Escola de Camins
Escola Tècnica Superior d'Enginyeria de Camins, Canals i Ports
UPC BARCELONATECH

Comparing physically-based with data-driven models for landslide susceptibility:
A case study in the Catalan Pyrenees

MASTER FINAL THESIS

Final Thesis developed by:
Manuel Antonio Álvarez Chaves

Directed by:
Prof. Dr. Vicente Medina
Prof. Dr. Marcel Hürlimann
Prof. Dr. Càrol Puig-Polo

Examiner:
Prof. Dr. Gerald Corzo Pérez (IHE-Delft)

Master of Science in:
Flood Risk Management

Barcelona, September 2021

Erasmus Mundus Programme in Flood Risk Management

Abstract

Because of their characteristics, landslides represent one of the most significant hazards in mountainous regions and are an important source of risk for people and infrastructure. To diminish their impacts, a landslide susceptibility assessment aims to identify areas where landslides can initiate and propagate.

In this research project, a physically-based (FSLAM) and four data-driven models (logistic regression, SVC, classification tree and random forest) were used to map landslide susceptibility for a case study area located in the Catalan Pyrenees. The results for all models were then compared in order to determine which performed best and under which conditions. The advantages and disadvantages of each model were also discussed as well as the limitations of their end products. Two cases for each of the models were proposed: the first one calculates landslide susceptibility in so-called "dry" conditions where no rainfall input is being considered, the second case considers two rainfall conditions registered during a 2013 event that triggered several landslides in the study area. From said event, a landslide inventory was collected and in this research project it was used to assess the performance of all models.

In terms of pure performance, data-driven models did better than the physically-based model. Using ROC AUC as the single metric to evaluate model performance, the logistic regression (0.806) and the random forest (0.853) scored higher than FSLAM (0.762). Nevertheless, the results of data-driven models are accompanied with a deal of uncertainty that make them less suitable for landslide susceptibility mapping. In terms of applicability, the decision tree (accuracy = 0.735) provided a simple tool that is easy to understand and apply which would make it very powerful in a decision-making process where stake-holders with many different backgrounds are involved.

After the main objectives of the project were achieved, several other tests using data-driven models were performed. These tests included testing the applicability of the logistic regression models to another area with similar characteristics to that area in which the models were trained, and using additional explanatory variables to obtain improvements in model performance. These tests offer insights which may guide future research objectives.

Keywords: landslide assessment, susceptibility mapping, physically-based models, FSLAM, data-driven models, logistic regression, support-vector classifier, classification tree, random forest

Acknowledgements

Although there is a single name in the cover, this thesis represents the collective effort of many people and the culmination of two years of experiences.

Firstly I would like express my gratitude towards my advisors:

Vicente, thank you for allowing me the time to learn so many things with this project as well as all the support you gave in regards to topics that were completely new to me. Also thank you for making every meeting so enjoyable, in many ways they were also a highlight of every week for me too.

Marcel, thank you for the direction you gave to this project and all the feedback given on our meetings and during the writing process.

Càrol, thank you for your encouraging words after all my presentations no matter how good, or bad, they were.

Moreover, I would like to kindly thank all others involved in the EnGeoModels research group at UPC. Claudia, Ona, Joaquín and Zizheng, thank you for patiently listening to me ramble week after week about data-driven models, performance metrics and other esoteric, and sometimes confusing, topics. And to the whole team, thank you for allowing me to participate and listen during your weekly meetings, all of our discussions have made this process a much more enriching experience.

I would also like to thank all the professors and lecturers who were involved in the M. Sc. programme. Particularly, I would like to mention professors Dimitri Solomatine, Biswa Bhattacharya and Allen Bateman. Throughout my life, I have always said that I have had very good teachers and after these two years, this statement continues to be true.

To all the people of FRM 8. Thank you for sharing these past two years with me. You all have had a great impact in my life and in the future, I hope we meet again as professionals, academics and leaders.

Angie, thank you for all the beers that we shared and for bringing the best in me for all the assignments in which we worked together.

Alejandra, thank you for all the joy you brought to my life these past two years. When signing up for this programme I didn't realize I would sharing this stage of my life with such an amazing person.

Por último, a mi mamá y a mi papá. Gracias por todo el apoyo y, a pesar de la distancia, por la compañía que me dieron estos dos años. Saber que siempre puedo contar con ustedes realmente es algo invaluable.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

Landslides are a general term used to describe the downslope movement of soil, rock, and organic materials under the effects of gravity. In more specific terms, a landslide in which the sliding surface is located within the upper layers of the terrain (typically at depths ranging from centimetres to a few meters) is called a shallow landslide. The triggering mechanism of this phenomena is often intense rainfall where water causes slope instability. Landslides and flooding are closely linked because both are related to precipitation, runoff, and the saturation of ground by water and it's cascading effects (Highland & Bobrowsky, 2008).

Because of their characteristics, landslides represent one of the most significant hazards in mountainous regions and are an important source of risk for people and infrastructure (Froude & Petley, 2018). To perform a proper risk assessment for shallow landslides, one must start with the susceptibility analysis. This study identifies prone areas where landslides can initiate and propagate (Fell et al., 2008). As an initial step, this study also provides results which are useful in land use planning and early warning systems.

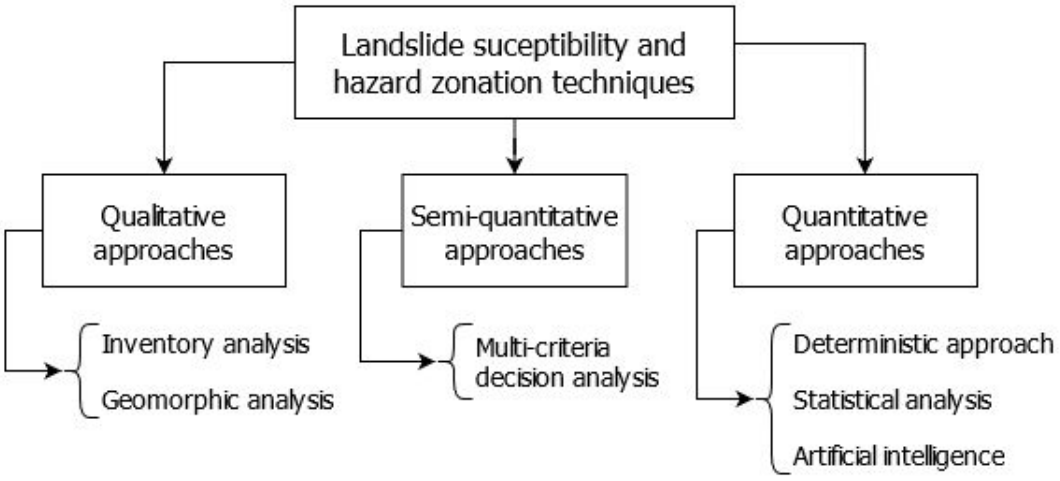

Figure 1.1: Aprroaches to landslide suscetibility assessment
Shano et al. (2020)

Figure 1.1 shows a summary of the different methods and techniques used for landslide susceptibility assessment and hazard zonation techniques. In recent times there has been a preference to use quantitative approaches such as numerical models based on physics, statistical models or data-driven models.

Recently Medina et al. (2021) have proposed an innovative physically-based model called FSLAM (Fast Shallow Landslide Assessment Model) where the factor of safety of a particular cell in a raster map is calculated through a geotechnical model based in the infinite slope analysis and the Mohr-Coulomb theory, as well as a hydrological model based on different contributions to the position of the water table. In their model a stochastic approach has been used to include the main input parameters of the model. Successful case studies on the application of the model have been concluded in the Principality of Andorra, and more recently in the Val d'Aran region located near the Pyrenees mountains in the north-east of Catalonia, Spain.

Data-driven models have been criticised for landslide susceptibility analysis because they ignore the complex physical processes involved in landslide initiation, but they have been proven accurate in several studies not only for landslides but also for the initiation of debris flows (Goetz et al., 2015; Zêzere et al., 2017; Zhang et al., 2019). In addition, recent developments have shown that integrating the results of physically-based and data-driven models rather than only using one of these approaches, can generate higher quality susceptibility maps (Sun et al., 2021).

1.2 Problem Definition

As mentioned previously, FSLAM is a new model based on the physical process that leads to the failure of a slope unit. As such, the model is reliant on the parameters used to describe the physical properties of the materials involved in the failure process. The model also operates at a regional scale and, at this scale, the determination of specific properties is a challenge in most cases (Cho, 2007). To overcome this challenge, FSLAM uses a stochastic approach where the values of said physical properties are included as ranges.

By contrast, data-driven models find relationships between the system state variables (input and output) without explicit knowledge of the physical behaviour of the system (Solomatine & Ostfeld, 2008). In this case, the explanatory or input variables can be of any type, be it numerical, categorical, etc. This presents an advantage when compared with a physically based model when, as an example, the type of soil is known to be colluvium but specific parameters, such as cohesion, internal friction angle or layer depth are much more difficult to determine.

1.3 Research Questions

The basis for the present study are the following research questions:

- Which data-driven model is best suited to determine the susceptibility of a raster land-unit to generate a shallow-landslide?
- Based on the insights offered by the data-driven model, which explanatory variables for landslides are the most important?

- Which are the advantages of a data-driven model over the physically-based model to determine landslide susceptibility? Which are its disadvantages?
- Can the input parameters used in FSLAM also serve as explanatory variables or predictors in a data-driven model?
- Can the data set used to train the data-driven model be extended using parameters derived from the basic data?
- How does the data-driven model perform when using different subsets of the available training data?
- When doing a performance comparison between the physically-based model FSLAM and a data-driven model, which one proves to be better and under which conditions?
- How does including the results or outputs of a physically-based model in the training data affect the results of the data-driven model?

1.4 Research Objectives

Main Objective

Compare the performance and insights of a a data-driven model with the physically-based model FSLAM in determining the susceptibility of each cell in a raster map regarding rainfall-induced shallow landslides.

Specific Objectives

1. Test different statistical and machine learning methods to determine which are most fit to describe the relationship between shallow landslide predisposition factors and their occurrence.
2. Analyse the process by which each specific data-driven models determines susceptibility.
3. Compare the results obtained by the data-driven model that proved best with the ones computed by FSLAM for the susceptibility assessment of a specific region located in the Catalan Pyrenees.
4. Evaluate improvements in performance obtained when using the outputs of FSLAM as explanatory variables for the data-driven model.
5. Explore the usage of a data-driven model in a different region with similar characteristics from where it was trained.

1.5 Innovation and Practical Value

Although both, physically-based models and data-driven models for landslide susceptibility assessment have been thoroughly explored in the literature, a direct comparison of both types within the same study area could not be found. Specifically in regards to data-driven models, the literature shows that there is still discussion to be had on which methods are best suited for this type of approach (Reichenbach

et al., 2018). This research project will add to this discussion by testing different methods in a new case study area to determine which are best, and determine if they follow or deviate from the current trends. In addition, exploring the application of these methods in the Pyrenees is completely new.

The research project will also directly benefits the overall FSLAM project. As mentioned previously, FSLAM is a new shallow-landslide susceptibility analysis model which is particularly attractive because of its simplicity, ease of use and short computational time. In particular, one of the key advantages of FSLAM is the stochastic approach applied to soil properties in order to consider their uncertainty. The determination of the soil properties at a regional scale is in many cases a challenge, and this approach makes it so the uncertainty of these parameters is properly considered in the assessment. A data-driven model might be friendlier in this case because in FSLAM, even though the user inputs one of the soil categories of the study area, the user still needs to input a range of values for the specific parameters of the geotechnical model. By contrast, for the data-driven model, the category would be sufficient.

1.6 Thesis Outline

This thesis contains a total of six chapters. The next chapters are laid out as follows:

Chapter 2 is an overview of the available literature on previous research publications that have used the FSLAM model as well as some of the relevant research publications related to the usage of statistical and data-driven models for landslide susceptibility.

Chapter 3 presents the case study area of Val d'Aran located in the Catalan Pyrenees as well as Berguedà, a county in Catalonia that will be used for a specific case application.

Chapter 4 lays out the methodology to train and test a data-driven model for landslide susceptibility. This chapter also presents the theoretical framework which FSLAM and the data-driven models use.

Chapter 5 presents the results obtained for both FSLAM and the different data-driven models.

Chapter 6 finally summarizes some of the main findings and answers the research questions. Further, this chapter also gives recommendations for future studies on this same line of research.

Chapter 2

Literature Review

2.1 The FSLAM model

Medina et al. (2021) - Fast physically-based model for rainfall-induced landslide susceptibility assessment at regional scale

The paper serves as various purposes. First: it introduces the Fast Landslide Susceptibility Assessment Model (FSLAM) and gives insight into how the model computes susceptibility. This procedure will be described in detail in following sections of this research project. Second: in the paper, all the input parameters are subjected to a sensitivity analysis in order to determine which ones have the most impact on the output. And third: the application of the model at a regional scale where the country of Andorra was used. The regional scales is defined as: "in the order of magnitude of 100 km^2", and Andorra comprises about 470 km^2 making it a good candidate for this kind of assessment.

One of the main advantages of FSLAM is the usage of a stochastic approach to define the key input parameters of cohesion (soil and root) and friction angle. It is notorously difficult to find accurate information about these parameters unless field work is performed and, at the regional scale, this is just not feasible. FSLAM takes a range of values for these three parameters and is able to compute not only a factor of safety (FS) but also a probability of failure (PoF). The PoF result is later used to determine whether a cell is unstable or stable where several thresholds are tested but the authors reach the conclusion that the threshold could be adjusted depending on the usage for the model. As an example, a low threshold would generate a lot of false positives but this result would be adequate for the usage within an early warning system. For the case of Andorra, a PoF of 50% threshold provided a good accuracy score of 0.70.

Hürlimann et al. (2021) - Impacts of future climate and land cover changes on landslide susceptibility: regional scale modelling of the Val d'Aran region (Pyrenees, Spain) ← (under review)

This is the second of the two papers on the application of the FSLAM model. This is a more applied research where the model was first calibrated using information of a landslide event that was triggered during 2013, and then used to analyse landslide

susceptibility in the near, mid and far-future. It is important to note that a landslide inventory of 392 entries was used in this study to calibrate the FSLAM model.

For the calibration process of the rainfall and CN related inputs, an additional module was added to FSLAM in order to compute runoff and compare the results with two discharge measuring stations within the study area. The geotechnical parameters were interpreted from geological maps because, as mentioned previously, the stochastic approach used for this inputs permits a more coarse assignment of these parameters. The accuracy of the simulation of the 2013 was assessed through a comparison between the inventory points and a sample of 5000 random points in the study area. The authors found that, for the calibrated model, the PoF increased around 60% for points in the landslide inventory, and a significant increase was not observed for the random sampled points. The final AUC score for the ROC curve of this model was 0.78.

As the overall result, the researchers found that for the future scenarios the daily rainfall expected for the 100 year return period will increase while a significant percentage of the land cover composed by grassland will change into forest area. This makes it so the destabilization effects generated by the increase in rainfall are compensated by the greater soil root cohesion by the forest trees. When the two impacts are assessed in the future scenarios, the overall stability conditions of region improve.

2.2 Data-driven models for landslides

Reichenbach et al. (2018) - A review of statistically-based landslide susceptibility models

A critical review of statistical methods for landslide susceptibility modelling and zoning is performed. The assessment is based on 565 peer-reviewed publications from 1983 to 2016. The authors find that there is significant heterogeneity in the data types and scales, modelling approaches and criteria for evaluating the models.

The authors use the susceptibility quality level (SQL) index introduced by Guzzetti et al. (2006) to rank the quality of publications. SQL is an index based on the information provided in the publication in regards to the degree of model fit, model prediction performance and error associated with the predicted susceptibility. Some of the key insights revealed in the study is that a significant amount (4.5%) of the published landslide susceptibility assessments are poor (SQL = 0) and that only a few (1.1%) reach the highest quality level (SQL = 7). Then, recommendations on the criteria to reach an adequate SQL index are given by the authors. These criteria emphasize the importance of evaluating the model fit and its prediction performance. The authors cite Peng et al. (2014) as a study reaching SQL = 7.

Another important insight revealed by the authors is that no single statistical method proves to be superior in all the research studies, although there are some common ones that appear in a significant proportion of the articles, such as the logistic regression. And finally, the authors point out that comparing and combining different modelling approaches seems to give the best susceptibility assessments.

Guzzetti et al. (2006) - Landslide hazard assessment in the Collazzone area, Umbria, Italy

For this particular study area an impressive inventory of 2787 landslides was prepared by analysing aerial photographs taken between 1941 and 1997, and field surveys from 1998 to 2004. The region that served as study area measures 79 km^2.

Because of the amount of information in the inventory, including dates and sizes of the landslides, the authors are also able to assess the probability of occurrence in a determined period and size of the landslides in order to finally compute hazard. Specifically for spatial occurrence, they use the statistical method of the discriminant analysis of 46 thematic variables.

The authors perform a very interesting assessment of the uncertainty of their chosen model. The final output of the model is a probability between unstable ($P = 1$) and stable ($P = 0$). They find that for units having intermediate values of probability ($P = 0.5$), the model is incapable of satisfactorily classifying the terrain and the obtained estimate is highly variable and depends on the sample. Further, they comment that for this particular probability, the results are unreliable.

Peng et al. (2014) - Landslide susceptibility mapping based on rough set theory and support vector machines: A case of the Three Gorges area, China

The most notable advancement in this research paper is the usage of a reduction algorithm to reduce the number of attributes used to estimate landslide susceptibility. From an initial list of 20 environmental parameters, the authors are able to identify 13 that are the most important to predict landslide susceptibility. This procedure was done using a rough set (RS) theory based algorithm as implemented in the RSES2 software (Bazan & Szczuka, 2005).

The authors then go ahead and train two different models using support vector machines (SVMs). For the testing set, the model trained using the reduced set of environmental predictors showed better performance than the model trained using all environmental predictors. In addition, the authors perform an uncertainty analysis modelled by the one done by Guzzetti et al. (2006) arriving at similar results.

Zêzere et al. (2017) - Mapping landslide susceptibility using data-driven methods

Different landslide susceptibility maps are created for the Silveira basin of 18.2 km^2 in the region of Lisbon, Portugal. The maps are created by varying the statistical method used to compute them, the basic terrain mapping unit used in the assessment and the selection of feature type to represent the actual landslides in the model. The statistical methods used by the authors were: the logistic regression, discriminant analysis and information value.

The authors found the the best model, based on AUC of the ROC curve, was one using the logistic regression statistical method and the census terrain units. This gives space to the claim by the authors that: "the model with the highest AUC ROC is not necessarily the best landslide susceptibility model". The model with the second highest ROC AUC based on the logistic regression and grid cell terrain

units, is then established as the best for the region and the authors explain that this particular model better considers the spatial component of landslide susceptibility.

Sun et al. (2021) - Exploring the impact of introducing a physical model into statistical methods on the evaluation of regional scale debris flow susceptibility

In this paper, the authors evaluate debris flow susceptibility at a regional scale for an area in north-east China comprising of 2625 km^2. The authors use several statistical methods to compute susceptibility, including the support vector machine (SVM), analytic hierarchy process (AHP) and Shannon entropy, with the SVM proving must successful in terms of perfomance based on the ROC AUC metric. Additionally, the authors establish a new approach by introducing a physical model into statistical methods. The combined model consists of two parts, the statistical model and TRIGRS which is a physically-based model.

The results obtained by integrating the models consider both the prediction result of the statistical method for debris flow susceptibility and the mechanism of debris flow initiation. The authors state that the performance of the integrated models is significantly better than that of the single statistical model and these integrated models are able to generate higher quality debris flow susceptibility maps.

As a measure of the improvement by the models, the simple SVM model obtained a ROC AUC score of 0.889 while the integrated model that uses the SVM and TRIGRS obtained a ROC AUC score of 0.922. An increase of 0.033 or 3.7% relative percent.

Chapter 3

Study Area

3.1 Val d'Aran

3.1.1 Location

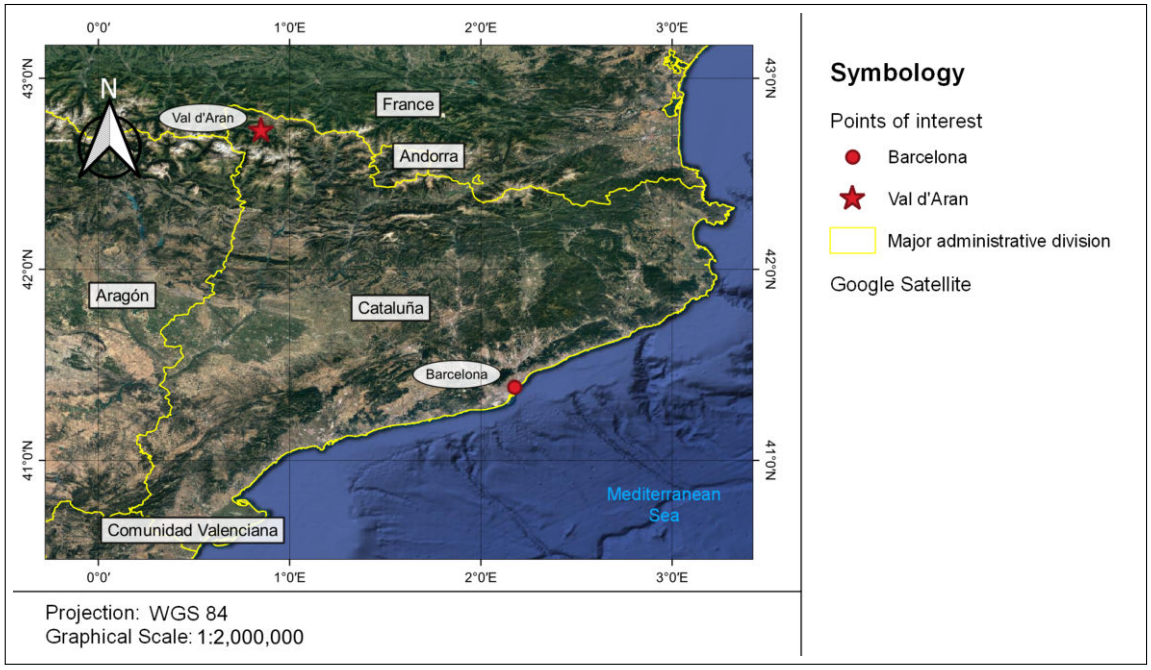

Figure 3.1: General location map

Located in the northern side of the Pyrenees, the valley of Aran or Val d'Aran is an administrative entity in Lleida, Catalonia, Spain. The study area was chosen mainly because of the amount of available information collected from previous studies (Hürlimann et al., 2021; Shu et al., 2019), as well as the fact that the application of data-driven modelling for the region is new.

Figure 3.1 shows the general location of Val d'Aran. As mentioned previously, the valley is located on the northern side of the Pyrenees which means that it is also the only region within Catalonia with a drainage network that flows into the Atlantic Ocean. Traditionally, it is also part of the Occitania historical and cultural region

which encompasses part of southern France, Monaco and some smaller regions in Italy.

3.1.2 The 2013 landslide episode

In this section, events that can be categorized as Multiple-Occurrence Regional Landslide Events or MORLE are described. The events happened in the Central Pyrenees during the 17th and 18th of June 2013. The causes of the events are a combination of i) extreme rainfall and ii) high meltwater volumes due to unusually heavy snowfall during the winter. The extreme weather also caused flooding in certain parts of the Pyrenees.

The total economic losses were estimated to be more than 100 million Euro and the Val d'Aran region was significantly affected. People had to be evacuated from certain villages. Several bridges and roads as well as individual houses were destroyed. The situation was described as "extremely severe and complicated" by a spokesperson of the County of Val d'Aran and, although economic losses were heavy, no human lives were lost (CatalanNews, 2013). The flood and landslide events led to a review of the scenarios being used as a measure to improve the existing disaster plans (Victoriano et al., 2016).

