatic attribute precipitation. Then, this was used in a statistical analysis to fit predictability models; namely, a GLM and a GAM non-linear model. Finally, these models were ran with known trigger mechanisms to find out theoretical evidence about the influence of predisposing factors.

Despite the problems that arose in the midst of this research such as limited, incomplete, or mistaken records of landslides since 1989, as well as outdated digital maps, the models showed high predictability performance. The predictions of landslide occurrence using GLM and GAM were compared using the values of AUC-ROC curves. The AUC from the GLM yielded an overall quantitative index of accuracy corresponding of 0.947 (AUC), while the GAM model yielded an overall quantitative

index of accuracy corresponding to 0.954 (AUC). Thus, the prediction of landslide occurrences can be achieved with higher certainty with GAM than with a GML. Moreover, from all the variables that were used in the GAM, elevation had a different smoothing effect that resulted in better predictability than the GLM.

The climatic attribute precipitation has a relatively high weight in every regression and is statistically significant in both models, which is consistent with the instances of high monthly precipitation due to the stratiform type of precipitation that increase the soil moisture saturation, and preconditions the occurrence of landslides. Elevation is the second estimator that in both models was found statistically significant explaining landslide occurrence in Imbabura. This might be explained by the influence of weather effects that are produced at high altitudes, low temperatures, high air and soil moisture, and wind at small spatial scales. Moreover, from all the types of land uses that may have an effect on landslide occurrence, agriculture and livestock land use showed a negative influence, compared to the other categories.

FURTHER RESEARCH

Since this study was focused on one province (i.e. Imbabura Province), it would be valuable to replicate it over the entire northern region of Ecuador to test the generalization of the cause-effect relationships here reported. The study was based on historical reports; it would therefore be beneficial to gather more landslide records with the use of orthophotos, particularly in inaccessible areas that have not been recorded historically. In this regard, it is also important to update the thematic maps in order to improve the quality of the input data and reduce the uncertainty in the model fitting. According to the influence of agriculture and livestock in landslide occurrence, it is imperative to determine the role that land use may play in contributing landslide occurrence given local and more specific descriptions in smaller areas. Lastly, the proposed statistical models can be used for scenario investigations on the likelihood of landslide occurrences as driven by climatic and environmental changes.

BIBLIOGRAPHY

Ali, S., Biermanns, P., Haider, R., & Reicherter, K. (2019). Landslide susceptibility mapping by using a geographic information system (GIS) along the China-Pakistan Economic Corridor (Karakoram Highway), Pakistan. *Natural Hazards and Earth System Sciences*, *19*(5), 999–1022. https://doi.org/10.5194/nhess-19-999-2019

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. https://doi.org/10.1016/j.geomorph.2004.06.010

Baby, P., Rivadeneira, M., Barragán, R., & Christophoul, F. (2013). Thick-skinned tectonics in the Oriente foreland basin of Ecuador. *Geological Society, London, Special Publications*, *377*(1), 59–76. https://doi.org/10.1144/SP377.1

Bendix, J., & Lauer, W. (1992). Die Niederschlagsjahreszeiten in Ecuador und ihre klimadynamische Interpretation (Rainy Seasons in Ecuador and Their Climate-Dynamic Interpretation). *Erdkund*, *46*(2), 118–134. Retrieved from http://www.jstor.org/stable/25646379

Borgatti, L., Vittuari, L., & Zanutta, A. (2010). Geomatic methods for punctual and areal control of surface changes due to landslide phenomena. In *Landslides: Causes, Types and Effects* (p. 417). Nova Science Publishers.

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

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

Brönnimann, C. S. (2011). *Effect of Groundwater on Landslide Triggering PAR* (Vol. 5236). École Polytechnique Fédérale De Lausanne.

Brunetti, M. T., Peruccacci, S., Rossi, M., Luciani, S., Valigi, D., & Guzzetti, F. (2010).

