Towards near real-time forecasting of rainfall-induced landslides

MAHNOOR AHMED
August, 2022

SUPERVISORS:
Asst. Prof. Dr. Luigi Lombardo
Asst. Prof. Dr. Hakan Tanyaș

Towards near real-time forecasting of rainfall-induced landslides

MAHNOOR AHMED
Enschede, The Netherlands, August, 2022

Thesis submitted to the Faculty of Geo-Information Science and Earth Observation of the University of Twente in partial fulfilment of the requirements for the degree of Master of Science in Geo-information Science and Earth Observation.
Specialization: Natural Hazard and Disaster Risk Reduction.

SUPERVISORS:
Asst. Prof. Dr. Luigi Lombardo
Asst. Prof. Dr. Hakan Tanyaș

THESIS ASSESSMENT BOARD:
Prof. Dr. Cees van Westen (Chair)
Assoc. Prof. Dr. Raphaël Huser (External Examiner, King Abdullah University of Science and Technology)

ABSTRACT

Frequent extreme events are expected to result in an increased frequency of landslide occurrences. The rapid growth of infrastructure development extends to terrain with unstable slope conditions creating a hazardous environment for communities. To address this problem, multiple studies have approached the investigation of slope stability via static slope susceptibility estimation to eventually construct a Landslide Early Warning System (LEWS). Landslide Hazard Assessment of Situational Awareness (LHASA) is the current real-time model running to visualise Rainfall-Induced Landslides (RIL) hazard on a global scale. However, these models do not include rainfall signal into the estimation of failure probabilities, thus neglecting the orographic effect.

This study attempts to address the interaction of rainfall with terrain characteristics with the use of a Generalized Additive Model (GAM) in a Bayesian framework. The framework allows the uncertainty estimation of the estimated probabilities in a unified model rather than two individual phases as used in LHASA among other susceptibility estimations. Moreover, to acquire rainfall information from satellite-derived products a model selection tool was implemented to identify a suitable antecedent rainfall window for the selected study site. To move away from separate rainfall-thresholds as used vastly in the literature, a method to identify a probabilistic threshold for warning signals was also explored. In addition, along with cross validation techniques, external validation became a possibility for this study due to the availability of a multi-temporal inventory for the Lower Mekong Region (LMR).

The findings of this study suggested 8 days to be a suitable antecedent rainfall window for the selected site in North-western Vietnam The results also show an overall good performance measured by the Area under the Curve (AUC) of Receiver Operating Characteristics (ROC) curves for multiple validation techniques. These routines included temporal, spatial, random in the spatio-temporal domain and external validation. The results of ROC curves obtained from these techniques ranged from somewhere between 0.6-0.9 for the AUC values. The lower performing models were assumed to be linked with the aspects of inventory quality that plays an important role for defining a landslide susceptibility model and terrain complexity for the external validation outputs. To set a warning and no-warning cut-off point, 0.0057 was calculated as an average which was taken from the optimal threshold range for all the inventories used. The final model was then translated into a visualisation tool via Google Earth Engine (GEE) applications.

In conclusion, the proposed framework offers to further improve the uncertainty estimation to increase accuracy and displays the possibility of accounting for rainfall-terrain interactions in the context of near-real time landslide susceptibility. The visualization tool has enormous room for improvement in terms of computation speed and exploring other statistical summaries generated from the model for various purposes.

ACKNOWLEDGEMENTS

This work marks an end to an era of learning, for which I have many supporters and guides to thank for. I would like to express my deepest gratitude to my supervisors Asst. Prof. Dr. Luigi Lombardo and Asst. Prof. Dr. Hakan Tanyas, for always being available to push me in the right direction with my research. Their experience and intelligence allowed me to go an extra mile with their valuable suggestions. Receiving consistent mentoring and encouragement has made me realize the importance of having supportive supervisors in your corner. A special thanks to the brilliant minds I connected with during my internship under the research group of Assoc. Prof. Dr. Raphaël Huser, a truly excellent and kind mentor at KAUST, who offered me immense support, guidance and knowledge. I'm extremely grateful to Dr. Rishikesh Yadav, for graciously helping me with most of the statistical problems in crucial moments of my research by so kindly taking out the time. A sincere thanks to Silius Mortensønn Vandeskog for being a desk away to solve my INLA errors. Ashok Dahal, who has not only been a great friend but an honest and most helpful mentor. I am deeply indebted to his support and care for all stages of this study. Truly grateful for Omar Zaman (Zaman Foundation) for making this possible.

Those who have helped me survive the 2 years spent at ITC deserve a special thanks. Sahara, Lynette, Diana and Dewi – the girls who have been there from the very start showing love and support. Sogol, my *wahid yaar*, for all her consolation and kindness. Komal, my number one, has selflessly been at my disposal for anything and everything. I very much appreciate the people around me who have been my support system during anxious and stressful times. Rubab, Aamna and Som, for keeping me sane, Abdul Rehman and Anzal for telling me '*Insha'Allah*, *hojaye ga yaar*', Abeer, Fateh, Ponz and Enzo for always offering their help in panic mode. Zealaf, Hira, Salar, Nithesh, Somnath ji and any of those I have missed – all of you matter, for bringing joy and being there whenever I needed you.

Last but not the least, my ever-loving and supporting family, without whom I would not have been here. I cannot begin to express my gratitude especially for my Ami, the woman who is my *noor* in life.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AIC	Akaike Information Criteria
ASTER	Advanced Spaceborne Thermal Emission and Reflection Radiometer
AUC	Area Under the Curve
CHIRPS	Climate Hazards Group InfraRed Precipitation with Station data
CI	Credible Interval
DEM	Digital Elevation Model
DIC	Deviance Information Criteria
EVI	Enhanced Vegetation Index
EWS	Early Warning System
FN	False Negative
FP	False Positive
FPR	False Positive Rate
GAM	Generalized Additive Model
GDEM	Global Digital Elevation Model
GEE	Google Earth Engine
GLC	Global Landslide Catalog
GLM	Generalised Linear Model
GPM	Global Precipitation Measurement
IMERG	Integrated Multi-satellitE Retrievals GPM
INLA	Integrated Nested Laplace Approximation
LEWS	Landslide Early Warning System
LHASA	Landslide Hazard for Situational Awareness
LMR	Lower Mekong Region
MODIS	Moderate Resolution Imaging Spectroradiometer
NASA	National Aeronautics and Space Administration
NDVI	Normalised Difference Vegetation Index
OBIA	Object Based Image Analysis
RIL	Rainfall-Induced Landslides
ROC	Receiver Operating Characteristics
SALaD	Semi-Automatic Landslide Detection
SRTM	Shuttle Radar Topographic Mission
SU	Slope Unit
TMPA	TRMM Multi-satellite Precipitation Analysis
TN	True Negative
TNR	True Negative Rate
TP	True Positive
TPR	True Positive Rate
TRMM	Tropical Rainfall Measuring Mission
USD	United States Dollar
WAIC	Watanabe Akaike Information Criteria

1. INTRODUCTION

1.1. General overview

Landslides have a smaller spatial extent in comparison to other natural hazards, such as floods, cyclones and earthquakes, but are still highly disastrous. Annual economic loss ranges between USD 1-5 billion for countries like Japan, United States and India (Hidayat et al., 2019). In the seven-year time period between 2004 and 2010, more than 2500 deadly landslides were reported with a death toll of over 32,000 (Petley, 2012). Looking towards the future, the disastrous impacts of landslides (damage, loss of life and economic losses) are expected to worsen with increasing frequency of extreme events caused by climate change (Hidayat et al., 2019). The interaction of dense population and landslide prone regions call for well-developed Early Warning Systems (EWS) to aid in disaster management at regional levels (Guzzetti et al., 2020; Hidayat et al., 2019).

An EWS can be defined as a system which is capable of providing meaningful and critical information to be communicated to organizations, authorities and the public. The expected response to an early warning is to act timely in order to reduce potentially devastating impacts. However, how 'early' is appropriate for a response to translate into effective action can vary with the type of disaster and its environmental setting. The purpose of EWS was laid out in terms of disaster prevention, reducing risk and avoid bearing heavy economic losses and casualties (Guzzetti et al., 2020). The initiation of the concept of risk reduction began with the United Nation's framework in the 1990, called the United Nations' International Decade for Natural Disaster Risk Reduction Early Warning Programme. This initiative continued to develop in the Hyogo Framework for Action (UNISDR, 2005) and also followed in the Sendai Framework for Disaster Risk Reduction (UNISDR, 2015) which is a running framework till 2030.

A rigorous observation of landslides being triggered by rainfall dates back to the 1970s with empirical information being collected in Japan, Hong Kong, New Zealand and United States of America (Guzzetti et al., 2020). One of the earliest scientific effort for issuing early warning and predicting landslides was introduced in Southern California after observing rainfall triggered debris flows causing damage, deaths and injuries (Campbell, 1975). The author also emphasized that occurrences of landslides are not solely the quantity of rainfall that initiates slope failures but a complex interaction of surface conditions that makes the land prone to sliding. Campbell (1975) put forward the importance of rainfall intensity-duration relationship and even proposed a local threshold, also advising regarding the necessary revision of such empirical numbers with changing geomorphology.

Landslide Early Warning System (LEWS), particularly for Rainfall-Induced Landslides (RIL), is an EWS which is specifically dedicated to landslides. Most of these systems are developed on a rainfall-threshold to indicate a possible slope failure. A rainfall-threshold is defined as a measure of precipitation, in relation to slope conditions, which when exceeded can trigger landslides in a given area (Guzzetti et al., 2008; Segoni et al., 2018). When hydrological conditions of the slope are known, these thresholds are considered a decent division between triggering and non-triggering levels. For instance, the updated threshold division includes a lower threshold, below which no landslides are expected to occur, and an upper threshold, above which landslides are highly likely to occur (Segoni et al., 2018). There are two main methods to define thresholds; statistically and physically-based approaches (Guzzetti et al., 2020). Physically-based approaches are process-based models which integrate detailed lithological, morphological and hydrological information of slopes to determine when and where a landslide can be expected by quantifying the amount of rainfall needed to trigger it. However, it is strenuous to gather this information over a large spatial extent, especially considering the financial constraints of equipment and expertise needed to obtain such data (Guzzetti et al., 2007).

On the other hand, statistical approaches e.g., Frequentist and Bayesian (Guzzetti et al., 2020), use historical landslide inventories and rainfall intensity-duration relationship to define thresholds which will trigger possible landslides in a given area. The intensity-duration thresholds are derived by determining the minimum amount of rainfall in a certain duration which results in landslide occurrences as indicated in past events (Guzzetti et al., 2007). These rainfall thresholds are then combined with the output of a landslide susceptibility model featuring predisposing factors to assign geographic units (grid cells, slope units) with a probability of landslide occurrence (Lee et al., 2008). The resulting information conveys the landslide hazard expected for a given area and its operational use in LEWS is expressed by categorizing the probability into classes, from low to very high (Wubalem, 2021).

Recently, the implementation of LEWS in different regions has experienced a rapid development, with many prototypes designed at different geographic scales (Guzzetti et al., 2020).

1.2. Background research in Vietnam

Several LEWS have been developed at different spatial scales (regional, national and global), with their status ranging from being at the design stage to being currently or previously operational and now dismissed. The majority of well-established regional LEWS are concentrated in the United States of America (USA) and Italy (Guzzetti et al., 2020).

In Vietnam, which is the study area selected for this work, most of the landslide research is limited to developing landslide susceptibility models rather than translating them into LEWS. Most of the focus has been directed to the hilly Northern province of Hoa Binh which lies Southwest of the capital city. The province receives over 80% of its annual precipitation in the months between May to October (Tien Bui et al., 2013, 2012, 2011). In addition, the threat from landslides in the Northern sector of Vietnam is increasing due to sprawling infrastructure towards mountainous terrains (Tien Bui et al., 2013). Generating susceptibility maps has thus become a common practice. However, these maps are mostly static or temporally stationary in nature and sometimes do not even undergo a validation process. Hence, these models cannot be considered reliable in their predictive nature (Chung and Fabbri, 2003). Aside from the study cases mentioned above, additional susceptibility assessments have been produced across several Vietnamese provinces relying on statistical methods (see, Tien Bui et al. 2011). These studies entail different methods to understand the effect of morphometric properties and the triggering factors. Among them, simpler routines such as expert based weighted maps or more objective logistic regression methods have been tested, as for example for the Hoa Binh province (Tien Bui et al., 2011). Moreover, other experiments have been run in the area including evidential belief functions and fuzzy logic models to estimate landslide susceptibility (Tien Bui et al., 2012). However, the primary weakness of such models is that the estimated spatial probabilities are purely based on terrain characteristics, leaving unaccounted the time-variant influence of the triggering factor (Tien Bui et al., 2011).

In addition to the static nature of the work produced in Vietnam, another element that requires attention is certainly the acquisition and use of high quality/completeness landslide inventories to train any model. In fact, the work of Tien Bui et al. (2011) already highlights that the lack of systematic nation-wide inventories, especially in rural areas, induce biased results.

Moving away from the focus on terrain-driven susceptibility (Tien Bui et al., 2012, 2011), a few studies approach methods to link the spatio-temporal patterns of the trigger to the resulting slope failures. For instance, Tien Bui et al. (2013) progressed on this by identifying rainfall as the main cause of slope instabilities in the province of Hoa Binh in the years from 1990 to 2010. Due to the tropical rainfall regime, responsible for persistent precipitation, the role of cumulative rainfall is considered to be significant. For example, a study in Vietnam introduced a 15-day cumulative rainfall threshold to develop landslide hazard maps (Tien Bui et al., 2013). These maps represented temporal and spatial probability estimates of landslide occurrences, thus defining 'where' and 'when' a slope failure can occur (Tien Bui et al., 2013). However, a number of misclassified areas reflected in the work of Tien Bui et al. (2013). The reason behind this was due to the fact that susceptible terrains and rainfall thresholds were estimated independently from each other. Therefore, no real interactions were allowed between the two factors resulting in flat areas being flagged with high alert warnings, solely based on exceedance of the rainfall threshold (Tien Bui et al., 2013). Thus, rainfall thresholds do not always produce reliable results and it is considered essential to move towards combining rainfall information with predisposing landscape characteristics.

Another level of complexity on the definition of LEWS is due to the spatial scale at which such systems are designed to operate. It is of crucial importance, for local administrations, to develop LEWS informed of soil hydrological characteristics as well as micro-climate patterns. Examples of site specific LEWS have in fact been designed and implemented using soil information in several regions in the North of Vietnam (Gian et al., 2017; Ha et al., 2020). An early warning and monitoring system was proposed for RIL using wireless sensor nodes drilled into the soil layer to determine slope instability (Gian et al., 2017). Such methods usually require in-situ measurements of rainfall (Gian et al., 2017) and water content in the soil (Ha et al., 2020). For the rainfall component, efforts have been made to operationalize LEWS using rain gauges and/or weather stations. The in-situ information acquired from the placement of these devices may provide accurate and temporally consistent information although they are discrete in their spatial extent.

On the other side of the spatial spectrum lie LEWS that operate over large landscapes, thus making field data acquisition often unfeasible. In such cases, the use of affordable and omnipresent rainfall estimated from satellite data are usually considered an optimal alternative. Their strength boils down to the coverage they guarantee, both in space and time, whereas their weakness is usually linked to the coarse spatial resolution of the data provided. Among the first examples of satellite-based rainfall estimates used to define EWS, Hong et al. (2007) paved the way for numerous improvements for the years to come.

1.3. Global Landslide Early Warning System

Hong et al. (2007) first proposed a framework to use geospatial datasets to assess probability of landslide occurrences at global scales, both due rainfall and earthquake triggers. A similar parallel study (Hong et al., 2006) focused on identifying region-based rainfall thresholds rather than a single global threshold, which inevitably led to the misrepresented hazard estimates. Based on this, a preliminary global framework was introduced by Kirschbaum et al. (2009), which developed an algorithm to build a near-real time global landslide susceptibility model using satellite rainfall estimates. Eventually, a regional predictive model capable of producing near-real time forecasts was developed for Central America (Kirschbaum et al., 2012), where warning levels were produced by intersecting rainfall intensity-duration thresholds and a static susceptibility map (methodology shown in Figure 1). Comparing results particularly for Central America from the regional model mentioned above, with global susceptibility described in Hong et al. (2007) showed the regional model performing better overall (Kirschbaum et al., 2012). This can be attributed to use of landslide inventory specific to Central America to support the model calibration and comparison by Kirschbaum et al. (2012) and thus, it was suggested to use regional information for a possible global LEWS using a similar framework for further extension (Kirschbaum et al., 2012). Some limitations in the work of Kirschbaum et al., (2012), were due to the coarse spatial resolution of Tropical Rainfall Measuring Mission (TRMM) Multisatellite Precipitation Analysis (TMPA) estimates, such as reserving the algorithm's optimum functionality for areas larger than 2500 km^2 as well as the use of a single extreme rainfall event introducing overestimation by the model.

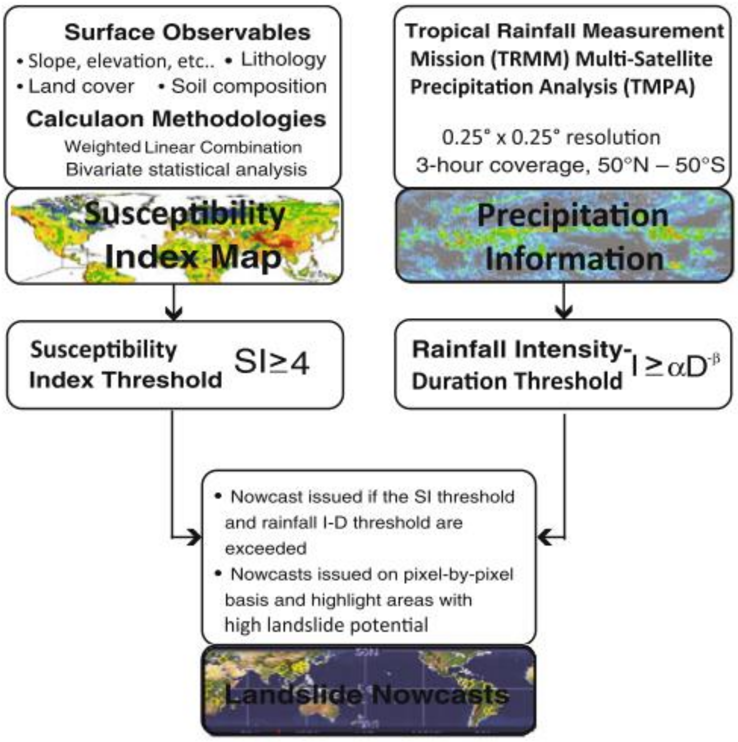

Figure 1 Framework for the proposed regional (extendable to global scale) landslide hazard nowcast algorithm (Kirschbaum et al., 2012).

Since then, a lot has changed in terms of data quality as well as modelling tools. These elements have allowed to develop LEWS even further, and recent advancements have given rise to the latest version of LEWS from NASA, this being referred to as Landslide Hazard for Situational Awareness (LHASA) (Kirschbaum and Stanley, 2018). Specifically, LHASA uses the Integrated Multi-satellitE Retrievals (IMERG) for Global Precipitation Measurement (GPM) and TRMM to provide a nowcast every thirty minutes (Kirschbaum & Stanley, 2018). This concept, derived from aforementioned methods (Hong and Adler, 2007; Kirschbaum et al., 2012), is currently an operational global LEWS. With respect to the previous version described above, LHASA relies on the NASA's Global Landslide Catalog (GLC; Kirschbaum et al., 2010) to continuously improve the data needed for model calibration. The steps followed to issue nowcasts are shown in Figure 2, indicating the use of an Antecedent Rainfall Index (ARI) to filter out the areas which will further be evaluated by the underlying susceptibility map. The regions falling in a high ARI and intersecting with moderate and high susceptibility will in turn issue a nowcast warning for a landslide hazard (Kirschbaum & Stanley, 2018). Though the global model did not predict well the landslide occurrences reported in the GLC, it did set a base for the possibility of integrating geospatial datasets to be used in a dynamic context.

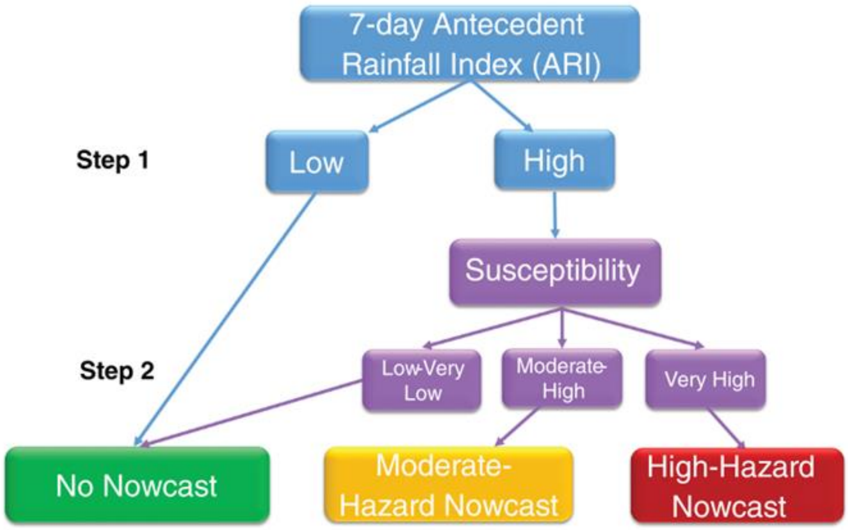

Figure 2 LHASA decision tree structure for issuing landslide hazard nowcasts (Kirschbaum and Stanley, 2018).

For this reason, LHASA was later modified in an attempt to solve the prediction misclassification by updating to its second version (see, Stanley et al., 2021). For the second version, two new dynamic predictors were introduced, these being the snow cover and soil moisture (Stanley et al., 2021). However, there were yet significant drawbacks prevailing in the resulting nowcasts, and some limitations were brought by the integration of two new dynamic variables. In addition to removing inventory information prior to the availability of the soil moisture data, the dynamic variables reduced the spatiotemporal domain by being specific in data-relevant areas. This also led to the exclusion of smaller Indonesian islands which did not fall under the coverage of soil moisture information. Moreover, LHASA version 2 nowcasts also displayed oddly high probabilities of landslide occurrences in some areas like the Northern Andes, and it simultaneously underestimated in other areas of the African region (Stanley et al., 2021).