After the episode, an inventory of 392 landslides was created through the interpretation of aerial photography, helicopter flights and field surveying. The inventory can be seen in Figure 3.2 on top of a cartographic map of the Val d'Aran region.

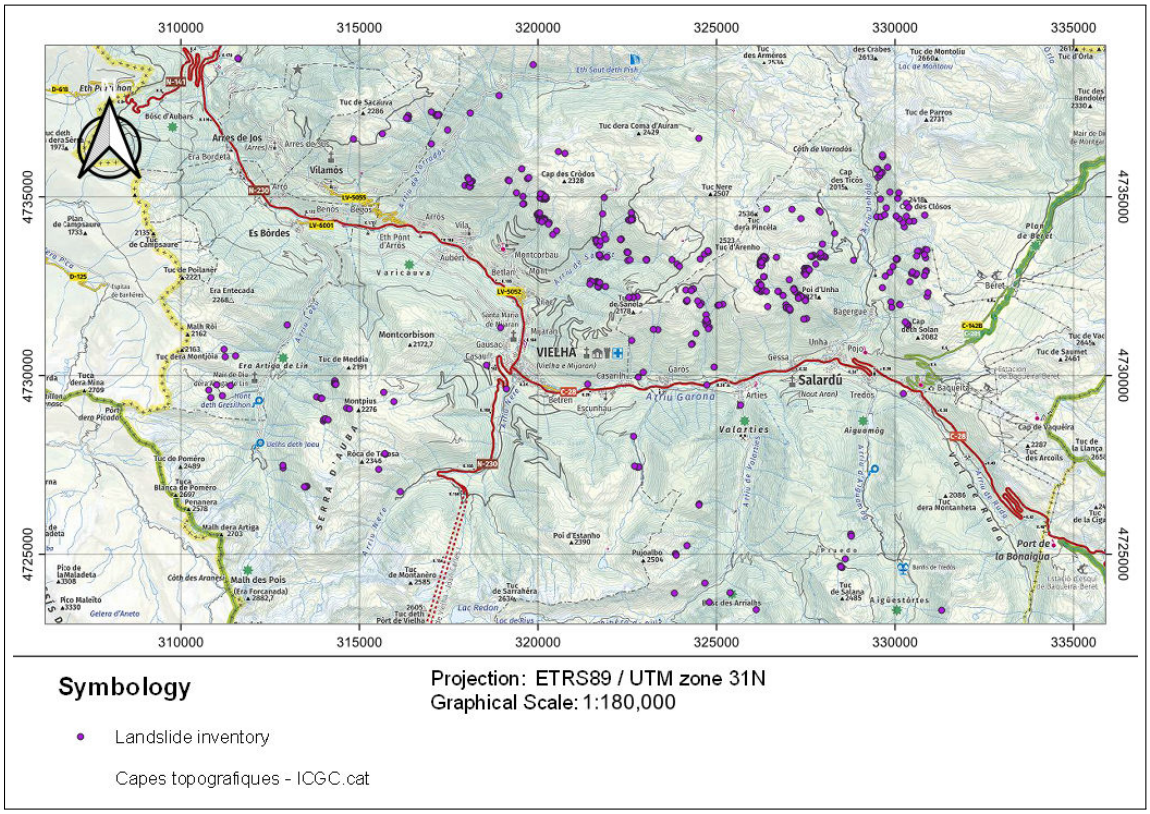

Figure 3.2: Landslide inventory map

High-quality landslide inventories are important to calibrate and validate models of

landslide susceptibility and to evaluate the performance of physical slope stability models (Shu et al., 2019). The inventory is even more relevant when preparing data-driven models because, in this particular case, this is an exercise in "supervised learning" where a function is calibrated in order to map an input to an output based on example input-output pairs. The inventory serves as the examples for the data-driven models to learn from.

3.1.3 2013 rainfall event

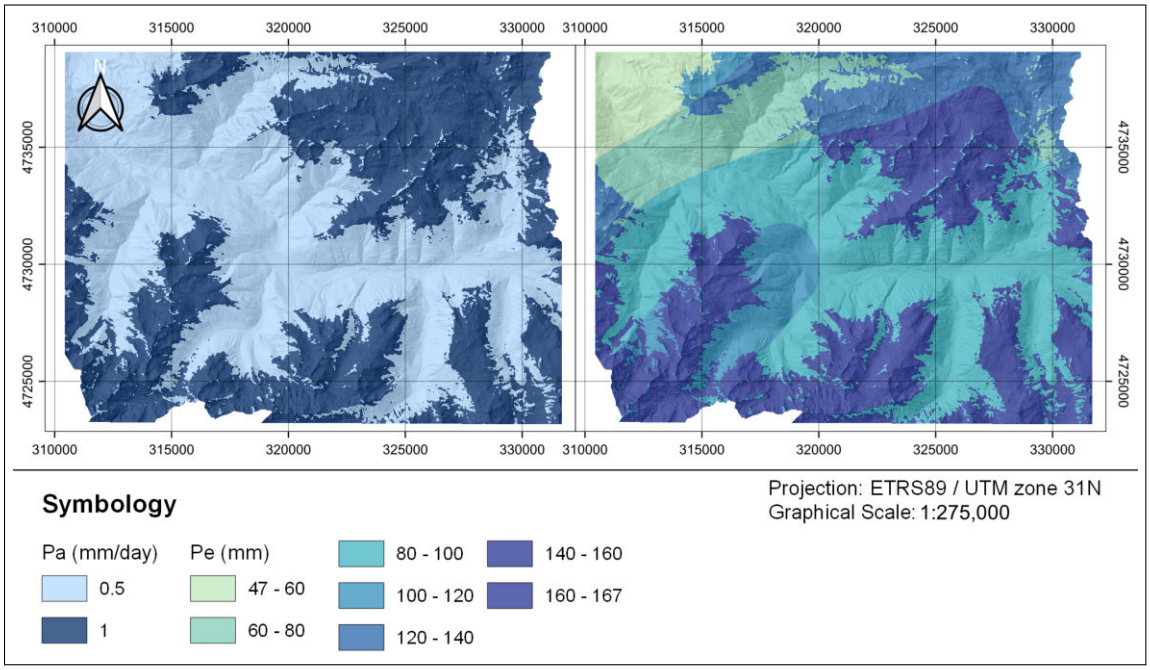

Figure 3.3: Antecedent rainfal (left) and event rainfall (right) maps

Even though the main triggering factor for the landslides was rainfall, the event was actually triggered by a combination of two exceptional factors:

1. Extreme rainfall: 124.7 mm in between the two days of which 101.2 mm fell on the 18th of June.
2. High meltwater due to the accumulation of snow (>300 cm measured at one station within the region).

The map shown in Figure 3.3 shows an approximation to the rainfall and snowmelt conditions that led to the landslide episode. These conditions were calibrated by Hürlimann et al. (2021) using FSLAM and the landslide inventory.

As it will be explained in the next chapters, FSLAM uses rainfall as a means to calculate the position of the water table. In order to account for the snow, the recorded rainfall map was modified into the maps shown in Figure 3.3. Antecedent rainfall (P_a) considers rainfall and snow, and shows a clear divide between areas that experienced snowmelt and areas that did not. The event rainfall (P_e) was estimated combining the observed rainfall at different weather stations within the study area and 60 mm are added in some parts to account for snowmelt.

3.1.4 Digital elevation model (DEM)

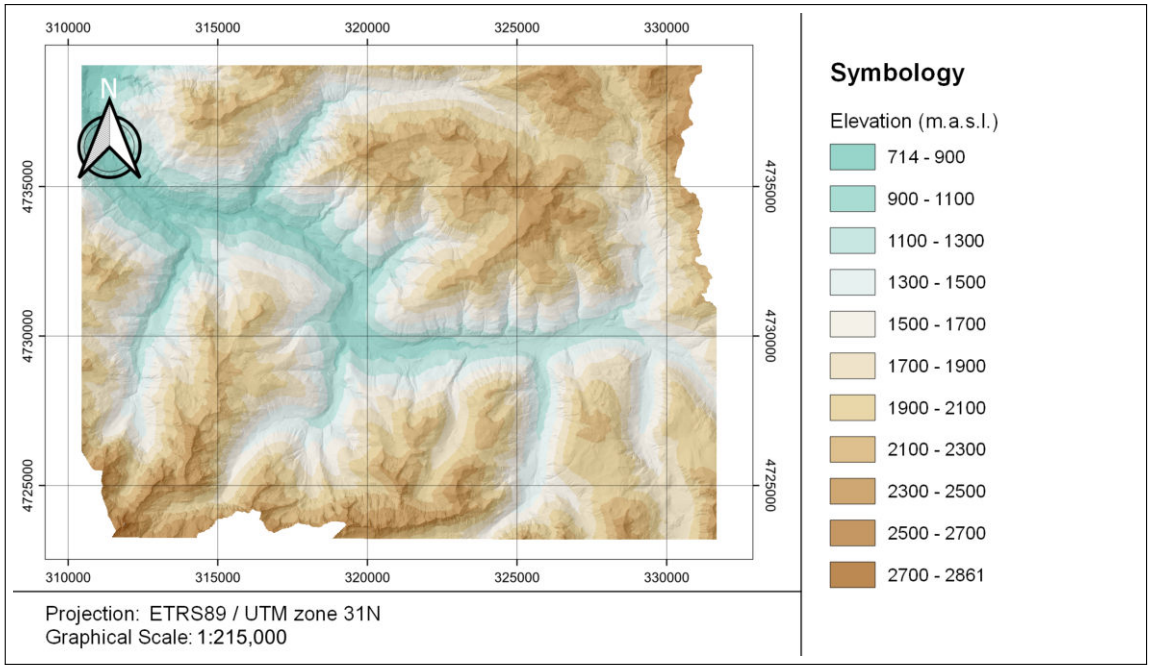

Figure 3.4: Digital elevation model (DEM) map

The DEM, shown in Figure 3.4, was obtained from the Institut Cartogràfic i Geològic de Catalunya (ICGC) and covers an extent 337.2 km^2 of the 620.5 km^2 that comprise the whole administrative area of Val d'Aran (54%). The DEM has a resolution of 5 m by 5 m and contains values for 13 486 326 cells, of these cells, 489 384 cells contain "no data" values which reduces the total area analyzed to 324.9 km^2.

3.1.5 Land use/land cover (LULC)

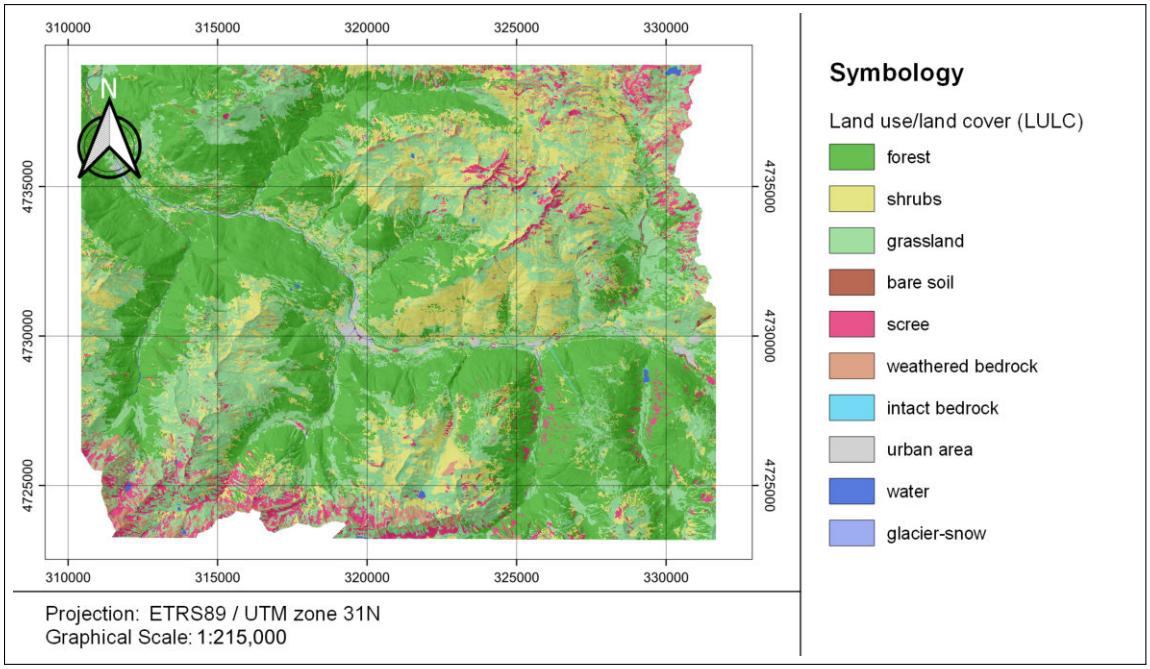

Figure 3.5: Land use/land cover (LULC) map

The LULC map, shown in Figure 3.5, was obtained from The Centre for Research on Ecology and Forestry Applications (CREAF) and dates from 2009. Although the original map contains more classes, it was reclassified to only show 10 different classes where forest (43.1%), grassland (30.8%) and shrubs (16.7%) are the most dominant.

Shu et al. (2019) previously used the landslide inventory to perform an analysis on how land use changes starting from 1946 may have affected landslide susceptibility of the region. Particularly, they highlight that a significant increase of forest and shrubs in the period mentioned might have reduced the overall landslide susceptibility of the region.

3.1.6 Soil type

The soil map, shown in Figure 3.6, was obtained from the ICGC and dates from 2017. Originally it was a geological map that was reclassified into 11 lithological categories in order to use their corresponding mechanical soil properties. In this map, the most dominant classes are phyllite-slate (28.5%), mudstone (18.1%) and colluvium (12.6%) but most landslides tend to occur on mudstone, limestone (8.9%) and colluvium (Hürlimann et al., 2021).

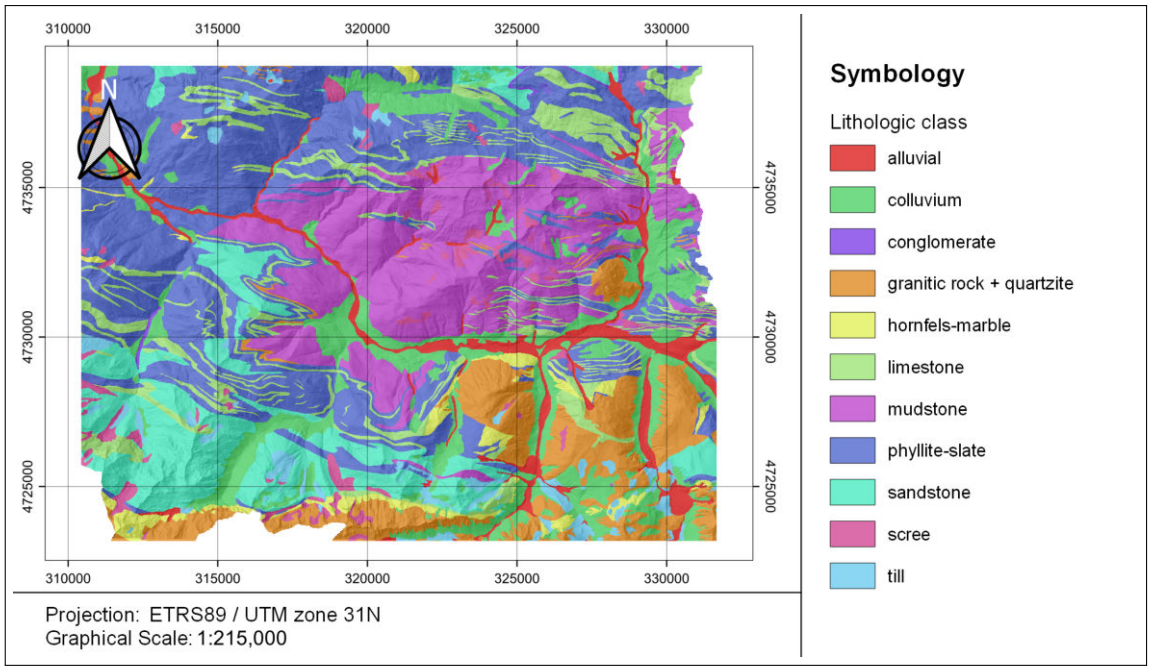

Figure 3.6: Soil type map

3.2 Berguedà

Due to their significant impact in the region, all throughout the Pyrenees there have been several counties that have served as study cases for landslide assessments. In the previous sections Val d'Aran was described but another of such regions is the county of Berguedà. Located 100 km southeast of Val d'Aran, Berguedà consists of the upper Llobregat river valley and the mountainous areas that surround it.

Through time, a chronology of landslides in the Berguedà region has been reconstructed from technical reports, field surveying and geomorphological analysis. Currently, a landslide inventory of 998 landslides has been collected from several sources (Baeza & Corominas, 2001; Corominas & Moya, 1999). The points identifying each landslide are shown in Figure 3.7.

In a similar fashion to the thematic maps presented for Val d'Aran, a DEM, LULC map and soil type map are also available for the Berguedà region. These are shown in Figures 3.8 and 3.9.

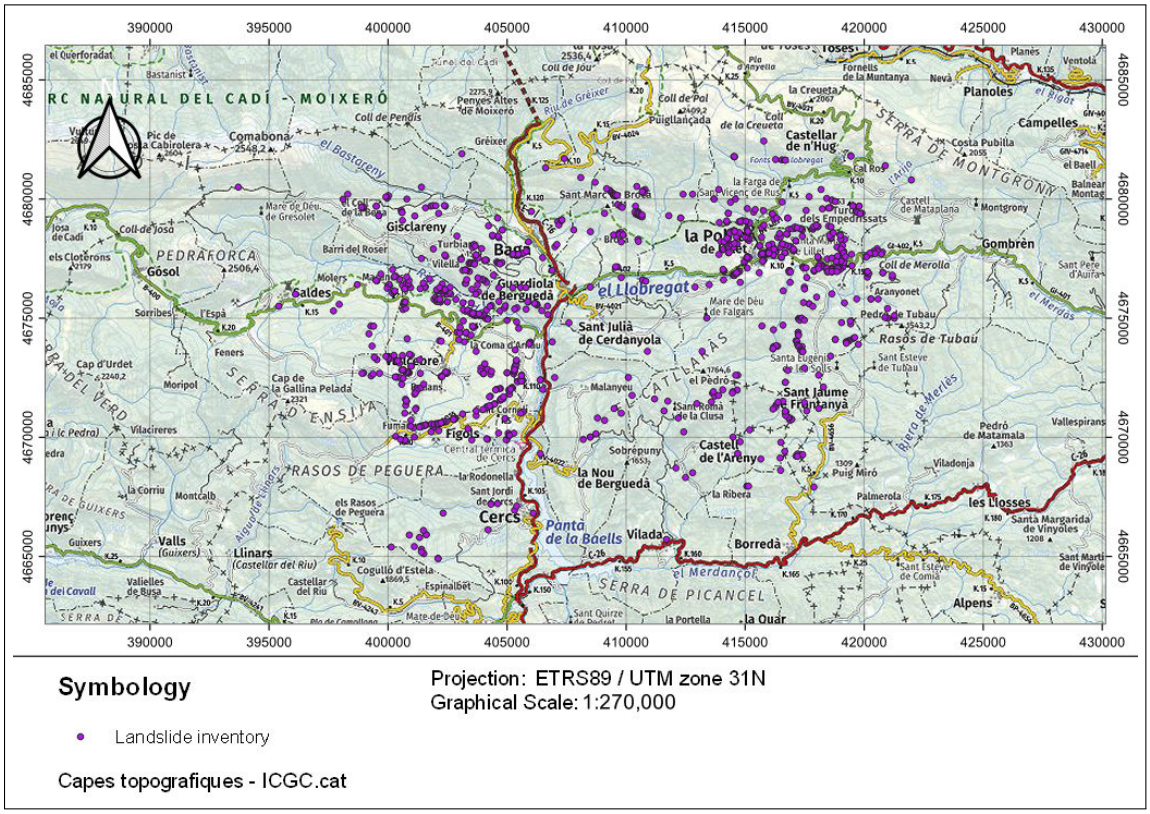

Figure 3.7: Landslide inventory for the Berguedà region

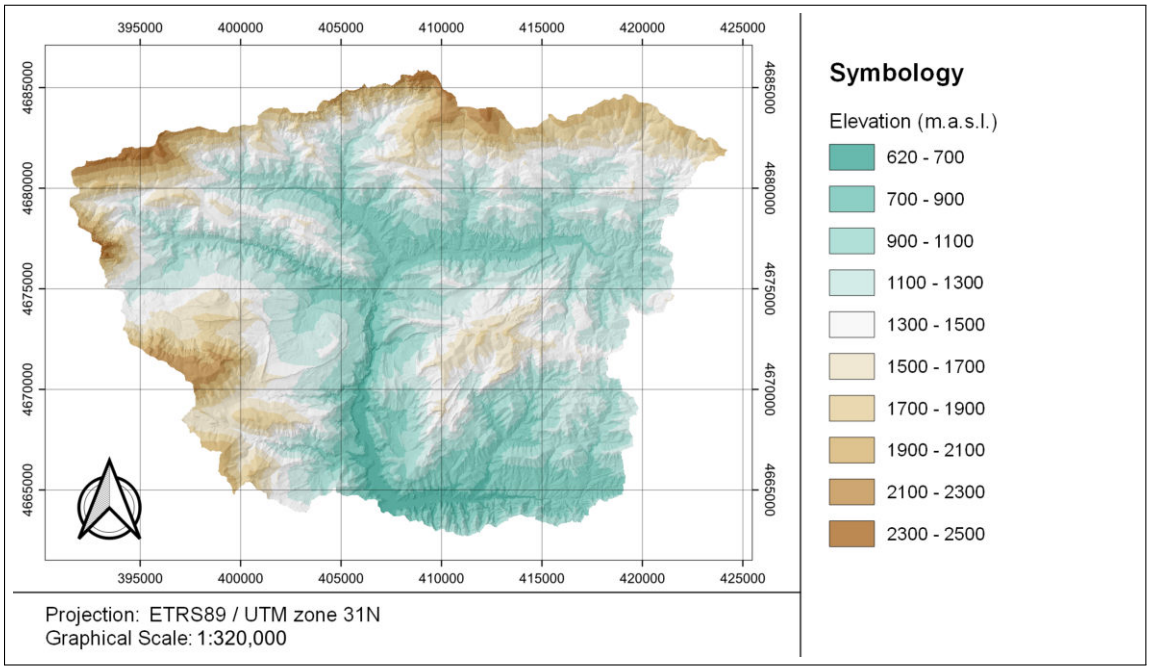

Figure 3.8: DEM for the Berguedà region

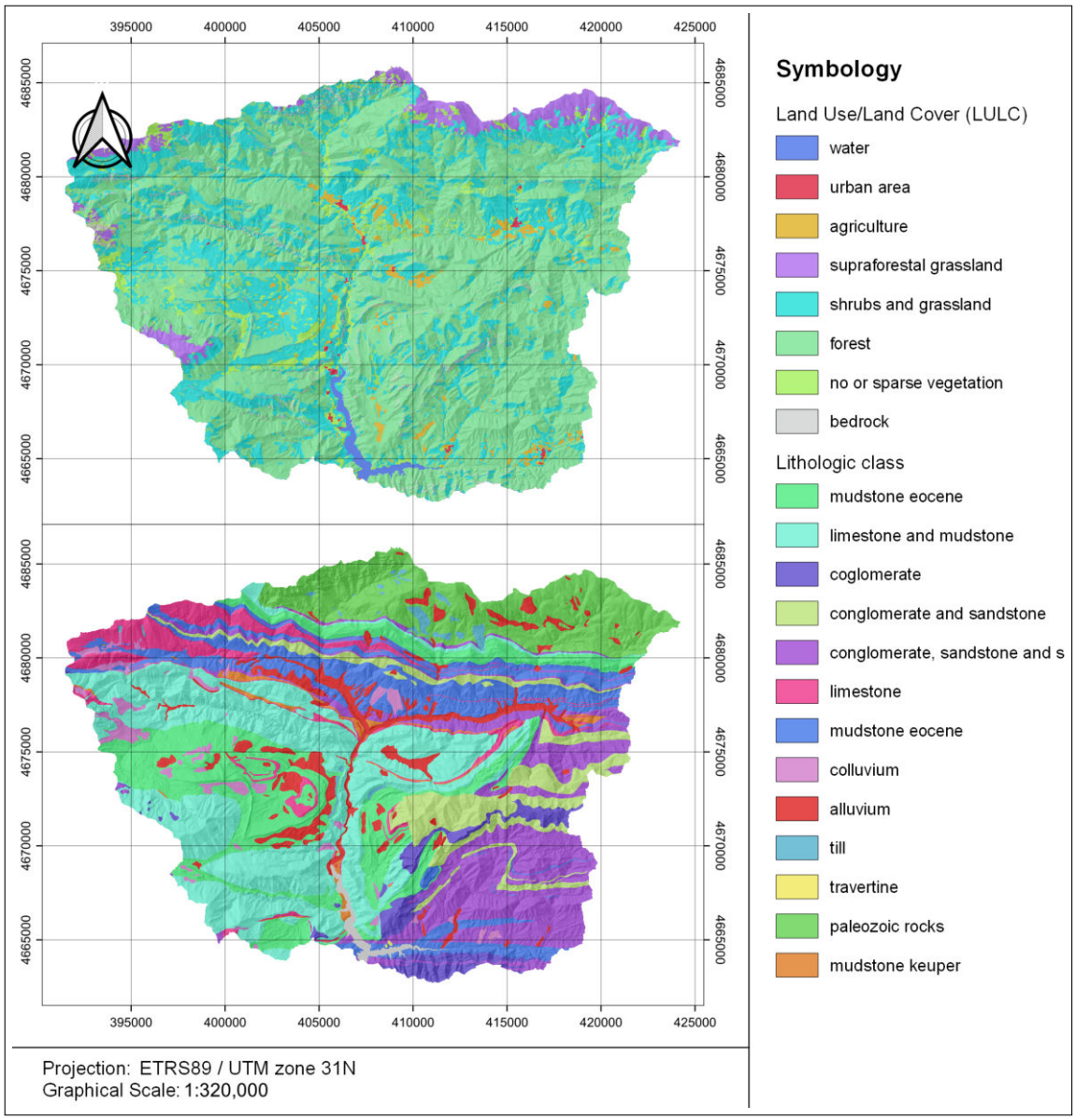

Figure 3.9: LULC (top) and Soil type (bottom) maps for the Berguedà region

Chapter 4

Tools and Methods

4.1 Theoretical framework

4.1.1 FSLAM

Developed at the Universitat Politècnica de Catalunya (UPC), FSLAM is a physically based model used to analyze landslide susceptibility at a regional scale. FSLAM works through the combination of a geotechnical model and a hydrological model.