Rainfall thresholds for the possible occurrence of landslides in Italy. *Natural Hazards and Earth System Sciences*, *10*(3), 447–458. https://doi.org/10.5194/nhess-10-447-2010

Bui, D. T., Tuan, T. A., Klempe, H., Pradhan, B., & Revhaug, I. (2016). Spatial prediction models for shallow landslide hazards : a comparative assessment of the efficacy of support vector machines , artificial neural networks , kernel logistic regression , and logistic model tree. *Landslides*, *13*(September 2014), 361–378. https://doi.org/10.1007/s10346-015-0557-6

Buitrón Vinueza, S. M. (2014). *Metodología y modelo para movimientos en masa utilizando técnicas de SIG y Teledetección*. Universidad San Francisco de Quito.

Bujang, M. A., Sa'at, N., Tg Abu Bakar Sidik, T., & Lim, C. J. (2018). Sample Size Guidelines for Logistic Regression from Observational Studies with Large Population : Emphasis on the Accuracy Between Statistics and Parameters Based on Real Life Clinical Data. *Malaysian Journal of Medical Sciences*, *25*(4), 122–130.

Burga Cholca, S. S. (2019). *Caracterización Litológica – Estructural y Evaluación de los Deslizamientos en la zona de Cachi, Pujilí, Provincia de Cotopaxi. Trabajo*. Universidad Central del Ecuador.

Campos Cedeño, A. F., Salas Guillén, P. A., Macias Ramos, J. L., Sinichenko, E. K., & Gritsuk, I. I. (2019). Estimation of the runoff of the hills of the city of Portoviejo-Ecuador to assess the degree of flooding in the region. *IOP Conference Series: Materials Science and Engineering 675*, 9. https://doi.org/10.1088/1757-899X/675/1/012020

Casella, G., Fienberg, S., & Olkin, I. (2013). An Introduction to Statistical Learning. In *Springer Texts in Statistics*. https://doi.org/10.1016/j.peva.2007.06.006

Changoluisa, J. A., & Pineda, L. E. (2020). *Climatological Reconstruction Of January 2016 Flood In Esmeraldas*. Yachay Tech University for Experimental Technological Research.

Chen, W., Pourghasemi, H. R., Panahi, M., Kornejady, A., Wang, J., Xie, X., & Cao, S. (2017). Spatial prediction of landslide susceptibility using an adaptive neuro-fuzzy

inference system combined with frequency ratio, generalized additive model, and support vector machine techniques. *Geomorphology*, *297*, 69–85. https://doi.org/10.1016/j.geomorph.2017.09.007

Columbia University. (2021). International Research Institute for Climate and Society global data library. Retrieved from https://iridl.ldeo.columbia.edu/

Crosta, G. B., & Frattini, P. (2003). *Distributed modelling of shallow landslides triggered by intense rainfall*. 81–93.

Cruden, D. M., & Varnes, D. J. (1996). Chapter 3: Landslide Types and Processes. In *Landslides: investigation and mitigation* (pp. 36–75). Washington, DC: Transportation research board, US National Research Council. Special Report 247.

Dalgaard, P. (2008). *Introductory Statistics with R* (Second). https://doi.org/10.1017/CBO9781107415324.004

Davis, J. C. (2002). *Statistics and Data Analysis in Geology* (Third). https://doi.org/10.1198/tech.2005.s338

Davis, J. C. (2005). *Statistics and Data Analysis in Geology* (Third). https://doi.org/10.1198/tech.2005.s338

DesInventar. (2020). DesInventar. Retrieved from DesInventar (LatinAmerican Network of Social Studies on Disaster Prevention, LA RED) website: https://db.desinventar.org/

Domínguez, R. Z. (2014). El deslizamiento de La Josefina "Tragedia Nacional." *Galileo*, (1), 87–98. Retrieved from https://issuu.com/ucuenca/docs/galileo201-206/98

Dykes, A. P., & Welford, M. R. (2007). Landslides in the Tandayapa Valley , northern Andes , Ecuador : implications for landform development in humid and tectonically active mountain ranges. *Landslides*, *4*(2), 177–187. https://doi.org/10.1007/s10346-006-0076-6

Eidt, R. C. (1969). The climatology of South America. *Biogeography and Ecology in South America*, 54–81.