2. RESEARCH PROBLEM

The frequency of extreme weather-related events (like extreme precipitation) is expected to increase in the future. This will likely contribute to an increased number of landslides and associated losses (Hidayat et al., 2019). The need to move towards a dynamic system which assesses the risk of landslides is defined by the community and geographic area, as well as the spatial extent considered. Rainfall triggered landslides are the frequently observed and devastating movements and it is essential to determine slope failures attributed to rainfall related weather events. However, developing rainfall thresholds is highly subjective due to high variations in climate and seasonality, along with their interaction with natural and anthropogenic conditions of a certain region (Kirschbaum & Stanley, 2018). Nonetheless, these thresholds are still widely contributing to landslide early warning systems (Segoni et al., 2018).

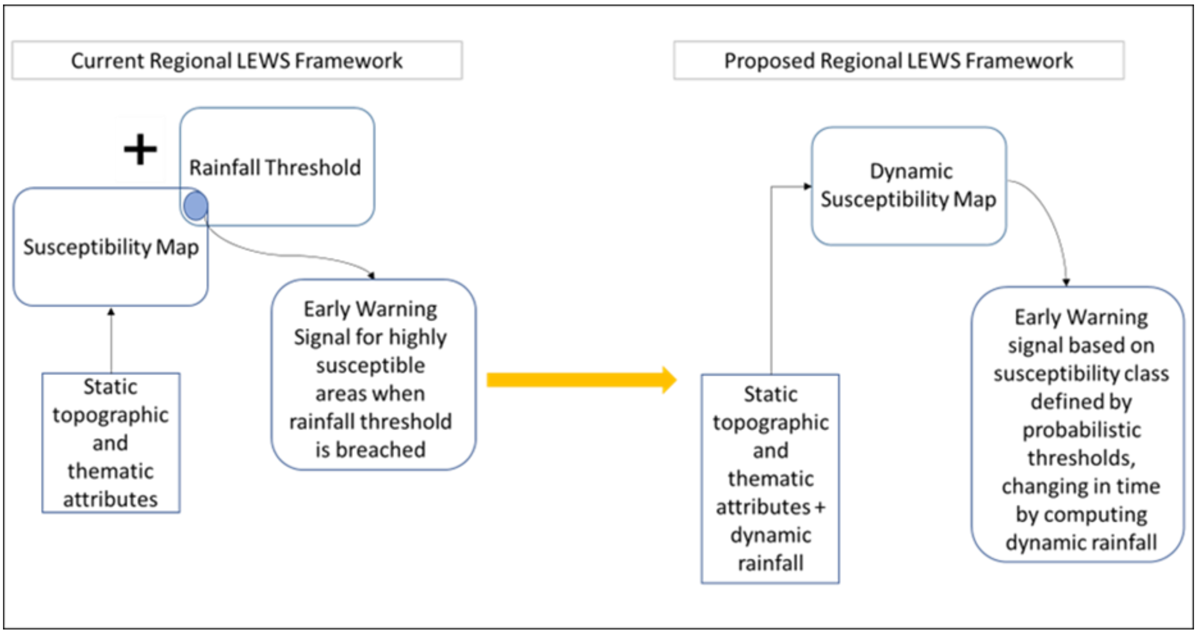

Figure 3 Conceptual framework for transitioning to a dynamic LEWS.

Overall, there are two main limitations in most of the available LEWS. These include lack of uncertainty estimation (Guzzetti et al., 2020) and the use of rainfall thresholds estimated independently from the landscape where they are applicable in reality. Therefore, an unreliable warning system can prove to be costly, especially if the warning level is underestimated.

The aim of this study is to propose the implementation of a LEWS where predisposing and triggering factors are both featured within the same model (Figure 3), which is capable of producing uncertainty estimates alongside the mean probabilistic response.

Typically, applications of statistical models are capable of predicting landslides over a geographic space by building landslide susceptibility models, i.e., a key part of developing a LEWS. Current implementation of

LEWS presents an empirically based rainfall threshold combined with statistically derived landslide susceptibility maps, in essence combining two outputs as one. This assumes that rainfall patterns are independent of terrain characteristics, which may not be a valid assumption. Due to the orographic effect, especially in mountainous zones and highlands, an influence of heavier rainfall exists in some parts of the terrain as compared to its surrounding areas (Adler et al., 2003; Gariano et al., 2017; Guzzetti et al., 2008; Kirschbaum et al., 2012; Nguyen et al., 2014). Hence, instead of using rainfall (the dynamic variable) and geomorphological covariates (static variables) independently, they can be jointly used in a modelling scheme to approach a LEWS. Thus, rainfall will play the role of a covariate, like any other, in the susceptibility model and develop a system which is integrational of dynamic and static components. Besides for integrating covariates' interactions, a model that features both predictors would also allow for a proper uncertainty estimation.

3. STUDY OBJECTIVES

3.1. Main objective

To explore the possibility of an operational regional alarm system for rainfall-induced landslides in near real time, using a statistical model that incorporates both rainfall estimates and terrain properties, expressed at the slope unit level for a test site in Northern Vietnam.

3.1.1. Sub-objectives

- To identify a suitable antecedent rainfall window as a covariate.
- To model dynamic behaviour of landslide probability as the spatio-temporal signal of the rainfall pattern varies.
- To transition from separate rainfall thresholds to unified probabilistic thresholds for alert levels.
- To translate the model into an interactive visualization tool via a cloud-platform for comprehensive display (to be possibly extended into a forecasting tool).

3.2. Research questions

1. What impacts the selection of a suitable antecedent rainfall window for a dynamic model? (Objective 1)
2. What is the behaviour of a model built as a dynamic rainfall-induced prediction system which does not rely on rainfall thresholds but uses rainfall estimates as a covariate in a statistical model? (Objective 2)
3. How efficiently can the model predict landslide probabilities if short-term rainfall estimates are plugged in the statistical model to visualize changing susceptibility with changing precipitation? (Objective 2)
4. What is the added value of a model that features rainfall estimates? (Objective 2)
5. What is the added value of a model which accounts for the uncertainty estimation? (Objective 2)
6. How can a suitable/optimal probabilistic cut-off be defined to separate alert levels for informed decision making? (Objective 3)
7. What is the capability of the model to provide landslide warning signals in the form of a local alarm system? (Objective 4)

3.3. Thesis outline

The structure of this thesis is organised in the following chapters. Chapter 4 describes the methodology and data used to approach the research problem addressed above. The tools are also explained to give a full overview of the methods. Chapter 5 describes the study area and the data available for that particular area which is essential for shaping the model building phase. Chapter 6 presents the obtained results followed by Chapter 7 where the results are discussed with respect to the research objectives and questions. The final section, Chapter 8 concludes the study by providing final remarks and recommendations as well as highlighting the limitations of this study.

4. METHODOLOGY AND DATA

To integrate rainfall in a statistical model along with other covariates, a multi-variate approach is used in the context of an additive model. This section provides an overview of the data collection as well as the methodology followed in this research, explaining the modelling framework used to obtain the output.

4.1. Research methodology

The research framework aimed to achieve in this study is described in the sections below.

4.1.1. Modelling framework

The approach in this research is based on Bayesian statistics and framed in the context of a binomial Generalized Additive Model (GAM). Most of the modelling procedure has been implemented in R (RStudio Team, 2022) and specifically using the R-INLA (Integrated Nested Laplace Approximation) package, which has recently become a standard for Bayesian inference (Rue et al., 2009). For additional details and accessibility, see https://www.r-inla.org/.

The aim is to develop a binary reference model based on multi-temporal event-specific inventories over the given study area. Here, a binomial GAM, which has been evaluated as a suitable method for landslide predictions (Goetz et al., 2015), will serve the purpose to handle linear as well as non-linear behaviours of the selected covariates with respect to landslide occurrences. The model will be built, as shown in Figure 4, such that it will integrate topographic and thematic variables carrying a static predisposing factor's signal, as well as rainfall estimates and Enhanced Vegetation Index (EVI). EVI is similar to Normalized Difference Vegetation Index (NDVI) and conveys a dynamic predisposing control on landslide occurrences. Any model framed in a Bayesian context will natively provide an uncertainty description of the constitutive elements of the model (Luo et al., 2021; Wagenmakers et al., 2008). Moreover, a reference predictive equation (1) can be estimated on the basis of landslide-event inventories triggered by heavy rain, and then the same equation can be used for nowcasting by removing the previous rainfall signal and plugging-in forecasted or current rainfall estimates (Luo et al., 2021). This will define a dynamic susceptibility whose patterns can be converted into maps as all or part of the predictor set change in time.

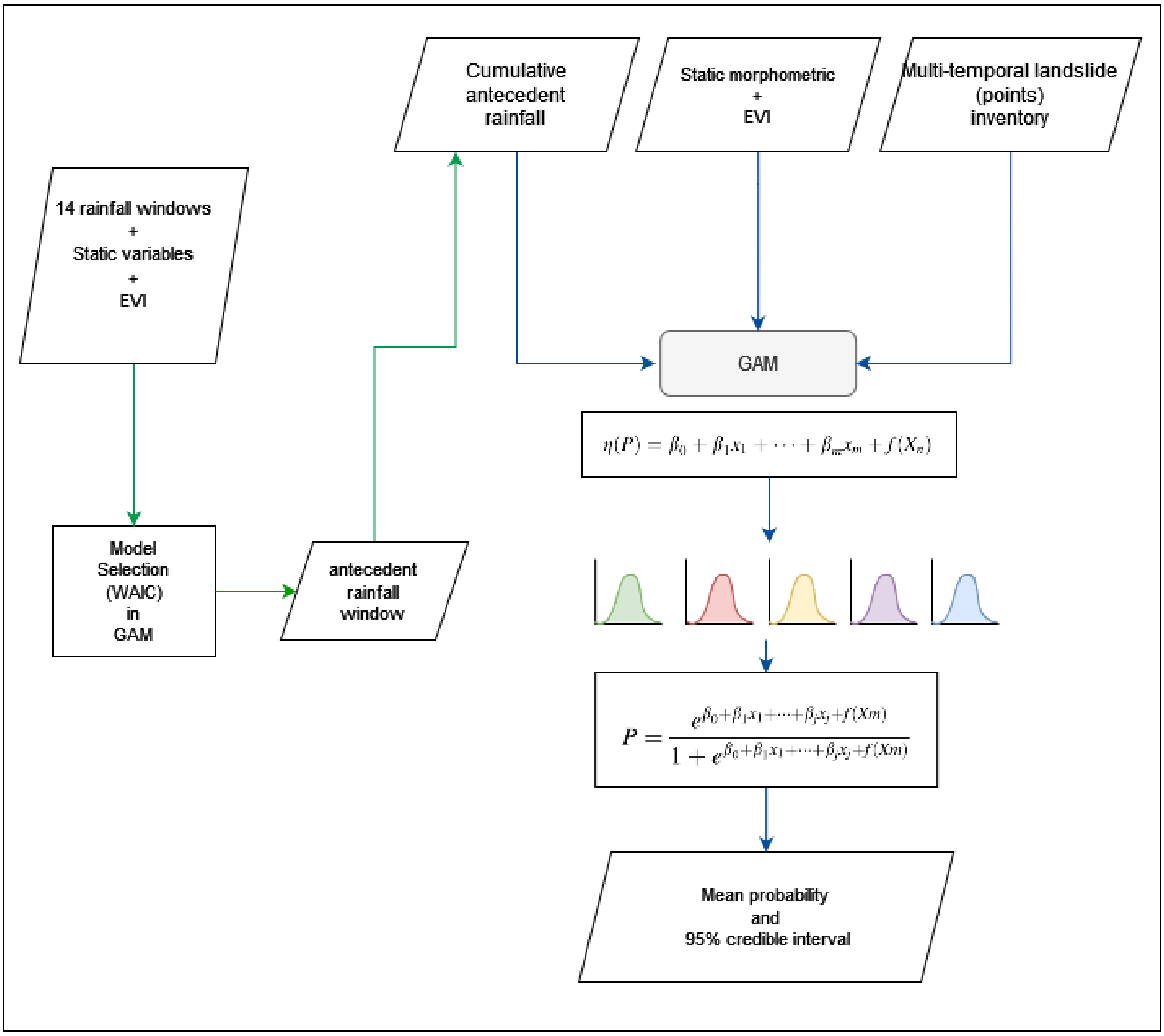

Figure 4 Workflow of the inputs and outputs of the basic model building phase. Green arrows represent steps followed before those led by blue arrows.

Moving toward the specifics of what a GAM implies, this model is an extension of the more common Generalized Linear Model (GLM). On the subject of landslide susceptibility modelling, a GLM holds the assumption that the behaviour of landslides presence/absence corresponds to a Bernoulli probability distribution (Brenning, 2008), whose unknown probability can be modelled through the following linear construction:

$$\eta(P) = \beta_0 + \beta_1 x_1 + \cdots + \beta_m x_m \text{ , (1)}$$

where P is the probability indicating presence of landslide within a mapping unit, β_0 is the global intercept, every β represents the regression coefficient for the covariates ($\mathcal{X}$), which are assumed to exert their effect linearly on unstable slopes, and η represents the logit link. The logit function accommodates the transformation of equation (1) which indicates linear combination of products between chosen covariates and respective coefficients, and shifts from modelled odds scale to the required and interpretable probability scale. In other words, the probability P can be recovered by inverting (1) as follows:

$$P = \frac{e^{\beta_0+\beta_1 x_1+\cdots+\beta_m x_m}}{1+e^{\beta_0+\beta_1 x_1+\cdots+\beta_m x_m}}$$, (2)

GAM is an extension of the approach presented above. A GAM allows one to include linear effects as well as nonlinear covariates' behaviours (Brenning, 2008; Muenchow et al., 2012). In this case, equation 1 can be expressed in its simplest form as follows:

$$\eta(P) = \beta_0 + \beta_1 x_1 + \cdots + \beta_m x_m + f(X_n)$$, (3)

where f expresses the nonlinear function of a covariate X which has n discrete classes.

4.1.2. Model scope

A Bayesian version of a binomial GAM retrieves a distribution (range of values) for each model component (Luo et al., 2021). This means that mean and credible intervals of probabilities can be obtained for every mapping unit.

In a LEWS based on a Bayesian version of a binomial GAM, forecasted rainfall can be plugged-in the model to follow the same process and return probability estimates by integrating rainfall as the dynamic covariate in the model. More dynamic variables can also be introduced, thus, EVI is also considered to be a time-variant property of this model. The removal of rainfall thresholds, as used in previous literature, and inclusion of rainfall estimates directly into the model give way to the visualization of a dynamic probabilistic LEWS, as a product of the varying rainfall patterns.

4.1.3. Model selection

One element that traditional LEWS heavily rely on is the concept of intensity-duration (Guzzetti et al., 2008, 2007; Hong et al., 2006; Kirschbaum et al., 2012). This notion implies that landslides occur in response to rainfall that can be discharged before the actual date of the failure. Hence, the model sought in this research should also be considered to account for this intensity-duration relation. For this reason, the daily cumulative antecedent rainfall will be integrated in the reference model. This will be done by summing the rainfall of the event day to the rainfall of *t* days prior to the event, where *t* ranges from 1 to 14 days. However, the choice of the antecedent rainfall window is not set in stone. The literature suggests different rainfall antecedent windows. LHASA version 1 integrated a rainfall threshold which was based on 7-day antecedent rainfall (Kirschbaum and Stanley, 2018) to explain the saturation in soils which served as a catalyst for slope failure on any day. Conversely, Tien Bui et al. (2013) uses 15 days. Therefore, to test the most suitable rainfall window for the study area, a model selection tool is necessary to be introduced.

In Bayesian framework, the Watanabe Akaike Information Criteria (WAIC) is often used as a model selection tool. A stand-alone WAIC value does not bring valuable information but in a relative comparison, it can be used to select the most appropriate predictor set (Whalen and Hoppitt, 2016). The most representative covariate subset of a larger group would lead to smaller WAIC values. Other similar metrics exist, like the older version of WAIC, Akaike Information Criterion (AIC) and Deviance Information Criterion (DIC). However, WAIC is preferred over its alternatives due to its ability to evaluate the model's fit by utilizing the full posterior distribution in Bayesian context, as compared to a single estimate (Watanabe, 2013).

Therefore, for this study, keeping the model parameters the same, a different rainfall window will be integrated and this plug-in will be moved over 14 different models carrying one cumulative antecedent rainfall window each, from 1 to 14 days. The WAIC for every model will be stored and compared to assess the lowest among all 14 models.

To corroborate the indications collected through the WAIC, a parallel temporal cross validation technique will also be added. The temporal validation technique refers to leaving one of the six inventories out for validation while calibrating on the remaining five. This will be done for all of the 14 models, and the Receiver Operator Characteristic (ROC) curve will then be used to assess the performance. This will help providing additional information and support the choice of the most suitable day to express the rainfall intensity-duration control on landslides.

4.1.4. Performance assessment

ROC curves can be used when evaluating the ability of a binary classifier to correctly identify the areas with and without landslide occurrence (Yang and Berdine, 2017; Zou et al., 2007). This tool plots True Positive Rate (TPR) against False Positive Rate (FPR), obtained from a confusion matrix (Table 1), in a curve which admits to a resultant Area Under Curve (AUC) to evaluate overall accuracy of the diagnostic test (Zou et al., 2007). A ROC curve and its AUC can both be used to assess the performance of an explanatory model as well as a predictive task (comparing the estimates to unknown data) (Zou et al., 2007). The performance for all 14 models were assessed by using the average AUC of the six temporal validation outputs for each model. AUC ranges between 0 and 1, where anything below 0.5 accounts to unacceptable discrimination, since 0.5 itself would mean the model does no better than what would be the output by chance (Hosmer et al., 2003). Values below 0.7 are also not preferably acceptable, however AUC in the range of 0.7 becomes acceptable and increases in excellency of discrimination accuracy as the AUC approaches 1 (Hosmer et al., 2003; Yang and Berdine, 2017; Zou et al., 2007).

Table 1 Confusion matrix to determine correctly and incorrectly classified points by a binary classifier.

		Observed	
		Presence	Absence
Predicted	Presence	TP	FP
	Absence	FN	TN

Table 1 graphically defines True Positive (TP), False Positive (FP), True Negative (TN) and False Negative (FN). From this confusion matrix, F1 score can be obtained further via measuring precision and recall (Goutte and Gaussier, 2005). Precision is the ratio of TP to the total predicted positives (the latter defined as TP + FP), and recall (or sensitivity) is the ratio of TP to the observed positives (the latter defined as TP + FN). F1 score is the harmonic mean of the precision and recall for a binary classifier. The value range of this metric is bounded between 0 and 1, with F1 score = 1, measuring high performance of a test. F1 score is considered to be a measure of good performance especially for unbalanced data. Thus, it will be used to explore the selection of a probabilistic cut-off. However, this metric may not be independently sufficient to identify a fair cut-off point, thus specificity or True Negative Rate (TNR) and sensitivity (TPR) will also be examined along with the F1 score (more information can be found in Annex 4).

4.1.5. Model validation

The validation of the model is an essential part of any predictive approach and determines the quality of the model when fed with unknown datasets (Chung and Fabbri, 2008, 2003; Lombardo and Tanyas, 2020; Remondo et al., 2003). The validation techniques in this research will be of multiple characteristics. Since the inventories used for this study represent the temporal domain of this model, temporal validation along with variations of spatial validation will be included.

Temporal validation will deliver the results of building the model with five out of six inventories and validating on the sixth inventory, which belongs to a different time step. This step will be repeated for validating on all of the six inventories one by one, to understand the nature of the model as well as the impact of the inventory richness.

Moreover, a sequential temporal validation techniques will also be explored. This means that the calibration starts by the first inventory in ascending temporal order and validate on the second one. This will snowball to moving the calibration input data in time, for instance calibrating on the first and second inventories while validating on the third, and then moving to calibrating on the first three inventories and validating on the fourth. The sequence will be carried forward till the first five inventories are used to build the model and the last inventory in time is used for validation. This type of temporal variety in the data is extremely useful since many susceptibility models are missing this component. Looking back and checking the quality of a model built to predict landslide occurrences is possible with future events occurring in the same region, which is the essence of the model's purpose.

One of the spatial validation technique is performing a 10-fold cross validation, where random sampling of mapping units is used to create subsets of testing and training datasets in a space-time dimension. This cross validation approach will allow the model to be trained on 90% of the area and tested on 10% of the remaining area. The 10 iterations will produce a map that has prediction values over the entire study area (Luo et al., 2021).

The second approach for spatial validation includes grid-based partitioning, where the study area is segregated into blocks through an overlaying lattice, therefore creating multiple spatial units with each including a number of mapping units. These grid blocks, and the corresponding mapping units within them, are used for cross validation which is in contrast with random sampling.

4.1.6. Mapping unit

Choosing the mapping unit addresses the spatial partition of the landscape, resulting in an individual unit upon which the failure probability will be assigned to once the model retrieves a fit. Among the choice of multiple types of spatial partitioning of the landscape that have been proposed and used in previous studies (Steger and Kofler, 2019), the predominant choice is represented by a grid-cell. The grid-cell is a type of mapping unit that uses a squared lattice covering the study area, with regular grids dividing the extent of the area as matched by the spatial resolution of the chosen Digital Elevation Model (DEM) (Reichenbach et al., 2018). Though, the fine resolution of such a mapping unit may be preferred to describe initiation points of landslide occurrences, for slope management methods, it may not be the most suitable unit to use, especially for implementing mitigation measures.

Early warnings are also an indication where slope stabilization measures need more attention. However, stabilization practices usually are implemented on the entirety of a slope rather than on a single grid-cell in practical application. Even so, a lattice of a coarser resolution can distort the natural morphology of downhill movement of material by simultaneously including ridges and streamlines in a single unit, as visually evident in Figure 5. To resolve this neglection of morphological properties, Carrara et al., (1991) proposed a coarser mapping unit which represents landslide behaviour in a more comprehensible manner with respect to the landscape. This unit is referred to as a Slope Unit (SU), and it is a delineation of space by streamlines and ridge lines, under the control of homogenous slope exposition (Ba et al., 2018; Carrara et al., 1991). The subdivision of geographic space in this manner implies that the process of failure in a given slope unit is independent of the failure mechanism in the neighbouring unit. This puts the decision maker at an advantage to focus on the instability of individual slopes for monitoring and early actions. Naturally, the coarser nature of slope units, as compared to a fine lattice of grid-cells, allows faster computation of complex data-driven models by using fewer objects to assign probabilities on. Moreover, a comparison of mapping units (slope units and grid-cells) based on a statistical analysis for landslide susceptibility resulted in better performance of the model using slope units due to their close relation with representation of the geographic environmental setting (Ba et al., 2018).