Geotechnical model

The geotechnical model used by FSLAM is based on the infinite slope analysis and the Mohr-Coulomb theory for failure. The model calculates a factor of safety based on the following equation:

$$FS = \frac{c_s + c_r}{g\rho_s z \cos\theta \sin\theta} + \left(1 - \frac{h}{z}\left(\frac{\rho_w}{\rho_s}\right)\right)\left(\frac{\tan\phi}{\tan\theta}\right) \tag{4.1}$$

Where:

c_s = soil cohesion in kPa
c_r = root cohesion depending on the land use in kPa
g = gravitational constant in $\frac{m}{s^2}$
ρ_s = soil density in $\frac{kg}{m^3}$
θ = slope angle in degrees
h = position of the water table in m
z = total soil depth in m
ρ_w = water density in $\frac{kg}{m^3}$
ϕ = soil friction angle in degrees

As mentioned previously, some of these parameters can be introduced into the model as ranges in order to perform a stochastic calculation of the FS and finally the PoF. Specifically the parameters that are introduced stochastically are cohesion and internal friction angle because they belong to separate terms in Equation 4.1 and therefore a Gaussian distribution can be obtained for FS if the distribution of

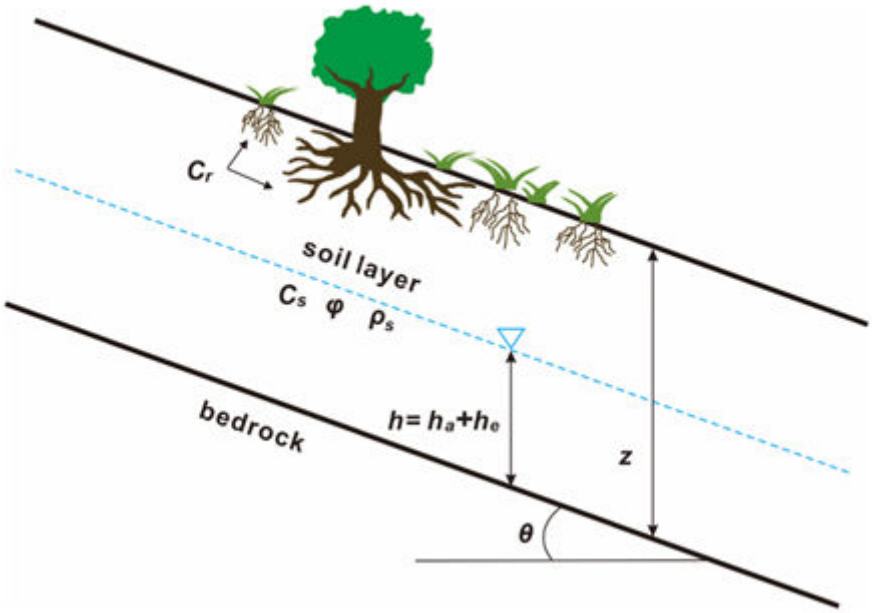

Figure 4.1: Schematic of the geotechnical model
Medina et al. (2021)

those two parameters is also supposed to be Gaussian (Medina et al., 2021). The probability distribution function is as follows:

$$\chi\left(\mu_{FS}, \sigma_{FS}^2\right) = \frac{\chi\left(\mu_c, \sigma_c^2\right)}{\gamma_s z \cos\theta \sin\theta} + \left(1 - \left(\frac{h}{z}\right)\left(\frac{\gamma_w}{\gamma_s}\right)\right)\left(\frac{\chi\left(\mu_{\tan\phi}, \sigma_\phi^2\right)}{\tan\theta}\right) \tag{4.2}$$

Where:

$$\mu_{FS} = \frac{\mu_{\tan\phi}}{D} + \frac{\mu_c}{A} \tag{4.3}$$

$$\sigma_{FS}^2 = \frac{\sigma_{\tan\phi}^2}{D^2} + \frac{\sigma_c^2}{A^2} \tag{4.4}$$

$$A = \frac{z\gamma_s \sin\left(2\theta\right)}{2} \tag{4.5}$$

$$D = \frac{\tan\theta}{1 - \left(\frac{h}{z}\right)\left(\frac{\gamma_w}{\gamma_s}\right)} \tag{4.6}$$

In these equations, χ represents the probability density function of the computed FS, μ the mean of a parameter and σ^2 the variance of a parameter. In addition, in the previous set of equations, the densities (ρ) have been substituted by the bulk weight (γ) of the materials in $\frac{N}{m^3}$.

Having obtained a distribution for FS as shown in Equation 4.2 allows for the calculation of PoF as the area under the curve that is below FS = 1. Depending on the uncertainty of the stochastic parameters, when comparing the results for two cells, one of them can have a higher FS while also having a bigger PoF as shown in Figure 4.2.

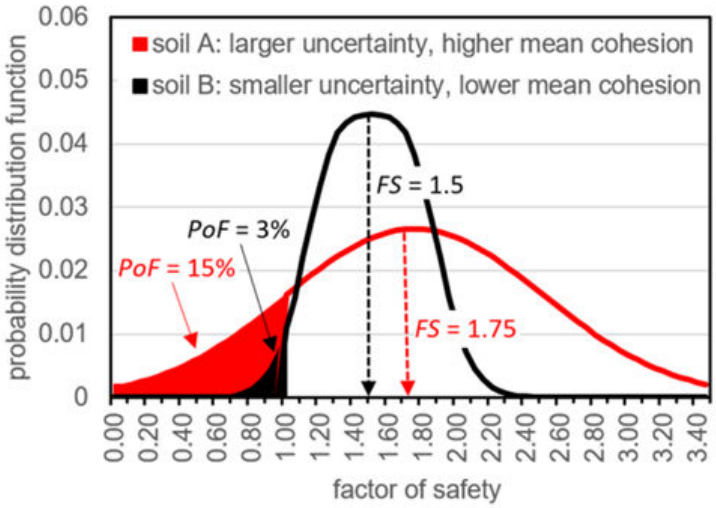

Figure 4.2: Probability distribution for the computed FS
Medina et al. (2021)

Hydrological model

As it can be seen in Figure 4.1, the position of the water table depends on the addition of two parameters.

$$h = h_a + h_e \tag{4.7}$$

These parameters are understood to be the contributions by lateral (h_a) and vertical (h_e) flows as shown in Figure 4.3. Lateral flow refers to the long term steady state contribution of groundwater to a specific cell during a previous time window, while vertical flow refers to the contribution by an event that occurs during a single day.

The contribution by lateral flow uses the following equation:

$$h_a = \left(\frac{a}{b}\right) \frac{q_a}{K \sin\theta \cos\theta} \left(\frac{\rho_w}{\rho_s}\right) \tag{4.8}$$

Where:

a = drainage area in m^2
b = width of the cell in m
q_a = effective antecedent recharge in mm/day
K = horizontal hydraulic conductivity in m/s

The contribution by vertical flow uses the following equations:

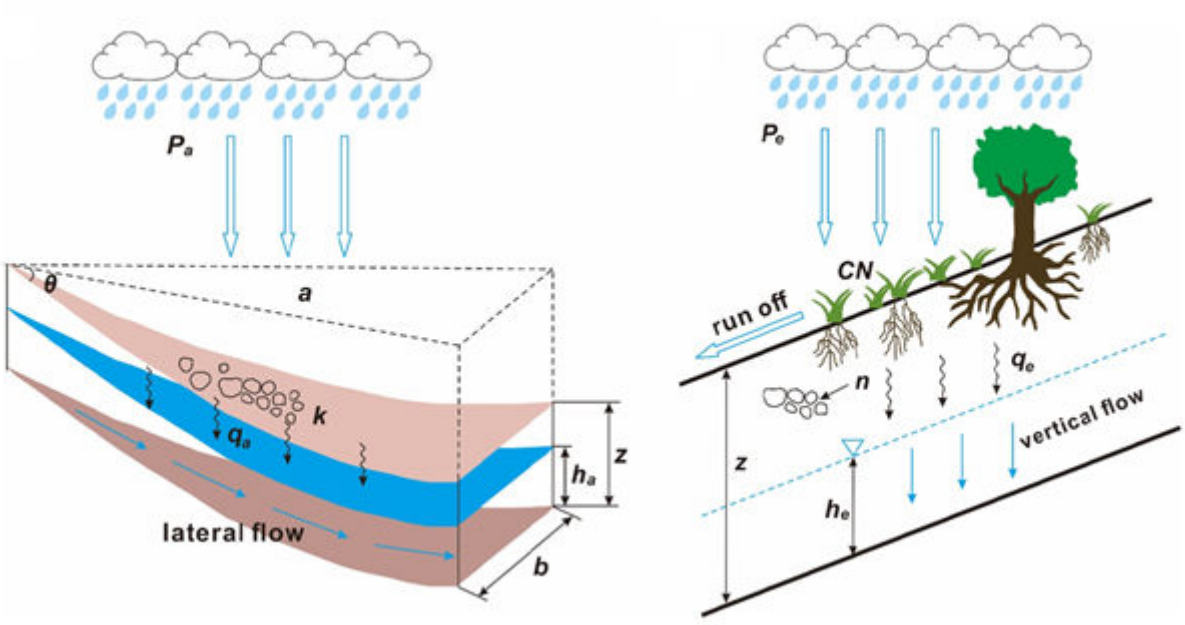

Figure 4.3: Schematic of the hydrological model
Medina et al. (2021)

$$q_e = P_e - \frac{\left(P_e - \left(\frac{5080}{CN} - 51\right)\right)^2}{P_e + 4\left(\frac{5080}{CN} - 51\right)} \tag{4.9}$$

$$h_e = \frac{q_e}{\eta} \tag{4.10}$$

Where:

q_e = efective event recharge in mm
P_e = event rainfall in mm
CN = curve number[1]
η = soil porosity in $\frac{m^3}{m^3}$

4.1.2 Data-driven classification methods

Classification

Classification is the process of predicting a qualitative response (Y) from a single or multiple observations (X). In the case of landslide susceptibility, this is a special type of classification where the outcome is binary, therefore the response value is always either 0 (False) or 1 (True), either the analyzed unit is susceptible (1) or not susceptible (0) to generate a landslide.

In general, the relationship that is being modeled is shown in Equation 4.11 and can be read as the probability of Y equals 1 given X. Because what is modeled is a

[1] as described in the rainfall-runoff model by the Soil Conservation Service (SCS) of the United States

probability, a threshold must be set in order for the model to decide to which class an observation belongs, usually this threshold is set at 0.5.

$$p = Pr(Y = 1|X) \tag{4.11}$$

Logistic regression

In logistic regression, p is modeled using the logistic function which has the following equation when the outcome variable has k predictors, x being the explanatory variable or variables.

$$p = \frac{e^{\beta_0+\beta_1 x_1+\cdots+\beta_k x_k}}{1 + e^{\beta_0+\beta_1 x_1+\cdots+\beta_k x_k}} \tag{4.12}$$

Where each β is a regression coefficient for each of the explanatory variables. With a little bit of manipulation, one can arrive to the following expression:

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k \tag{4.13}$$

Equation 4.13 shows that the logistic function is a form of a linear function in logistic space. In addition, on the left side of the equation are the logarithm of the odds, or the ratio of the number of events that produce a True response to the number that do not.

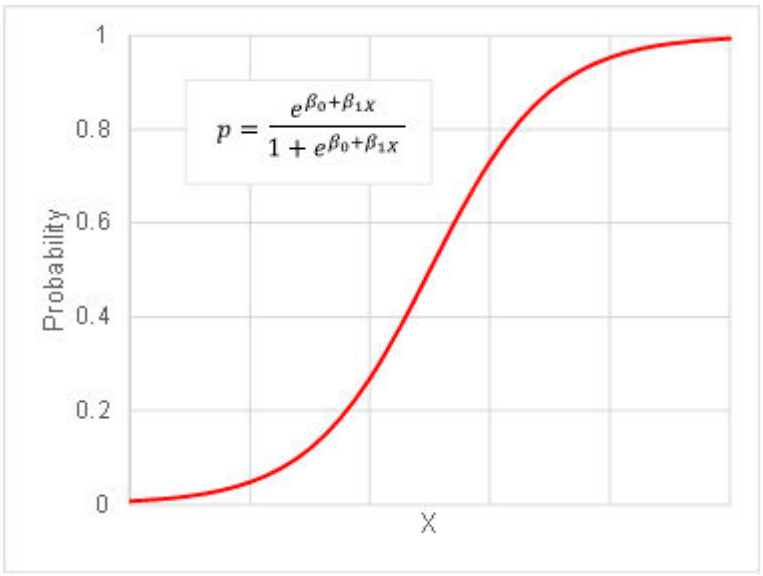

Figure 4.4: General form of the logistic curve for one observation

Support vector machines

A support vector machine (SVM) constructs a hyper-plane or set of hyper-planes in high-dimensional space, which can be used for regression or classification. This method is a generalization of the maximal margin classifier. In Figure 4.5 a data

set is being divided by the maximal margin classifier where the black line represents the plane that maximally separates each subset and the distance between the black line and dotted lines represents the maximal margin.

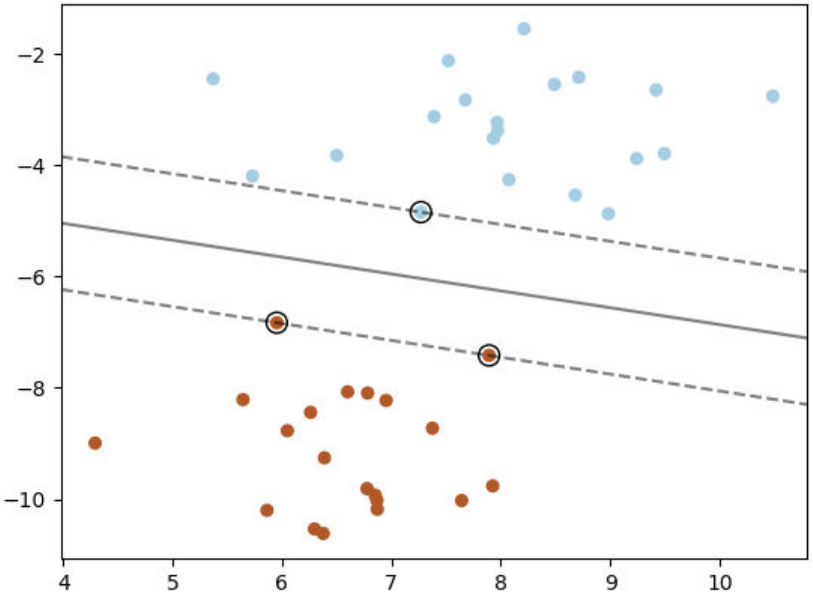

Figure 4.5: Example of the maximal margin classifier in two dimensional space Pedregosa et al. (2011)

The main issue with the maximal margin classifier is that the method doesn't work if there is no hyperplane which perfectly separates the observations. To allow for the misclassification of observations, the support vector classifier (SVC) is used, this method is also known as the soft margin classifier.

The SVC allows for some observations to be on the wrong side of the hyperplane. The optimal hyperplane can be found by solving the following optimization problem. In this case $x_1, ..., x_n$ are the observation variables and $y_1, ..., y_n$ are the class labels.

$$MAX_{\beta_0,\beta_1,...,\beta_p,\epsilon_1,...,\epsilon_n,M} M \tag{4.14}$$

$$\sum_{n=1}^{p} \beta_j^2 = 1 \tag{4.15}$$

$$y_i \left(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip}\right) \geq M(1 - \epsilon_i) \tag{4.16}$$

$$\epsilon_i \geq 0, \qquad \sum_{i=1}^{n} \epsilon_i \leq C \tag{4.17}$$

Where M is the margin space that separates the observations (space between dotted lines in Figure 4.5, β are the coefficients that define the hyperplane, ϵ are slack

variables that allow for observations to be on the wrong side of the hyperplane, and C is a nonnegative tuning parameter. In general Equation 4.16 is the most important equation because it defines the hyperplane that separates the data.

Classification trees

Tree based methods involve segmenting the observation space into a number of simple regions or nodes. Each region is associated with a statistical descriptor of the observations that fall within, typically the mean is used. The set of rules used to describe the segmented observation space can be summarized in what is called a decision tree. Regression and classification trees are types of decision trees where the predicted value for the former is a number while the latter is a class. Tree-based methods are simple and useful for interpretation, however they are typically not competitive with other methods for supervised learning such as those described earlier in this chapter (James et al., 2013).

In order to grow a classification tree, the process of recursive binary splitting is applied. In this approach, observations are divided into two subsets or branches. Selecting the value (s) that divides both branches is done by minimizing a measure of the observations within each branch. Such measures can be the classification error rate, entropy or the Gini index. The Gini index (G) is defined as a measure of total variance across all classes (K) and can be calculated using the following equation:

$$G = \sum_{k=1}^{K} \hat{p}_{mk} \left(1 - \hat{p}_{mk}\right) \tag{4.18}$$

Where $\hat{p_{mk}}$ represents the proportion of observations in the mth subset that belong to the kth class. The Gini index can also be seen as a measure of "purity" where small values indicate that a node contains mainly observations belonging to the same class.

From the previous description, one can assume that, as more branches are introduced, the Gini index of every terminal node in the decision tree will become smaller. This is an issue because trees with more terminal nodes are biased towards the observed data used to create them and perform poorly when used in other data sets. Because of this, the process of minimal cost-complexity pruning is introduced in order to reduce the number of branches and terminal nodes.

Cost-complexity pruning adds a weight (α) to the measure used to evaluate the tree depending on the number of terminal nodes that the tree has. For this process the misclassification rate (E) is used instead of the Gini index.

$$E = 1 - \max_{k} \left(\hat{p}_{mk}\right) \tag{4.19}$$

So the optimization function in order to find the appropriate weight can be written as:

$$MIN \quad E_{\alpha} = E + \alpha \left|T\right| \tag{4.20}$$

Where E_α is the adjusted misclassification rate, E is the misclassification rate of the complete tree and T is the number of terminal nodes in the tree.

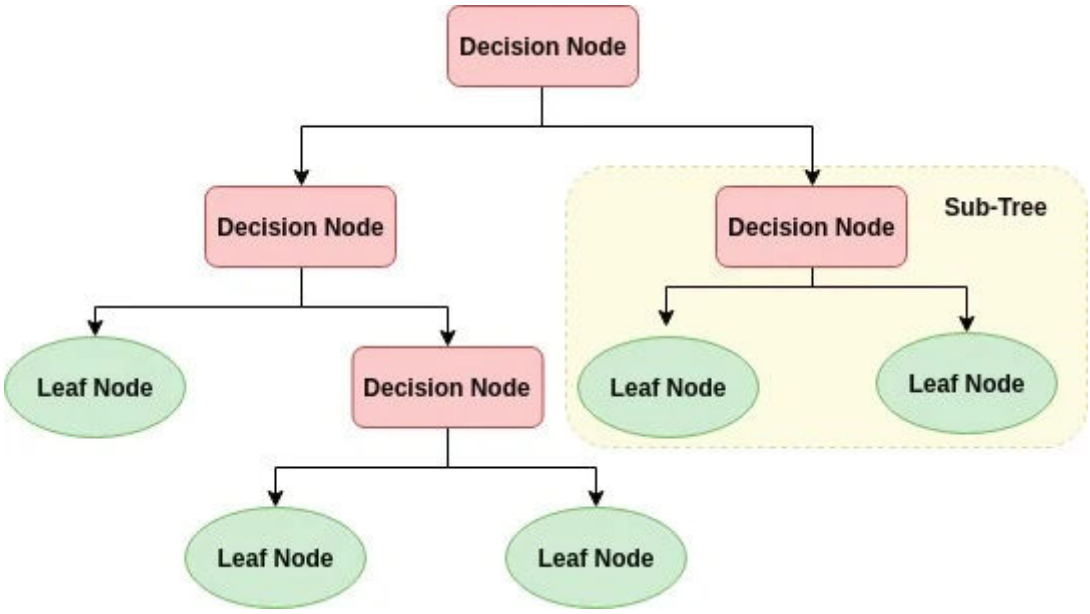

Figure 4.6: General form of the decision tree

Random forests

A process called bootstrapping consists of the creation of multiple copies of the original data set by sampling with replacement. Random forests are made up of a committee of multiple decision trees that have been prepared using bootstrapped data sets (Hastie et al., 2009). The process of splitting is also different than with regular decision trees. When building a decision tree for a random forest, each time a split or branch is considered, a random sample of observations is chosen as split candidates from the full set of observations. This random sample is also a subset of the total observations where typically the square root of the total observations is used (taken as the smallest integer value that is bigger than or equal to the square root).

Because random forests operate as a committee, detailed analysis of each of the members of the committee can lead to deeper insights about the observations or predictor variables. In these applications, seldom are predictor variables equally important. In most cases, only a small group of them are relevant and the others can be excluded without significantly affecting the performance of a model. In random forests, feature importance can be analysed by counting the number of times (and depth within a tree) that a particular feature is used to split the observations and create new branches.

4.2 Methodology

4.2.1 Overview

The analysis, preparation and evaluation of the data-driven models was done using the Python programming language (Van Rossum & Drake, 2009) through the

JupyterLab platform which is an extension of Jupyter Notebooks (Thomas et al., 2016). Specifically, the models were built using the scikit-learn package (Pedregosa et al., 2011). To prepare the data for scikit-learn, the DataFrame object in the pandas package was used (Reback et al., 2021). For the computation of the explanatory variables derived from the DEM, geo-processes from the SAGA-GIS software were used (Conrad et al., 2015).