Emck, P. (2007). *A Climatology of South Ecuador with special focus on the Major*

Andean Ridge as Atlantic-Pacific Climate Divide. Erlangen-Nürnberg University.

Farinango, C., & Geovanny, E. (2014). *Estudio geológico del Paleógeno en la cordillera Occidental Septentrional del Ecuador, provincias de Carchi e Imbabura*. Escuela Politécnica Nacional.

Fustos, I., Abarca-del-Rio, R., Moreno-Yaeger, P., & Somos-Valenzuela, M. (2020). Rainfall-Induced Landslides forecast using local precipitation and global climate indexes. *Natural Hazards*, *102*(1), 115–131. https://doi.org/10.1007/s11069-020-03913-0

Gad Provincial de Imbabura. (2018). *Plan de desarrollo de ordenamiento territorial de la provincia de Imbabura, 2015-2035*.

Garcia-chevesich, P., Wei, X., Ticona, J., Mart, G., Zea, J., Garc, V., … Krahenbuhl, R. (2021). The Impact of Agricultural Irrigation on Landslide Triggering : A Review from Chinese , English , and Spanish Literature. *Water*, *13*(10), 1–17.

Goetz, J. N., Guthrie, R. H., & Brenning, A. (2011). Geomorphology Integrating physical and empirical landslide susceptibility models using generalized additive models. *Geomorphology*, *129*(3–4), 376–386. https://doi.org/10.1016/j.geomorph.2011.03.001

Gorsevski, P. V, Gessler, P. E., & Elliot, W. J. (2006). *Spatial Prediction of Landslide Hazard Using Logistic Regression and ROC Analysis*. *10*(3), 395–415.

Gutscher, M., Malavieille, J., Lallemand, S., & Collot, J. (1999). Tectonic segmentation of the North Andean margin: impact of the Carnegie Ridge collision. *Earth and Planetary Science Letters*, *168*, 255–270.

Hand, D. J. (2007). *Principles of data mining*. https://doi.org/10.2165/00002018-200730070-00010

Hastie, T., & Tibsgirani, R. (1986). Generalized Additive Models. *Statistical Science*, *1*(3), 297–310.

Havenith, H. B., Strom, A., Calvetti, F., & Jongmans, D. (2003). Seismic triggering of landslides. Part B: Simulation of dynamic failure processes. *Natural Hazards and Earth System Science*, *3*(6), 663–682. https://doi.org/10.5194/nhess-3-663-2003

Highland, L. M., & Bobrowsky, P. (2008). The landslide Handbook - A guide to understanding landslides. *US Geological Survey Circular*, (1325), 147. https://doi.org/10.3133/cir1325

Huang, Y., Wang, Y., & Cui, X. (2019). Differences between Convective and Stratiform Precipitation Budgets in a Torrential Rainfall Event. *Advances in Atmospheric Sciences*, *36*(5), 495–509. https://doi.org/10.1007/s00376-019-8159-1

Hungr, O., Leroueil, S., & Picarelli, L. (2013). The Varnes classification of landslide types: an update. *Landslides*, *11*(2), 167–194. https://doi.org/10.1007/s10346-013-0436-y

IIGE. (1979). *Geologic maps (Otavalo, Ibarra, Pacto). Scale 1 100.000.* INIGEMM, Britanic Mission, and IGM.