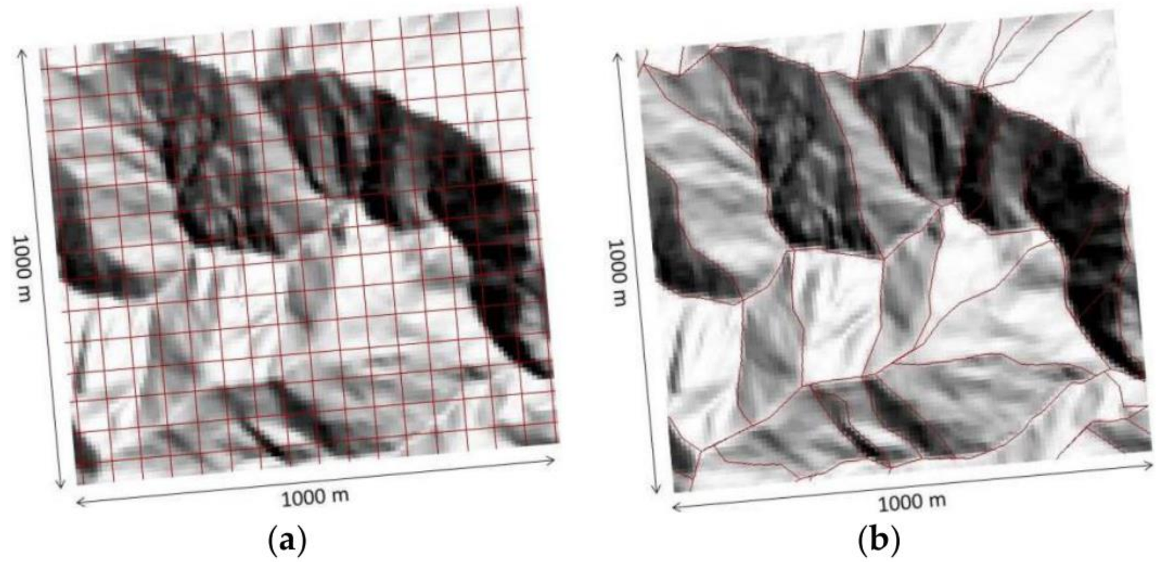

Figure 5 Comparison of landscape partitioning showing grid-cell (a) and slope units (b). Source: (Liu et al., 2018).

To set the mapping unit for the analysis of this study, the generation of slope units will be done using the software r.slopeunits (Alvioli et al., 2016). All SUs that coincides with a landslide point(s) will be assigned a ‘presence’ status; indicating that a landslide is present in this slope unit for the events under consideration. Meanwhile, any SU that does not include intersecting landslide points will receive a status of ‘absence.’

4.2. Data preparation

Google Earth Engine (GEE), a cloud-based platform, is a freely accessible tool that consists of various geospatial datasets of varying spatial and temporal resolutions. GEE behaves as a source holding a repository of aerial, satellite and ground-based data infused with remotely sensed data as well as built-in algorithms to manage big geo-data. The services available through this platform provide usability in analysis which requires earth observation data in understanding morphological processes (Kumar and

Mutanga, 2018). The initiation of GEE in 2010 provided great ease by tackling the long processing time on a personal computer and the ability to skip downloading heavy satellite imagery data. In addition, Earth Engine also allows visualizing the results of any methods and processes into a cloud-based app. Such apps can access satellite data in real time from the Earth Engine Catalog that contains pre-processed as well as raw datasets, providing users with a variation of data types (Kumar and Mutanga, 2018).

Predicting landslide occurrences is not straightforward as it requires estimating relationships of factors with respect to slope failures. There exists a complex interaction of slope condition as well as external influences which act together to make the slope susceptible to failure. Assessment of landslide susceptibility depends on the availability of information of predisposing factors and the trigger, which is not always known or estimated to explain the failure event (Lombardo et al., 2020). Models requiring rich temporal data can be supported by the multiple petabyte of geospatial (along with socioeconomic) repository provided by Earth Engine, with vast datasets available within a unified platform (Gorelick et al., 2017).

There exist quite a few satellite-derived rainfall products at different temporal and spatial resolution. IMERG aims to provide rainfall information every 30 min using sensor information from GPM and Tropical Rainfall Measurement Mission (TRMM) at nearly 11 kilometres of spatial resolution (Tang et al., 2020). Though LHASA used the above-mentioned information for its landslide warnings, it is still a coarse resolution as compared to the Climate Hazards group Infrared Precipitation with Stations (CHIRPS). CHIRPS (Funk et al., 2015), with daily and monthly estimates, fills gaps in datasets which cannot offer low latency and fine spatial resolution integrated with station data for richness. For the aforementioned reasons, CHIRPS has the highest spatial resolution in gridded precipitation satellite datasets, and a 2 day latency period which is reserved for blending station data to retrieve the first product (Funk et al., 2015). Table 2 shows the description of the aforementioned datasets used in this study.

Table 2 Description of the major datasets used in this study.

Data	Source	Resolution	Description	Type
NASA SRTM DEM	(Farr et al., 2007)	~30 meters	Quasi-global data, which uses other datasets to produce a more complete and void-filled product	Source for extracting static covariates
CHIRPS Daily	(Funk et al., 2015)	0.05°/ ~5.5 kilometres	Near-global rainfall dataset, which incorporates satellite imagery with in-situ station data.	Dynamic (trigger) covariate
MODIS Aqua Vegetation Indices	(Didan, 2015)	250 meters	A product that provides NDVI and EVI. Where EVI is more sensitive to dense vegetation and removes residual contamination in the atmosphere.	Dynamic covariate

4.2.1. Generation of slope units

The software r.slopeunits (Alvioli et al., 2016) was used to generate slope units from a medium-high resolution SRTM DEM. The software was mainly utilized to optimize the mapping unit for the purpose of terrain division performed with optimal parameters in the context of landslide susceptibility modelling

(Alvioli et al., 2016). The minimum area of an SU was set to 40,000m² to dissect the region in a reliable manner. A circular variance of 0.5, with a large flow accumulation threshold of 80,000m² was used to identify most of the slope units. The clean size was set to 20,000m² to remove unrealistic and oddly small subdivisions of the terrain which can be built upon plains or homogenous slopes (Alvioli et al., 2016). The iteration number was set to 20, to obtain the SU as per the input requirements with the reduction factor being 10.

4.2.2. Covariates

The input variables of the model which will be implemented in this research will inherently be of two relevant characteristics: static and dynamic. The static covariates are assumed to be time invariant, whereas the dynamic characteristics will change over space and time.

4.2.2.1. Time-invariant factors

Among many of the available datasets in GEE Data Catalog, for much of the literature addressing landslide susceptibility assessment, topographic information is extracted from global DEMs. The most common DEMs include the Shuttle Radar Topography Mission (SRTM) and Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) Global Digital Elevation Model (GDEM). The SRTM, specifically a Digital Terrain Model (DTM), is considered a medium-high resolution DEM (Brock et al., 2020). The SRTM version available accessible from the GEE Catalog is the third version of the product which uses the data collected in 2000 from the original mission by NASA combined with Italian and German Space Agencies (Farr et al., 2007). This version of the product has been corrected for voids using other existing DEMs like ASTER GDEM2, and hence, was considered a fair choice to use in extracting information for the static covariates.

The static covariates were extracted and aggregated to SU level, by using the mean and standard deviation of all pixel values contained within a slope unit. The covariates (summarized in Table 3) will be used to inform the model regarding the static, or temporally-invariant signal of the landscape with respect to its tendency to initiate landslides.

Table 3 Static covariates, used in this study, described by their characteristics expressed in the landscape.

Covariate (unit of measurement)	Description
Elevation (m)	Elevation is used as a proxy for its orographic effect on rainfall (Görüm, 2019).
Slope Steepness (m)	Slope steepness is one of the most intuitive factors to be used in landslide susceptibility. This covariate represents the balance between stabilizing forces and gravitational influence which, if in a weak balance, can easily be tipped over by external influences brought by rainstorms (Wu and Sidle, 1995).
Planform Curvature	Planform (or planar) curvature dominates the divergence of forces acting upon the slope as well as the direction of surface runoff (Heerdegen and Beran, 1982; Ohlmacher, 2007). Depending on its value range, planar curvature can indicate a planar, convergent or divergent terrain.

Profile Curvature	Profile curvature runs parallel to the highest degree of slope steepness and controls the forces acting upon it (Heerdegen and Beran, 1982; Ohlmacher, 2007). This curvature can be convex, concave or linear, contributing to the acceleration or deceleration of overland flows.
Eastness	Sine of aspect, a cyclic DEM-derived variable, contributes to the surface conditions by defining the direction of the slope. Ranging from -1 to 1 (representing slopes facing Westwards to Eastwards, respectively) indicates the influence on slopes brought by solar radiation and winds (Leempoel et al., 2015).
Northness	The cosine of aspect indicates the exposition of a given slope towards the North (values =1) or South (values =-1) directions. Depending on the hemisphere under consideration, this factor can indicate slopes that are exposed to sunlight for longer duration within a day, therefore also conveying information related to soil moisture conditions (Epifânio et al., 2014).
Internal Relief (m) (also referred to as relief hereafter)	Internal relief can have a strong correlation with the landslide activity (Görüm, 2019; Qiu et al., 2018). Relief is the difference between the elevation in a given pixel and the mean elevation in a specific neighbourhood, here defined with a 1km radius from each pixel in the study area. Its values are usually interpreted in terms of potential energy, a fundamental component of landslide dynamics.

In addition to these covariates, the roundness index of each SU was also included, this being calculated as the maximum length inscribable within a SU polygon divided by the square root of the SU area. Moreover, a measure of SU length was considered by using the maximum distance between any two pixels within a SU. These two measures were tested for collinearity (estimating the Pearson correlation coefficient between the two) and although the term 'SU length' appears in both, the correlation (see Annex 3 for details) between the two was estimated to show weak correlation (Schober and Schwarte, 2018) to justify the inclusion of both in the model.

4.2.2.2. Time-variant factors

The process consists of building a reference model capable of estimating the effect of rainfall onto the landslide scenario so that the predictive equation is used to project future landslide susceptibility patterns at varying precipitation amounts. However, the literature is rich of studies that indicate a dual effect of the rainfall on slope failures (Segoni et al., 2018). Prolonged rainfall before the day of the landslide initiation contributes to increased pore pressures and overall weight of the masses hanging on a given slope (Guzzetti et al., 2008; Segoni et al., 2018).

The aim of this study revolves around modelling landslides in relation to a rainfall event, which brings us to the triggering factor. To retrieve the trigger information, daily maximum rainfall estimates have been extracted from CHIRPS. This dataset has a resolution 0.05°, with daily rainfall aggregates available (Funk et al., 2015). Among analogous products, CHIRPS offers the highest spatial resolution and an overall good performance in the South-eastern Asian sector (Tang et al., 2020). For this reason, it will be used in this work to support the analyses expressed at a regional level.

The reference model will examine the rainfall signal with respect to landslide occurrences by making initial use of cumulative antecedent daily rainfall. These will be passed separately and for a maximum window of 14 days prior to the landslide occurrence date. This is done to select the most representative intensity-duration rainfall window.

An additional dynamic covariate introduced in this model is the Enhanced Vegetation Index (EVI). The EVI was taken for three weeks prior to the event date from the Moderate Resolution Imaging Spectroradiometer (MODIS) AQUA satellite that has a return period of 16 days with 250 meters spatial resolution (Didan, 2015). The covariate was used to describe the presence of vegetation cover 21 days prior to the event date as a proxy for root-strength (Wu and Sidle, 1995).

5. AREA OF INTEREST

To test the concept of a unique model, a study area is chosen influenced by the availability of available inventory information. This section describes the study area chosen for this study and the supporting landslide inventory acquired.

5.1. Study area

Among the countries most vulnerable to storms, Vietnam is heavily affected by natural disasters such as flooding and landslides causing damage to livelihood as well as the economy (Tien Bui et al., 2012). The lack of available and rich datasets concerning landslide inventories in developing countries like Vietnam makes it difficult to carry out susceptibility studies (Tien Bui et al., 2013). NASA's open data portal of Global Landslide Catalog was initiated to create a record for rainfall-triggered landslides globally (Kirschbaum et al., 2010). For the years 2003, 2007 and 2008, proportion of reported landslides were the second highest, following one of the highest proportion of fatalities (Kirschbaum et al., 2010). Vietnam, among other countries of Southeast Asia show that the peaking casualties are reported between June and November (Kirschbaum et al., 2010) which is also the rainy season that causes fatal landslides in Vietnam among other countries of the Lower Mekong Region (LMR, Amatya et al., 2022). The selected study area (see Figure 6) is located in the North-Western region of Vietnam, and encompasses the cluster of landslides detected for the LMR, among some events located towards central and Southern Vietnam (Amatya et al., 2022). The North-Western area of Vietnam is constituted of mountainous terrain, where the elevation can go higher than 3000 meters and is associated with steep slopes even exceeding 60^{o}. There are 15 districts of Vietnam that fall under the study area where landslides have occurred in the recent years (Amatya et al., 2022). The study area (over 59,000 km^2 wide) does not have the highest population density but is on the extreme end of the poverty index in Vietnam (Bangalore et al., 2019).

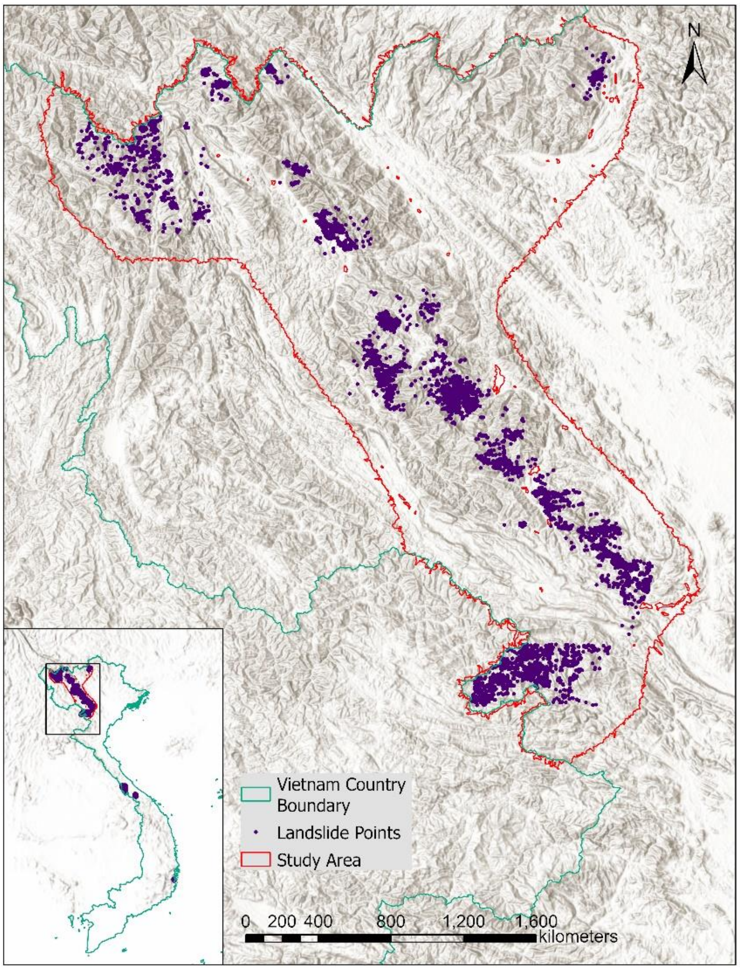

Figure 6 Study Area in Vietnam showing the landslides detected country wide and within the study area.

5.2. Inventory

An important prerequisite for estimating landslide susceptibility is the availability of multi-temporal inventories with adequate quality and completeness (Harp et al., 2011; Petschko et al., 2013). Quality and completeness are measures of accuracy and of how representative of the actual landslide scenario an inventory is. However, in certain areas worldwide, landslide inventories are scarce to begin with. This is mostly due to the fact that some countries lack the resources for record-keeping of landslide occurrences. This in turn makes it difficult to generate landslide susceptibility estimates and monitor the evolution of the susceptibility patterns in time. Hence, a unique opportunity was provided by NASA's implementation of LHASA to the LMR (Amatya et al., 2022). The interest in the area thus required for rainfall-triggered landslides to be mapped, knowing that their occurrence is predominantly destructive towards livelihood and economy during the rainy season, falling in the second half of the year (Amatya et al., 2022).

To implement LHASA to the LMR, a Semi-Automatic Landslide Detection (SALaD) system, which is an Object-Based Image Analysis (OBIA) approach, equipped with a change detection module (SALaD-CD). SALaD-CD was therefore built with the aim of generating event-based inventories for RIL over the LMR (Amatya et al., 2022). The LMR includes five countries namely, Myanmar, Laos, Thailand, Vietnam and Cambodia. Most of the landslide inventories were identified in the territory of Vietnam.

The area affected by landslides was initially determined on news media and official reports dated between 2009 to 2020. Within these areas landslide were then mapped using pre- and post-event satellite imagery from PlanetScope and RapidEye (Amatya et al., 2022). The choice between the constellations of satellite imagery was influenced by the availability and coverage of desired area in time as well as the common problem of cloud cover in satellite imagery. National scale landslide inventories do exist for Vietnam already, although they are not freely accessible or available for public use. NASA introduced a pre-processing step where images are adjusted by performing radiometric normalization and co-registration. Nevertheless, this pre-processing still cannot completely solve the limitations due to cloud disturbance, hence part of the resulting inventories may not be associated with a specific date of occurrence. Normalized Difference Vegetation Index (NDVI) is used as a discriminant measure to automatically detect landslides from pre- and post-images. Notably, falsely classified areas such as barren land or agricultural areas were removed manually to clean the inventory. The final output was translated into point data rather than polygons because landslide information in the GLC, the input for the global LHASA model, are landslide points. These were extracted from the landslide surface by using NASADEM to mark the highest elevation along the perimeter as the initiation point (Amatya et al., 2022).

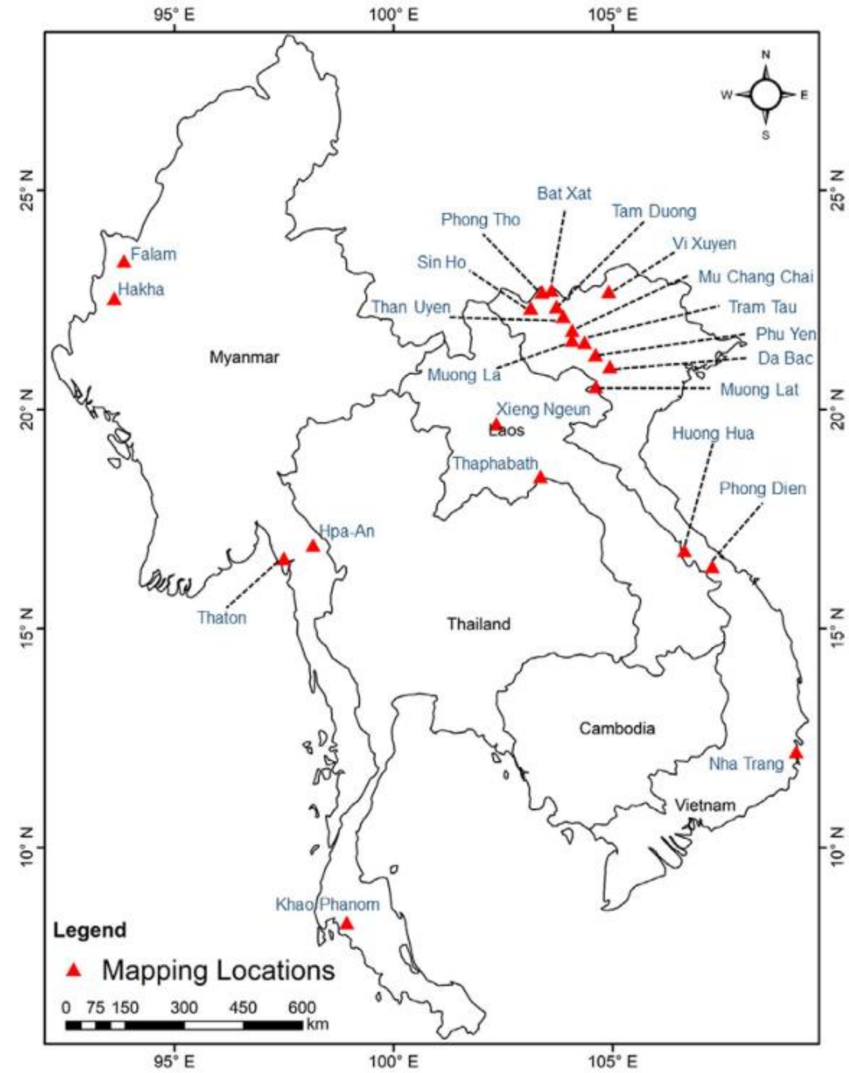

Figure 7 Mapped landslide points in the Lower Mekong Region (Amatya et al., 2022).

As shown in Figure 7, 15 locations were identified for Vietnam, with events spanning from 2017 to 2020, in the months of June to November. The inventories in Vietnam spread over the whole country, but are mainly concentrated in the North Western sector. Unlike Myanmar and Laos, where fewer inventories were identified and were spread quite far apart, Vietnam was considered to be a suitable choice with multiple dated inventories. To reduce computation time, the study area was reduced to the cluster of the inventories concentrated in the upper region of the elongated country boundary (as defined in Figure 8) while the events which occurred in the Southern region were stored separately as independent validation data sets. Table 4 shows the inventories with their indexing and the number of landslide points contained for each event date in the territory of Vietnam. As visible from the table, the dates marked for each event can have a range of uncertainty. For example, inventory 2 and 6 can be events of any of the six days of their corresponding date ranges. This can make it difficult to identify the precipitation estimates which in reality triggered the landslides. However, these inventories are still useful in informing the model of the study area's susceptibility, especially inventory 6 since it contains a significant number of landslide initiation points.

Table 4 Landslide inventory with identified dates and corresponding landslide initiation points.