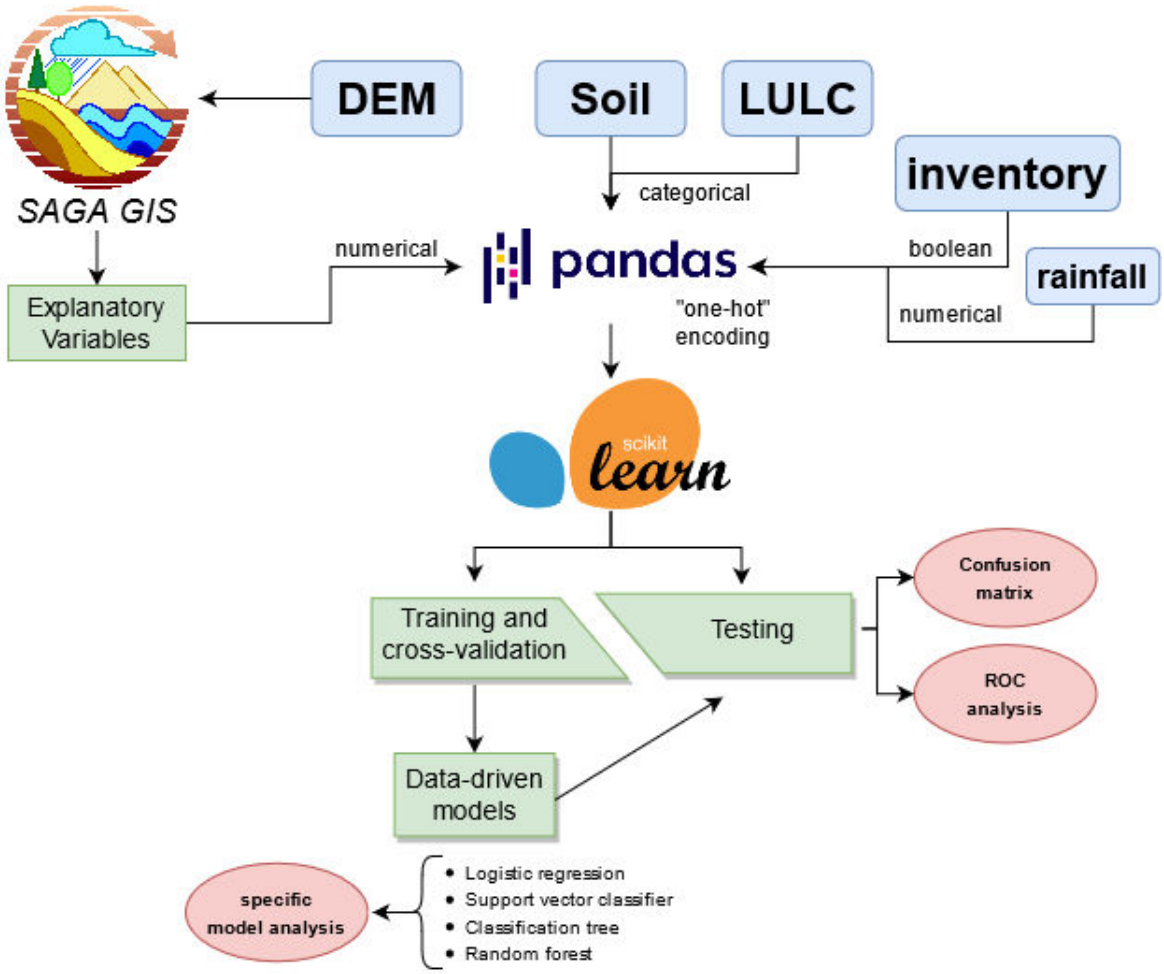

Figure 4.7: Workflow for the preparation and evaluation of data-driven models

Figure 4.7 shows a general overview of the workflow for preparing and evaluating the data-driven models used in the project. Inputs are shown in blue, intermediary steps are shown in green and final outputs are shown in red.

From the DEM, six explanatory variables were computed. These new variables as well as the antecedent and event rainfall of a sampled raster cell are collected within a DataFrame object as numerical variables. Soil and LULC are collected as categorical variables and "one-hot" or dummy encoded within the DataFrame. Dummy encoding refers to the process of coding a categorical variable into dichotomous variables, or variables stored in multiple fields as True or False. In addition, if a cell is marked by a point in the landslide inventory, an attribute of True for susceptibility is collected in the DataFrame, otherwise, the cell is marked as False.

The DataFrame was used by scikit-learn to separate the data into training and testing sets, train and validate the model using the training data, and finally test

the model with the testing data. The final outputs are performance metrics of each of the models and specific results depending on the features that each model has.

4.2.2 Explanatory variables

As mentioned previously, six explanatory variables are calculated from the DEM using the geoprocessing toolbox available within SAGA GIS. These six variables are:

1. Slope: the angle between the surface of the terrain and a horizontal plane measured in degrees.
2. Aspect: the direction to which the surface of the terrain faces measured in degrees.
3. Profile curvature: curvature parallel to the slope. Is a dimensionless ratio that indicates if the terrain surface is convex (positive) or concave (negative).
4. Planform curvature:curvature perpendicular to the slope. Is a dimensionless ratio that indicates if the terrain surface is laterally convex (positive) or laterally concave (negative).

A graphic representation of the curvature variables can be seen in Figure 4.8. The specific algorithm used to calculate these first four variables was described by Zevenbergen and Thorne (1987).

5. Flow accumulation: the total area that drains to a particular cell measured in square meters calculated using a D8 flow direction algorithm as described by Tarboton (1997).
6. Topographic wetness index (TWI): dimensionless ratio that serves to quantify topographic control of hydrological processes. Can be calculated using Equation 4.21 where: a is the drainage area of the specific cell (flow accumulation) and b is the local slope.

$$TWI = \ln\left(\frac{a}{\tan(b)}\right) \tag{4.21}$$

Apart from these six, seven additional explanatory variables were also used. In total, thirteen explanatory variables were collected.

7. LULC class: the LULC map shown in 3.5 contains 10 distinct classes which represent different types of land use or land cover. An index has been assigned to each of the classes.
8. Soil class: the soil map shown in Figure 3.6 contains 11 distinct classes which represent different types of soil. An index has been assigned to each of the classes.
9. Antecedent rainfall: average daily precipitation during the previous 30 days before the analyzed condition. Measured in mm/day.
10. Event rainfall: total daily rainfall associated with the landslide event. Measured in mm.

Table 4.1: LULC classes

Index	**Class**
1	forest
2	shrubs
3	grassland
4	bare soil
5	scree
6	weathered rock
7	intact bedrock
8	urban area
9	water
10	glacier-snow

Table 4.2: Reclassified soil classes

Index	**Class**
1	alluvial
2	colluvium
3	conglomerate
4	granitic rock + quartzite
5	hornfels-marble
6	limestone
7	mudstone
8	phyllite-slate
9	sandstone
10	scree
11	till

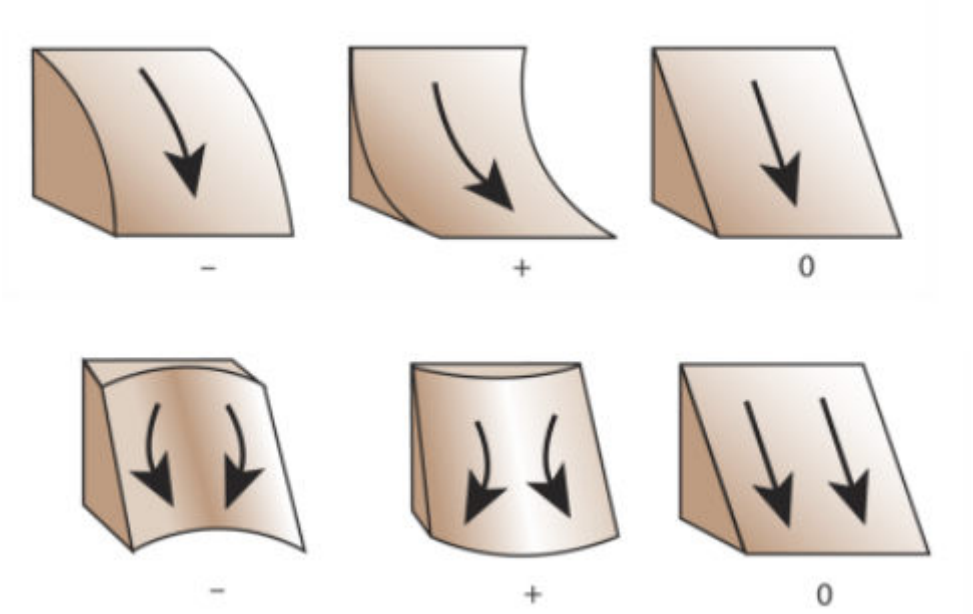

Figure 4.8: Profile curvature (top row) and planform curvature (bottom row) ESRI (2014)

11. Event probability of failure: PoF of cells as the result of the combined effects of antecedent and event rainfall.
12. Dry probability of failure: PoF of cells where the water table has been set at its lowest possible position (the soil column is dry).
13. Saturated probability of failure: PoF of cells where the water table has been set at its highest possible position (the soil column is completely saturated).

The last three explanatory variables had been previously computed by FSLAM (Hürlimann et al., 2021).

Different explanatory variables were used in different models as is shown in Table 4.3. For all cases, explanatory variables 1 through 8 were used. Because rainfall is sometimes difficult to obtain, variables 9 and 10 were used as a particular case for all models. This last point allows two cases to be evaluated: the first which is a static or "dry" case, and the second which considers the effects of antecedent and event rainfall as failure conditions. Lastly, variables 11 through 13 were used to evaluate how outputs from FSLAM affect the performance of the best model when they are added as inputs.

4.2.3 Sampling

As mentioned in Section 3.1.2, for the 2013 landslide episode 392 landslides where identified and collected to form an inventory, the spatial distribution of the inventory is also shown in Figure 3.2. Figure 4.9 shows the output of the code written to collect the sample. It is to be noted that even though the identified landslides were 392, only 391 are collected in the sample. This is because two of the landslides in the inventory were separated at a distance of less than five meters. Because five meters is the resolution of the DEM, these two landslides are collected as just one observation.

Table 4.3: Explanatory variables used in different models

Cases		
Dry	**2013 Event**	**Model coupling**
Morphometric variables	Morphometric variables	Morphometric variables
LULC class	LULC class	LULC class
Soil class	Soil class	Soil class
	Antecedent rainfall	Event PoF
	Event rainfall	Dry PoF
		Saturated PoF

Additionally, this is a balanced sample, that is to say, an equal amount observations for susceptible and non-susceptible cells was used. The final DataFrame for the data-driven models had 782 observations.

```
True cells:  391
False cells:  12987031
391  have been sampled from the false cells.
Therefore, the final DataFrame for the data-driven model has: 782 observations.
```

[7]:

	lat	long	elevation	slope	aspect	curv_plan	curv_prof	facc	twi	soil	lulc	rain_ant	rain_event	pof_event	pof_dry	pof_sat	response
144779	4738872.5	311597.5	790.799988	13.145604	327.057495	0.025769	-0.010359	1.397940	4.673251	2.0	3.0	0.5	49.0	0.0000	0.0000	1.0000	1.0
152975	4738862.5	311617.5	794.770020	18.441339	307.935059	-0.117006	-0.044545	2.352183	6.514342	2.0	3.0	0.5	49.0	0.0000	0.0000	1.0000	1.0
289083	4738697.5	319872.5	1709.359985	43.968288	173.810425	0.000644	0.018706	2.889302	6.688886	8.0	1.0	0.5	70.0	0.8803	0.3543	0.1197	1.0
993416	4737832.5	318912.5	1285.219971	47.644943	194.897537	0.076285	0.069331	0.000000	-11.605381	8.0	1.0	0.5	69.0	0.6894	0.5387	0.1381	1.0
1362535	4737377.5	317017.5	1734.319946	39.489166	182.852234	-0.000287	-0.003942	3.122216	7.382731	8.0	1.0	0.5	67.0	0.6695	0.0711	0.3305	1.0

Figure 4.9: First five rows of the collected sample DataFrame

In addition to the explanatory variables presented in Section 4.2.2, the properties of latitude and longitude in the UTM zone 31N coordinate system were also collected, allowing for the sample to be presented in a map as shown in Figure 4.10

4.2.4 Train-test split

The training dataset contains examples used to fit the parameters of a specific classifier. Because this is an exercise in supervised learning, each observation is paired with its corresponding output. During the learning process, the parameters of the classifiers are adjusted in order to best match the observations with their corresponding output. The performance of the classifier is evaluated by comparing the predicted outputs with the actual values for all observations.

The code used to split the sample into the training and testing sets is as follows:

```
# Split X and y into training and testing sets
from sklearn.model_selection import train_test_split
X_train, X_test,
y_train, y_test
  = train_test_split(X, y, test_size=0.25, random_state=0)
```

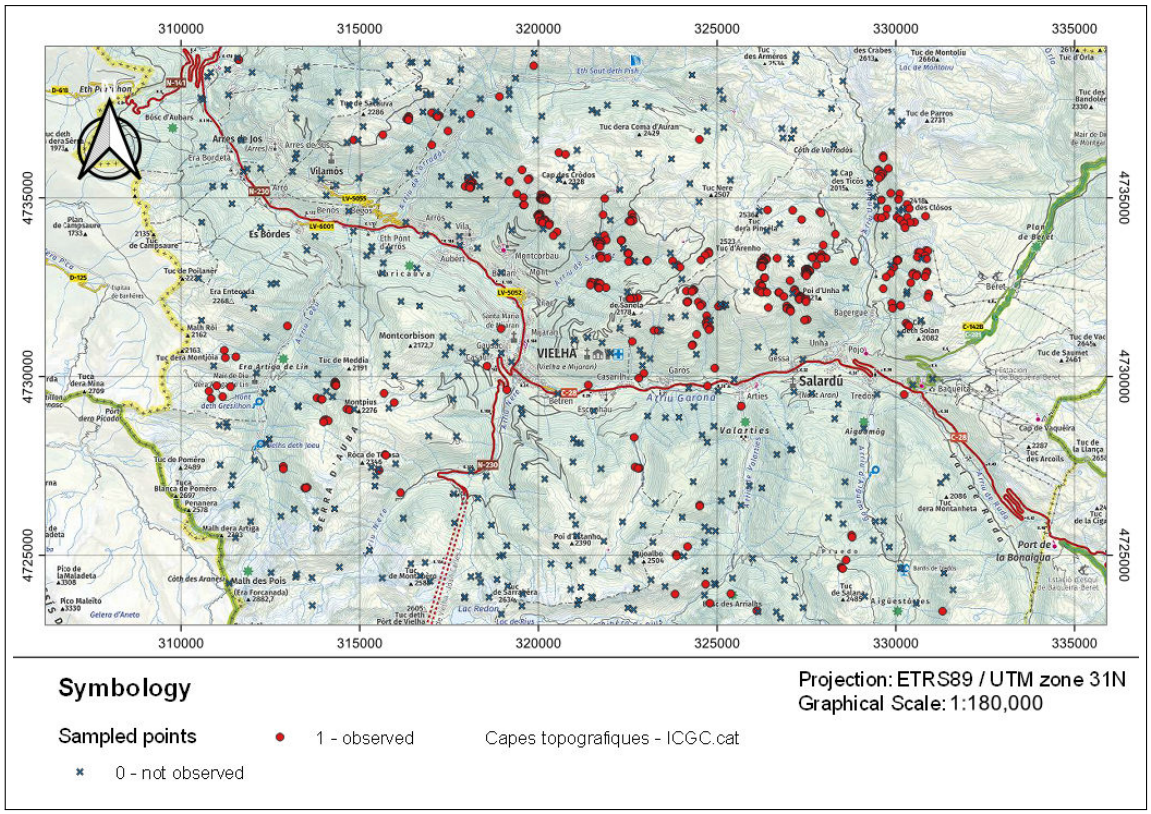

Figure 4.10: Training data or landslide sample map

Here scikit-learn (sklearn) is being used to perform the task of randomly splitting the sample. As is shown in Figure 4.11, a split of 75% of the data is being used to train the models while 25% of the data is used to test the models. The testing data can also be seen as "out of sample" data and is used to evaluate the performance of the model.

4.2.5 Cross-validation

In order to accurately assess the performance of the model during training, the process of 5-fold cross-validation is used. In general for k-fold cross-validation, the original sample is randomly partitioned into k equal sized subsamples. Of the k subsamples, a single subsample is left out of the training process and used to test the model. This process is repeated k times, always leaving out one of the samples. The final performance metric during training is reported as the average of the performance across all the folds. As an example, the subsamples are shown in Figure 4.11 for k = 5.

Hyper parameter tuning

In addition to being used to evaluate the performance of the model during training, the validation dataset is also used for hyper-parameter tuning. As implemented in scikit-learn, each model has a specific set of parameters that affect it's performance. Adjustment of these parameters will lead to better or worse performance of the model during training. One must also not forget that perfect performance during training is undesirable because it means that the model is overfitting.

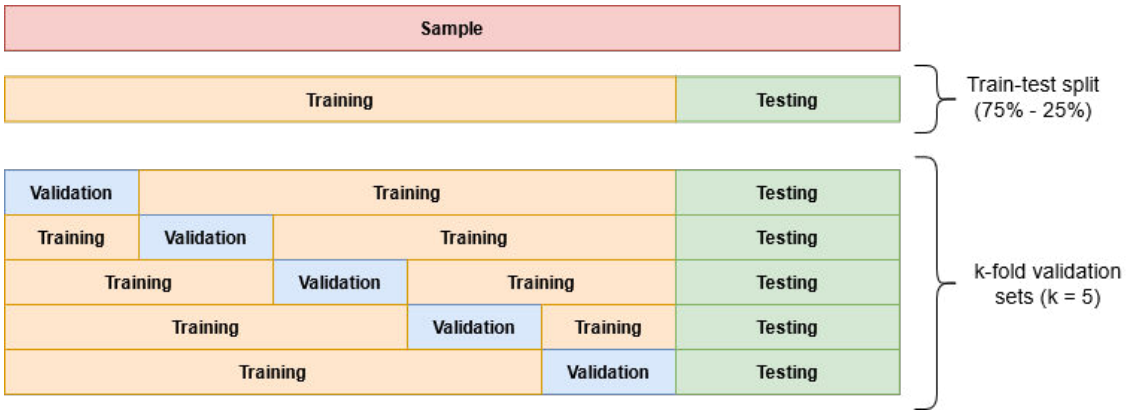

Figure 4.11: Partitions of the sample data

As an example, the following codes performs 5-fold cross-validation and hyper-parameter tuning for the logistic regression:

```
from sklearn.linear_model import LogisticRegression
from sklearn.model_selection import GridSearchCV
logreg = LogisticRegression()
param_grid = {'penalty' : ['l1', 'l2'],
              'C' : np.logspace(-4, 4, 20)}
clf = GridSearchCV(logreg, param_grid = param_grid,
      scoring = 'roc_auc', cv = 5)
best_logreg = clf.fit(X_train, y_train)
```

The third and fourth lines create an instance of the logistic models and the ranges of hyperparameters that have been chosen to tune the model. In the case of the logistic regression, two types of penalty functions are being evaluated as well as 20 values for "C" the regularization parameter. In this case, the regularization term is added to the optimization function that finds the coefficients of the logistic regression in order for discourage the model if it finds a very good but complex solution. More complex solutions lead to overfitting so simpler solutions are preferred because they lead to models that are better at generalizing.

The fifth line create an instance of a GridSearch object. This object creates a cross-validated model for each of the possible combinations of hyperparameters given in the parameter grid. Because the hyperparameter grid contains 20 parameters and 5-fold cross-validation is being used, in total, 200 models are trained. The model with the highest performance in the "ROC AUC" metric is then saved, along with its performance metrics, as the best model.

4.2.6 Testing

In this last process, the trained and tuned data-driven model is used to predict probabilities and classes for the testing data set. The predicted classes are computed from the predicted probabilities by setting a threshold in order to decide to which of classes a cell belongs. Usually this threshold is set at 0.5 where raster cells with a predicted probability below or equal to 0.5 are classified as False (or not susceptible to generate a shallow landslide in our specific case) or True when the predicted probability is more than 0.5.

From these results, specific performance metrics can be computed for each of the models. These metrics are then used to compare data-driven models between themselves and also to choose the best data-driven model and compare it with FSLAM.

4.3 Evaluation metrics

In order to evaluate the models, the confusion matrix, metrics derived from the confusion matrix and the receiver operator charactersitic (ROC) curve are used.

4.3.1 Confusion matrix

The confusion matrix is a specific contingency table that allows for the visualization of the performance of a model. The rows of the matrix account for the actual values the model tried to predict while the columns account for the predictions made by the model. Figure 4.12 shows the typical layout of a confusion matrix.

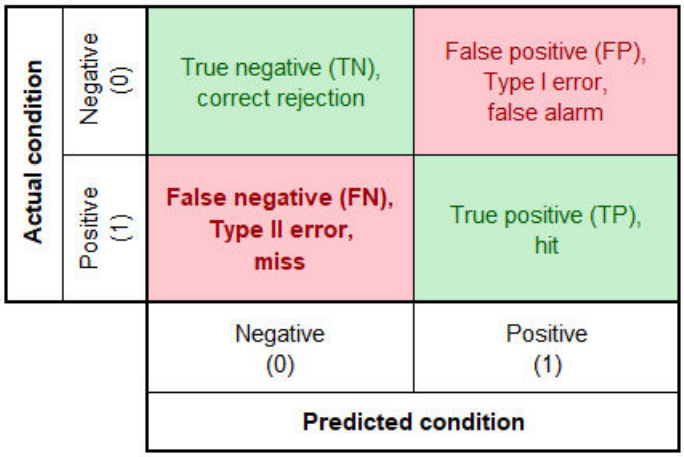

Figure 4.12: Generic layout of a confusion matrix

The confusion matrix is a very powerful tool because it allows for the computation of different performance metrics depending on the model evaluated. The terms and metrics derived from the confusion matrix are described below.

P = number of real positive cases
N = number of real negative cases
TP = number of true positives or hits
TN = number of true negatives or correct rejections
FP = number of false positives or false alarms (Type I error)
FN = number of false positives or misses (Type II error)

Type I and II errors are very important in landslide assessment. A Type I error or a false alarm means that a cell an area is declared landslide susceptible, limiting its possibilities for development. While a Type II error or miss would mean that an area is declared non-susceptible, allowing for its development but a landslide and its ensuing disaster still occur.

Typically, classification models metrics focus on the capacity of the model to correctly predict positives and negatives, so the accuracy (ACC) of the model is reported. Accuracy is the ratio between correctly predicted cases and the total number of cases. This is a number between 0 and 1, and values closer to 1 indicate better performance.

Because of the importance of the two types of error for landslide assessment, the false positive rate (FPR) and the false negative rate (FNR) are also reported. FPR is the ratio between false positives and total number of negative cases. FNR is the ratio between false negatives and the total number of positive cases. Both are number between 0 and 1 where values closer to 0 indicate better performance.

Finally, as a measure of overall performance from the confusion matrix, the Matthews correlation coefficient (MCC) is used (Matthews, 1975). MCC takes into account true and false positives and negatives and is considered a balanced measure of performance. MCC is a number between -1 and 1 where 1 represents perfect classification, 0 indicates performance no better than a random predictor and -1 indicates total disagreement between the actual and predicted responses.

$$ACC = \frac{TP + TN}{P + N} = \frac{TP + TN}{TP + TN + FP + FN} \tag{4.22}$$

$$FPR = \frac{FP}{N} = \frac{FP}{FP + TN} \tag{4.23}$$

$$FNR = \frac{FN}{P} = \frac{FN}{FN + TP} \tag{4.24}$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)\,(TP + FN)\,(TN + FP)\,(TN + FN)}} \tag{4.25}$$

4.3.2 ROC curve

The ROC curve is a plot that illustrates performance metrics of a binary classifier as the discrimination threshold is varied (Fawcett, 2006). As mentioned previously, the weakness of the performance metrics described in Section 4.3.1 is that the confusion matrix depends on a threshold that is set. Performance of a model will vary depending on the discrimination threshold; the ROC curve permits an assessment of the performance of the model across all possible discrimination thresholds.