Iverson, R. M. (2000). Landslide triggering by rain infiltration. *Water Resources Research*, *36*(7), 1897–1910. https://doi.org/10.1029/2000WR900090

Jaillard, E, Lapierre, H., Ordoñez, M., Álava, J. T., Amórtegui, A., & Vanmelle, J. (2009). Accreted oceanic terranes in Ecuador: Southern edge of the Caribbean Plate? *Geological Society Special Publication*, *328*, 469–485. https://doi.org/10.1144/SP328.19

Jaillard, Etienne, Hérail, G., Monfret, T., Díaz-Martínez, E., Baby, P., Lavenu, A., & Dumont, J. F. (2000). Tectonic evolution of the Andes of Ecuador, Peru, Bolivia and northernmost Chile. *Tectonic Evolution of South America*, 481–558. https://doi.org/10.13140/RG.2.1.4349.5525

JAXA. (2020). ALOS Global Digital Surface Model "ALOS World 3D - 30m (AW3D30)." Retrieved from https://www.eorc.jaxa.jp/ALOS/en/aw3d30/index.htm

Kuhn, M., & Johnson, K. (2013). Applied predictive modeling. In *Applied Predictive Modeling*. https://doi.org/10.1007/978-1-4614-6849-3

Larsen, M. C., & Simon, A. (1993). A rainfall intensity-duration threshold for landslides in a humid- tropical environment, Puerto Rico. *Geografiska Annaler, Series A*, *75 A*(1–2), 13–23. https://doi.org/10.1080/04353676.1993.11880379

Lazzari, M., Piccarreta, M., & Manfreda, S. (2018). The role of antecedent soil moisture conditions on rainfall-triggered shallow landslides. *Natural Hazards and Earth System Sciences Discussions*, (December), 1–11. https://doi.org/10.5194/nhess-2018-371

Lebras, M., Megard, F., Dupuy, C., & Dostal, J. (1987). Geochemistry and tectonic setting of pre-collision Cretaceous and Paleogene volcanic rocks of Ecuador. *Geological Society of America Bulletin*, *99*(4), 569–578. https://doi.org/10.1130/0016-7606(1987)99<569:GATSOP>2.0.CO;2

Lee, S., & Min, K. (2001). Statistical analysis of landslide susceptibility at Yongin, Korea. *Environmental Geology*, *40*(9), 1095–1113. https://doi.org/10.1007/s002540100310

Lee, S., & Predhan, B. (2007). Landslide hazard mapping at Selangor , Malaysia using frequency ratio and logistic regression models. *Landslides*, *4*, 33–41. https://doi.org/10.1007/s10346-006-0047-y

Leonarduzzi, E., Molnar, P., & McArdell, B. W. (2017). Predictive performance of rainfall thresholds for shallow landslides in Switzerland from gridded daily data. *Water Resources Research*, *53*(8), 6612–6625. https://doi.org/10.1002/2017WR021044

Li, Z., Chen, M., Gao, S., Hong, Z., Gourley, J. J., & Yang, H. (2020). Cross-Examination of Similarity , Difference and Deficiency of Gauge , Radar and Satellite Precipitation Measuring Uncertainties for Extreme Events Using Conventional Metrics and Multiplicative Triple Collocation. *Remote Sensing*, *12*(8), 18. https://doi.org/10.3390/rs12081258

Lombardo, L., & Mai, P. M. (2018). Presenting logistic regression-based landslide susceptibility results. *Engineering Geology*, *244*(July), 14–24. https://doi.org/10.1016/j.enggeo.2018.07.019

López Cevallos, F. D. (2004). *Utilización de un SIG para establecer zonas de afectación por amenazas naturales: sismos, erupciones volcánicas y deslizamientos: posibles consecuencias en la salud de la población en la parroquia Tababela.* Universidad San Francisco de Quito.