Year	Inventory	Date	Landslides Points
2017	1	2nd-3rd August	2014
	2	23rd-28th August	99
	3	10th-11th October	3944
2018	4	23rd-24th June	1310
	5	3rd August	302
	6	27th August-1st September	1641

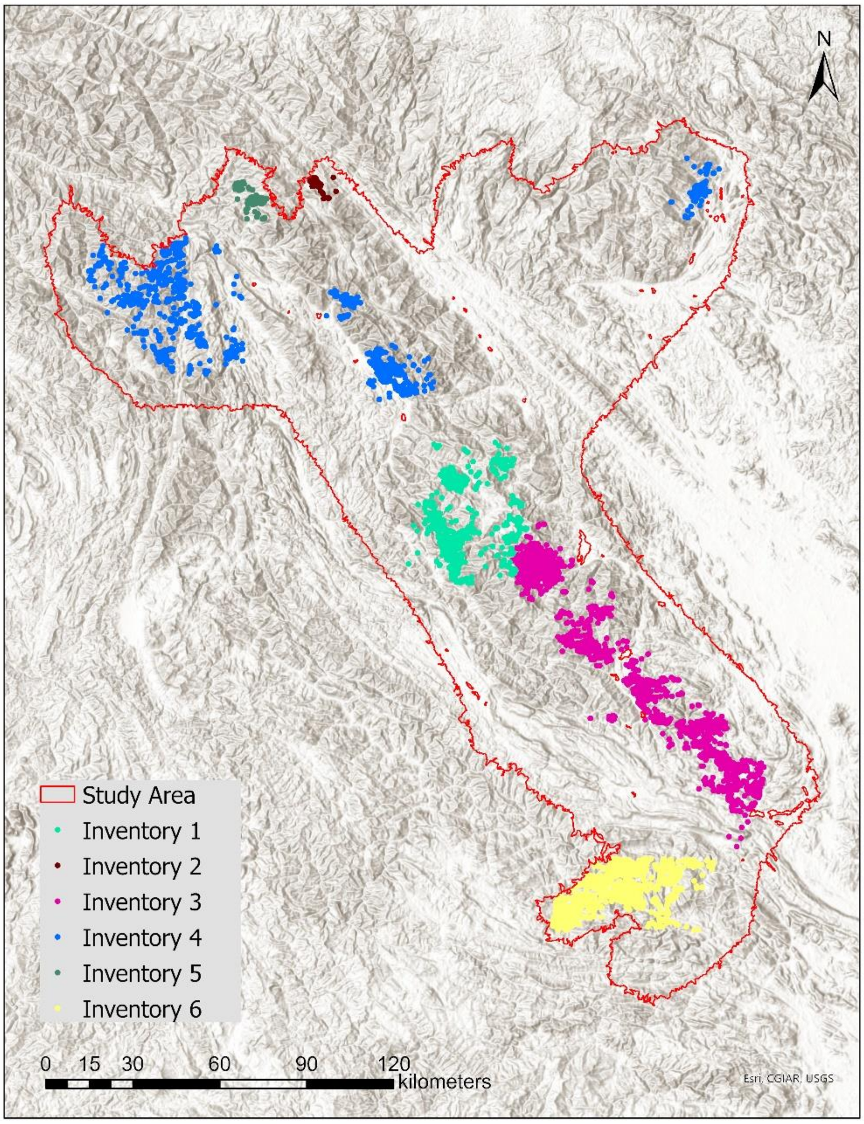

Figure 8 Colour-coded landslide inventories spread in the study area.

6. RESULTS

The following sections will provide the description of results obtained in this study. The outputs included are from the processes mainly followed as per the workflow displayed in Figure 4.

Section 6.1. describes the results obtained from the model selection tool and the corresponding selection of an antecedent rainfall window.

Section 6.2. presents the output of the reference model with its covariates' effects. It also shows the obtained probabilistic estimates of landslides, combined and individually for the six inventories.

Section 6.3. exhibits the results from multiple validation routines with performance metrics.

Section 6.4. conveys the output of F1 score after setting quantile thresholds for comparing the probability estimated predicted with the observed binary input.

6.1. Model selection

Table 5 shows the list of the first 10 models, ordered by ascending WAIC and associated AUC values. The lowest WAIC is attributed to one and two days of antecedent rainfall. These windows may be too small to explain the effect of antecedent rainfall in a common situation and especially here where the inventory dating reports errors over several days. Notably, AUC values are mostly above 0.8. The highest AUC value is displayed for the 8-day antecedent rainfall, with no individual AUC of a temporal validation dataset being lower than 0.75, which is also accompanied by the third-lowest WAIC. To consider the uncertainty of event date and looking at the highest AUC in Table 5, 8-day antecedent rainfall window was selected as a suitable window to be integrated as the dynamic covariate of rainfall in the reference model to be used further in this study.

Table 5 First 10 models, ranked by ascending order of WAIC with respective AUC values obtained by temporal validation

Antecedent Rainfall Days	WAIC	Average AUC
2	23492.17	0.82752
1	23636.73	0.82740
8	**23685.13**	**0.83115**
3	23735.00	0.82988
9	23790.42	0.81752
4	23824.17	0.82922
7	23834.70	0.82992
14	23843.29	0.82418
10	23849.71	0.82815
11	23862.83	0.82880

6.2. Reference model

A GAM was fitted to the six inventories combined together for building a reference spatio-temporal model. The inventories were of different qualities in terms of number of landslide points, area covered and temporal accuracy. However, as per the initial concern over Vietnam's rainy season which initiates these landslides, these inventories seemingly aligned with the time of high precipitation in June to November (Amatya et al., 2022). A set of covariates as described in Section 4.2.2.1., were extracted from a DEM (finer resolution), and then aggregated to SU level (coarser resolution) with the mean (m) and standard deviation (sd) to describe the variation within the mapping unit and how it contributes towards the slope stability. A GAM was then used to explain the linear as well as non-linear behaviours of the selected covariates with respect to slope failures (Luo et al., 2021) highlighting their role as either negative, positive or negligible. For this study, the non-linear covariates were modelled as ordinal variables, with adjacent class dependence, which means they were extracted from a continuous property and converted to an ordinal one to model its nonlinear behaviour. A first order Random Walk (RW1) was used to derive individual regression coefficients for every class constraining the adjacent class dependence through a spline function (Bakka et al., 2018). With three covariates fed to the model as random effects (non-linear), and the rest of the covariates as fixed effects (linear), a binomial version of a Bayesian GAM was used to build the reference model. Before building the model, the linear covariates were rescaled to have mean zero unit variance to express the covariates on the same unitless scale.

The reference model estimates a distribution of regression coefficients for every fixed effect as well as for every class of each of the random effects. This distribution can then be used to highlight the estimated uncertainty around the mean effect of the variables. Figure 9 presents the regression coefficient distributions for each fixed effect, summarized through the mean values and relative width of the 95% credible interval (CI hereafter). Usually, non-significant covariates with a mean regression value close to zero should be removed as the information they convey is negligible. In this study, this was the case for three covariates, preliminarily removed from the reference model, these being: standard deviation of relief and Eastness as well as the mean of the planar curvature.

6.2.1. Linear effects

As shown in Figure 9, the distribution of positive and negative effects varies among the covariates in the model. A significant and positive contribution is observed for elevation (sd), Eastness (m), planar curvature (sd), Northness (m and sd), antecedent rainfall and EVI (sd), with relative CI of varying width. Among them, the highest contributors are planar (sd), Northness (mean and sd) as well as the 8-day cumulative antecedent rainfall. Conversely, negatively contributing covariates include elevation (m), slope (sd), profile curvature (m and sd), EVI (m) and roundness index of the slope unit.

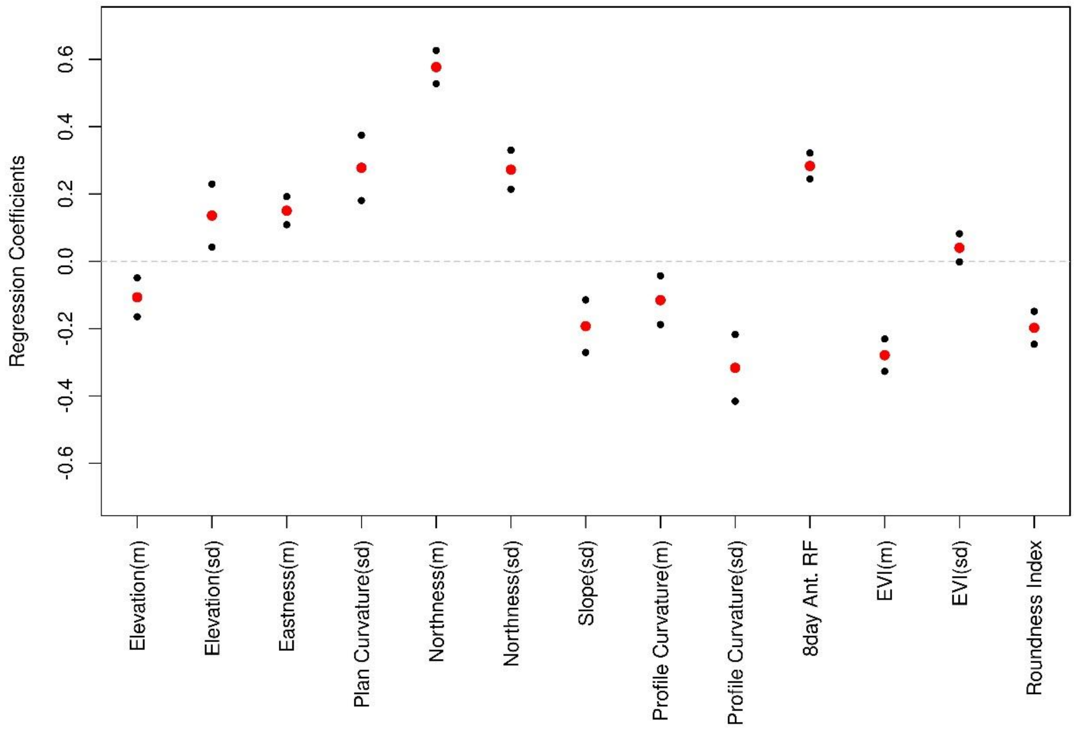

Figure 9 Regression coefficients with 95% credible interval corresponding to both significant and non-significant covariates of the final model.

6.2.2. Nonlinear effects

Among the choice of non-linear covariates, the slope steepness (mean) effect is presented in Figure 10, through its posterior distribution. There, a clear nonlinear trend emerges with a negative contribution towards slope failure marginally affecting the estimated susceptibility up to steepness values of 25 degrees. This threshold marks a change in coefficient sign. The increase in relative contribution is shown to keep on increasing linearly up to 35 degrees, after which an asymptote is reached.

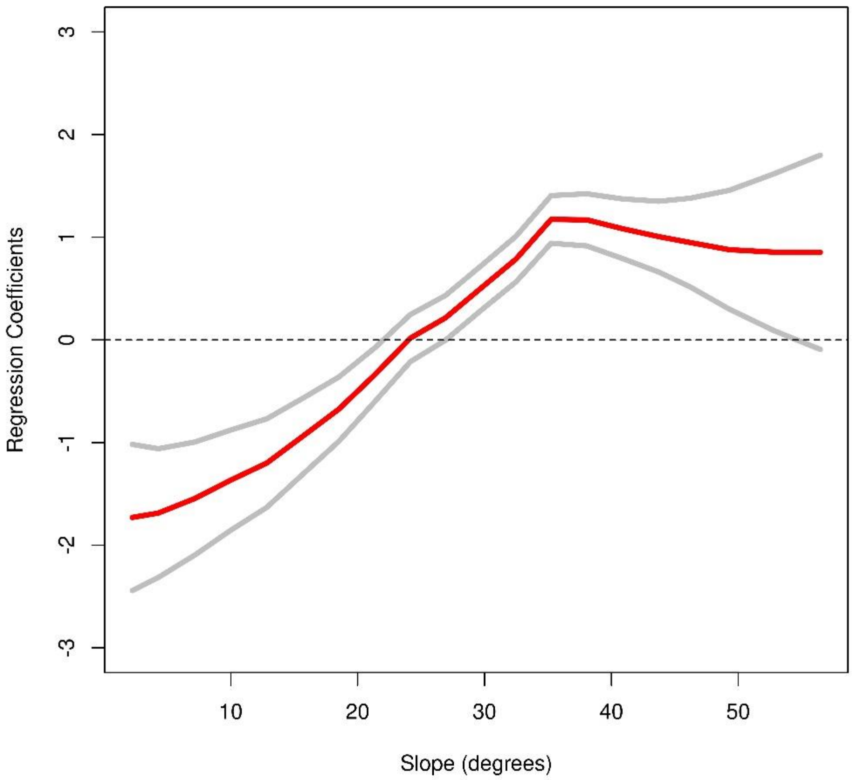

Figure 10 Posterior distribution of mean Slope (expressed in degrees) as a non-linear effect. Red line expresses the mean behaviour while the grey ones highlight the 95% CI.

The posterior distribution of mean relief (Figure 11) displays a rather changing trend along its range of values, with a similar behaviour as compared to the slope steepness. They both show a starting and ending point where uncertainty is larger and an initial negative contribution that transitions to positive (here at 500m of elevation change) before flattening out or slightly declining.

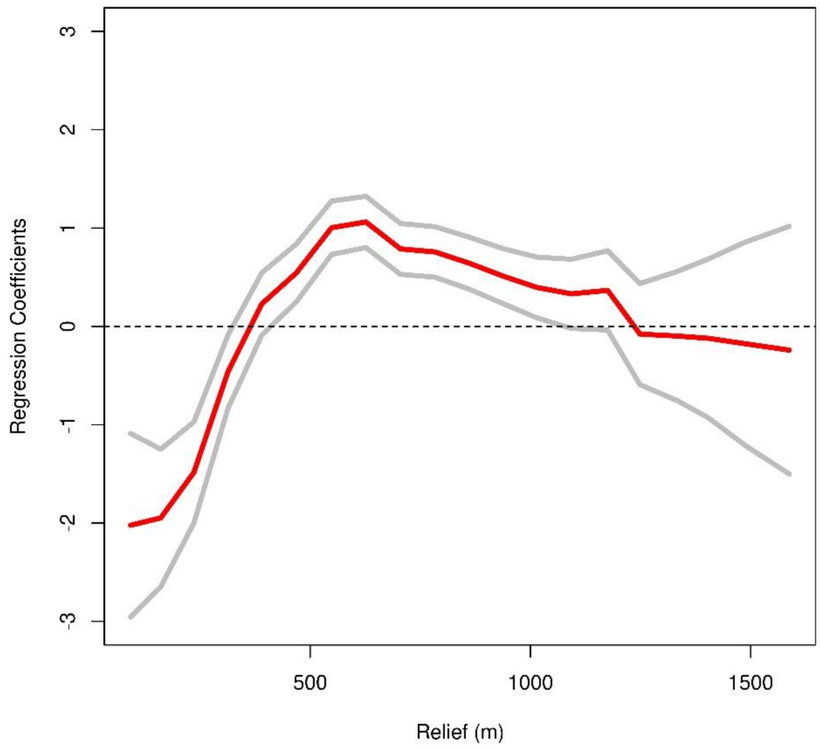

Figure 11 Posterior distribution of mean Relief (expressed in meters) as a non-linear effect. Red line expresses the mean behaviour while the grey ones highlight the 95% CI.

The effect of the maximum distance within a SU (Figure 12) begins with a negative and rather narrow CI after which the contribution to the susceptibility becomes increasingly positive and stabilizes at around 6000m, after which the CI becomes very large as compared to the remaining covariate value range.

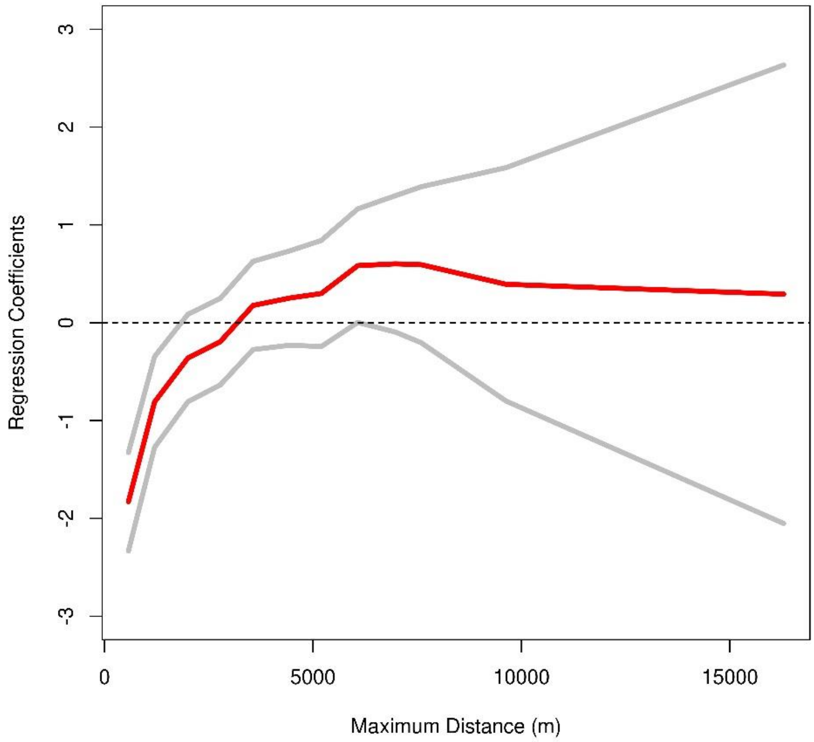

Figure 12 Posterior distribution of Maximum Distance (expressed in meters) as a non-linear effect. Red line expresses the mean behaviour while the grey ones highlight the 95% CI.

6.2.3. Dynamic susceptibility overview

The spatio-temporal reference model is dissected in its most representative maps in the figures below. Figure 13 provides a general overview of the susceptibility patterns across the study area and for the individual events considered. A matrix map of six inventories with their mean probabilities and width of 95% CI shows the higher probabilities shifting for each inventory as influenced by the concentration of landslide occurrences in space for that specific event.

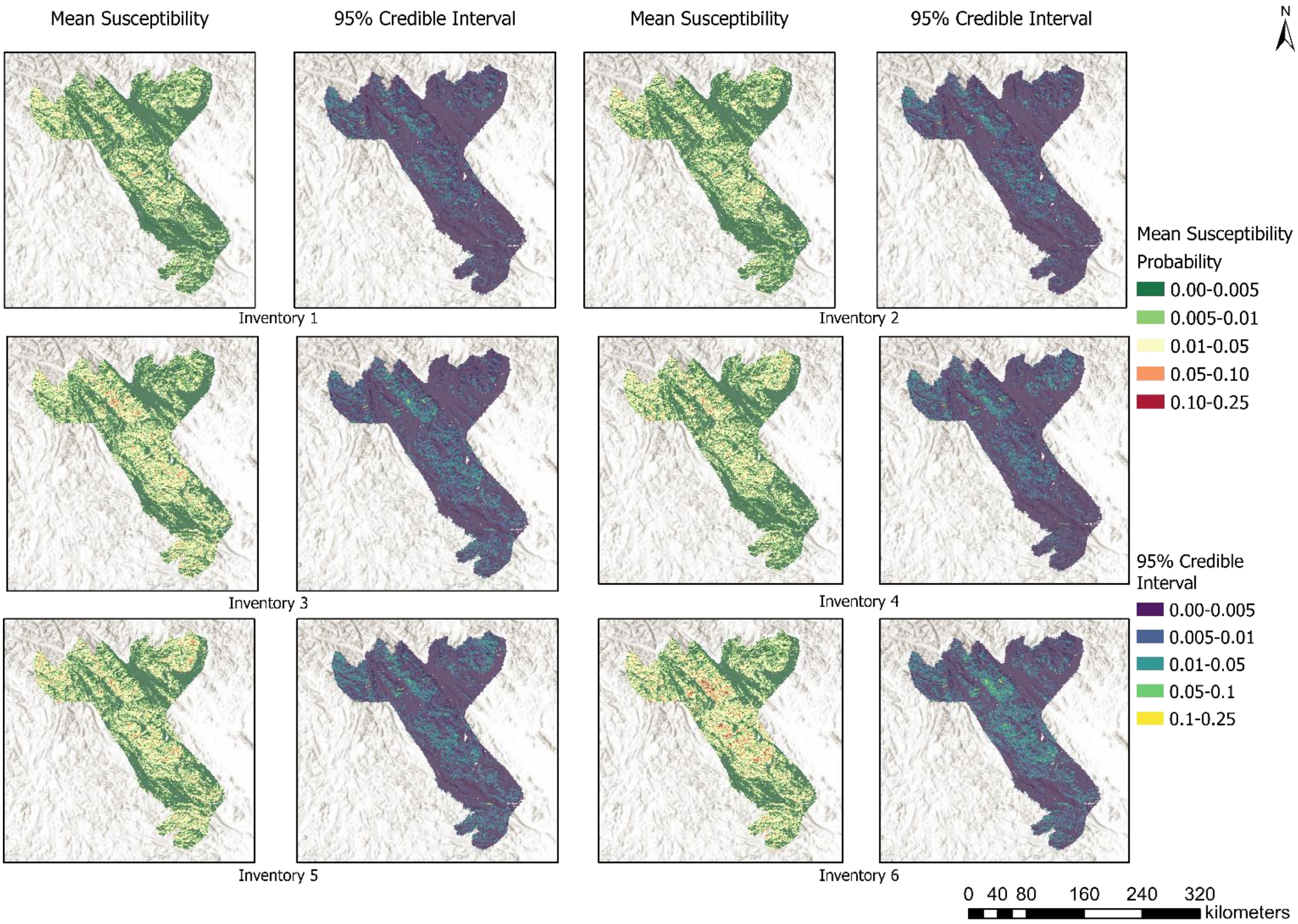

Figure 13 Individual susceptibility retrieved for every inventory with the width of 95% credible intervals.

Specifically, two statistical moments were chosen, these being the mean of the mean probability of landslide occurrence (Figure 14) and its variation measured in two-times the standard deviation (Figure 15) of the mean susceptibility per SU. Figure 14 provided a close-up of the study area and the mean of the mean probabilities (as values) as obtained from each of the six inventories. Relatively higher probabilistic estimates are seen along the mountainous terrain, and their corresponding variation in Figure 15.