In order to plot the ROC curve, an additional metric in the form of the true positive rate (TPR) has to be computed. TPR is the ratio between true positives and the total number of positive cases. It can also be computed using the previously established FNR. As a performance metric, it takes values between 0 and 1 where being closer to 1 is better.

$$TPR = \frac{TP}{P} = \frac{TP}{TP + FN} = 1 - FNR \tag{4.26}$$

In a way, the ROC curve is simultaneously displaying the two types of errors for all possible thresholds. Type I error in the x-axis with the FPR and Type II error in the y-axis through the TPR.

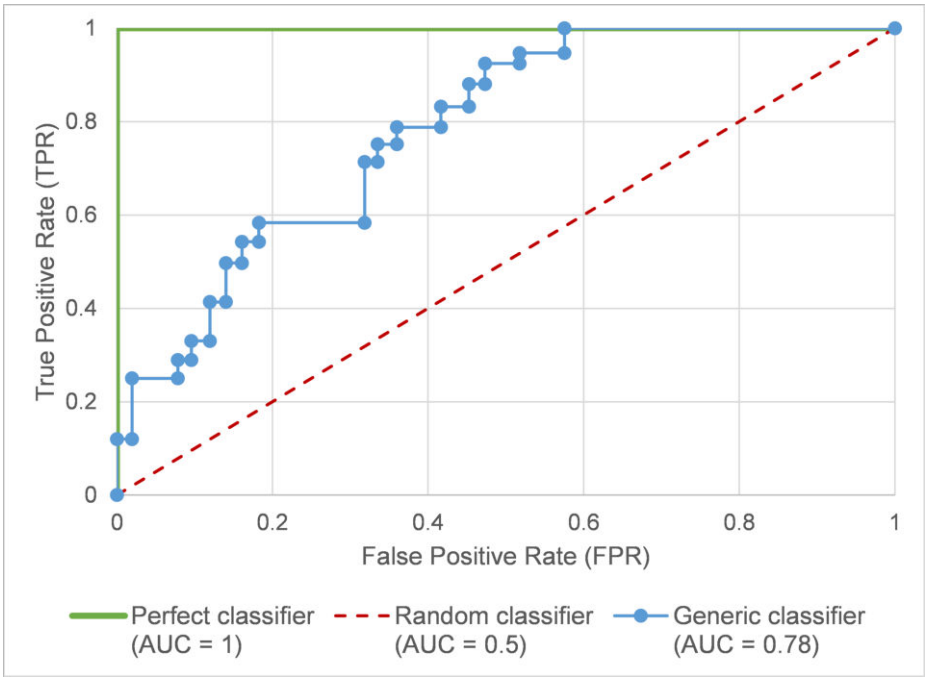

Figure 4.13: Generic ROC curve

The overall performance of a classifier, summarized over all possible thresholds is given by the area under the curve (AUC). An ideal ROC curve will have a single point where FPR = 0 and TPR = 1, thus the AUC is equal to 1. Further, for any random classifier FPR will be equal to TPR across all thresholds and thus AUC is equal 0.5.

Chapter 5

Results and Discussion

In this chapter the results obtained by the physically based model FSLAM and the different data-driven models are presented. The results are presented in two cases: the first case doesn't consider neither antecedent rainfall nor event rainfall as explanatory variables and the second case does. The results are presented in these two cases due to the fact that, in literature, data-driven models are used to compute landslide susceptibility mainly in a static condition; that is to say, the explanatory variables used are commonly static in time (Reichenbach et al., 2018). Another important consideration is that data-driven models typically use historical landslide inventories that have been collected throughout the years while, in the case of Val d'Aran, an inventory for a single landslide episode is being used.

5.1 FSLAM

FSLAM serves as a starting point and baseline to assess the performance of all data-driven models. For this model, the case that doesn't consider rainfall can be seen as the condition in which the soil column is completely devoid of water or *dry*. This condition is computed by FSLAM through Equations 4.1 and 4.2 by setting $\frac{h}{z} = 0$. This condition is also able to detect *Unconditionally Unstable* cells or cells that have a high PoF and will fail regardless of the height of the water table. The second case which considers rainfall is taken as the calibrated model as prepared by Hürlimann et al. (2021) and PoF after considering the increases in the water table by the antecedent and event rainfalls is taken as the final result for this case.

The ROC curves in Figure 5.1 show clear dissimilar performance between the two conditions. The 2013 event ROC AUC is equal to 0.762 which should be taken as the final performance metric for FSLAM in this region. The dry condition shows a lower performance mainly due to the fact that the inventory being used to calibrate is the result of a landslide episode triggered mainly by snowmelt and rainfall. As a general statement, for a region like Val d'Aran the condition in which there is no water in the soil column is unrealistic so this result should be taken only as reference. The ROC curve of the *POF Dry* case also shows significant variation across thresholds making the results of any possible confusion matrix very sensible to the threshold set to distinguish non-susceptible and susceptible cells. This also shown in Figure 5.2 where both cases show significantly different performance metrics mainly because of

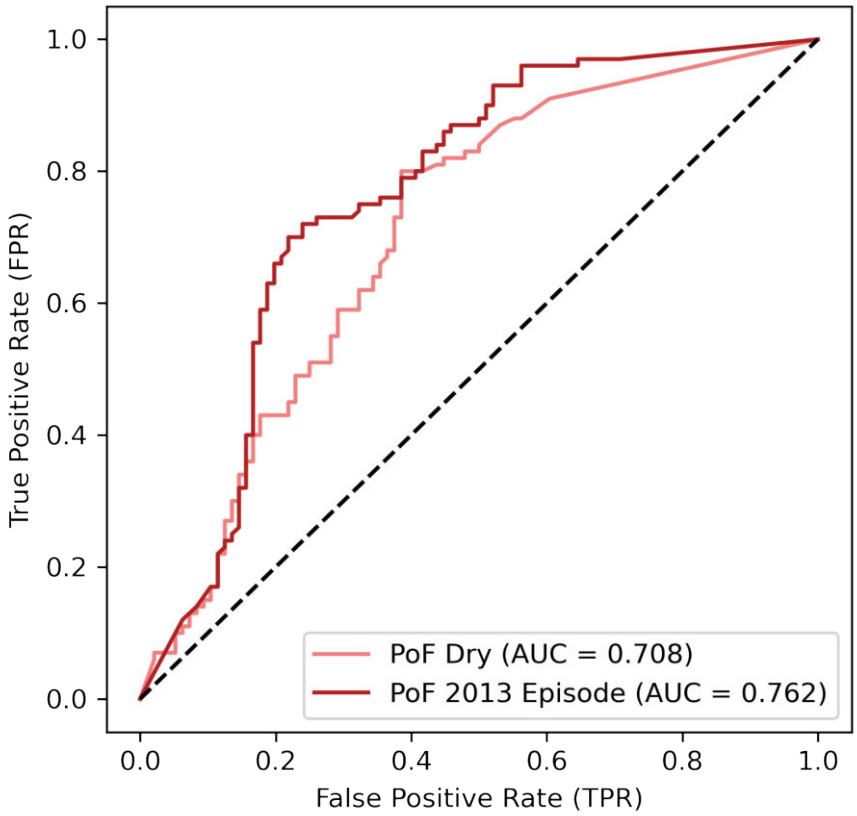

Figure 5.1: ROC curves for FSLAM

the sensitivity of the *POF Dry* case to a discrimination threshold.

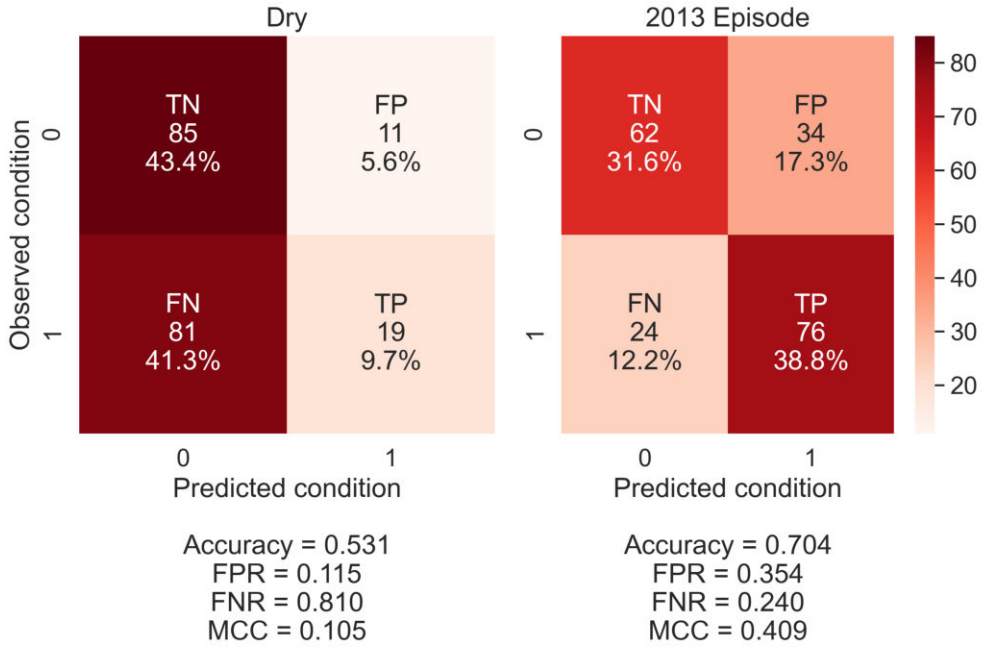

Figure 5.2: Confusion matrices for FSLAM

5.2 Logistic regression

The first statistical method to be tested is the logistic regression. As explained in Section 4.1.2, the logistic regression is similar to linear regression in the logarithmic space. An optimization problem will find adequate coefficients for the explanatory variables in order to obtain the best performance during the training phase.

As discussed in Section 4.2.5, the optimization problem implemented in scikit-learn to find adequate parameters for the logistic regression has a series of hyperparameters that need to be adjusted mainly to improve performance while also taking into account that overfitting the model during training must be avoided. The main hyperparameter of the logistic regression is C which serves as a regularization parameter. For the case of no rainfall $C = 0.61585$, while for the case that considers rainfall $C = 29.76351$. This means that the case which uses rainfall is a more complex model that requires heavier penalties to avoid overfitting. This is also shown in Table 5.1 and Table 5.2 where the latter uses more of the explanatory variables to make a prediction, therefore it is a more complex model. The final performance scores are shown in Figure 5.3 where the model that considers rainfall performed better overall by a slight margin. Figure 5.3 also shows the increase of performance as the hyperparameters are adjusted to find better coefficients that more suitably match the training data and the spread in performance across all five validation folds (shaded area).

In the testing set, performance is increased for both cases of the logistic regression, this is shown in Figure 5.4. This means that both models have avoided the pitfalls of overfitting. An even better result, is that the model which considers no rainfal shows better performance in the testing dataset. This is important because it confirms that simpler models tend to be better a generalizing than more complex models.

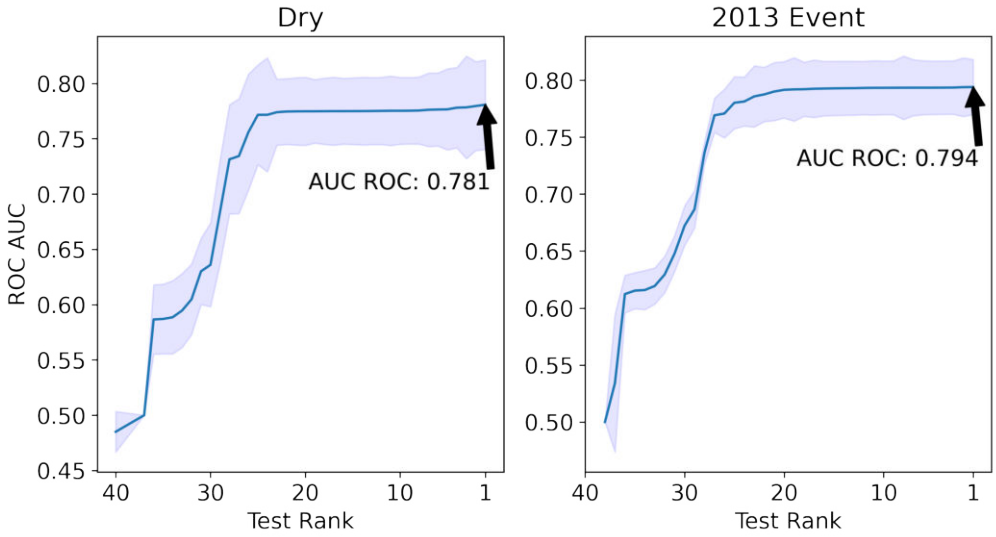

Figure 5.3: Ranking and performance of logistic models during the training phase

Although adding rainfall as an explanatory variable increased the performance of the model during training, when making predictions in "out of sample" data the more complex model performed slightly worse. This is also shown in Figure 5.5, where the model that considers rainfall predicted a higher number of false negatives and there FNR increased. This is worrying because, as mentioned in Section 4.3.1, FNR is associated with Type II error which in the case of landslides means that an area was predicted as not susceptible when in reality it might be capable of generating a landslide.

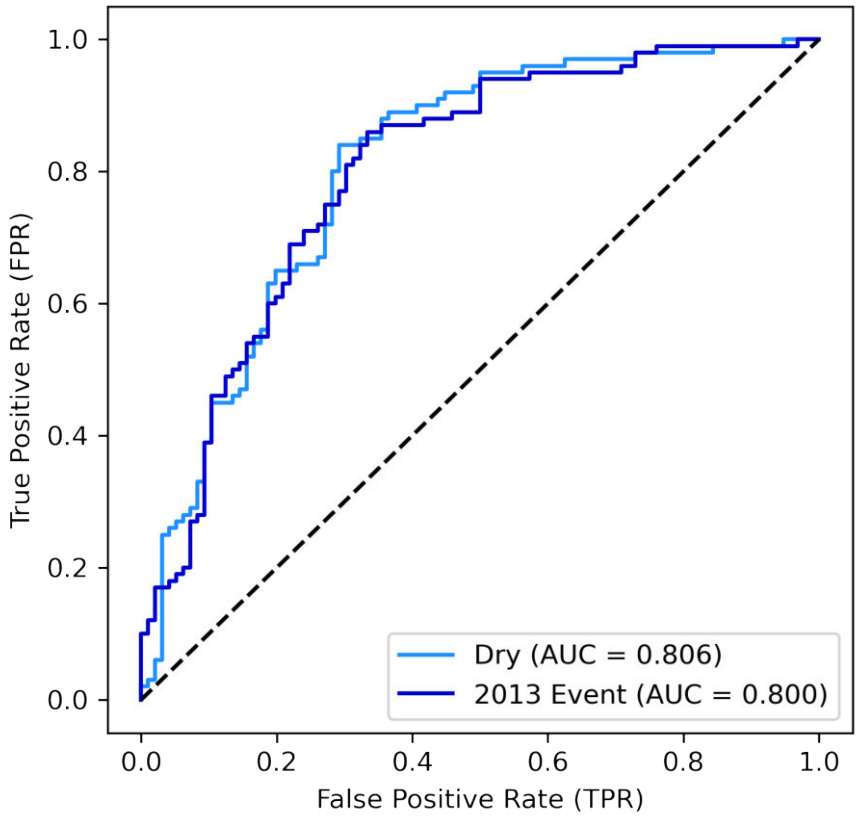

Figure 5.4: ROC curves for the logistic models

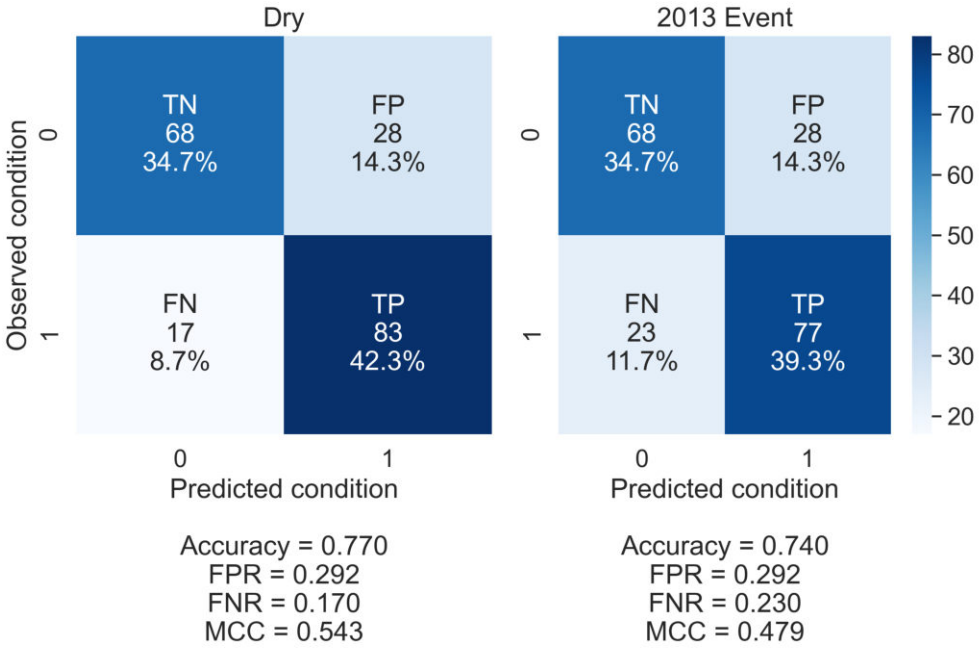

Figure 5.5: Confusion matrices for the logistic models

As a final, more in-depth, analysis, the coefficients for both cases of the logistic models as well as p-values for each coefficient are shown in Table 5.1 and Table 5.2. The tables are ranked in descending order using the absolute z-score of each explanatory variable. z-score was computed as the ratio between the estimator's mean coefficient and its standard error, and p-values are the probability to obtain a result that satisfies the null-hypothesis, with the null-hypothesis being that that a certain explanatory variable is not relevant in explaining landslide susceptibility. An explanatory variable is considered relevant only when it can be used to predict landslide susceptibility with 95% confidence ($p < 0.05$). To estimate these coefficients and their corresponding mean, standard error, z-scores and p-values, cross-validation across 100-folds of the training data was used considering the already adjusted hyperparameters for best performance.

When comparing the rankings of explanatory variables one must notice the switch that happens to the land uses of *urban area* and *bare soil* when the 2013 rainfall is considered. When rainfall is not considered, these two land uses are not statistically significant when making a prediction. In contrast, when rainfall is considered, *urban area* becomes the most important variable when predicting landslide susceptibility. Intuitively, this makes sense. When rainfall is considered as the triggering mechanism, one would not expect areas that have been covered in roofs, pavement and concrete to generate landslide and this is proven by the coefficient having a negative sign, meaning that it reduces the computed probability. A similar analysis can be done for *bare soil* where the coefficient is positive, therefore areas with this type of land use will be more susceptible to landslides when rainfall is considered. The same happens for the soil categories of *conglomerate* and *till* but their explanation is less intuitive and visual.

From the variables derived from the DEM, *slope* is really the only one that is relevant when making a prediction even though *aspect* is considered by the model that doesn't take into account rainfall and, *profile curvature* and *flow accumulation* are used by the model that takes into account rainfall.

Both models agree that five categories of *soil, mudstone sandstone, granitic rock + quartzite, colluvium* and *limestone*, and two categories of land use, *forest* and *grassland*, are statistically significant. Both models also agree that *TWI* and *planform curvature* have no statistical relevance when making predictions, as well as the five land uses of *scree, weathered rock, intact bedrock, water* and *glacier-snow*; and the soil category of *phyllite-slate*.

Table 5.1: Coefficients of the logistic regression model for the Dry case

Explanatory Variable	**Coefficient Mean**	**Coefficient Std. Error**	**Z-Score**	**P-Value**
soil_7.0: mudstone	1.067385	0.023105	46.196907	0.00000
slope	0.048418	0.001102	43.953031	0.00000
lulc_1.0: forest	-1.001251	0.027576	36.308878	0.00000
soil_9.0: sandstone	-1.504874	0.045763	32.883967	0.00000
lulc_3.0: grassland	0.518134	0.022576	22.950719	0.00000
soil_4.0: granitic rock + quartzite	-1.092249	0.059073	18.489752	0.00000
soil_2.0: colluvium	0.430253	0.031423	13.692295	0.00000
soil_6.0: limestone	0.225259	0.033071	6.811482	0.00000
soil_5.0: hornfels-marble	-0.285079	0.082732	3.445813	0.00028
aspect	-0.000239	0.000096	2.488322	0.00642
facc	-0.035788	0.017726	2.019032	0.02174
soil_1.0: alluvial	-0.044902	0.056739	0.791375	0.21436
twi	0.001732	0.002821	0.614132	0.26956
soil_10.0: scree	-0.001465	0.011652	0.125763	0.44996
curv_plan	0.000000	0.000000	NaN	NaN
curv_prof	0.000000	0.000000	NaN	NaN
lulc_2.0: shrubs	0.000000	0.000000	NaN	NaN
lulc_4.0: bare soil	0.000000	0.000000	NaN	NaN
lulc_5.0: scree	0.000000	0.000000	NaN	NaN
lulc_6.0: weathered rock	0.000000	0.000000	NaN	NaN
lulc_7.0: intact bedrock	0.000000	0.000000	NaN	NaN
lulc_8.0: urban area	0.000000	0.000000	NaN	NaN
lulc_9.0: water	0.000000	0.000000	NaN	NaN
lulc_10.0: glacier-snow	0.000000	0.000000	NaN	NaN
soil_3.0: conglomerate	0.000000	0.000000	NaN	NaN
soil_8.0: phyllite-slate	0.000000	0.000000	NaN	NaN
soil_11.0: till	0.000000	0.000000	NaN	NaN

Table 5.2: Coefficients of the logistic regression model for 2013 Event case

Explanatory Variable	**Coefficient Mean**	**Coefficient Std. Error**	**Z-Score**	**P-Value**
lulc_8.0: urban area	-4.54649	0.10627	42.78244	0.00000
slope	0.04916	0.00116	42.37931	0.00000
lulc_4.0: bare soil	3.35507	0.09936	33.76681	0.00000
soil_7.0: mudstone	0.93828	0.02782	33.72682	0.00000
rain_ant	-3.73607	0.11577	32.27149	0.00000
soil_9.0: sandstone	-1.91055	0.06026	31.70511	0.00000
lulc_1.0: forest	-1.45692	0.04600	31.67217	0.00000
rain_event	0.01608	0.00080	20.10000	0.00000
soil_4.0: granitic rock + quartzite	-1.43823	0.07324	19.63722	0.00000
lulc_3.0: grassland	0.58758	0.04319	13.60454	0.00000
soil_2.0: colluvium	0.47128	0.03522	13.38103	0.00000
soil_6.0: limestone	0.41346	0.03763	10.98751	0.00000
soil_10.0: scree	-1.01183	0.09313	10.86471	0.00000
soil_11.0: till	0.58232	0.06004	9.69887	0.00000
soil_3.0: conglomerate	-1.88070	0.19942	9.43085	0.00000
soil_5.0: hornfels-marble	-1.11465	0.13363	8.34132	0.00000
soil_1.0: alluvial	-0.77378	0.09439	8.19769	0.00000
curv_prof	5.23727	0.66975	7.81974	0.00000
lulc_2.0: shrubs	-0.13790	0.04653	2.96368	0.00152
facc	-0.05409	0.02613	2.07003	0.01922
aspect	-0.00013	0.00010	1.30000	0.09680
curv_plan	-0.20523	0.18122	1.13249	0.12871
twi	0.00502	0.00471	1.06582	0.14325
lulc_6.0: weathered bedrock	0.01563	0.03761	0.41558	0.33886
lulc_5.0: scree	0.00463	0.04819	0.09608	0.46173
lulc_7.0: intact bedrock	0.00000	0.00000	NaN	NaN
lulc_9.0: water	0.00000	0.00000	NaN	NaN
lulc_10.0: glacier-snow	0.00000	0.00000	NaN	NaN
soil_8.0: phyllite-slate	0.00000	0.00000	NaN	NaN

5.3 Support vector classifier

The support vector classifier is the second method tested. As explained in Section 4.1.2, the SVC tries to find the widest possible separating margin between the output categories. This is different than the logistic regression in which a likelihood function is being optimized to best fit the data. SVC is a more complex method than the logistic regression but it was chosen to be tested because it is known to be a very good "out of the box" classifier, so hyperparameter tuning is not as crucial and this method offers an alternative to the logistic regression.