Lyyn, H. (2014). *Landslide Types and Processes, a fact sheet*. Retrieved from
http://pubs.usgs.gov/fs/2004/3072/

Ma, T., Li, C., Lu, Z., & Bao, Q. (2015). Geomorphology Rainfall intensity – duration
thresholds for the initiation of landslides in Zhejiang Province , China.
Geomorphology, *245*, 193–206. https://doi.org/10.1016/j.geomorph.2015.05.016

Mackinnon, M. J., & Puterman, M. L. (1989). Collinearity in generalized linear models.
Communications in Statistcs, Theory and Methods, *18*(9), 3463–3472.
https://doi.org/10.1080/03610928908830102

Marra, F., Morin, E., Peleg, N., Mei, Y., & Anagnostou, E. N. (2017). Intensity-
duration-frequency curves from remote sensing rainfall estimates: Comparing
satellite and weather radar over the eastern Mediterranean. *Hydrology and Earth
System Sciences*, *21*(5), 2389–2404. https://doi.org/10.5194/hess-21-2389-2017

McCall, G. J. H., Laming, D. J. C., & Scott, S. C. (1992). *Geohazards: natural and
manmade* (First). Chapman & Hall.

Molina, A., Govers, G., Poesen, J., Hemelryck, H. Van, Bièvre, B. De, & Vanacker, V.
(2008). Environmental factors controlling spatial variation in sediment yield in a
central Andean mountain area. *Geomorphology*, *98*(3–4), 176–186.
https://doi.org/10.1016/j.geomorph.2006.12.025

Morán-Tejeda, E., Bazo, J., López-Moreno, J. I., Aguilar, E., Azorín-Molina, C.,
Sanchez-Lorenzo, A., … Vicente-Serrano, S. M. (2016). Climate trends and
variability in Ecuador (1966-2011). *International Journal of Climatology*, *36*(11),
3839–3855. https://doi.org/10.1002/joc.4597

Morán-Tejeda, E., López-Moreno, J. I., & Beniston, M. (2013). *The changing roles of
temperature and precipitation on snowpack variability in Switzerland as a function
of altitude*. *40*(January), 2131–2136. https://doi.org/10.1002/grl.50463

Motrenko, A., Strijov, V., & Weber, G. (2014). Sample size determination for logistic
regression. *Journal of Computational and Applied Mathematics*, *255*, 743–752.
https://doi.org/10.1016/j.cam.2013.06.031

Nerini, D., Zulkafli, Z., Wang, L. P., Onof, C., Buytaert, W., Lavado-Casimiro, W., &
Guyot, J. L. (2015). A comparative analysis of TRMM-rain gauge data merging

techniques at the daily time scale for distributed rainfall-runoff modeling applications. *Journal of Hydrometeorology*, *16*(5), 2153–2168. https://doi.org/10.1175/JHM-D-14-0197.1

Ochoa, A., Pineda, L., Willems, P., & Crespo, P. (2014). Evaluation of TRMM 3B42 (TMPA) precipitation estimates and WRF retrospective precipitation simulation over the Pacific-Andean basin into Ecuador and Peru. *Hydrology and Earth System Sciences Discussions*, *11*(1), 411–449. https://doi.org/10.5194/hessd-11-411-2014

Ohlmacher, G. C. (2007). *Plan curvature and landslide probability in regions dominated by earth flows and earth slides*. *91*, 117–134. https://doi.org/10.1016/j.enggeo.2007.01.005

Ohlmacher, G. C., & Davis, J. C. (2003). *Using multiple logistic regression and GIS technology to predict landslide hazard in northeast Kansas , USA*. *69*, 331–343. https://doi.org/10.1016/S0013-7952(03)00069-3

Pourghasemi, H. R., & Rossi, M. (2017). Landslide susceptibility modeling in a landslide prone area in Mazandarn Province, north of Iran: a comparison between GLM, GAM, MARS, and M-AHP methods. *Theoretical and Applied Climatology*, *130*(1–2), 609–633. https://doi.org/10.1007/s00704-016-1919-2