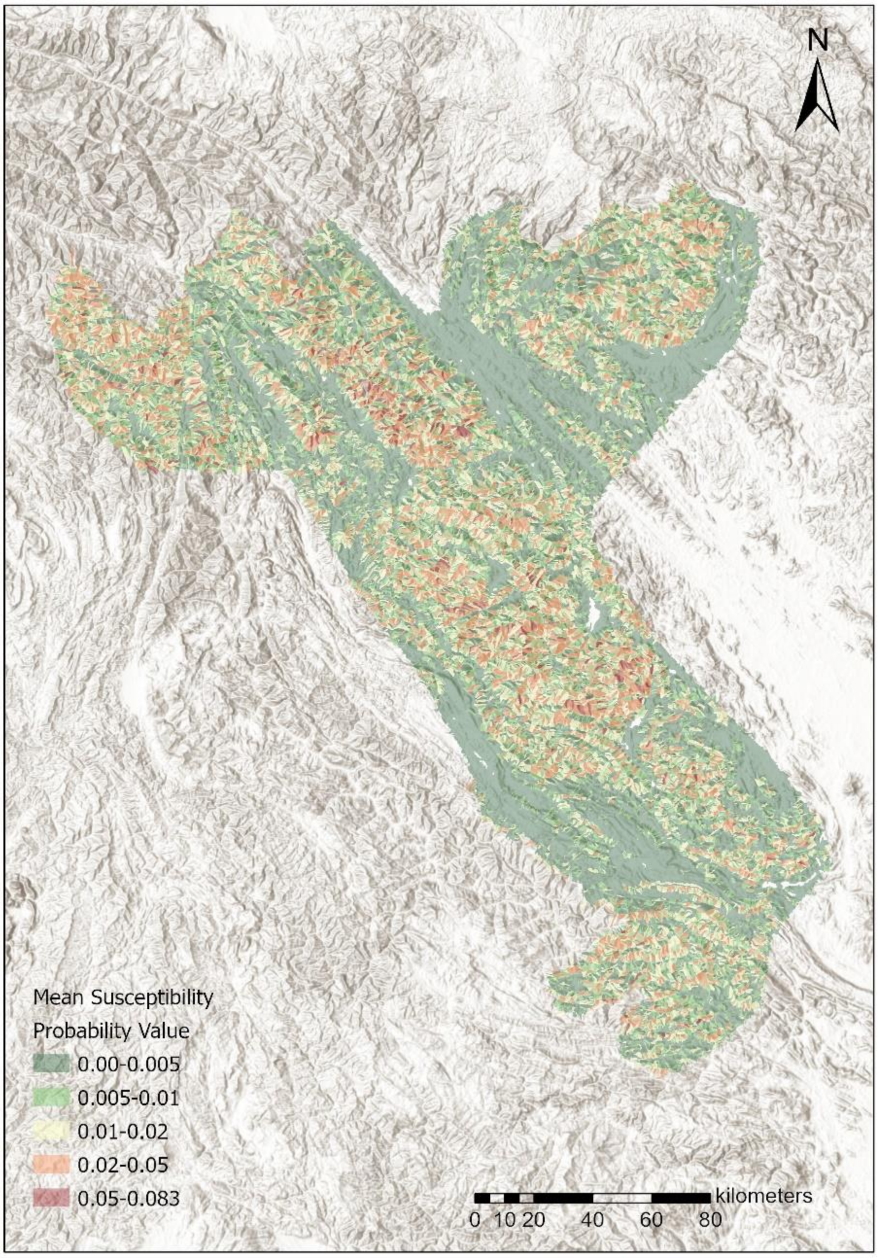

Figure 14 Mean susceptibility of the mean probability estimates of six inventories.

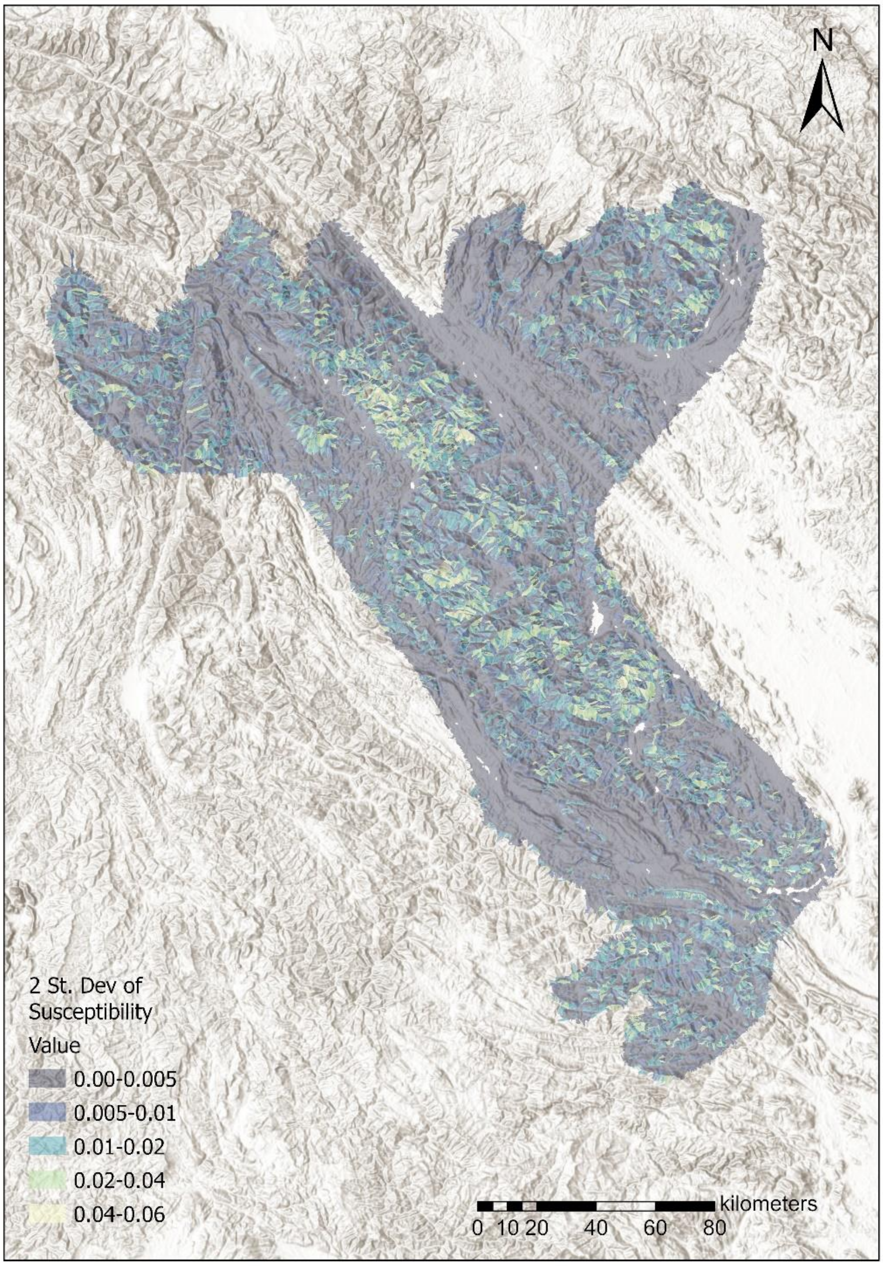

Figure 15 Two times standard deviation calculated from the six mean susceptibility maps.

To complement the above maps, the maximum susceptibility out of the mean susceptibility has also been extracted out of the spatio-temporal domain under consideration. The combination of the three (Figure 14, 15 and 16) is meant to provide an overview of the susceptibility dynamics, with SU that have been estimated to be susceptible on average (Figure 14) across the six events, how much the variation is expected to be around the mean behaviour (Figure 15) and what is the worst case scenario (Figure 16) that has been reached per SU. Notably, the reason why these summary statistics are referred to a mean estimate is due to the use of a Bayesian model, whose output per single event is summarized via a mean probability of landslide occurrence and the uncertainty estimated around it.

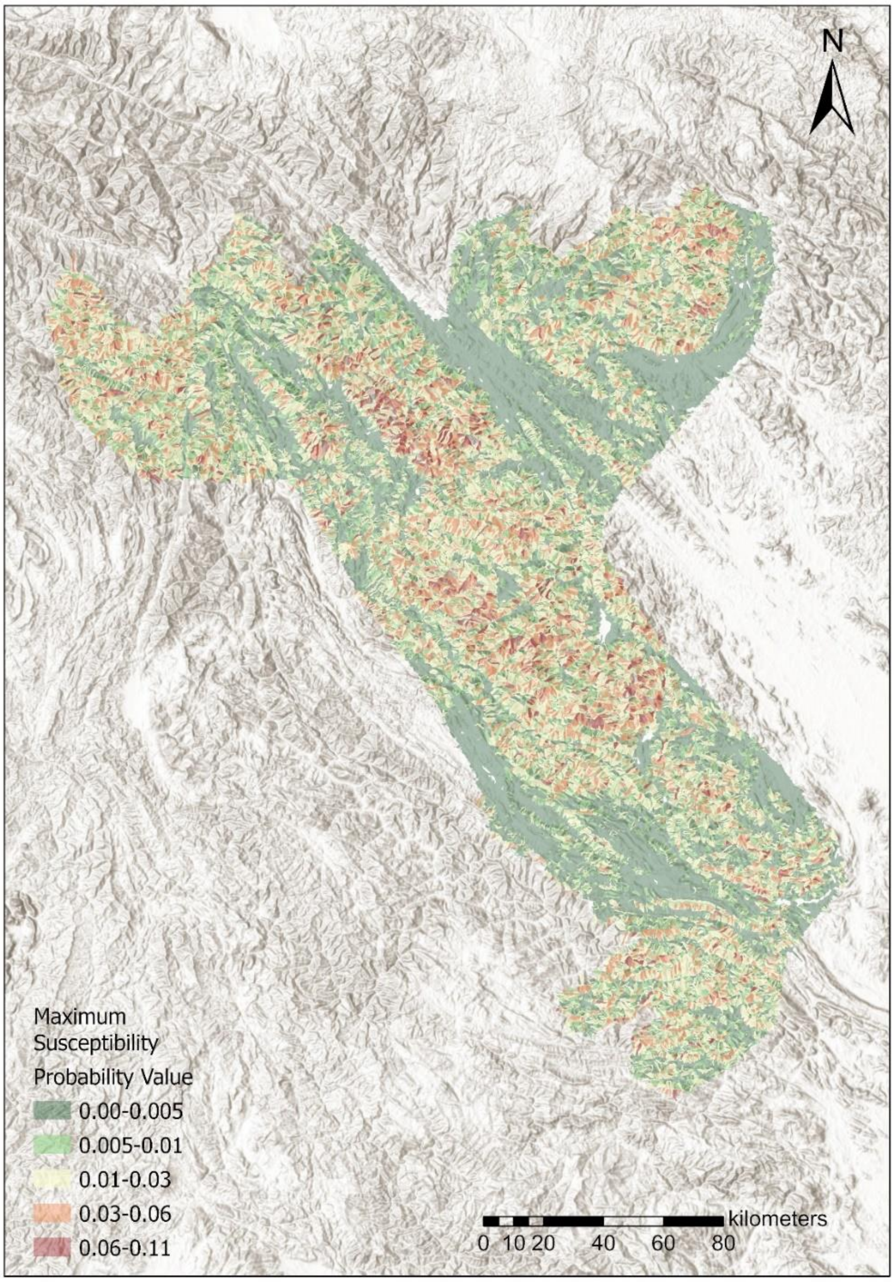

Figure 16 Maximum susceptibility of each slope unit among calibration susceptibilities of the six inventories.

6.3. Validation

Performing validation on an unknown dataset is essential to test the predictive capacity of any given model. In this research, multiple validation techniques have been implemented and summarized in the following sections.

6.3.1. Temporal validation

One of the rich qualities of the dataset prepared by Amatya et. al., (2022) for the LMR is the availability of multi-temporal inventories for different countries. This allowed to implement two suites of temporal cross-validations. The first one featured a calibration step on a specific inventory and validated it on the subsequent ones. This operation was then repeated by incorporating data as the time series moves forward. In other words, the second step features a calibration on the basis of data combined from the first and second inventory only to be validated on the third. Similarly, the third step used the first, second and third inventory for calibration and validated on the fourth one. This operation was sequentially implemented from the first to the last inventory in the time series.

Figure 17 shows the ROC curves obtained as a result of the validation performance and Table 6 assists in clarifying the data used and associated AUC values. There, an AUC range between 0.665 and 0.8962 can be observed.

Table 6 Assistive information on sequential temporal validation with inventory partitioning for calibration and validation.

ID/Step	Inventories used for calibration	Inventory used for validation	AUC
1	1	2	0.7670
2	1,2	3	0.6646
3	1,2,3	4	0.8936
4	1,2,3,4	5	0.8962
5	1,2,3,4,5	6	0.7519

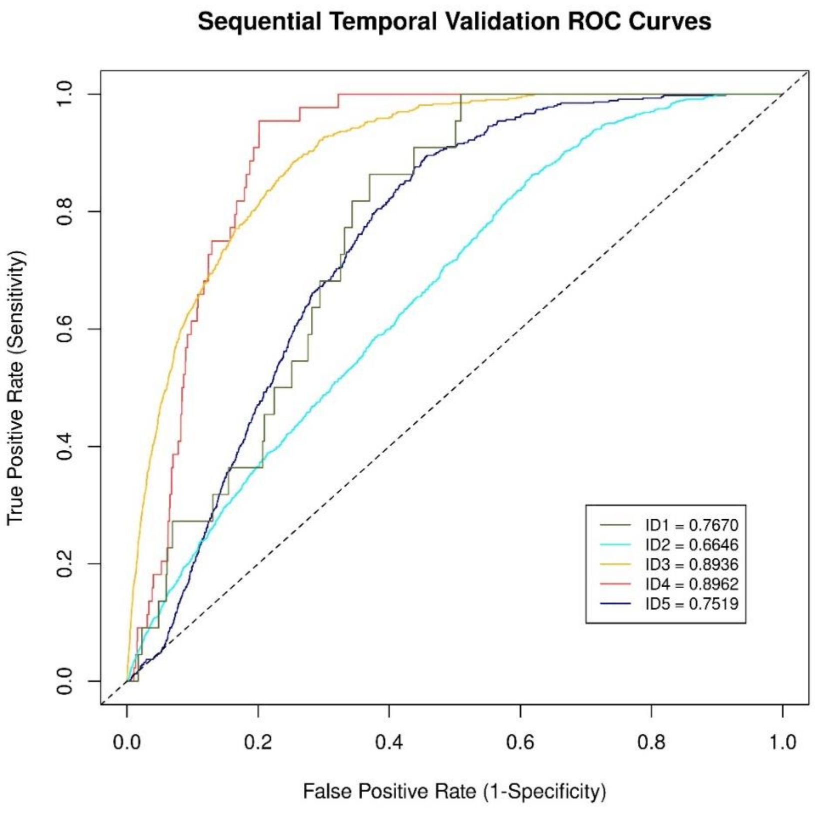

Figure 17 ROC curves with AUC for sequential temporal validation.

A complementary view of the temporal prediction capacity was achieved through a leave-one-time-out scheme. This means that the second temporal cross-validation was performed by calibrating each time on five out of the six inventories and validating on the missing one. This operation was then repeated six times covering the whole spatio-temporal domain. Notably, in all of the six cross-validations, every output exceeded the value of 0.75 for the AUC as shown in Figure 18.

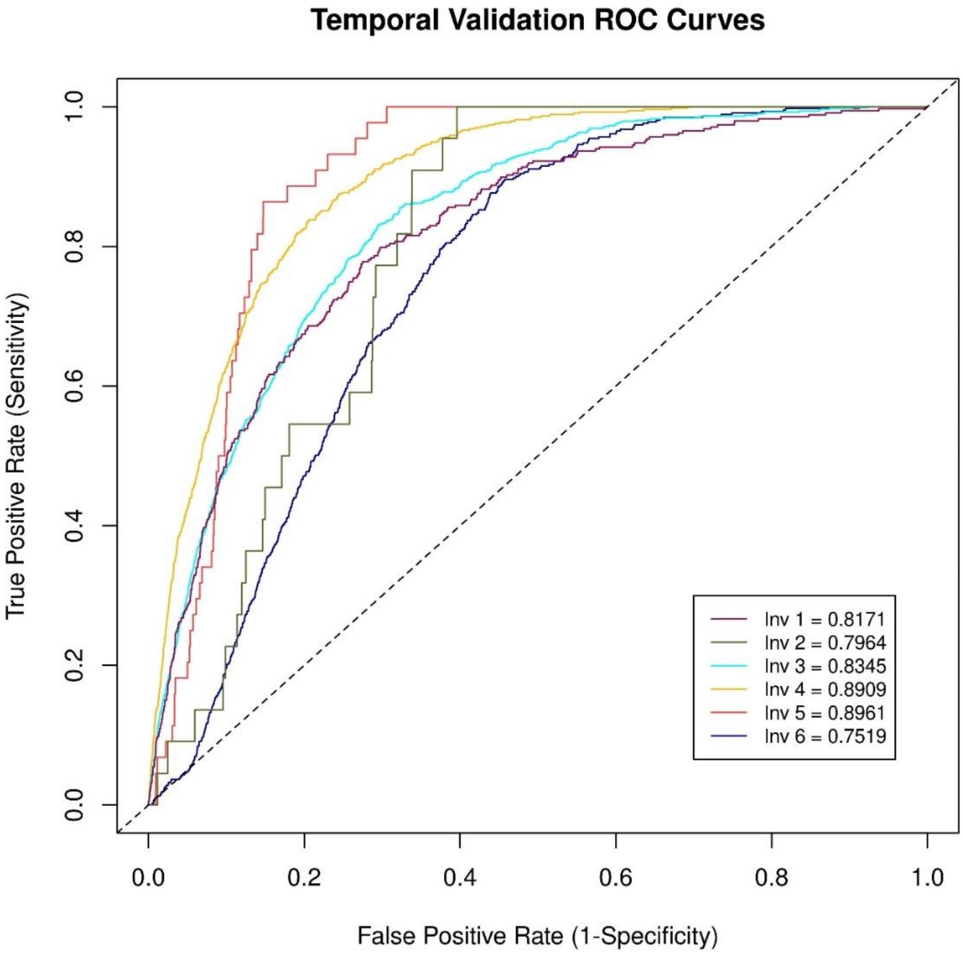

Figure 18 ROC curves with AUC values shown for leave-one-out temporal validation. The legend shows label of the inventory, with its AUC, used as validation dataset.

6.3.2. Spatial validation

In analogy to the temporal validation case, the same testing procedure can be imagined across the geographic space. Instead of considering each time replicate for validation, one can subdivide the landscape into spatial subsets, calibrating on a single or cluster of them and validation on the complementary subset. For this study, the spatial distinction is made by dividing the study area into a lattice, and selecting spatial subsets according to each grid element. This implies that every time a spatial subset is extracted so are the temporal replicates for that specific grid block. Notably, the generated lattice included 19 grid-cells. Though the lattice itself is made of equal-sized gird blocks, the underlying study area did not homogeneously cover each grid element of the lattice. As a result, some grid cells contained more SU than others. In addition to this, landslides also did not cover the whole study area homogeneously. In turn, this meant that some grid cells had more presence data than others and at times no landslide presence data was contained within a grid cell at all. For this reason, aggregating more than one lattice element was required, and the choice of the best aggregation between adjacent grids has also been explored here.

Figure 19 visualizes the lattice overlayed on the study area along with the distribution of landslide points inside. Table 7 shows the block number(s), as referred in Figure 19, that were used for validation along with their corresponding AUC. The highlighted rows are displaying color-coded model performance AUC, where red means an unacceptable value (AUC < 0.7), yellow indicates a good performance (0.7 < AUC < 0.8), blue dictates extremely good performance (0.8 < AUC < 0.9) and green is reserved for excellent performances (AUC > 0.9). Only a few blocks show AUC values below 0.7, which include block 12 and 18. The blocks showing excellent performance are mainly concentrating on the Northeast side of the study area (3,4,5,9,10) but also isolated blocks like block 16 gives an exceptional performance.

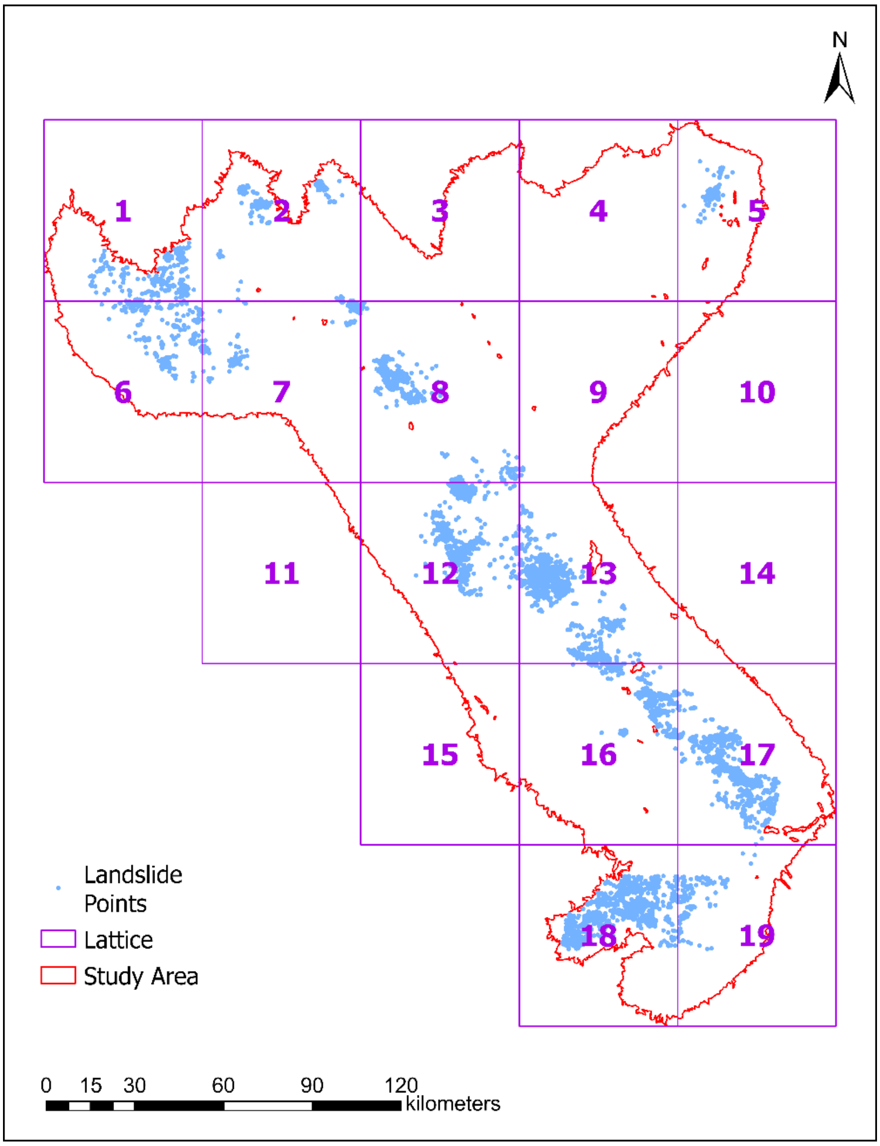

Figure 19 Lattice visualised over the study area for spatial validation technique.

Table 7 AUC values shown for ROC curves extracted using block(s) spatial validation. Red: low AUC, Yellow: acceptable AUC, Blue: good AUC, Green: excellent AUC.

Validating Block	AUC	Validating Block	AUC
1	0.7675	13	0.7565
2	0.8174	13+14	0.7693
3	0.9204	15+16	0.9291
4+5	0.9002	16	0.9195
5	0.9286	17	0.869
6	0.7698	18	0.6236
7	0.7624	19	0.8316
8	0.8356	18+19	0.6406
9	0.9313	18+16	0.7858
9+10	0.9589	3+4+9+10+11+14+15	0.9084
11+12	0.7292	16+18+19	0.7351
12	0.6885	12+8+7	0.7638

6.3.3. 10-fold random cross-validation

A third validation routine integrated a 10-fold random cross validation in space and time. This means that the whole spatio-temporal domain has been divided into ten random subsets, each representing 10% of the whole data. Then, the calibration relied on nine subsets at a time whereas the validation performance was measured on the remaining data (one subsequent subset).