Even though hyperparameter tuning is not as important, the same process for training the logistic regression was followed to train the SVC. The kernel function is the most important hyperparameter used by the SVC.

```
param_grid = [{'kernel': ['rbf'], 'gamma': [0.1,0.5,1,2,5,10], 'C':
                                    [0.1, 1, 10, 100, 1000]},
              {'kernel': ['linear'], 'C': [0.1, 1, 10, 100, 1000]},
              {'kernel': ['poly'], 'degree' : [2,3,4,5], 'C': [0.1,
                                    1, 10, 100, 1000]}]
```

For the SVC, three different kernel functions were tested. The purpose of the kernel function is to transform the data. This new form increases the dimensionality of the problem but in these higher dimension, finding the hyperplane that separates the different categories of data is easier. In a sense, it is adding information to the problem.

As with the logistic regression C serves as the regularization parameter. For the case with no rainfall considered, a linear kernel function with a regularization parameter of $C = 0.1$ was found to be best. For the case that uses rainfall, a a polynomial kernel function of the 2nd degree with a regularization parameter of $C = 1000$ was best.

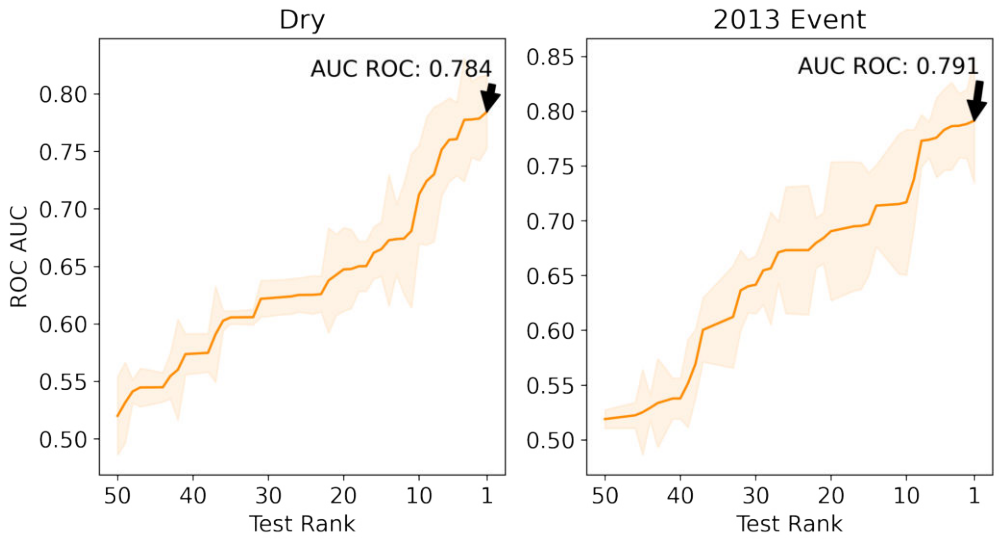

Figure 5.6: Ranking and performance of the SVC models during the training phase

As with the logistic regression, in training the more complex model that uses rain shows better performance. Also similarly to the logistic regression, the model with

no rainfall uses a linear kernel function with a low regularization parameter, while the model that uses rainfall uses a polynomial kernel function with a high regularization parameter.

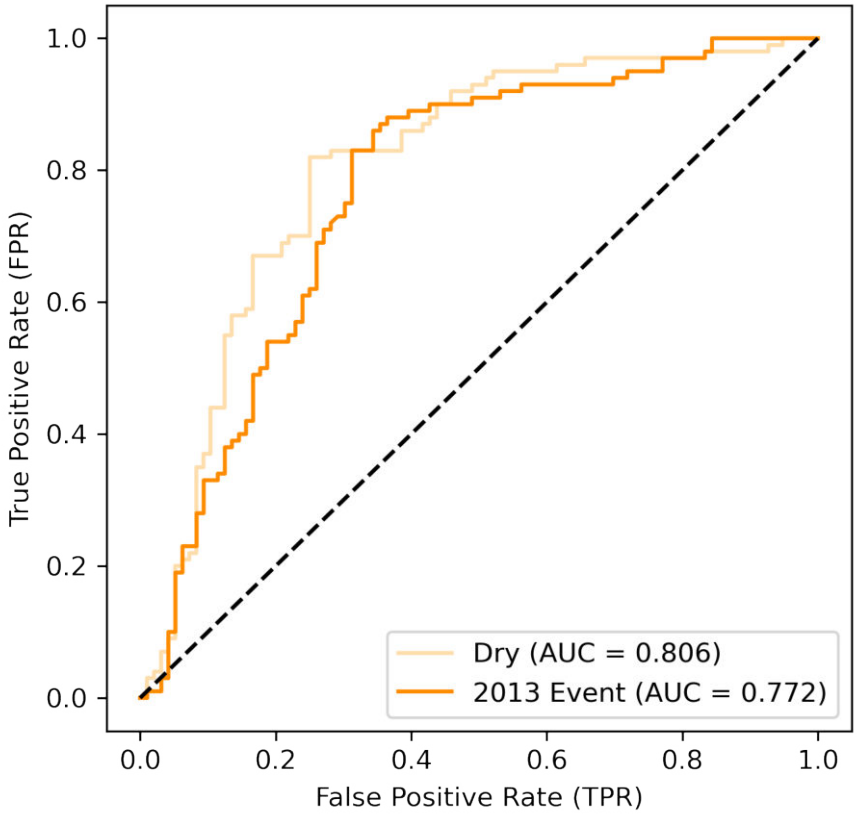

Figure 5.7: ROC curves for the SVC models

In testing, results are again very similar to the logistic regression. As shown in Figure 5.7, the less complex model which doesn't use rainfall in its inputs is better at generalizing and thus it obtains a better score in testing. The confusion matrix in Figure 5.8 shows that the model that uses rainfall improved a little bit by reducing the number of false positives but there is a more significant increase in false negatives.

In the end, because of the similar results between the logistic and SVC models, the former would be preferred because they are more approachable and permit deeper insight into the decision-making process as shown with the analysis of Table 5.1 and Table 5.2. Finally, in terms of pure performance, both methods outperform the physically-based model.

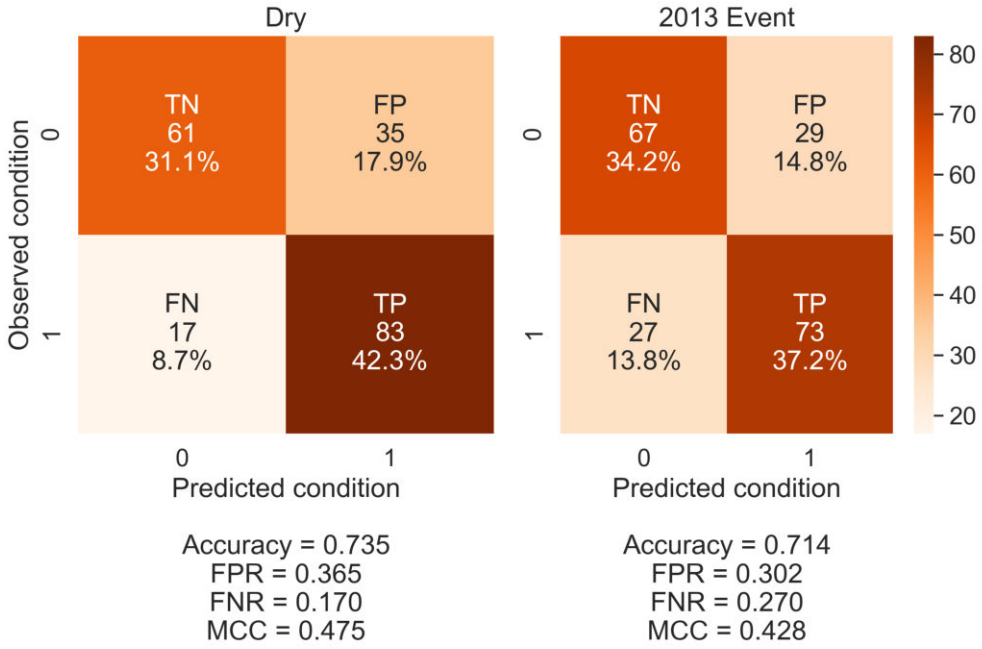

Figure 5.8: Confusion matrices for the SVC models

5.4 Classification tree

To prepare Tree based models, a similar process to the one followed for the logistic regression and the SVC is used. 5-fold cross-validation is again used in order for the results to not be highly sample dependant. To select an adequate CCP α that maximizes performance in terms of accuracy, cross-validation is run through a range of values for α. This process is plotted in Figure 5.9 and α is selected as the value which achieves the highest accuracy.

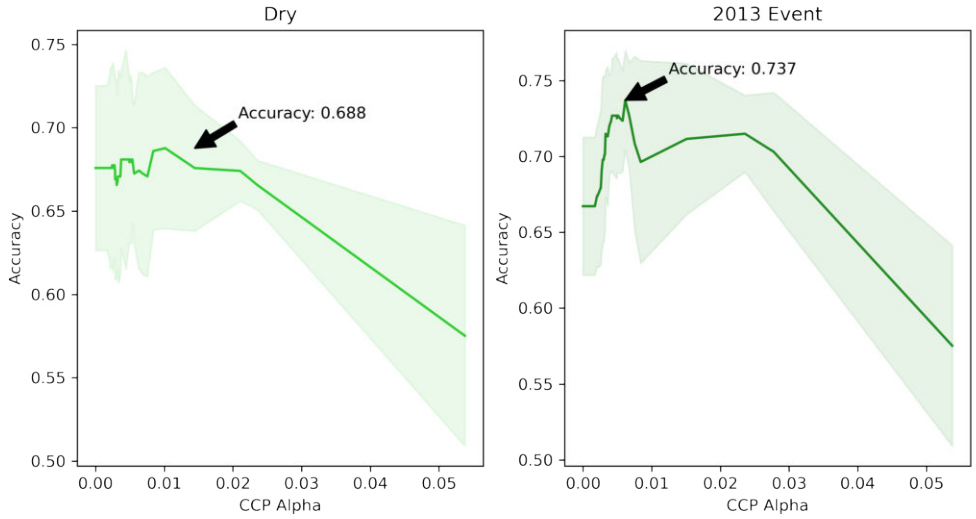

Figure 5.9: Performance of the Tree models during the training phase

During the training process, the best α for the decision tree that doesn't use rainfall

is 0.01 for an accuracy of 0.688. For the decision tree that uses rainfall, α is selected as 0.006 for an accuracy of 0.737. This is a similar result as the one obtained for regularization in the logistic regression and SVC models. A higher α leads to more pruning and smaller, less complex, decision trees. In this case, the addition of rainfall creates a more complex model that requires less pruning in order to achieve its highest performance.

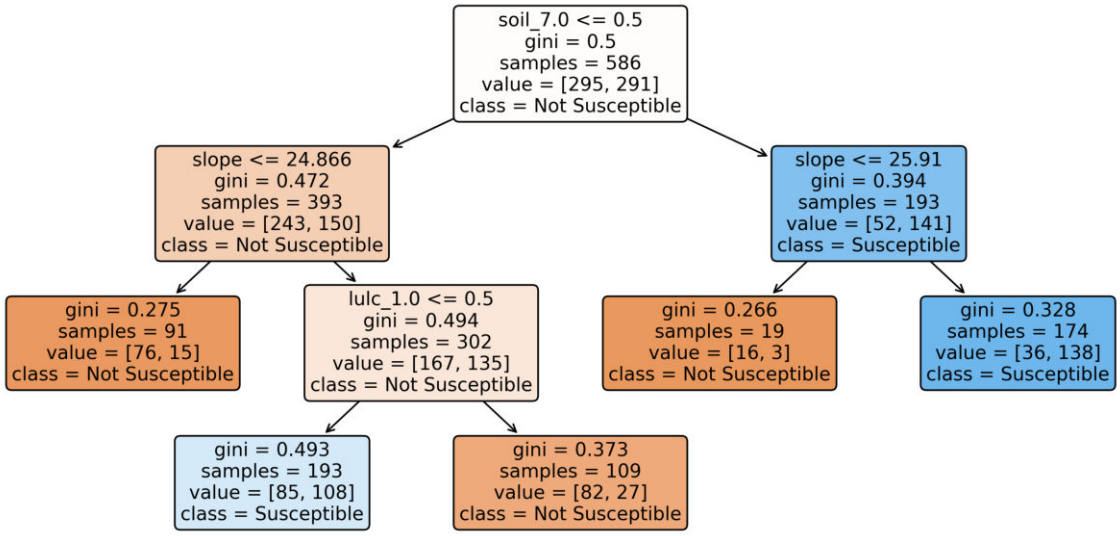

Figure 5.10: Classification Tree for the Dry case

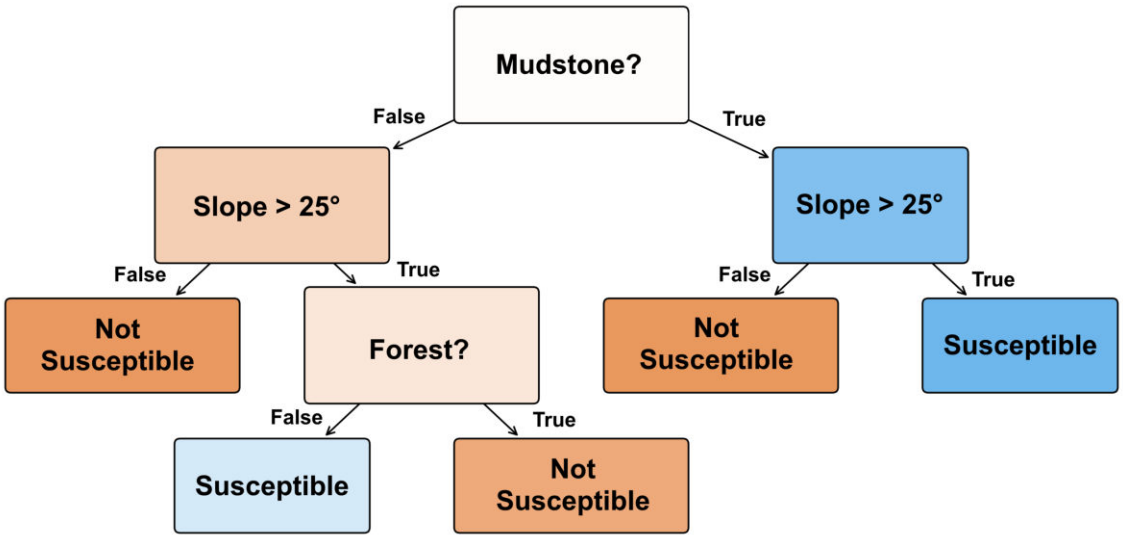

Figure 5.11: Simplified Classification Tree for the Dry case

Figure 5.10 and Figure 5.11 show the decision tree for the no rainfall considered case. The first one shows the decision tree as an output of scikit-learn where the decision rule, Gini purity, samples in this branch, number of data points belonging to each class and the class that comprises the majority of the samples in the branch are shown in each of the tree nodes. The second one shows only the decision rules and the final classes in order to make the decision tree. The motivation to simplify the decision tree is that decision-makers are not necessarily trained to understand

the specifics of these models and a decision tree as shown in Figure 5.10 might seem intimidating and therefore not very useful in the decision-making process. By contrast, the tree in Figure 5.11 is simple, easy to understand and to follow, which makes it more likely to be used.

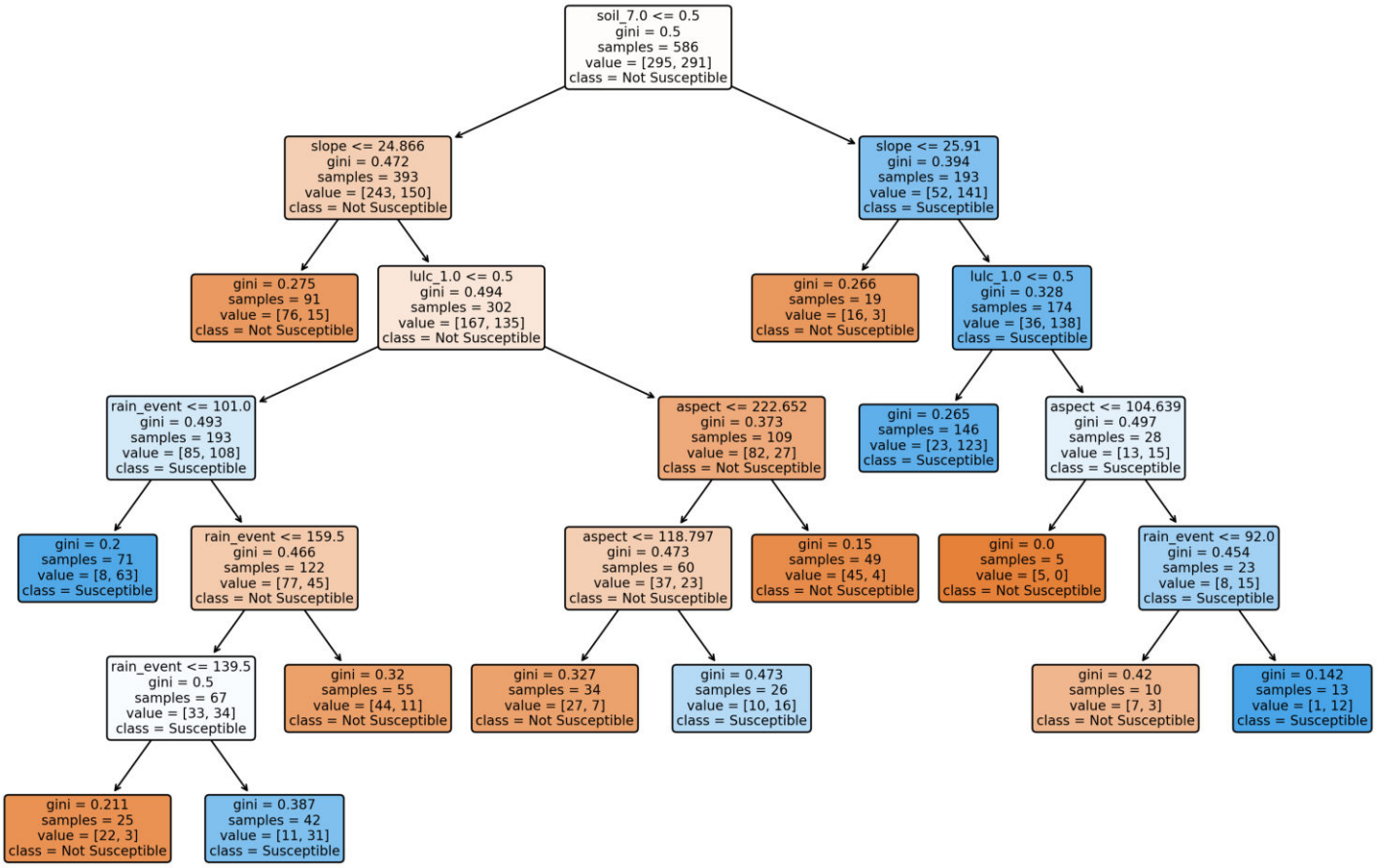

Figure 5.12: Classification Tree for the 2013 Event case

Figure 5.12 shows the decision tree for the case that uses rainfall. As more explanatory variables are added, the size of trees tends to grow because they will consider more inputs in order to achieve a higher performance. This tree tends to be unwieldy especially for decisions which involve the value of event rainfall. In the lower left portion of the tree a series of branches consider different thresholds for event rainfall. Intuitively, one would suppose that higher event rainfall would lead a decision path that classifies an area as *Susceptible* but this is not the case. If the value for event rainfall is above 159.5 mm, the area is classified as *Not Susceptible*. This serves as a reminder that these methods mostly focus on performance, and are not the best when used to give insights to the underlying physical process.

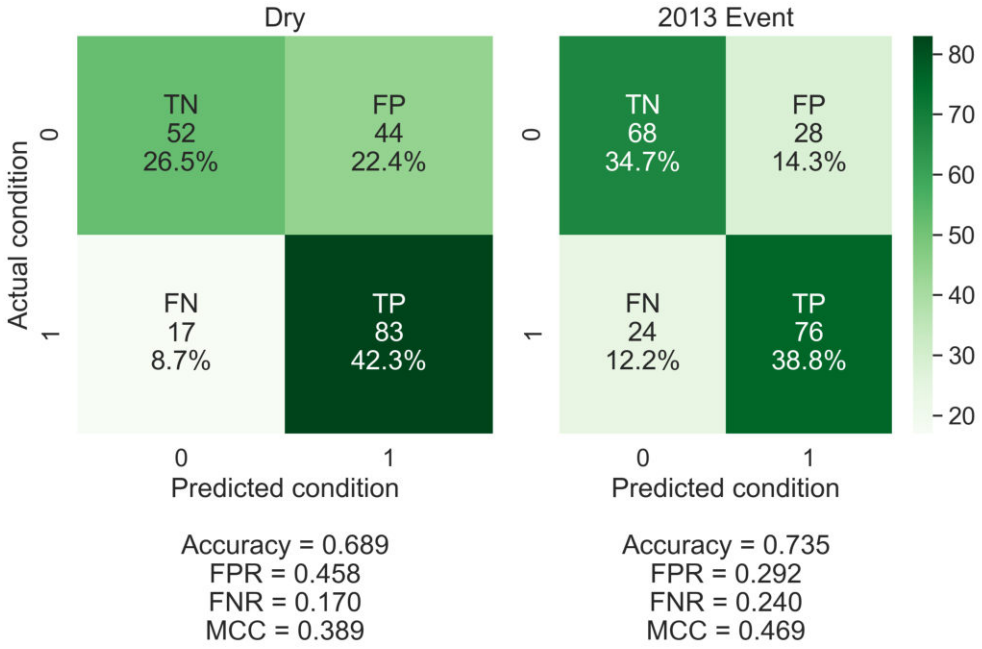

Figure 5.13: Confusion matrices for the Tree models

5.5 Random forest

As with the previous models, a similar approach was used to train the random forest models. This process is more similar to the logistic regression and the SVC because to define the maximum depth for the ensemble of trees, a single hyperparameter has to be adjusted instead of testing different values of α for cost-complexity pruning. The list of hyperparameters tested is shown in the following block of code.

```
param_grid =  {'n_estimators': [400, 450, 500, 550],
               'criterion':['gini','entropy'],
               'max_depth': [15, 20, 25],
               'max_features': ['auto','sqrt', 10],
               'min_samples_leaf': [2,3],
               'min_samples_split': [2,3]}
```

The parameter that defines the depth of the trees is *max_depth* and because of the random sampling that is used to train the random forest, each tree is allowed to have much higher depths than those presented in Section 5.4 for single classification trees. Other important hyperparameters are: *n_estimators* which defines the number of trees in the ensemble, *criterion* which defines which type of impurity measure will be used to create the splits within the trees, *max_features* which us number of features to consider when looking for the best split (auto meaning the squared root of the total amount of features), *min_samples_leaf* and *min_samples_split* which define the minimum number of samples required to define a leaf or a split within each tree.