Prokešová, R., Alžbeta, M., Tábořík, P., & Snopková, Z. (2012). Towards hydrological triggering mechanisms of large deep-seated landslides. *Landslides*, *10*(3). https://doi.org/10.1007/s10346-012-0330-z

Radbruch-Hall, D. H., & Varnes, D. J. (1976). Landslides — Cause and effect. *Bulletin of the International Association of Engineering Geology*, *13*(1), 205–216. https://doi.org/10.1007/bf02634797

Roperch, P., Mégard, F., Laj, C., Mourier, T., Clube, T. M., & Noblet, C. (1987). Rotated oceanic blocks in Western Ecuador. *Geophysical Research Letters*, *14*(5), 558–561.

Ruiz, G. M. H. (2002). *Exhumation of the northern Sub-Andean Zone of Ecuador and its source regions: a combined thermochronological and heavy mineral approach* (ETH). https://doi.org/https://doi.org/10.3929/ethz-a-004489528

Salazar, A. (2017). *Análisis de la distribución espacil y temporal de los deslizamientos*

en la cuenca del río salado y zonas aledañas al volcán Reventador, desde 1987.
Escuela Politecnica Nacional.

Schmidt, J., Turek, G., Clark, M. P., Uddstrom, M., & Dymond, J. R. (2008).
Probabilistic forecasting of shallow , rainfall-triggered landslides using real-time numerical weather predictions. 349–357.

Segovia Puente, F. X. (2017). *Zonificación y evaluación de amenazas por deslizamientos y caídas de roca en el cantón Guano, provincia del Chimborazo, escala 1:25000*. Universidad Central del Ecuador.

Sistema Nacional de Información. (2014). Archivos de Información Geográfica.
Retrieved from https://www.sni.gob.ec/coberturas

SNGRE. (2021). Servicio Nacional de Gestión de Riesgos y Emergencias. Retrieved from https://www.gestionderiesgos.gob.ec/

Soto-Cordero, L. (2019). *Earthquake and Tectonic Processes in the Mid-Atlantic U.S. Passive Margin and the Ecuador Subduction Zone*. Lehigh University.

Soto, J., Palenzuela, J. A., Galve, J. P., Luque, J. A., Azañón, J. M., Tamay, J., & Irigaray, C. (2019). Estimation of empirical rainfall thresholds for landslide triggering using partial duration series and their relation with climatic cycles. An application in southern Ecuador. *Bulletin of Engineering Geology and the Environment*, *78*(3), 1971–1987. https://doi.org/10.1007/s10064-017-1216-z

Spikings, R., Cochrane, R., Villagomez, D., Van der Lelij, R., Vallejo, C., Winkler, W., & Beate, B. (2015). The geological history of northwestern South America: From Pangaea to the early collision of the Caribbean Large Igneous Province (290-75 Ma). *Gondwana Research*, *27*(1), 95–139. https://doi.org/10.1016/j.gr.2014.06.004

Springman, S., Kienzler, P., Casini, F., & Askarinejad, A. (2009). Landslide triggering experiment in a steep forested slope in Switzerland. *Proceedings of the 17th International Conference on Soil Mechanics and Geotechnical Engineering: The Academia and Practice of Geotechnical Engineering*, *2*(October), 1698–1701. https://doi.org/10.3233/978-1-60750-031-5-1698

Toro Álava, J., & Jaillard, E. (2005). Provenance of the Upper Cretaceous to upper Eocene clastic sediments of the Western Cordillera of Ecuador: Geodynamic

implications. *Tectonophysics*, *399*(1-4 SPEC. ISS.), 279–292. https://doi.org/10.1016/j.tecto.2004.12.026

Torres Alves, G. (2018). *Application of Bayesian Networks to Estimate River Discharges in Ecuador* (Delft University of Technology). Retrieved from https://repository.tudelft.nl/islandora/object/uuid%3A6579ad40-1339-44b6-b9d4-6cae90190819?collection=education