Figure 20 displays ten ROC curves, one for each random 10-fold cross-validation, showing similar performances. The relative range of the AUC values is shown in Figure 21. Here it can be seen that the minimum and maximum AUC values do not exhibit a large variation, both of them being representative of near-excellent predictive performance.

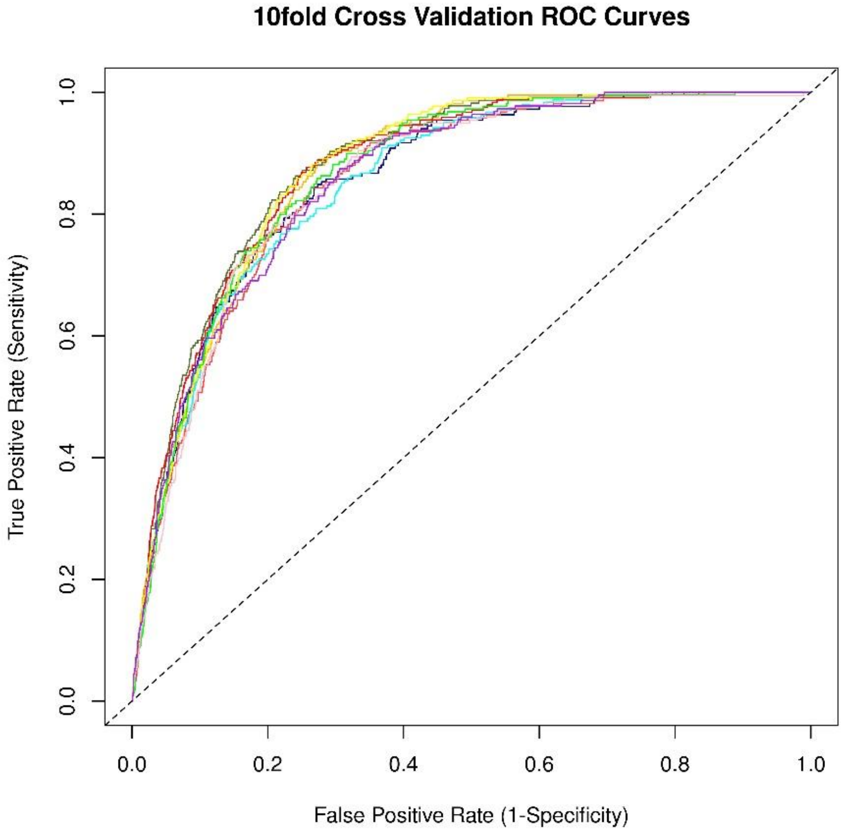

Figure 20 ROC curves for 10-fold cross validation via random sampling.

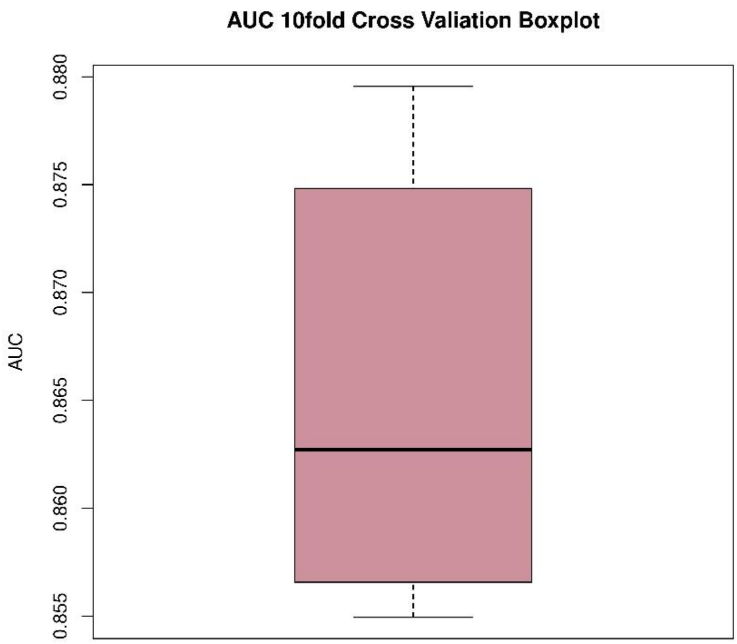

Figure 21 Boxplot visualization for the range of AUC values (derived from Figure 20) for 10-fold random cross validation.

6.3.4. External validation

Taking advantage of multiple mapped sites in the Lower Mekong Region, the model was tested over an external dataset. The area of the Vietnam as a country covers a large latitude range across which the landscape significantly varies. As a result, the available landslide inventories also vary across space, occupying most of the North Vietnam and being locally present in fewer specific locations to the West and South of the LMR. Specifically, five external (i.e., not falling within the study site) testing sites were selected (Figure 22, out of which 3 fell within the South of Vietnam and 2 sites fell in the neighbouring country of Laos. Table 8 holds the information on the events for which the landslides were mapped in the external sites. The aim behind the external validation is to test the applicability of the model in nearby geographical areas for the possibility of extending the concept of this model outside the setting of the study area. This procedure is commonly referred to model transferability (Cama et al., 2017; Loche et al., 2022).

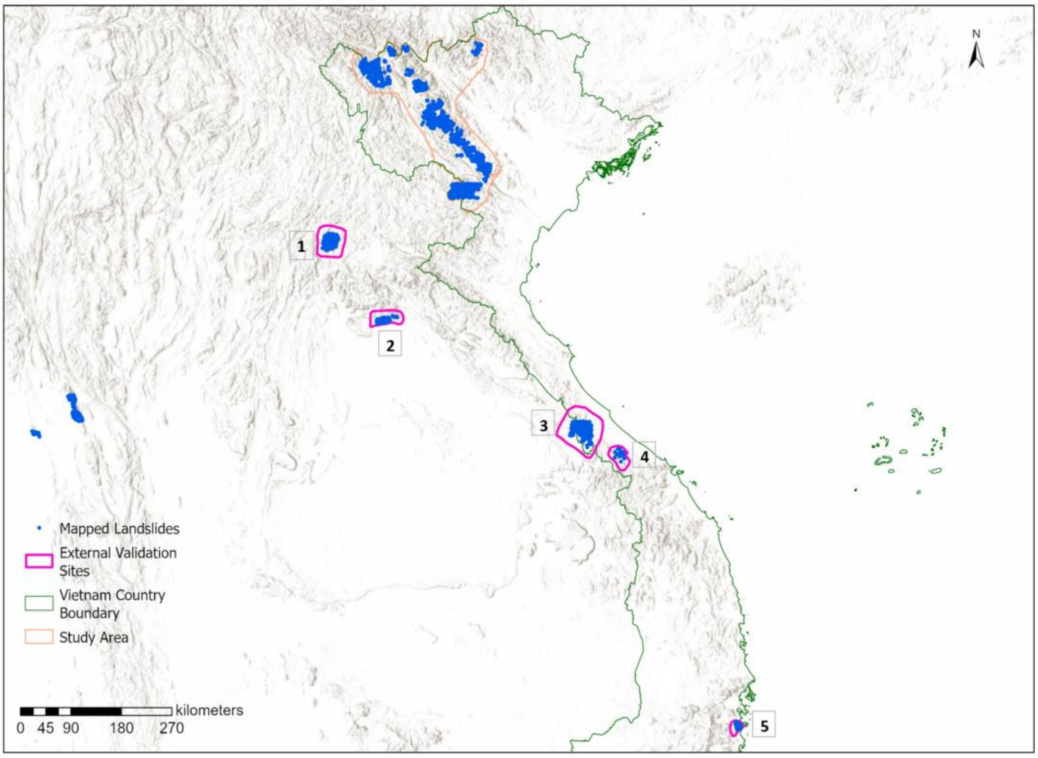

Figure 22 Sites selected for external validation.

Table 8 Supporting table for landslide information on the chosen sites for external validation.

Validation site	Country	Event date
1	Laos	30th August 2018
2	Laos	11th September 2015
3	Vietnam	18th October 2020
4	Vietnam	12th October 2020
5	Vietnam	18th November 2018

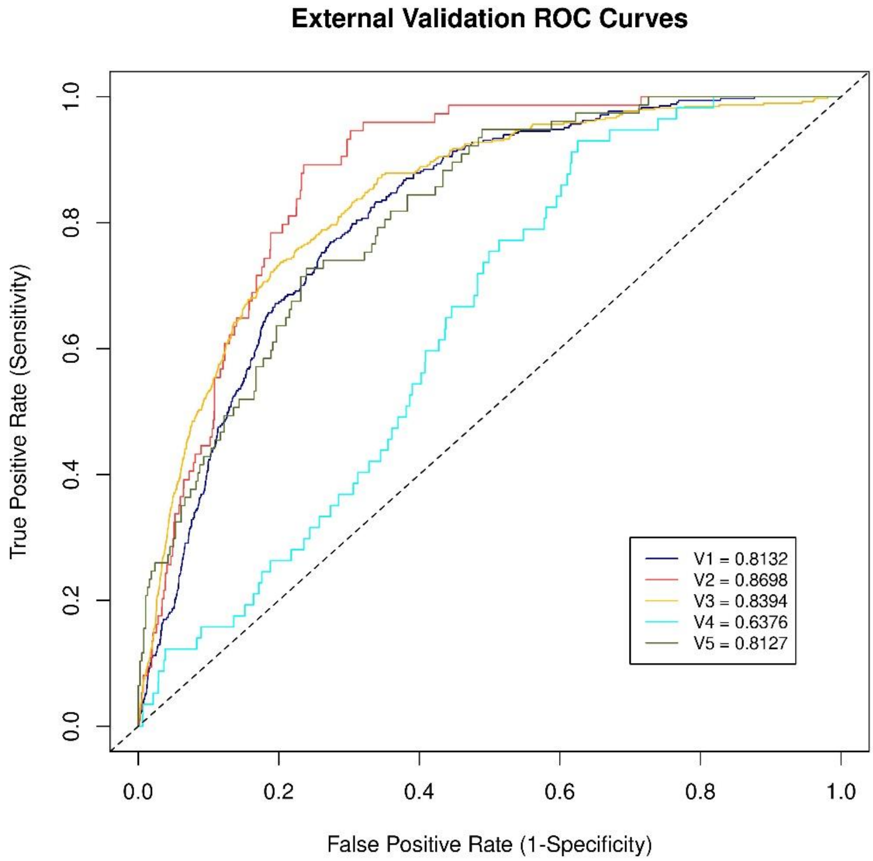

Figure 23 ROC curves with AUC values labelled with external site chosen for validation.

Figure 23 reports the ROC curves and associated AUCs obtained from the external validation. Four of the validations showed predictive performances well in line with those measured at the test site. However, only site 4 returned poor performance with an AUC of 0.6376.

6.4. Probability thresholds

In case of the leave-one-out temporal validation scheme, the retrieved probabilistic estimates were compared with the observed information by setting different quantile thresholds to assess the TNR, TPR and F1 score. The probability thresholds were set at every 0.05th quantile (5th percentile). The threshold here defines the point at which the corresponding probabilistic estimate would indicate the division into a binary signal. F1 scores, specificity and sensitivity are displayed in Figure 24, with a panelled structure according to the landslide inventory under consideration.

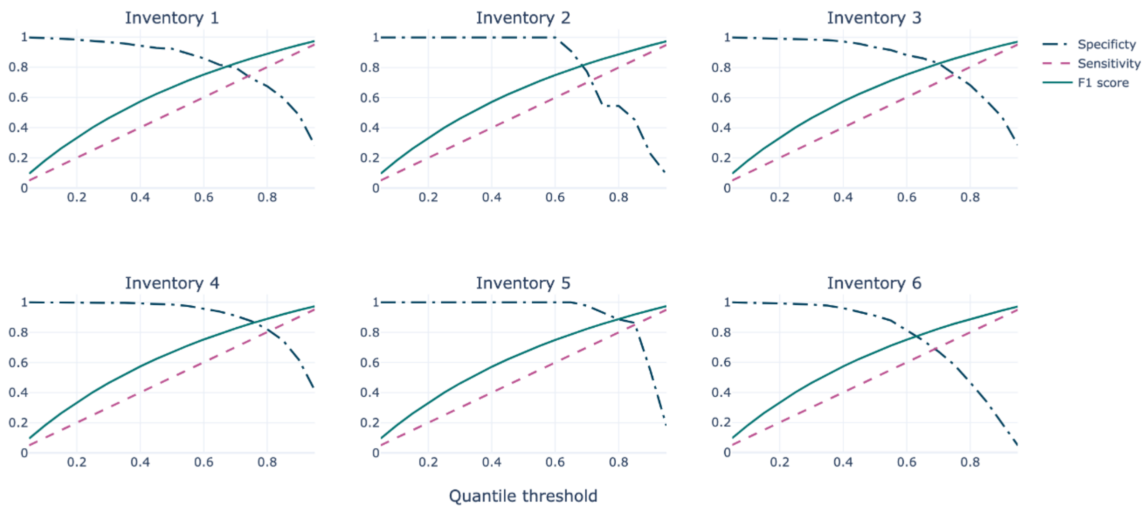

Figure 24 Plots for each of the six inventories displaying quantile threshold on the x-axis and range of specificity, sensitivity and F1 score on the y-axis. Notably, the each panel is labelled with the inventory used for validation (excluded from the training phase).

This information shows the decline of specificity with the increasing F1 score in all of the six inventories used for this method. This means that the TNR is compromised by choosing a high F1 score, with the intersection of TNR and TPR being nearly at a similar value for all six scenarios.

7. DISCUSSION

7.1. Model selection

Objective 1: To identify a suitable antecedent rainfall window as a covariate.

Research Question 1: What impacts the selection of a suitable antecedent rainfall window for a dynamic model?

For Northern Vietnam, the most suitable antecedent window for cumulative rainfall estimates is selected to be 8 days. This window selection opens up two important considerations. One is that in a dataset with an uncertainty in the landslide date as wide as 6 days (Table 4), any small window would have questioned the reliability of such timespan. In addition to this, the same measure LHASA relies on is of 7 days antecedent to the landslide occurrence, a timespan comparable to the one selected in this work.

For the produced landslide inventory for the Lower Mekong Region, satellite imagery was used to trace back events, where cloud cover is an obstacle for the usability of such images. The inventories were sometimes accurately marked with a single date but some of them, also which were used in this research study, had a date error of up to six days relating to the spatial event.

It is worth noting that the lowest WAIC is associated with one and two antecedent days. However, given the limited accuracy of the inventory dates, this could not be used because of considerations related to a larger dating error as compared to the one-day and two-day windows mentioned above. For these reasons, the next best WAIC was chosen (8-days). Moreover, Table 5 shows 8-days cumulative rainfall to be the most reasonable choice when combined with the support of choice brought by AUC (highest average AUC). It includes the 6-day date error introduced by some inventories, and behaves as a proxy to the rest of the models lying in the first half of all models (maximum antecedent window being 9-days for the first 7 models). This may indicate that the best timespan to aggregate rainfall falls in this range and that after the ninth day the model quality could relatively be lower given the inventory dates are compromised.

7.2. Generalized Additive Model

Transitioning from rainfall thresholds to integrating the rainfall signal into a dynamic susceptibility model provided interesting results. The interpretability was kept high as per any statistical models, the performance classified as suitable in most of the validation routines and uncertainty propagation was respected as compared to a traditional situation where susceptibility and rainfall thresholds are part of different modelling protocols.

Research Question 2: What is the behaviour of a model built as a dynamic rainfall-induced prediction system which does not rely on rainfall thresholds but uses rainfall estimates as a covariate in a statistical model?

The contextual use of terrain properties and rainfall characteristics suitably explained the distribution of landslide occurrences in space and time. In the course of this work, most interpretation efforts have been given to few covariates out of the large set used in the model. This was a choice to keep the description concise and because out of all the covariates slope steepness (Hong et al., 2007; Hung et al., 2016), rainfall (Segoni et al., 2018) and vegetation cover are indeed the most common ones used in the literature.

Linear covariates

The rainfall signal used for this model holds a strong positive relation with landslide occurrences (Figure 9). The same is valid for mean Northness, a continuous property indicating slope exposition towards the North for values equal to 1 and progressively indicating exposition to the South as values decrease up to -1. The indication of a strong positive influence of Northness could be interpreted in terms of sunlight exposure and thus in terms of soil moisture conditions. The Northern hemisphere to which the study area belongs exposes for a longer time South-facing slopes to the sunlight. Therefore South-facing slopes may be drier whereas the North facing one may be wetter and therefore starting with a greater moisture when the incoming clouds would discharge the rainfall. However, this is one of the possible interpretations of the Northness, as the aspect in general, is known to potentially reflect other site conditions. For instance, geological beddings, or the relation between strata direction and dipping with respect to the slope direction , thus being intrinsically linked to the aspect.

The mean of Enhanced Vegetation Index (EVI) is shown (Figure 9) to behave negatively. In turn, this indicates that the less dense the vegetation is, the more susceptible a given slope would be. A study in the upper Lo River catchment (Northern Vietnam) shows that more than 50% of the observed landslides are distributed in areas with no or new vegetation cover such as younger forests (Hung et al., 2016). Similar to the Northness case though, no single and clear interpretation of the EVI can be made based on this without in-situ information. In fact, dense or scarce vegetation is known to influence slope stability depending on the root-structure plants build underground. Some plants develop roots which drill straight through the soil column and into the bedrock, thus potentially adding root strength at the base of a theoretical sliding surface. Conversely, roots that radially spread out into the soil column may add weight to an already potentially unstable mass (Forbes and Broadhead, 2013; Hung et al., 2016).

The mean profile curvature also contributed to the decrease the expected susceptibility (on a marginal basis). This means the convex slopes would be more likely to host a failure. This may be attributed to the ability of concave slope on a vertical direction to retain more incoming water from locations above. Therefore, concave slopes would tend to retain moisture while convex slopes would contribute to wash off any overland flows, possibly holding less moisture as well as mobilizing unconsolidated materials (Daniel et al., 2021) making them more prone to landslides.

Nonlinear Covariates

The slope angle is one of the most essential variables to explain distribution of landslides (Hung et al., 2016). For this reason and for reasons of conciseness, only the random effect estimated for the slope steepness will be interpreted here.

The slope steepness starts with a negative sign and rapidly increases its contribution up to approximately 22 degrees, after which steepness, the contribution transitions to a positive sign as shown in Figure 10. Interestingly, 22 degrees is a critical steepness value reported in a number of empirical experiments for debris flows landslides (Iverson, 1997). And, being the landslide inventory used in this work made of impulsive event, it is reasonable to assume that debris flows may contribute to a significant part of the failures NASA mapped for the LMR (Amatya et al., 2022). After the 22 degree threshold, the slope steepness contribution keeps on increasing almost linearly up to 35 degrees (Figure 10), after which the regression coefficients estimated per each steepness class essentially reach a plateau.

Aside from the mean slope steepness behaviour described above, the uncertainty around it is also an interesting element to report and interpret. In fact, only the central portion of the plot exhibits a very narrow credible interval, whereas low and high steepness values are associated with a larger uncertainty. This may indicate that the model is less capable of discriminating slope steepness effects along the two tails of the distribution, where likely some exception exists and data may also be less representative of these classes. This is also the case for mean relief (Figure 11), however, maximum distance within a slope unit only expresses a narrow CI at smaller relief values as a possible influence of less representative data.

The range of values among different covariates can vary for different topographic information chosen to calibrate the initial model, thus it has to be kept under consideration that a localized study area may not reduce the uncertainty interval as a larger data sample possibly could.

<u>Research Question 3: How efficiently can the model predict landslide probabilities if short-term rainfall estimates are plugged in the statistical model to visualize changing susceptibility with changing precipitation?</u>

In the case of sequential temporal validation, the AUC of validating inventory 3 is oddly low as compared to other time-steps as shown in Table 6. This could be explained by the quality of inventory 2, given that it holds the lowest number of landslide points (only 99) among the other events listed in Table 4 along with, an uncertainty of the event date up to 6 days. Notably, the 99 landslide points mentioned above do not imply that 99 slope units are labelled with a presence condition (22 SUs were given a 'presence' status after aggregation for inventory 2).

Overall, predicting inventory 3 by using inventory 2 (along with inventory 1) as calibration information may have caused the low performance retrieved for this specific case. Since the quality of the inventory as input data affects the resulting model (Gaidzik and Ramírez-Herrera, 2021), it may also reflect in its prediction capabilities. In fact, landslides are much more numerous in inventory 3, as compared to inventory 1 and 2 combined, and inventory 3 has a much more precise date in comparison to inventory 2 (Table 4).

Similarly, the results for leave-one-out temporal validation method (shown in Figure 18) display relatively lower AUC values for inventory 2 and 6. The AUC values are still in the acceptable performance range but compared to their counterparts, the date error for these two inventories may be responsible for a lower performance overall. Hence, the accuracy and quality of inventories is observed to directly affect the predictive ability of the model and its potential to be extended and used over an unknown dataset or even in near-real time. Looking at the AUC values from validating on inventory 4 and 5, the performance is exceptional as the AUC reached close to 0.9 (Figure 18). This may indicate that the model's predictive power remains high as the model is fed with rich data, being able to define the relationship of terrain attributes along with the rainfall one effectively, since these inventories have low or no date error (Table 4). However, the AUC shifts to a lower value (still acceptable though) for validating inventory 1 in analogy, again showing that the validation dataset used in this step is associated with the lowest date accuracy of 6 days, recalling Table 4. Overall the performance for every validation cycle was classified between acceptable and even excellent.

Spatial cross validation using a lattice, in Figure 19, displayed an overall good performance. However, there were some validation blocks which show AUC value below 0.7 (namely block 12 and 18) in Table 7. Grid-cell 18, with AUC value closer to 0.6 than 0.7, has a dense landslide presence with little corresponding area of no landslides. Moreover, the area covered by the grid-cell is quite low as compared to blocks 13 and 17, which also have many landslide points but more area covered in their respective blocks (Figure 19). Therefore, it may be due to the unbalanced data in block 18 showing more presences than absences, contrary to the calibrated model as a whole, that the model specifically does not perform well in this gird cell's validation routine. Nonetheless, the model was considered to be performing well as per AUC values obtained from the random 10-fold cross validation routine displayed in Figure 20 and 21.