Although the list of hyperparameters is larger than in other cases, Figure 5.14 shows that overall performance is very similar all across the board. Random forests appear to be not very sensitive to the ranges of hyperparameters defined. The hyperpa-

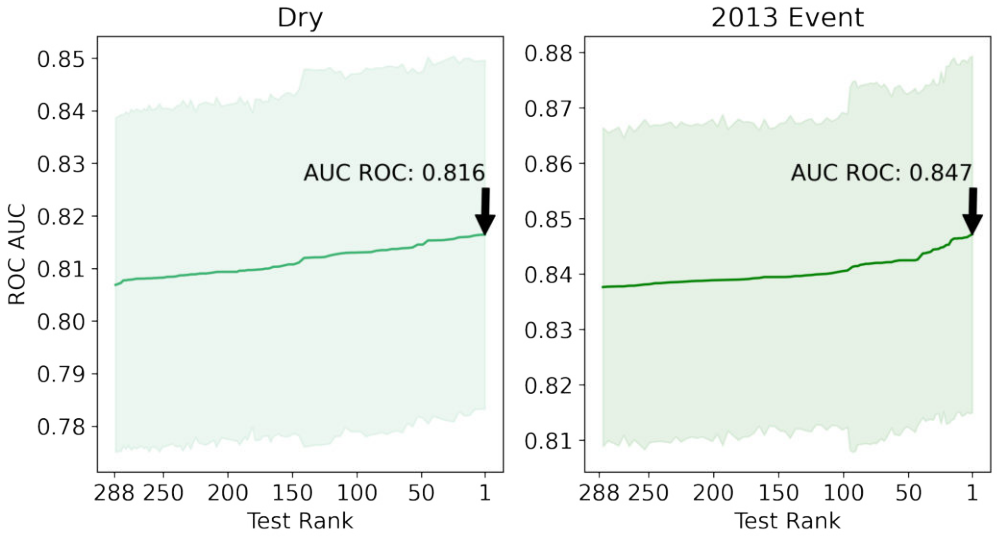

Figure 5.14: Performance of the Random Forest models during the training phase

rameters used for the model that doesn't consider rainfall are: *criterion* = *gini*, *max_depth* = 20, *max_features* = *auto*, *min_samples_leaf* = 3, *min_samples_split* = 2, *n_estimators* = 500 for a final ROC AUC of 0.816. The hyperparameters used for the model that considers rainfall are: *criterion* = *entropy*, *max_depth* = 25, *max_features* = 10, *min_samples_leaf* = 2, *min_samples_split* = 2, *n_estimators* = 450 for a final ROC AUC of 0.847.

In testing, across all thresholds random forests outperform the logistic regression and the SVC as shown inf Figure 5.15. Final AUC ROC scores for the models are 0.837 for the model that doesn't consider rainfall and 0.853 for the model that considers rainfall. It is interesting to note also that for these particular models, the model that uses rainfall again is the more complex one by having a larger *max_depth* hyperparameter but is performing better when generalizing than the simpler model.

As stated in Section 4.1.2, random forests permit an analysis of "feature importance" by counting the number of times that features appear in each of trees in the ensemble and their position within the tree. Figure 5.16 shows the 10 most important features for both cases of the models. This analysis is similar to the one performed for the logistic regression in Table 5.1 and Table 5.2. As with the previous analysis, the land use classes of forest and grassland, as well as the soil class of mudstone prove to be of high relevance to make predictions. Another interesting point is that class variables are found to be of more relevance in the model that doesn't consider rainfall while continuous features such as the variables derived from the DEM are more relevant to the model that considers rainfall. This is important especially for regions where the classification of soil categories is difficult. In these cases, a random forest that only uses features derived from a DEM and rainfall data might prove to be a good compromise. Finally the confusion matrices for each model is shown in Figure 5.17.

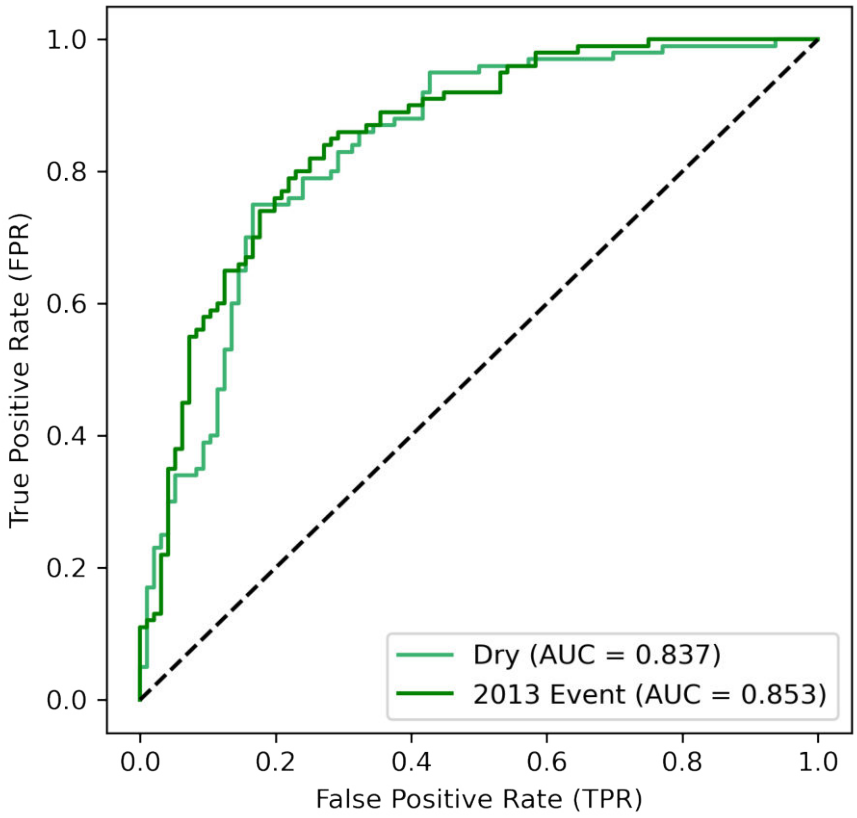

Figure 5.15: ROC curves for the Random Forest models

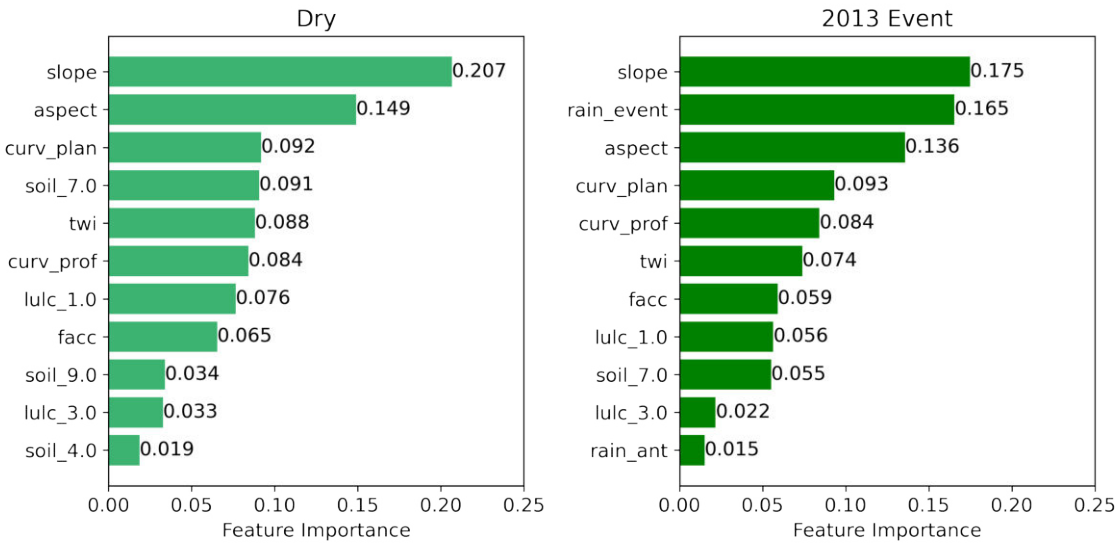

Figure 5.16: Feature importance for the Random Forest models

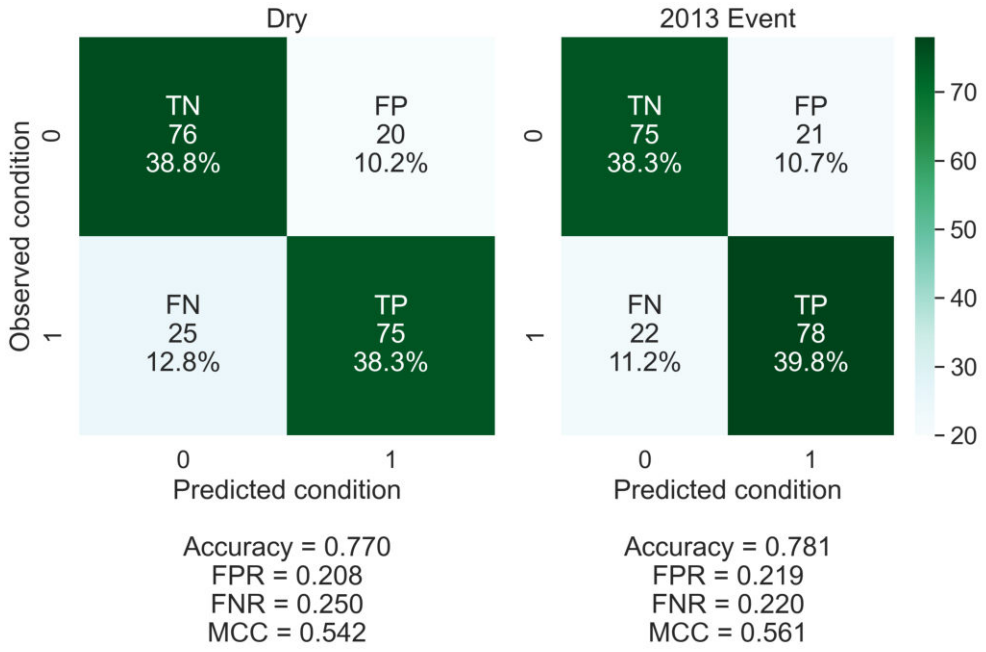

Figure 5.17: Confusion matrices for the Random Forest models

5.6 Performance Summary

In the previous sections, four different types of data-driven models were shown to have been tested in order to assess their performance at predicting landslide susceptible areas in the Val d'Aran region. Every model showed and overall measure of good performance and most of them, the exception being the simple classification tree that doesn't consider rainfall, outperformed FSLAM, the physically-based model.

For the logistic regression and the SVC, a simpler model that didn't take into account rainfall, performed better when generalizing to "out of sample" data. The addition of more information decreased the FPR or the number of false alarms (Type I error) while at the same time increasing the FNR or the number of misses (Type II error). In a real-life application, a miss would have grave consequences such as the approval of a development project in land that is landslide susceptible and thus should be avoided, which means that simpler data-driven models, such as the logistic regression or the classification tree, would be preferred.

Rule-based models, such as the decision tree and random forest showed improvements when the amount of data used to train them was increased. In the case of the simple decision tree, better performance comes at the cost of interpretability and simplicity. The example used shows that the less complex model uses only one category of soil, one category of land use and the value of the slope to determine whether an area is susceptible to landslides or not. This simplicity and interpretability would be useful for a decision-maker who would value a simpler model that can be use as a "rule of thumb". Although the more complex model is harder to read and interpret, it still uses only the slope, aspect, one category of land use and the total event rainfall which makes it highly useful in a real-life situation for a quick or preliminary assessment. The explanatory variables highlighted in these section as well as properties such as the "feature importance" of the random forest also can serve focus the efforts when collecting data by showing which explanatory variables are more important for the models to make a prediction.

In the end, even though the random forest showed the best performance when comparing metrics for all data-driven models, the logistic regression is chosen to do a more in-depth comparison with FSLAM. This choice was made mainly because the logistic regression is the model that is more typically used in the literature to perform landslide assessment when data-driven models are used (Reichenbach et al., 2018). In addition, with the information provided in Table 5.1 and Table 5.2, the model is highly interpretable and easy to understand.

Table 5.3: Performance summary for all models

Model	**Case**	
	Dry	**2013 Event**
FSLAM	ROC AUC = 0.708 Accuracy = 0.531 FPR = 0.115 FNR = 0.810 MCC = 0.105	ROC AUC = 0.762 Accuracy = 0.704 FPR = 0.354 FNR = 0.240 MCC = 0.409
Logistic regression	ROC AUC = 0.806 Accuracy = 0.770 FPR = 0.292 FNR = 0.170 MCC = 0.543	ROC AUC = 0.800 Accuracy = 0.740 FPR = 0.292 FNR = 0.230 MCC =0.479
SVC	ROC AUC = 0.813 Accuracy = 0.755 FPR = 0.323 FNR = 0.170 MCC = 0.514	ROC AUC = 0.772 Accuracy = 0.714 FPR = 0.302 FNR = 0.270 MCC = 0.428
Classification tree	ROC AUC = - Accuracy = 0.689 FPR = 0.458 FNR = 0.170 MCC = 0.389	ROC AUC = - Accuracy = 0.735 FPR = 0.292 FNR = 0.240 MCC = 0.469
Random forest	ROC AUC = 0.837 Accuracy = 0.770 FPR = 0.208 FNR = 0.250 MCC = 0.542	ROC AUC = 0.853 Accuracy = 0.781 FPR = 0.219 FNR = 0.220 MCC = 0.561

5.7 Comparison of Methods

As previously stated, a more in-depth comparison between the physically-based and data-driven approach to landslide susceptibility was performed using FSLAM and the logistic regression models presented in Section 5.2. In order to do the analysis, a new sample was collected from the data of Val d'Aran. This new sample contains the same 391 positive values that belong to the landslide inventory but the negative values have been increased from 391 (balanced sample) to 5000. This means that any further analysis performed might be skewed towards the ability of the models to predict negatives, but the ROC curve shown in Figure 5.18 shows similar performance to those obtained by the models in Section 5.1 and Section 5.2.

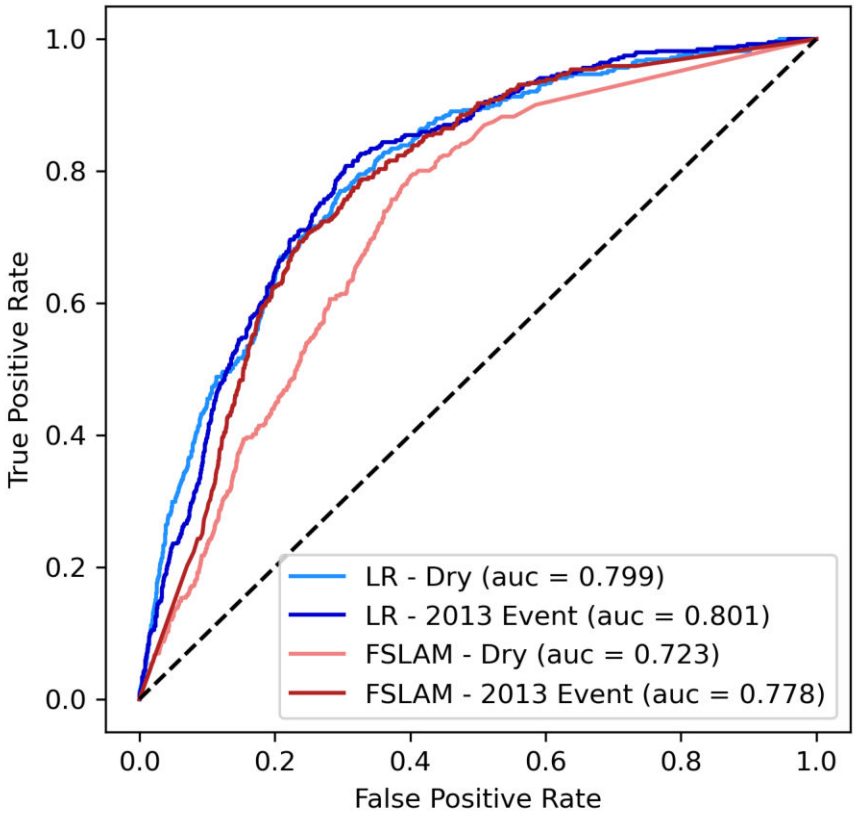

Figure 5.18: ROC curves for FSLAM and the logistic models

Figure 5.19 shows the different accuracy and MCC performance metrics obtained by the models for a range of decision thresholds form 0 to 1. It can be noted how highly dependant the performance of the data-driven model is in regards to the set decision threshold. Decision thresholds at both extremes show very poor performance while being highest when the middle point is selected. By contrast, FSLAM shows an almost horizontal curve meaning that performance is very similar independently of the threshold selected. A similar analysis describe Figure 5.20. If the decision threshold is set too high, the data-driven model will mostly report negatives which leads to an increased FNR and therefore more misses (Type II error). Similarly, when the threshold is set too low, the model will report more positives, increasing FPR and therefore the number of false-alarms (Type I error). By contrast FSLAM only tends to increase the number of misses as higher thresholds are set.

Why is this? Figure 5.19 and Figure 5.20 show the inherently different approach that both types of models use to determine stability or susceptibility. In Figure 5.19, *balanced accuracy* is presented instead of *accuracy* because the number of negative observations in the sample has been increased. *Balanced accuracy* accounts for the disparity between positive and negative observations in the sample.

FSLAM computes probability of failure as the area under the FS distribution that is below 1 as shown in Figure 4.2. Depending on the physical characteristics of the material present in a cell, FSLAM might calculate and FS distribution that is way higher than 1 and for which the lower tail of the probability density function (PDF) of FS never reaches values below 1. On the contrary, certain cell might have very poor physical characteristics that makes it impossible to obtain a PoF of less than 1. This is why performance for FSLAM is consistent across the range of thresholds. FSLAM is very good at determining which cells are generally unstable and stable and is able to discriminate these families.

By contrast, data-driven models are much more limited in the space given to decide if a cell is susceptible to landslides or not. As an example, say that a cell starts with a probability to be landslide susceptible of 0.5 and the data-driven model uses the information in the explanatory variables to push it, little by little, towards it being landslide susceptible or not. It is much more difficult for a data-driven model to obtain the same separation in families as FSLAM is able to determine. This claim is also supported by Figure 5.21.

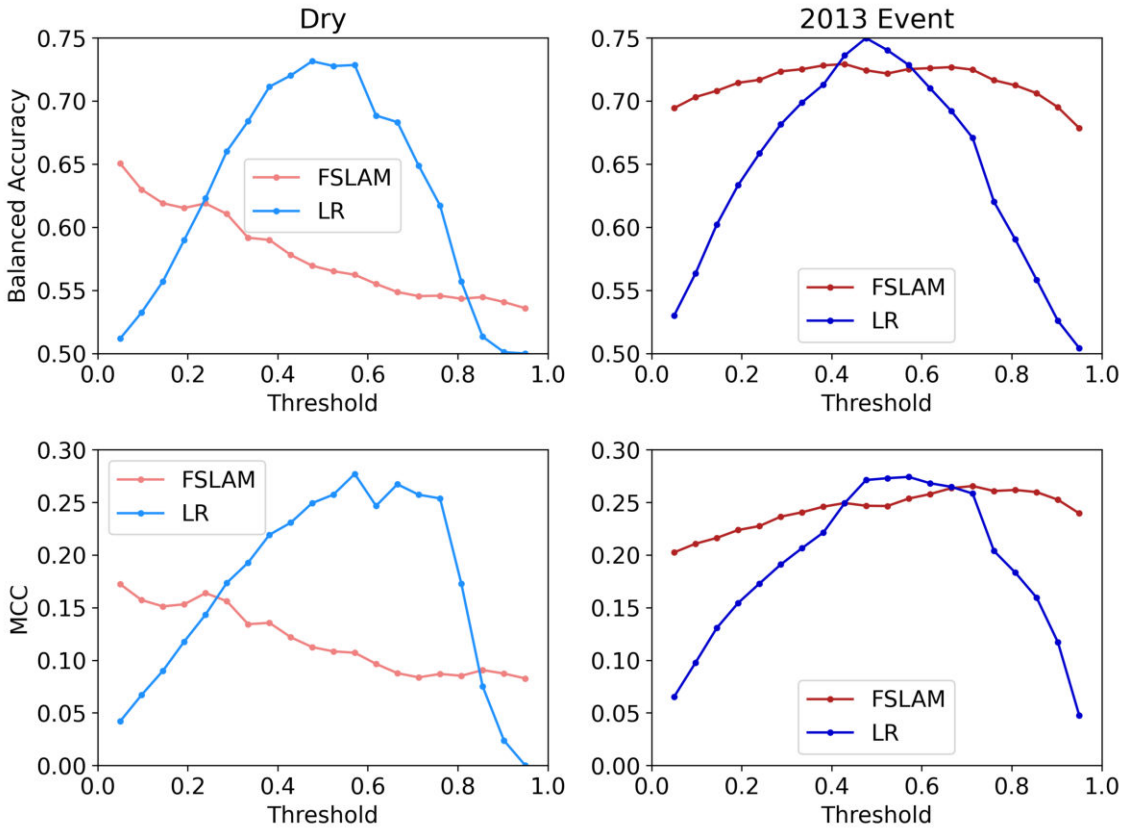

Figure 5.19: Comparison of balanced accuracy and MCC across all thresholds

Table 5.4 and Figure 5.21 give an idea on the distribution of predicted probabilities

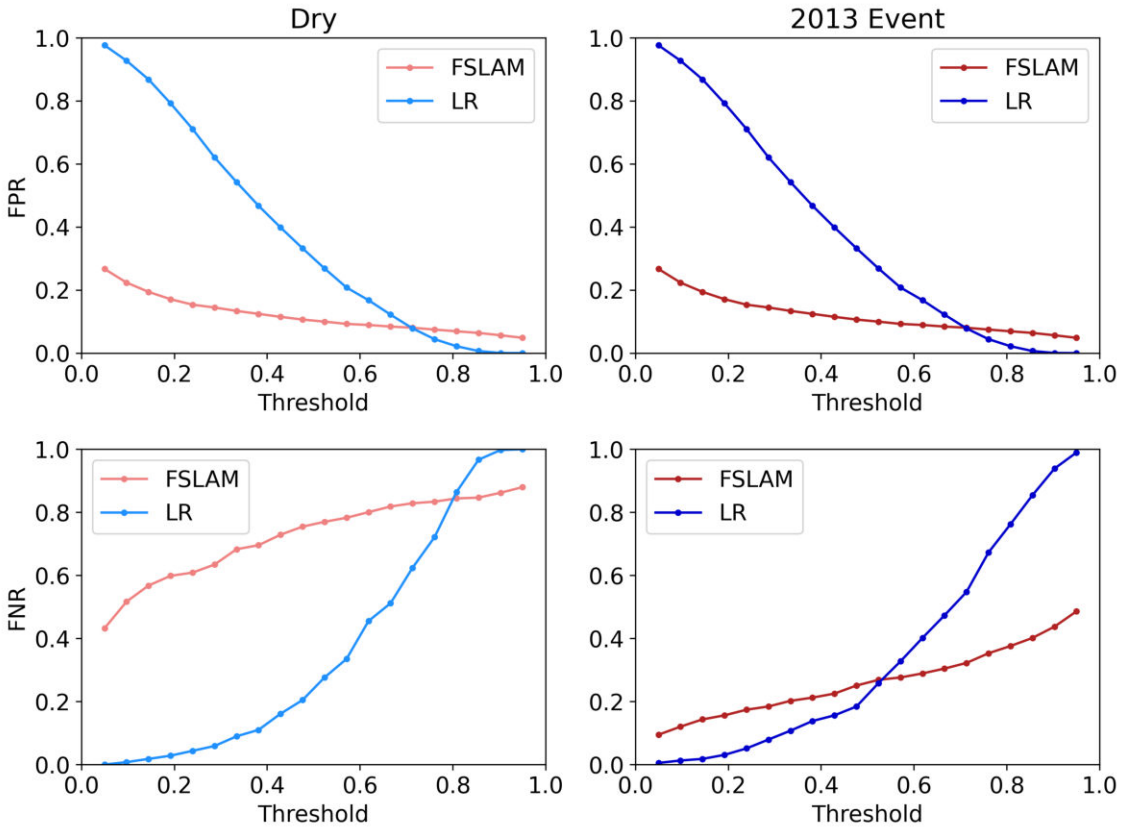

Figure 5.20: Comparison of FPR and FNR across all thresholds

as computed by FSLAM and the logistic regression models. Probabilities computed by FSLAM tend to accumulate in the extremes of the distribution, with a higher amount close to 0. Intuitively this makes sense because the amount of cells that are susceptible to landslides tend to be but a small fraction of the terrain as a whole. Meanwhile, probabilities computed by the logistic models are distributed all across the range of values between 0 and 1, with probabilities being skewed to the right (positive) for points inside the landslide inventory and skewed to the left (negative) for random points. Although the skewness of the computed values by the logistic regression model is the desired outcome, a more clear separation would be a more useful result. Additionaly, in FSLAM, it can be seen that the predicted probability for points in the landslide inventory change drastically when rainfall is considered, while for random points there is also an increase in predicted probability but it is less drastic than that shown by the inventory points. This sensitivity to the rainfall inputs also distinguishes FSLAM from the logistic regression models when analysing rainfall induced shallow landslides.