Tsai, F., Hwang, J. H., Chen, L. C., & Lin, T. H. (2010). Post-disaster assessment of landslides in southern Taiwan after 2009 Typhoon Morakot using remote sensing and spatial analysis. *Natural Hazards and Earth System Science*, *10*(10), 2179–2190. https://doi.org/10.5194/nhess-10-2179-2010

Ulloa, J., Ballari, D., Campozano, L., & Samaniego, E. (2017). Two-step downscaling of TRMM 3b43 V7 precipitation in contrasting climatic regions with sparse monitoring: The case of Ecuador in tropical South America. *Remote Sensing*, *9*(7), 1–23. https://doi.org/10.3390/rs9070758

Urgilez Vinueza, A., Robles, J., Bakker, M., Guzman, P., & Bogaard, T. (2020). Characterization and hydrological analysis of the guarumales deep-seated landslide in the tropical ecuadorian andes. *Geosciences (Switzerland)*, *10*(7), 1–24. https://doi.org/10.3390/geosciences10070267

Vallejo Cruz, C. (2007). *Evolution of the Western Cordillera in the Andes of Ecuador (Late Cretaceous-Paleogene)*. ETH.

Varnes, D. (1978). Chapter 2: Slope Movement Types and Processes. In *Landslides: Analysis and control* (pp. 11–33). Transportation Research Board.

Vennari, C., Gariano, S. L., Antronico, L., Brunetti, M. T., Iovine, G., Peruccacci, S., … Guzzetti, F. (2014). Rainfall thresholds for shallow landslide occurrence in Calabria, southern Italy. *Natural Hazards and Earth System Sciences*, *14*(2), 317–330. https://doi.org/10.5194/nhess-14-317-2014

Villacís, M., Vimeux, F., & Denis, J. (2008). *Analysis of the climate controls on the isotopic composition of 250 m , Ecuador*. *340*, 1–9. https://doi.org/10.1016/j.crte.2007.11.003

Vorpahl, P., Elsenbeer, H., Märker, M., & Schröder, B. (2012). How can statistical

models help to determine driving factors of landslides? *Ecological Modelling*, *239*, 27–39. https://doi.org/10.1016/j.ecolmodel.2011.12.007

Werner, E. D., & Friedman, H. P. (Eds.). (2010). *Landslides: causes, types and effects*. New York: Nova Science Publishers.

Wieczorek, G. F. (1996). Chapter 4: Landslide Triggering Mechanisms. In *Transportation Research Board Special Report. Landslides: Investigation and Mitigation.*

Wood, S. N. (2017). *Generalized Additive Models* (Second). Bristol: Chapman and Hall/CRC.

Yesilnacar, E., & Topal, T. (2005). *Landslide susceptibility mapping : A comparison of logistic regression and neural networks methods in a medium scale study , Hendek region (Turkey)*. *79*, 251–266. https://doi.org/10.1016/j.enggeo.2005.02.002

Zerathe, S., Lacroix, P., Jongmans, D., Marino, J., Taipe, E., Wathelet, M., … Tatard, L. (2016). Morphology, structure and kinematics of a rainfall controlled slow-moving Andean landslide, Peru. *Earth Surface Processes and Landforms*, *41*(11), 1477–1493. https://doi.org/10.1002/esp.3913

Zêzere, J. L., Trigo, R. M., & Trigo, I. F. (2005). Shallow and deep landslides induced by rainfall in the Lisbon region (Portugal): Assessment of relationships with the North Atlantic Oscillation. *Natural Hazards and Earth System Science*, *5*(3), 331–344. https://doi.org/10.5194/nhess-5-331-2005

Zhang, K., Wang, S., Bao, H., & Zhao, X. (2019). Characteristics and influencing factors of rainfall-induced landslide and debris flow hazards in Shaanxi Province, China. *Natural Hazards and Earth System Sciences*, *19*(1), 93–105. https://doi.org/10.5194/nhess-19-93-2019