The prediction of the model that was tested through multiple validation routines and a good predictive power was noticed even when testing over unknown data. External validation mainly sets forth a good performance with respect to its AUC values in Figure 22 but an oddly low AUC value among others is realized in site 4 (see, Table 8 for information on this external site). However, it is to be clarified here that the reference model was built in relation to the terrain properties of a relatively localized study site which has an impact on rainfall patterns in relation to its orography. The orographic effect can be localized and thus the relationship defined in this model may not be applicable for estimating in a more complex or different geographic setting (Adler et al., 2003). The terrain and rainfall patterns differ even within Vietnam, due to the changes in orography and monsoon seasons (Nguyen et al., 2014), hence, the model may not always be replicable in a broader spatial context. This, along with the ability of satellite-based rainfall estimates to resolve rainfall patterns in relation to the orographic effects (Kirschbaum et al., 2012), could possibly explain the source of low AUC for the external validation in site 4.

<u>Research Question 4: What is the added value of a model that features rainfall estimates?</u>

Current applications aimed at predicting landslide occurrences are essentially split into two types. The first one depicts the static information of locations prone to landslides independently of the given trigger. The second one explores rainfall data and the intensity-duration relation that exist between rainfall and landslide distributions. The drawback of this system is to consider these two elements separately. The added value of this work primarily relies in demonstrating that the rainfall can be integrated into the probabilistic analyses at the core of any susceptibility model. This in turn also makes sure that the uncertainty of these two elements correctly propagates into the dynamic probability estimates.

The most important achievement is therefore expressed by moving away from decoupled rainfall threshold and susceptibility models and then moving towards a unified assessment where alert levels can be provided on the basis of probability thresholds. The thresholds can then be described as ones obtained from a unified model which does not decouple rainfall thresholds from terrain susceptibility; and provides a fuller representation of landslide probabilistic estimates.

<u>Research Question 5: What is the added value of a model which accounts for the uncertainty estimation?</u>

The inclusion of mean behaviour alongside with its uncertainty levels offers multiple opportunities both in terms of interpretation and implementation of further analytical steps. In fact, reporting credible bands for the probabilistic estimates of landslide occurrences improves the understanding of how reliable a prediction is. For instance, a highly susceptible but also highly uncertain slope unit may be a much worse target for stabilization investments as compared to a highly susceptible and less uncertain slope unit.

As for the uncertainty estimated for the regression coefficients of each covariate in the model, this is extremely important to enable statistical simulations. In this work, the simulation phase in time has been limited to the use of mean regression coefficients. However, an even better simulation routine can be envisioned by solving thousands of randomly generated regression coefficients within the range defined in the previous uncertainty estimation phase. This routine can then produce statistical summaries out of those thousand results which would in turn be more detailed and accurate.

7.3. Threshold optimization for probabilistic estimates

Objective 3: To transition from separate rainfall thresholds to unified probabilistic thresholds for alert levels.

<u>Research Question 6: How can a suitable/optimal probabilistic cut-off be defined to separate alert levels for informed decision making?</u>

Moving away from rainfall thresholds and towards probability thresholds requires defining an optimal cut-off after which any slope unit would be reliably considered susceptible. The terms optimal and reliable here are quite complex though. In fact, multiple choices can be made for a threshold that produces, for instance, two alert levels, 'No-Warning' and 'Warning'. In fact, the choice of a suitable threshold value can influence the correct prediction of susceptible and non-susceptible locations (TP and TN, respectively) or mistakenly depicting them (FP and FN). As a result, it is important to reflect on what element a cut-off probability choice this study may want to minimize or focus on the holistic performance of the model. For instance, it may be of interest to minimize the number of FP as they have the effect of developing mistrust in the user with respect to the early-warning system. Conversely, it is possible to accept the FP to some degree because they provide a conservative classification, and it may be more suitable to minimize instead the number of FN, which may be case because such errors are usually associated to heavy losses.

For this reason, here the F1 score was selected as this is a type of performance metric that works very well in classification problems as it accounts for the four elements (model hits and misses) mentioned above. To achieve a balance between the TPR and TNR, both essential in this study for the aforementioned reasons, a combination of information from all six inventories is shown in Figure 25.

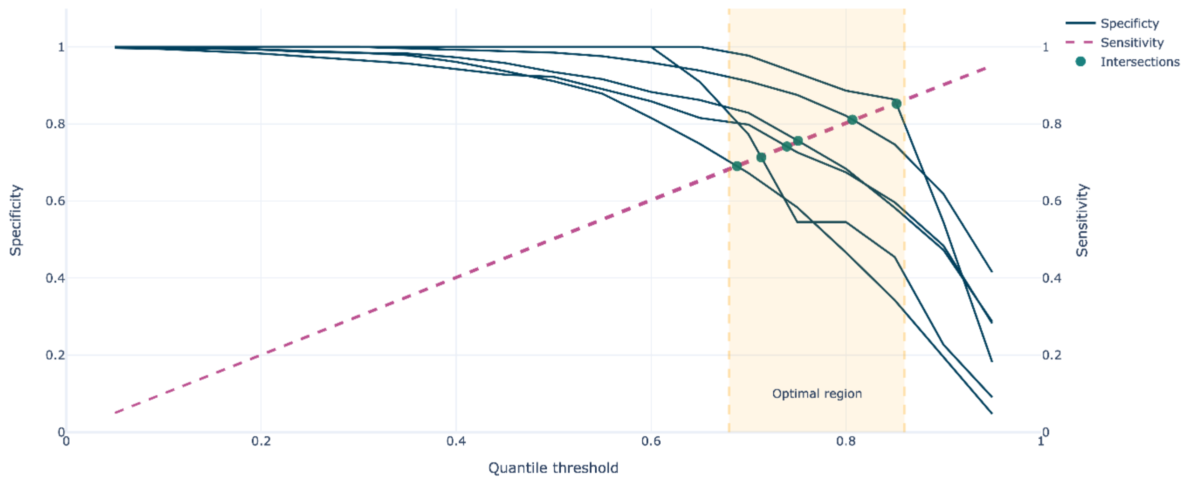

Figure 25 Combination of six inventories used at every 0.05th quantile for selecting a threshold to define probabilistic cut-off.

There, the trade-off between TNR and TPR is preferred to be minimized since the trend displayed in Figure 24 (and tables in Annex 5) show TNR declining as TPR and F1 score increases. Figure 25 visualises the intersection of specificity (TNR) and sensitivity (TPR) values in the optimal quantile threshold section, where both values are near or above 0.7 and correspond to a quantile threshold that produces an F1 score above 0.8. The range of the optimal threshold is roughly between 0.7-0.85th quantile of the probability estimates. The values at these quantile thresholds can then be combined, and an average of the predicted probability can be considered as the optimal cut-off point. External validation outputs could also be used for the testing of probability cut-offs, but since the model at this stage was predominantly used with the same study area without any transferability purpose, the local threshold is preferred to a more general one.

7.4. Web-based vizualisation tool

Objective 4: To translate the model into an interactive visualization tool via a cloud-platform for comprehensive display (to be possibly extended into a forecasting tool).

Research Question 7: What is the capability of the model to provide landslide warning signals in the form of a local alarm system?

The corresponding probabilities of the optimal quantile threshold (as discussed in section 7.3) were used to set a cut-off point for identifying slopes with a warning signal. Since the original calibrated model used six different inventories, the probabilities obtained were also of the same number, thus an average of those probabilities was used to set the probability cut-off, this being 0.0057. Using the GEE platform, with its features of creating a visual demonstration of the model through an interactive application, a close to real-time possibility is visualized. Since CHIRPS is used as the rainfall input for this model, it does not provide estimates in real-time in GEE. However, it currently has the best spatial resolution and it is also equipped with bias-correction using ground-station data. In the future, there is certainly a possibility of using a different rainfall input rather than CHIRPS, although the spatial resolution and accuracy may differ with different rainfall products. As a result, the model will need to be re-evaluated.

Currently, the platform built by integrating the mean regression coefficients obtained via the reference GAM, displays mean susceptibility with the option of viewing it with the date of choice. The date selection initiates data to be retrieved for the respective dynamic covariates, rainfall and EVI (m and sd), which are combined to the static ones to produce probability values for each SU. The date selection relies heavily on the availability of CHIRPS information thus, the choice can only be made up to approximately a month before the current date. The output is dynamic in nature and takes some time to load due to a heavy computations over a large dataset. A screen capture is shown in Figure 26, highlighting the interface of the application. This opens a realm of possibilities to visualise a landslide susceptibility model, integrational of rainfall, over a given region and even with its uncertainty levels. The next development may explore other rainfall radar measurement instead. For instance GPM not only provides data for the past but it also offers a 3-hourly rainfall forecast. As a result, this could move beyond rainfall forecast product and be extended to generate landslide forecast services.

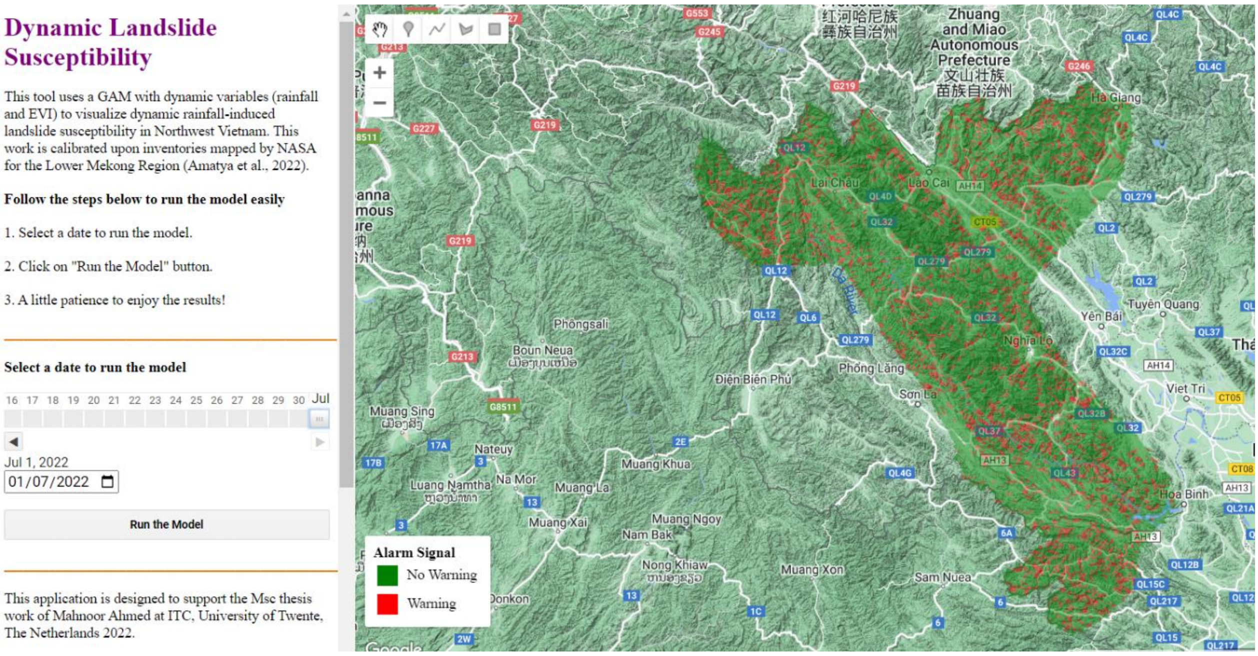

Figure 26 Interface of a possible forecasting tool displaying dynamic mean susceptibility for the study area.

The link to access the app personalised for this study is provided below:
Application for Dynamic Landslide Susceptibility for Northwest Vietnam.

8. CONCLUSION AND RECOMMENDATIONS

The conclusive remarks are provided below as per the author's perspective of this study, combined with limitations and future recommendations.

There were some limitations of this work which have been briefly highlighted in the sections above, thus this section will also discuss possible recommendations. A significant limitation observed in data acquisition as well as validation routines was the inventory quality. The lack of available, or freely available complete inventories are still rare for such studies. Moreover, the inventory was constructed on point data, rather than polygons, limiting point data's ability to inform the model of the area/size of the landslide connecting SU shape indices with possible failures.

The method mainly focuses on integrating rainfall in a modelling pipeline where landslide occurrence probabilities are dynamically estimated each day over a test site in Vietnam. There are several other influencing factors which were not considered due to the nature of the work but still are contributing factors towards landslide susceptibility. Such factors include; distance to roads, land use and land cover which can also be a dynamic influence, and fault lines as a factor of the geomorphological characteristics. These may also be explored in the future although the main novelty the research setup tested here would remain unchanged. Landslide probabilities can be produced in a dynamic manner through proper space-time statistical modelling. Moreover, current cloud computing environments such as Google Earth Engine are possible platforms where the analytical framework mentioned above can be transformed into an operational tool.

Surely, the use of CHIRPS limits the conversion of this model in a forecasting sense, however, as space-time applications in landslide science are at their infancy stage, much can still be improved. Still in the context of cloud computing, one could extend the model tested here by integrating GPM measurements instead of CHIRPS for calibration and then use the 3-hourly GPM forecast to forecast landslide occurrences. Such a development is around the corner although considerations will need to be given in terms of data spatial resolution and accuracy. In addition, a recent release of forecasted estimates for CHIRPS supported by precipitation forecasts of National Centers for Environmental Prediction (NCEP) Global Ensemble Forecast System version 12 (GEFS v12). This advancement has been made in particular for early warning systems with the CHIRPS forecasted data available for 15 days ahead with the similar high spatial resolution (Harrison et al., 2022). However, this dataset is not yet available on GEE Catalog since its release in June 2022, but it can be expected to be made accessible via GEE in the near future.

Notably, high computation power is required for the processing in the back-hand of the web-app in reasonable time. However, this thesis work relied on a free Google Earth Engine account and much more could be done extending the license to a professional one instead.

Further developments can be made towards uncertainty estimation. In this research, this aspect has not been explored due to time constraints. However, the vision for the future could welcome the full use of the Bayesian framework the model was based on. As a result, it is possible to simulate hundreds or thousands of susceptibility maps sampling from the regression coefficient distributions and render a new map next to the mean shown in this study. This map will illustrate the expected uncertainty around the mean and therefore express the reliability of the mean estimates. This would be particularly important for authorities and decision makers implementing slope stability measures.

LIST OF REFERENCES

Adler, R.F., Huffman, G.J., Chang, A., Ferraro, R., Xie, P.P., Janowiak, J., Rudolf, B., Schneider, U., Curtis, S., Bolvin, D., Gruber, A., Susskind, J., Arkin, P., Nelkin, E., 2003. The version-2 global precipitation climatology project (GPCP) monthly precipitation analysis (1979-present). J. Hydrometeorol. 4, 1147–1167. https://doi.org/10.1175/1525-7541(2003)004<1147:TVGPCP>2.0.CO;2

Alvioli, M., Marchesini, I., Reichenbach, P., Rossi, M., Ardizzone, F., Fiorucci, F., Guzzetti, F., 2016. Automatic delineation of geomorphological slope units with r.slopeunits v1.0 and their optimization for landslide susceptibility modeling. Geosci. Model Dev. 9, 3975–3991. https://doi.org/10.5194/GMD-9-3975-2016

Amatya, P., Kirschbaum, D., Stanley, T., 2022. Rainfall-induced landslide inventories for Lower Mekong based on Planet imagery and a semi-automatic mapping method. Geosci. Data J. 00, 1–13. https://doi.org/10.1002/gdj3.145

Ba, Q., Chen, Y., Deng, S., Yang, J., Li, H., 2018. A comparison of slope units and grid cells as mapping units for landslide susceptibility assessment. Earth Sci. Informatics 11, 373–388. https://doi.org/10.1007/S12145-018-0335-9/FIGURES/9

Bakka, H., Rue, H., Fuglstad, G.A., Riebler, A., Bolin, D., Illian, J., Krainski, E., Simpson, D., Lindgren, F., 2018. Spatial modeling with R-INLA: A review. Wiley Interdiscip. Rev. Comput. Stat. 10:e1443. https://doi.org/10.1002/WICS.1443

Bangalore, M., Smith, A., Veldkamp, T., 2019. Exposure to Floods, Climate Change, and Poverty in Vietnam. Econ. Disasters Clim. Chang. 3, 79–99. https://doi.org/10.1007/S41885-018-0035-4

Brenning, A., 2008. Statistical Geocomputing combining R and SAGA: The Example of Landslide susceptibility Analysis with generalized additive Models, in: SAGA - Seconds Out. Hamburger Beiträge zur Physischen Geographie und Landschaftsökologie, pp. 23–32.

Brock, J., Schratz, P., Petschko, H., Muenchow, J., Micu, M., Brenning, A., 2020. The performance of landslide susceptibility models critically depends on the quality of digital elevation models. Geomatics, Nat. Hazards Risk 11, 1075–1092. https://doi.org/10.1080/19475705.2020.1776403

Cama, M., Lombardo, L., Conoscenti, C., Rotigliano, E., 2017. Improving transferability strategies for debris flow susceptibility assessment: Application to the Saponara and Itala catchments (Messina, Italy). Geomorphology 288, 52–65. https://doi.org/10.1016/J.GEOMORPH.2017.03.025

Campbell, R.H., 1975. Soil Slips, Debris Flows, and Rainstorms in the Santa Monica Mountains and Vicinity, Southern California. U.S. Geol. Surv. Prof. Pap. 851 851, 51 pages.

Carrara, A., Cardinali, M., Detti, R., Guzzetti, F., Pasqui, V., Reichenbach, P., 1991. GIS techniques and statistical models in evaluating landslide hazard. Earth Surf. Process. Landforms 16, 427–445. https://doi.org/10.1002/ESP.3290160505

Chung, C.J., Fabbri, A.G., 2008. Predicting landslides for risk analysis - Spatial models tested by a cross-validation technique. Geomorphology 94, 438–452. https://doi.org/10.1016/j.geomorph.2006.12.036

Chung, C.J.F., Fabbri, A.G., 2003. Validation of Spatial Prediction Models for Landslide Hazard Mapping. Nat. Hazards 30, 451–472. https://doi.org/10.1023/B:NHAZ.0000007172.62651.2B

Daniel, M.T., Ng, T.F., Abdul Kadir, M.F., Pereira, J.J., 2021. Landslide Susceptibility Modeling Using a Hybrid Bivariate Statistical and Expert Consultation Approach in Canada Hill, Sarawak, Malaysia. Front. Earth Sci. 9. https://doi.org/10.3389/FEART.2021.616225/BIBTEX

Didan, K., 2015. MYD13Q1 MODIS/Aqua Vegetation Indices 16-Day L3 Global 250m SIN Grid V006 [Data set] [WWW Document]. NASA EOSDIS L. Process. DAAC. https://doi.org/10.5067/MODIS/MYD13Q1.006

Epifânio, B., Zêzere, J.L., Neves, M., 2014. Susceptibility assessment to different types of landslides in the

coastal cliffs of Lourinhã (Central Portugal). J. Sea Res. 93, 150–159. https://doi.org/10.1016/j.seares.2014.04.006
Farr, T.G., Rosen, P.A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L., Seal, D., Shaffer, S., Shimada, J., Umland, J., Werner, M., Oskin, M., Burbank, D., Alsdorf, D.E., 2007. The shuttle radar topography mission. Rev. Geophys. 45. https://doi.org/10.1029/2005RG000183
Forbes, K., Broadhead, J., 2013. The role of trees and forests in the prevention of landslides and rehabilitation of landslide-affected areas in Asia, Food and Agricultural Organization of the United Nation. Bangkok.
Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A., Michaelsen, J., 2015. The climate hazards infrared precipitation with stations-a new environmental record for monitoring extremes. Sci. Data 2. https://doi.org/10.1038/sdata.2015.66
Gaidzik, K., Ramírez-Herrera, M.T., 2021. The importance of input data on landslide susceptibility mapping. Sci. Reports 11. https://doi.org/10.1038/s41598-021-98830-y
Gariano, S.L., Rianna, G., Petrucci, O., Guzzetti, F., 2017. Assessing future changes in the occurrence of rainfall-induced landslides at a regional scale. Sci. Total Environ. 596–597, 417–426. https://doi.org/10.1016/j.scitotenv.2017.03.103
Gian, Q.A., Tran, D.T., Nguyen, D.C., Nhu, V.H., Tien Bui, D., 2017. Design and implementation of site-specific rainfall-induced landslide early warning and monitoring system: a case study at Nam Dan landslide (Vietnam). Geomatics, Nat. Hazards Risk 8, 1978–1996. https://doi.org/10.1080/19475705.2017.1401561
Goetz, J.N., Brenning, A., Petschko, H., Leopold, P., 2015. Evaluating machine learning and statistical prediction techniques for landslide susceptibility modeling. Comput. Geosci. 81, 1–11. https://doi.org/10.1016/J.CAGEO.2015.04.007
Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sens. Environ. 202, 18–27. https://doi.org/10.1016/J.RSE.2017.06.031
Görüm, T., 2019. Tectonic, topographic and rock-type influences on large landslides at the northern margin of the Anatolian Plateau. Landslides 16, 333–346. https://doi.org/10.1007/S10346-018-1097-7
Goutte, C., Gaussier, E., 2005. A Probabilistic Interpretation of Precision, Recall and F-Score, with Implication for Evaluation. Lect. Notes Comput. Sci. 3408, 345–359. https://doi.org/10.1007/978-3-540-31865-1_25
Guzzetti, F., Gariano, S.L., Peruccacci, S., Brunetti, M.T., Marchesini, I., Rossi, M., Melillo, M., 2020. Geographical landslide early warning systems. Earth-Science Rev. 200. https://doi.org/10.1016/J.EARSCIREV.2019.102973
Guzzetti, F., Peruccacci, S., Rossi, M., Stark, C.P., 2008. The rainfall intensity-duration control of shallow landslides and debris flows: An update. Landslides 5, 3–17. https://doi.org/10.1007/s10346-007-0112-1
Guzzetti, F., Peruccacci, S., Rossi, M., Stark, C.P., 2007. Rainfall thresholds for the initiation of landslides in central and southern Europe. Meteorol. Atmos. Phys. 98, 239–267. https://doi.org/10.1007/s00703-007-0262-7
Ha, N.D., Sayama, T., Sassa, K., Takara, K., Uzuoka, R., Dang, K., Van Pham, T., 2020. A coupled hydrological-geotechnical framework for forecasting shallow landslide hazard—a case study in Halong City, Vietnam. Landslides 17, 1619–1634. https://doi.org/10.1007/S10346-020-01385-8/FIGURES/15
Harp, E.L., Keefer, D.K., Sato, H.P., Yagi, H., 2011. Landslide inventories: The essential part of seismic

landslide hazard analyses. Eng. Geol. 122, 9–21. https://doi.org/10.1016/j.enggeo.2010.06.013
Harrison, L., Landsfeld, M., Husak, G., Davenport, F., Shukla, S., Turner, W., Peterson, P., Funk, C., 2022. Advancing early warning capabilities with CHIRPS-compatible NCEP GEFS precipitation forecasts. Sci. Data 9, 1–13. https://doi.org/10.1038/s41597-022-01468-2
Heerdegen, R.G., Beran, M.A., 1982. Quantifying source areas through land surface curvature and shape. J. Hydrol. 57, 359–373. https://doi.org/10.1016/0022-1694(82)90155-X
Hidayat, R., Sutanto, S.J., Hidayah, A., Ridwan, B., Mulyana, A., 2019. Development of a landslide early warning system in Indonesia. Geosci. 9. https://doi.org/10.3390/GEOSCIENCES9100451
Hong, Y., Adler, R., Huffman, G., 2007. Use of satellite remote sensing data in the mapping of global landslide susceptibility. Nat. Hazards 43, 245–256. https://doi.org/10.1007/s11069-006-9104-z
Hong, Y., Adler, R., Huffman, G., 2006. Evaluation of the potential of NASA multi-satellite precipitation analysis in global landslide hazard assessment. Geophys. Res. Lett. 33. https://doi.org/10.1029/2006GL028010
Hong, Y., Adler, R.F., 2007. Towards an early-warning system for global landslides triggered by rainfall and earthquake. Int. J. Remote Sens. 28, 3713–3719. https://doi.org/10.1080/01431160701311242
Hosmer, D.W., Lemeshow, S., Sturdivant, R.X., 2003. Applied Logistic Regression, Third. ed, Wiley Series in Probability and Statistics. John Wiley and Sons Inc. https://doi.org/10.2307/2532419
Hung, L.Q., Van, N.T.H., Duc, D.M., Ha, L.T.C., Van Son, P., Khanh, N.H., Binh, L.T., 2016. Landslide susceptibility mapping by combining the analytical hierarchy process and weighted linear combination methods: a case study in the upper Lo River catchment (Vietnam). Landslides 13, 1285–1301. https://doi.org/10.1007/S10346-015-0657-3/TABLES/8
Iverson, R.M., 1997. The physics of debris flows. Rev. Geophys. 35, 245–296. https://doi.org/10.1029/97RG00426
Kirschbaum, D., Stanley, T., 2018. Satellite-Based Assessment of Rainfall-Triggered Landslide Hazard for Situational Awareness. Earth's Futur. 6, 505–523. https://doi.org/10.1002/2017EF000715
Kirschbaum, D.B., Adler, R., Hong, Y., Hill, S., Lerner-Lam, A., 2010. A global landslide catalog for hazard applications: Method, results, and limitations. Nat. Hazards 52, 561–575. https://doi.org/10.1007/S11069-009-9401-4/TABLES/3
Kirschbaum, D.B., Adler, R., Hong, Y., Kumar, S., Peters-Lidard, C., Lerner-Lam, A., 2012. Advances in landslide nowcasting: Evaluation of a global and regional modeling approach. Environ. Earth Sci. 66, 1683–1696. https://doi.org/10.1007/s12665-011-0990-3
Kirschbaum, D.B., Adler, R., Hong, Y., Lerner-Lam, A., 2009. Evaluation of a preliminary satellite-based landslide hazard algorithm using global landslide inventories. Nat. Hazards Earth Syst. Sci. 9, 673–686. https://doi.org/10.5194/nhess-9-673-2009
Kumar, L., Mutanga, O., 2018. Google Earth Engine applications since inception: Usage, trends, and potential. Remote Sens. 10. https://doi.org/10.3390/rs10101509
Lee, C.T., Huang, C.C., Lee, J.F., Pan, K.L., Lin, M.L., Dong, J.J., 2008. Statistical approach to storm event-induced landslides susceptibility. Nat. Hazards Earth Syst. Sci. 8, 941–960. https://doi.org/10.5194/NHESS-8-941-2008
Leempoel, K., Parisod, C., Geiser, C., Daprà, L., Vittoz, P., Joost, S., 2015. Very high-resolution digital elevation models: Are multi-scale derived variables ecologically relevant? Methods Ecol. Evol. 6, 1373–1383. https://doi.org/10.1111/2041-210X.12427
Liu, Y.H., Li, D.H., Chen, W., Lin, B.S., Seeboonruang, U., Tsai, F., 2018. Soil Erosion Modeling and Comparison Using Slope Units and Grid Cells in Shihmen Reservoir Watershed in Northern Taiwan. Water 10. https://doi.org/10.3390/W10101387
Loche, M., Alvioli, M., Marchesini, I., Bakka, H., Lombardo, L., 2022. Landslide susceptibility maps of Italy: Lesson learnt from dealing with multiple landslide types and the uneven spatial distribution of the national inventory. Earth-Science Rev. 232.

https://doi.org/10.1016/J.EARSCIREV.2022.104125
Lombardo, L., Opitz, T., Ardizzone, F., Guzzetti, F., Huser, R., 2020. Space-time landslide predictive modelling. Earth-Science Rev. https://doi.org/10.1016/j.earscirev.2020.103318
Lombardo, L., Tanyas, H., 2020. Chrono-validation of near-real-time landslide susceptibility models via plug-in statistical simulations. Eng. Geol. 278. https://doi.org/10.1016/J.ENGGEO.2020.105818
Luo, L., Lombardo, L., van Westen, C., Pei, X., Huang, R., 2021. From scenario-based seismic hazard to scenario-based landslide hazard: rewinding to the past via statistical simulations. Stoch. Environ. Res. Risk Assess. 1–22. https://doi.org/10.1007/s00477-020-01959-x
Muenchow, J., Brenning, A., Richter, M., 2012. Geomorphic process rates of landslides along a humidity gradient in the tropical Andes. Geomorphology 139–140, 271–284. https://doi.org/10.1016/J.GEOMORPH.2011.10.029
Nguyen, K.C., Katzfey, J.J., McGregor, J.L., 2014. Downscaling over Vietnam using the stretched-grid CCAM: Verification of the mean and interannual variability of rainfall. Clim. Dyn. 43, 861–879. https://doi.org/10.1007/S00382-013-1976-5
Ohlmacher, G.C., 2007. Plan curvature and landslide probability in regions dominated by earth flows and earth slides. Eng. Geol. 91, 117–134. https://doi.org/10.1016/J.ENGGEO.2007.01.005
Petley, D., 2012. Global patterns of loss of life from landslides. Geology 40, 927–930. https://doi.org/10.1130/G33217.1
Petschko, H., Bell, R., Leopold, P., Heiss, G., Glade, T., 2013. Landslide inventories for reliable susceptibility maps in lower Austria, in: Margottini, C., Canuti, P., Sassa, K. (Eds.), Landslide Science and Practice. Springer, Berlin, Heidelberg , pp. 281–286. https://doi.org/10.1007/978-3-642-31325-7_37/TABLES/2
Qiu, H., Cui, P., Regmi, A.D., Hu, S., Zhang, Y., He, Y., 2018. Landslide distribution and size versus relative relief (Shaanxi Province, China). Bull. Eng. Geol. Environ. 77, 1331–1342. https://doi.org/10.1007/S10064-017-1121-5/TABLES/1
Reichenbach, P., Rossi, M., Malamud, B.D., Mihir, M., Guzzetti, F., 2018. A review of statistically-based landslide susceptibility models. Earth-Science Rev. 180, 60–91. https://doi.org/10.1016/J.EARSCIREV.2018.03.001
Remondo, J., González, A., Díaz de Terán, J.R., Cendrero, A., Fabbri, A., Chung, C.J.F., 2003. Validation of landslide susceptibility maps; examples and applications from a case study in northern Spain. Nat. Hazards 30, 437–449. https://doi.org/10.1023/B:NHAZ.0000007201.80743.fc
RStudio Team, 2022. RStudio: Integrated Development Environment for R. RStudio.
Rue, H., Martino, S., Chopin, N., 2009. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. J. R. Stat. Soc. Ser. B (Statistical Methodol. 71, 319–392. https://doi.org/10.1111/J.1467-9868.2008.00700.X
Schober, P., Schwarte, L.A., 2018. Correlation coefficients: Appropriate use and interpretation. Anesth. Analg. 126, 1763–1768. https://doi.org/10.1213/ANE.0000000000002864
Segoni, S., Piciullo, L., Gariano, S.L., 2018. A review of the recent literature on rainfall thresholds for landslide occurrence. Landslides 15, 1483–1501. https://doi.org/10.1007/S10346-018-0966-4
Stanley, T.A., Kirschbaum, D.B., Benz, G., Emberson, R.A., Amatya, P.M., Medwedeff, W., Clark, M.K., 2021. Data-Driven Landslide Nowcasting at the Global Scale. Front. Earth Sci. 9. https://doi.org/10.3389/FEART.2021.640043/BIBTEX
Steger, S., Kofler, C., 2019. Statistical Modeling of Landslides, in: Spatial Modeling in GIS and R for Earth and Environmental Sciences. Elsevier, pp. 519–546. https://doi.org/10.1016/b978-0-12-815226-3.00024-7
Tang, G., Clark, M.P., Papalexiou, S.M., Ma, Z., Hong, Y., 2020. Have satellite precipitation products improved over last two decades? A comprehensive comparison of GPM IMERG with nine satellite and reanalysis datasets. Remote Sens. Environ. 240. https://doi.org/10.1016/J.RSE.2020.111697

Tien Bui, D., Lofman, O., Revhaug, I., Dick, O., 2011. Landslide susceptibility analysis in the Hoa Binh province of Vietnam using statistical index and logistic regression. Nat. Hazards 59, 1413–1444. https://doi.org/10.1007/s11069-011-9844-2

Tien Bui, D., Pradhan, B., Lofman, O., Revhaug, I., Dick, O.B., 2012. Spatial prediction of landslide hazards in Hoa Binh province (Vietnam): A comparative assessment of the efficacy of evidential belief functions and fuzzy logic models. CATENA 96, 28–40. https://doi.org/10.1016/J.CATENA.2012.04.001

Tien Bui, D., Pradhan, B., Lofman, O., Revhaug, I., Dick, Ø.B., 2013. Regional prediction of landslide hazard using probability analysis of intense rainfall in the Hoa Binh province, Vietnam. Nat. Hazards 66, 707–730. https://doi.org/10.1007/S11069-012-0510-0/FIGURES/12

UNISDR, 2015. Sendai Framework for Disaster Risk Reduction 2015 - 2030. United Nations Off. Disaster Risk Reduct. 37.

UNISDR, 2005. Hyogo Framework for Action 2005-2015: Building the Resilience of Nations and Communities to Disasters, in: World Conference on Disaster Reduction. Kobe, Hyogo, Japan.

Wagenmakers, E.-J., Lee, M., Lodewyckx, T., Iverson, G.J., 2008. Bayesian Versus Frequentist Inference, in: Bayesian Evaluation of Informative Hypotheses. Springer, New York, NY, pp. 181–207. https://doi.org/10.1007/978-0-387-09612-4_9

Watanabe, S., 2013. A Widely Applicable Bayesian Information Criterion. J. Mach. Learn. Res. 14, 867–897. https://doi.org/10.48550/arxiv.1208.6338

Whalen, A., Hoppitt, W.J.E., 2016. Bayesian model selection with Network Based Diffusion Analysis. Front. Psychol. 7. https://doi.org/10.3389/FPSYG.2016.00409/BIBTEX

Wu, W., Sidle, R.C., 1995. A Distributed Slope Stability Model for Steep Forested Basins. Water Resour. Res. 31, 2097–2110. https://doi.org/10.1029/95WR01136

Wubalem, A., 2021. Landslide susceptibility mapping using statistical methods in Uatzau catchment area, northwestern Ethiopia. Geoenvironmental Disasters 8. https://doi.org/10.1186/S40677-020-00170-Y

Yang, S., Berdine, G., 2017. The receiver operating characteristic (ROC) curve. Southwest Respir. Crit. Care Chronicles 5. https://doi.org/10.12746/SWRCCC.V5I19.391

Zou, K.H., O'Malley, A.J., Mauri, L., 2007. Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models. Circulation 115, 654–657. https://doi.org/10.1161/CIRCULATIONAHA.105.594929

9. ANNEX

Annex 1.

Table 9 Complete list of the 14 antecedent rainfall windows with associated average AUC as ordered by ascending WAIC used for model selection.

Antecedent Rainfall Days	WAIC	Average AUC
2	23492.17	0.82752
1	23636.73	0.82740
8	23685.13	0.83115
3	23735.00	0.82988
9	23790.42	0.81752
4	23824.17	0.82922
7	23834.70	0.82992
14	23843.29	0.82418
10	23849.71	0.82815
11	23862.83	0.82880
6	23874.51	0.82635
5	23878.86	0.82603
13	23879.47	0.82780
12	23882.69	0.81555

Annex 2. Regression coefficients of linear covariates with their 95% confidence interval, including the removed covariates, namely standard deviation of Eastness and relief as well as mean planform curvature.

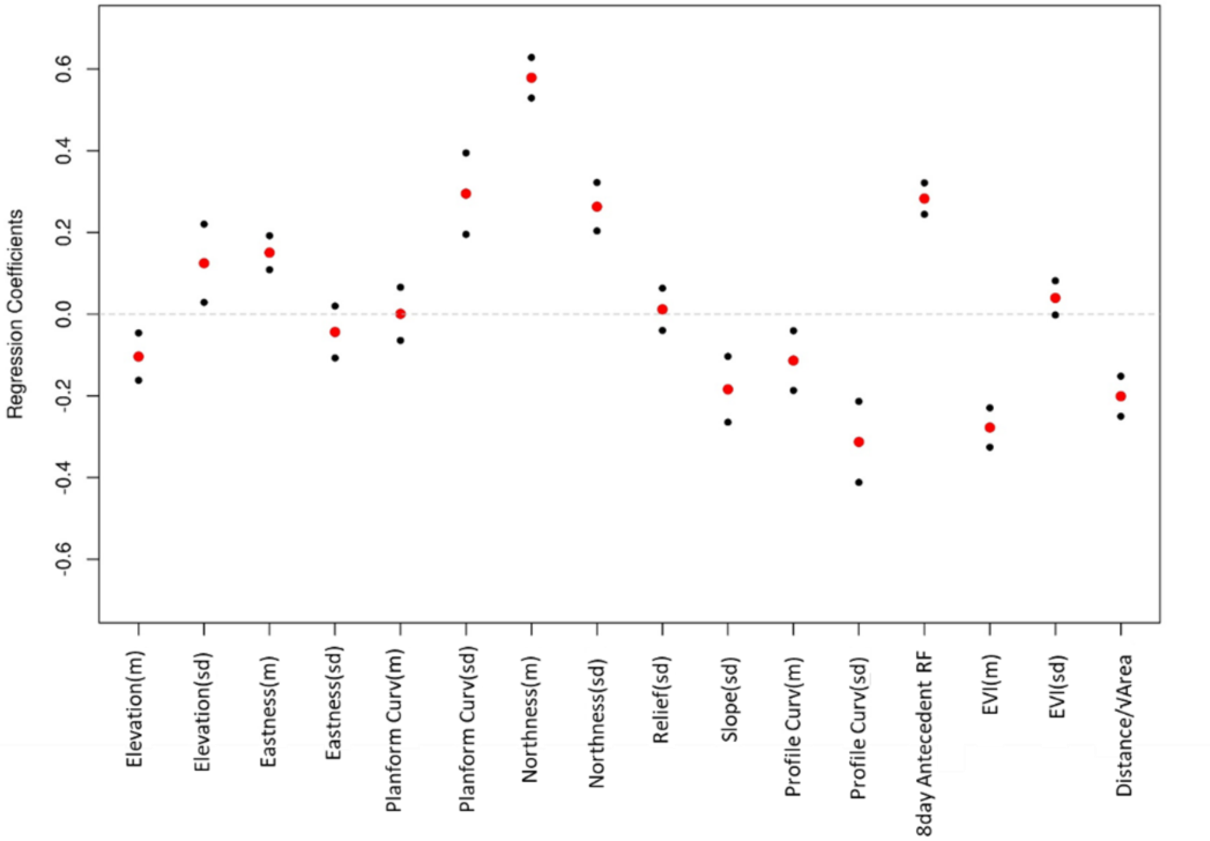

Figure 27 GAM including the non-significant linear covariates which were removed for further processing. Distance/√Area is referring to the roundness index.

Annex 3. Pearson's correlation for maximum distance in a SU and roundness index.

Pearsons correlation coefficient	0.4697871
95% credible interval	0.4669353-0.4714329
df	461998
p-value	2.2x10e16

Annex 4. The calculation method for TPR, TNR, FPR and precision for accuracy tests.

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

TNR = Specificity = $$\frac{TN}{FP + TN}$$

FPR = (1-Specificity) = $$\frac{FP}{FP + TN}$$

Precision = $$\frac{TP}{TP + FP}$$

Annex 5. Tables showing the values of F1 score with TPR and TNR at every quantile threshold for all six inventories named as: spec_*inventorynumbe*r, sens_*inventorynumber*, F1_*inventorynumber*

The optimal threshold is highlighted and the corresponding probability value is also identified.

Table 10 Following list of tables supporting the probability cut-off

Threshold	F1_1	Spec1	Sens1	Probability Value
0.05	0.095624	0.997118	0.050213	
0.1	0.182519	0.994236	0.100427	
0.15	0.261805	0.988473	0.150627	
0.2	0.334459	0.982709	0.200827	
0.25	0.40126	0.974063	0.251014	
0.3	0.462903	0.965418	0.301202	
0.35	0.519966	0.956772	0.351389	
0.4	0.572902	0.942363	0.40155	
0.45	0.622176	0.927954	0.451711	
0.5	0.668207	0.92219	0.501911	
0.55	0.711108	0.89049	0.551994	
0.6	0.751321	0.85879	0.602077	
0.65	0.789026	0.815562	0.652108	
0.7	0.824646	0.798271	0.702256	0.002725
0.75	0.857942	0.726225	0.752156	
0.8	0.889485	0.674352	0.802147	
0.85	0.919192	0.596542	0.852021	
0.9	0.947168	0.48415	0.901739	
0.95	0.973292	0.282421	0.951052	

Threshold	F1_2	Spec2	Sens2	Probability Value
0.05	0.095264	1	0.050014	
0.1	0.181865	1	0.100029	
0.15	0.260934	1	0.150043	
0.2	0.333413	1	0.200057	
0.25	0.400091	1	0.250071	
0.3	0.46164	1	0.300086	
0.35	0.518628	1	0.3501	
0.4	0.571545	1	0.400114	
0.45	0.620812	1	0.450129	
0.5	0.666794	1	0.500143	
0.55	0.709808	1	0.550157	
0.6	0.750134	1	0.600171	
0.65	0.787984	0.909091	0.65016	
0.7	0.823591	0.772727	0.700135	0.0039983
0.75	0.857134	0.545455	0.750084	
0.8	0.888886	0.545455	0.800099	
0.85	0.918892	0.454545	0.850087	
0.9	0.947278	0.227273	0.900036	
0.95	0.974235	0.090909	0.950012	

Threshold	F1_3	Spec3	Sens3	Probability Value
0.05	0.096204	1	0.050533	
0.1	0.18353	0.997537	0.10104	
0.15	0.26317	0.993842	0.151533	
0.2	0.336158	0.992611	0.202053	
0.25	0.403173	0.986453	0.25252	
0.3	0.465051	0.98399	0.303027	
0.35	0.522329	0.982759	0.353546	
0.4	0.575354	0.972906	0.403974	
0.45	0.624623	0.958128	0.45435	
0.5	0.670462	0.934729	0.504633	
0.55	0.713391	0.916256	0.554969	
0.6	0.753424	0.883005	0.605148	
0.65	0.791172	0.862069	0.655458	
0.7	0.826533	0.828818	0.705636	
0.75	0.859398	0.757389	0.755408	0.004099
0.8	0.890397	0.683498	0.805153	
0.85	0.9194	0.582512	0.85461	
0.9	0.946772	0.472906	0.903974	
0.95	0.971903	0.286946	0.952525	

Threshold	F1_4	Spec4	Sens4	Probability Value
0.05	0.095875	1	0.050351	
0.1	0.182978	1	0.100702	
0.15	0.262461	1	0.151053	
0.2	0.335282	1	0.201405	
0.25	0.402244	1	0.251756	
0.3	0.464028	1	0.302107	
0.35	0.521172	0.996276	0.352432	
0.4	0.574215	0.992551	0.402757	
0.45	0.623599	0.990689	0.453095	
0.5	0.669642	0.985102	0.503407	
0.55	0.712666	0.975791	0.553693	
0.6	0.752925	0.959032	0.603926	
0.65	0.790701	0.938547	0.654133	
0.7	0.826185	0.910615	0.704288	
0.75	0.859574	0.875233	0.754391	
0.8	0.890956	0.821229	0.804363	0.00552527
0.85	0.920479	0.746741	0.854191	
0.9	0.948046	0.61825	0.90364	
0.95	0.973659	0.41527	0.952565	

Threshold	F1_5	Spec_5	Sens_5	Probability Value
0.05	0.09529	1	0.050029	
0.1	0.181913	1	0.100057	
0.15	0.260999	1	0.150086	
0.2	0.333492	1	0.200114	
0.25	0.400183	1	0.250143	
0.3	0.461741	1	0.300172	
0.35	0.518738	1	0.3502	
0.4	0.571662	1	0.400229	
0.45	0.620934	1	0.450257	
0.5	0.666921	1	0.500286	
0.55	0.709939	1	0.550314	
0.6	0.750268	1	0.600343	
0.65	0.788152	1	0.650372	
0.7	0.823791	0.977273	0.700387	
0.75	0.857378	0.931818	0.75039	
0.8	0.889099	0.886364	0.800392	
0.85	0.919119	0.863636	0.850408	0.01454389
0.9	0.94738	0.545455	0.900255	
0.95	0.974165	0.181818	0.950075	

Threshold	F1_6	Spec_6	Sens_6	Probability Value
0.05	0.095784	1	0.050301	
0.1	0.182789	0.997831	0.100589	
0.15	0.262212	0.997831	0.15089	
0.2	0.334939	0.993492	0.201165	
0.25	0.401821	0.989154	0.25144	
0.3	0.463533	0.984816	0.301715	
0.35	0.520635	0.978308	0.351977	
0.4	0.573547	0.960954	0.402174	
0.45	0.622741	0.937093	0.452331	
0.5	0.668625	0.911063	0.502476	
0.55	0.711487	0.878525	0.552581	
0.6	0.751432	0.815618	0.602503	
0.65	0.788915	0.748373	0.652399	0.003176
0.7	0.824125	0.672451	0.702243	
0.75	0.857226	0.583514	0.752009	
0.8	0.888294	0.466377	0.801604	
0.85	0.917634	0.342733	0.851161	
0.9	0.945275	0.195228	0.900574	
0.95	0.971494	0.047722	0.949986	