Table 5.4: Summary statistics for predicted probabilities

Model	Predicted Probablities				
	Min	Max	Average	Median	Std. Dev.
LR - Dry	0.02	0.92	0.40	0.38	0.05
LR - 2013 Event	0	1	0.39	0.35	0.06
FSLAM - Dry	0	1	0.13	0	0.08
FSLAM - 2013 Event	0	1	0.34	0.11	0.16

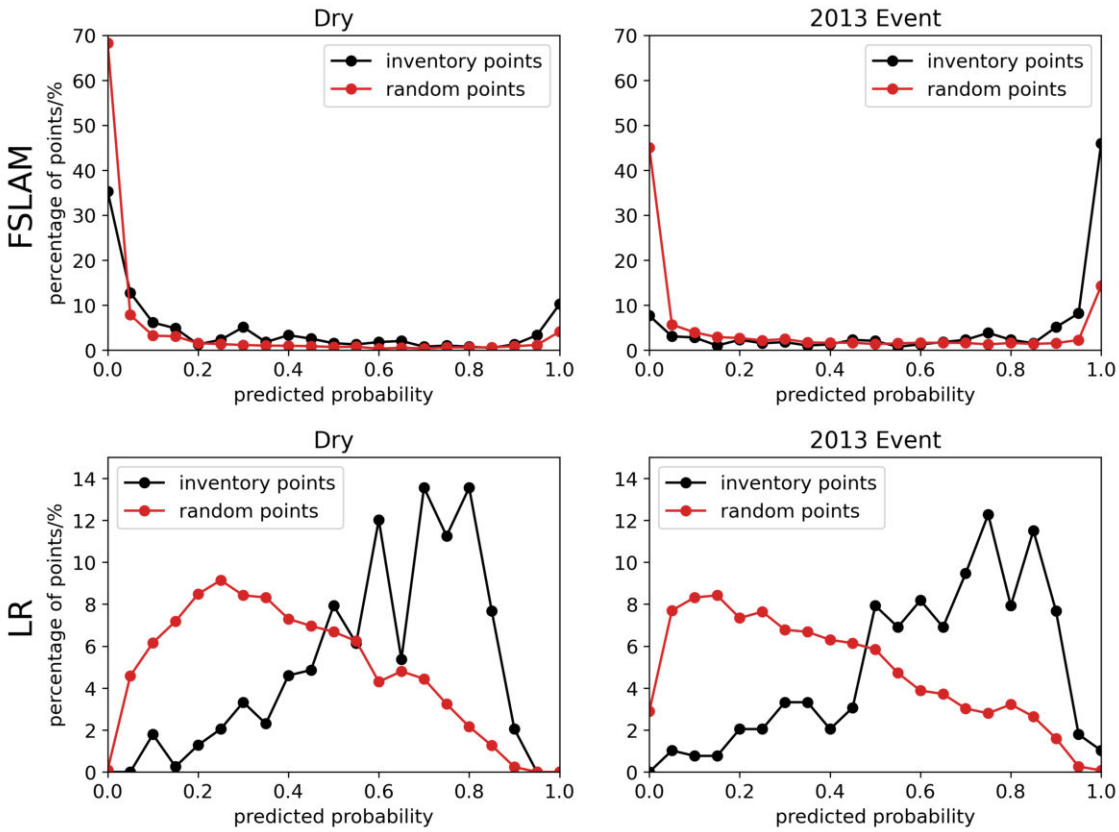

Figure 5.21: Curves on the predicted probabilities for all models

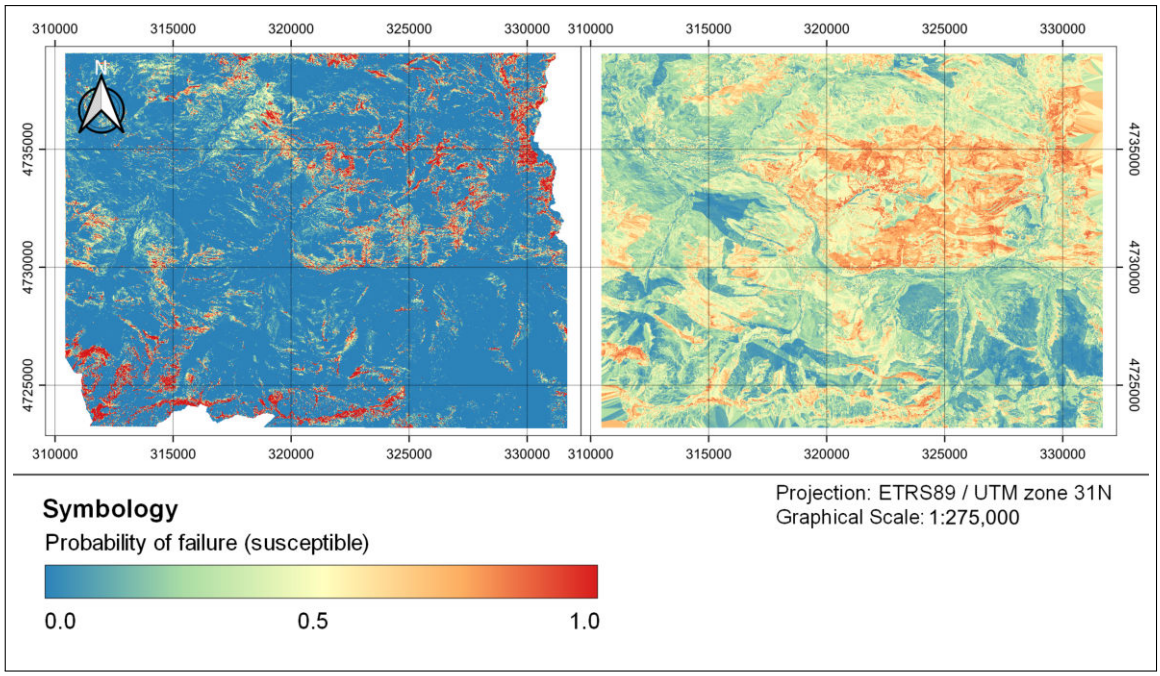

Figure 5.22: Predicted probability maps for FSLAM (left) and the logistic regression model (right) in the Dry case

As it has been shown in previous sections, data-driven models in general tend to be better than FSLAM when only performance metrics are considered. If performance is the main goal and requirement of a model, one might opt for a data-driven model, but in most cases pure performance is not the goal. Instead, landslide assessment techniques aim to map a probability of failure or susceptibility as spatial information. The maps produced by both types of models, shown in Figure 5.22 and Figure 5.23 show the biggest difference in both approaches. The maps produced by FSLAM give greater confidence in the areas that have been determined to be stable and unstable while the maps produced by the logistic regression models show many areas that are close to the defined threshold of 0.5, therefore are much more uncertain.

The difference is less drastic when class maps are shown, see Figure 5.24 and Figure 5.25. The results for the maps the consider the rainfall event in the case of FSLAM and rain for the logistic model are very similar and a user might consider them as equal. This is an issue because these maps are not able to communicate the underlying uncertainty associated to each of the models.

A final comparison is shown in Figure 5.26 where the predicted probabilities for the maps shown in Figure 5.22 and Figure 5.23 are shown as curve for cumulative density functions (CDFs). This result is similar to Figure 5.21. FSLAM is better at determining which cells have a PoF of 0 or 1, so for the PoF of 0 the CDF for FSLAM starts at 70% for the Dry case and 45% for the 2013 Event case. Also both curves for FSLAM have a big jump in the final transition between predicted probabilities of 0.9 and 1. In contrast, the CDFs for the logistic regression show a wider range in the percentage of cells calculated for each predicted probability.

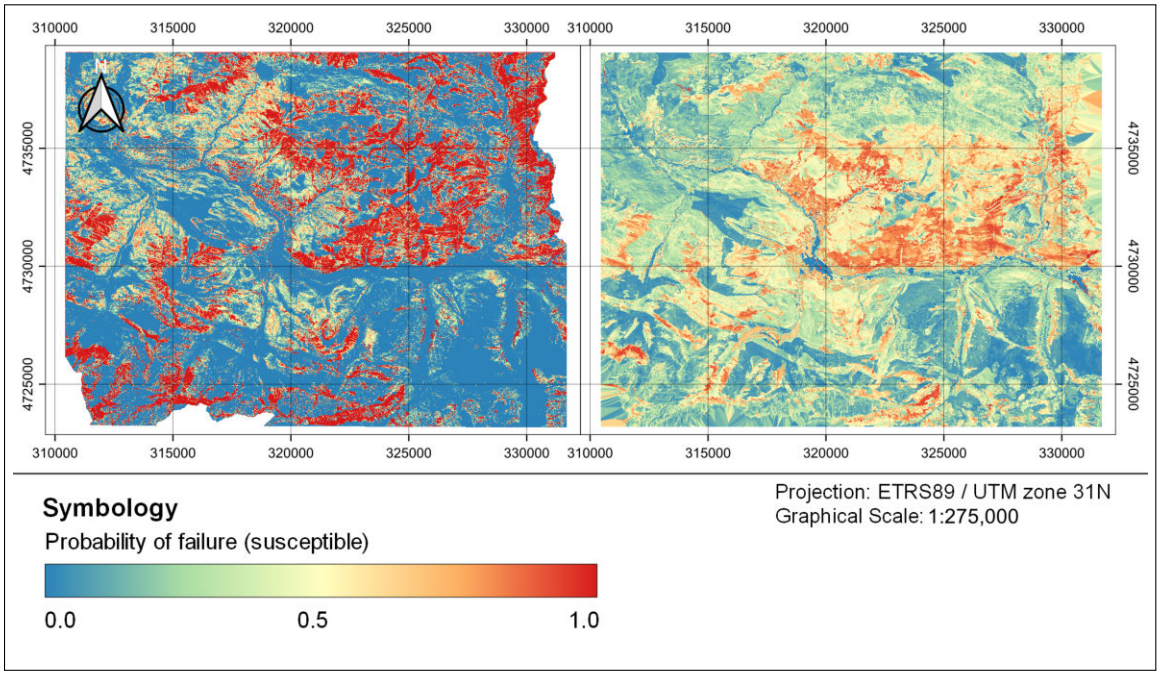

Figure 5.23: Predicted probability maps for FSLAM (left) and the logistic regression model (right) in the 2013 Event case

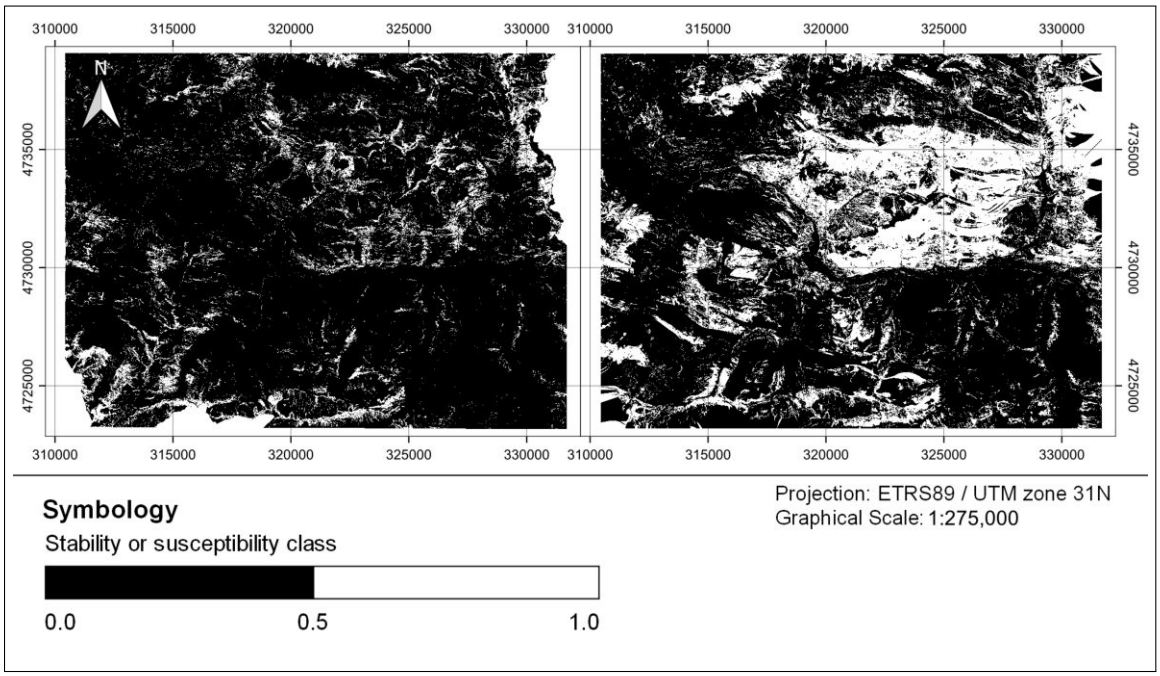

Figure 5.24: Predicted classes maps for FSLAM Dry (left) and the logistic regression model with no rainfaill (right)

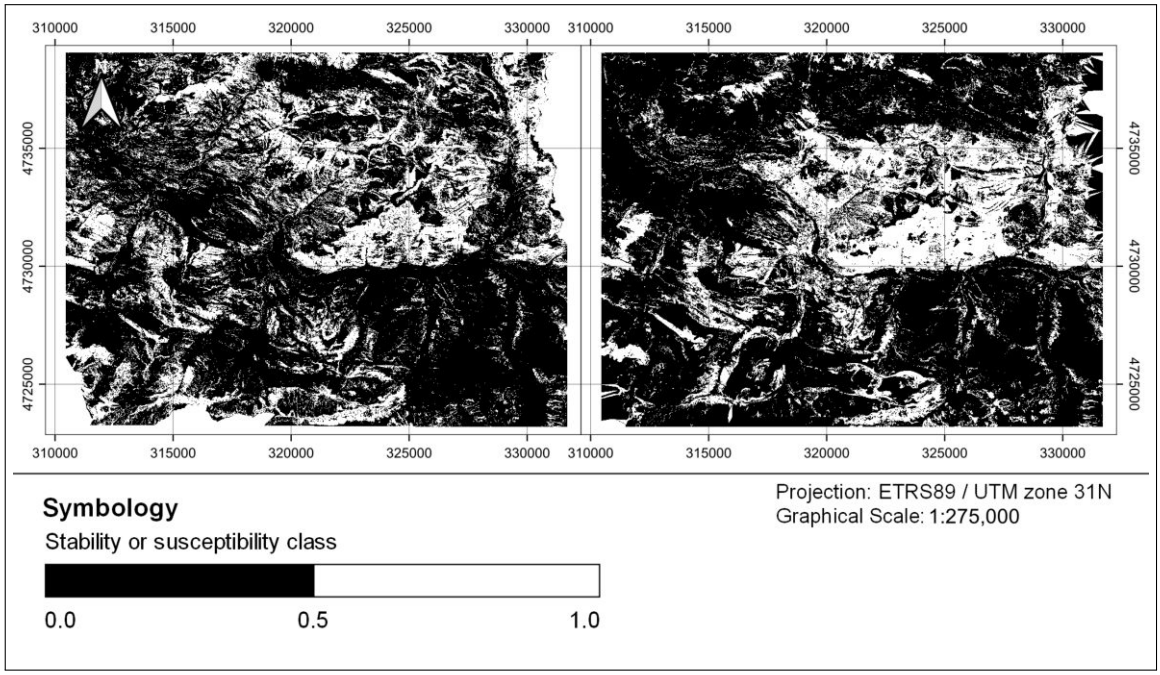

Figure 5.25: Predicted classes maps for FSLAM Event (left) and the logistic regression model with rainfaill (right)

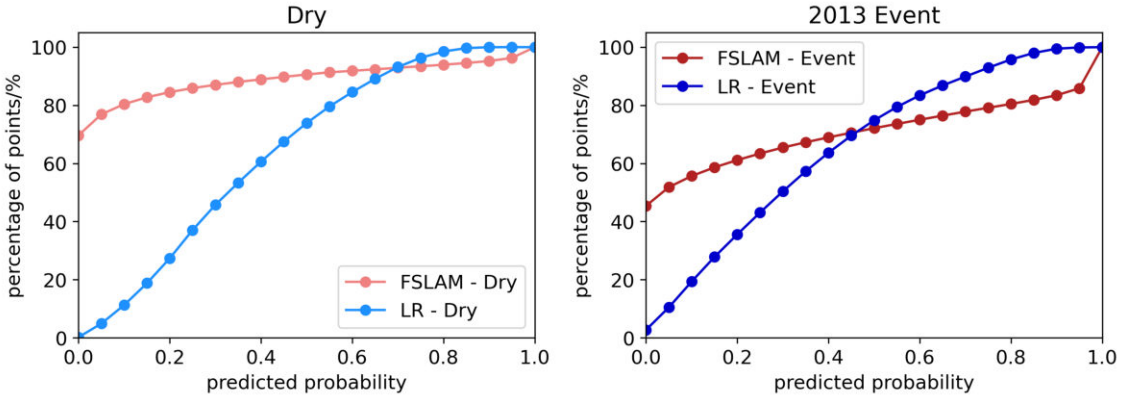

Figure 5.26: Cumulative density curves of the predicted probability maps

5.8 Model coupling

The final task with the logistic regression models is to assess their improvement when the outputs of FSLAM are used as explanatory variables. As explained in Section 4.1.1, FSLAM provides outputs of PoF when the soil column is completely dry and completely saturated for each cell in the model. These outputs can be used to determine which cells are "Unconditionally Unstable" and "Unconditionally Stable" respectively. Additionally, the output of PoF for the event for which FSLAM was calibrated using the landslide inventory can also be used as an explanatory variable.

Table 5.5 and Table 5.6 show how each of the new inputs affects the performance in comparison to the basic logistic regression models. In terms of AUC, PoF when the soil columns of every cell in the model are completely saturated seems to be the only output that significantly increases the performance of the models.

As in the previous section, if the objective of the model is to obtain the highest performance possible, it would seem worth it to first use FSLAM to compute PoF under fully saturated conditions and then use this output as an input parameter for a data-driven model. Otherwise, this would not be ideal because the increase in performance is slight at most and when using the data-driven models you would lose the nice separation between stability families that FSLAM is able to obtain.

Table 5.5: Improvements in model performance for the Dry case

Model	**AUC ROC**	**ACC**	**FPR**	**FNR**	**MCC**
No Rain	0.806	0.770	0.292	0.17	0.543
No Rain + PoF_Dry	0.806	0.770	0.292	0.17	0.543
No Rain + PoF_Event	0.818	0.750	0.271	0.23	0.500
No Rain + PoF_Sat	0.824	0.735	0.312	0.22	0.470
No Rain + PoF_Dry + PoF_Event	0.819	0.750	0.271	0.23	0.500
No Rain + PoF_Dry + PoF_Sat	0.824	0.735	0.312	0.22	0.470
No Rain + PoF_Event + PoF_Sat	0.820	0.719	0.302	0.26	0.438
No Rain + PoF_Dry + PoF_Event + PoF_Sat	0.823	0.730	0.302	0.24	0.459

Table 5.6: Improvements in model performance for 2013 Event case

Model	AUC ROC	ACC	FPR	FNR	MCC
Rain	0.800	0.740	0.292	0.23	0.479
Rain + PoF_Dry	0.807	0.730	0.312	0.23	0.459
Rain + PoF_Event	0.807	0.724	0.260	0.29	0.450
Rain + PoF_Sat	0.823	0.740	0.302	0.22	0.480
Rain + PoF_Dry + PoF_Event	0.810	0.709	0.292	0.29	0.418
Rain + PoF_Dry + PoF_Sat	0.826	0.745	0.292	0.22	0.490
Rain + PoF_Event + PoF_Sat	0.822	0.724	0.281	0.27	0.449
Rain + PoF_Dry + PoF_Event + PoF_Sat	0.824	0.745	0.271	0.24	0.489

5.9 Extrapolation

Usually, extrapolation of data-driven models is not recommended because of their reliance on the sample used as training data. Nonetheless, the capabilities of the logistic model of Val d'Aran to determine landslide susceptibility in the Pyrenees was tested by using the region of Berguedà as another study area. In this case, Berguedà is truly "out of sample" data for the data-driven model to make predictions on.

As it can be seen in Figure 3.5, Figure 3.6 and Figure 3.9, the classes in the thematic maps of Soil and LULC are different for both regions. This is mainly due to the fact that both data sets were reclassified by different users from the same original source. In order to test the data-driven models, classes in Figure 3.9 were reclassified as shown in Tables 5.7 and 5.8.

Because no input of rainfall was considered, only the *Dry* case models were applied. Taking advantage of the reclassification previously described, the logistic model and the decision tree were tested.

In the ROC analysis shown in Figure 5.27, the logistic model had a reduction in performance of 0.10 when compared with the performance obtained in Val d'Aran for the testing phase as shown in Figure 5.4. Still, the AUC metric that the model performs better than a random classifier. The performance of the model degrades when the specific threshold of 0.5 is set as decision criteria. According to the confusion matrix shown in Figure 5.28, the model has an accuracy just a little bit better than a random classifier. More worryingly, the computed FNR for the model is 0.801 which means four out of five cells classified as not susceptible would be misses and therefore susceptible to landslides. Because of the effects of Type-II error in landslide susceptibility assessment, this would be absolutely unacceptable. For other thresholds, FNR decreases while accuracy increases, meaning that in this specific application, the threshold of 0.5 does not lead to the best performance of the model.

The decision tree 5.11 was also applied in Berguedà to a much higher performance when compared to the logistic model as shown in Figure 5.29. This is a case where a more complex model, such as the logistic regression, is worse than a simple model

Table 5.7: Reclassification of the Berguedà soil types

Class Berguedà	**Class Val d'Aran**
alluvium	alluvial
colluvium	colluvium
conglomerate conglomerate and sandstone conglomerate, sandstone and shale	coglomerate
limestone and mudstone limestone travertine	limestone
mudstone gaumnian mudstone eocene mudstone keuper paleozoic rocks	mudstone
till	till

when generalizing to new data. Heavier regularization of the logistic regression model might lead to improved performance across the Pyreness region. Ideally, a third landslide inventory from a different region in the Pyrenees could be introduced in order to assess the performance of models based on training sets coming from different areas within the same region (the Pyrenees). Although this application shows promise, more data and testing is required before deeper insight can be found.

Table 5.8: Reclassification of the Berguedà LULC classes

Class Berguedà	**Class Val d'Aran**
forest	forest
shrubs and grassland agriculture	shrubs
supraforestal grassland	grassland
no or sparse vegetation	bare soil
bedrock	intact bedrock
urban area	urban area
water	water

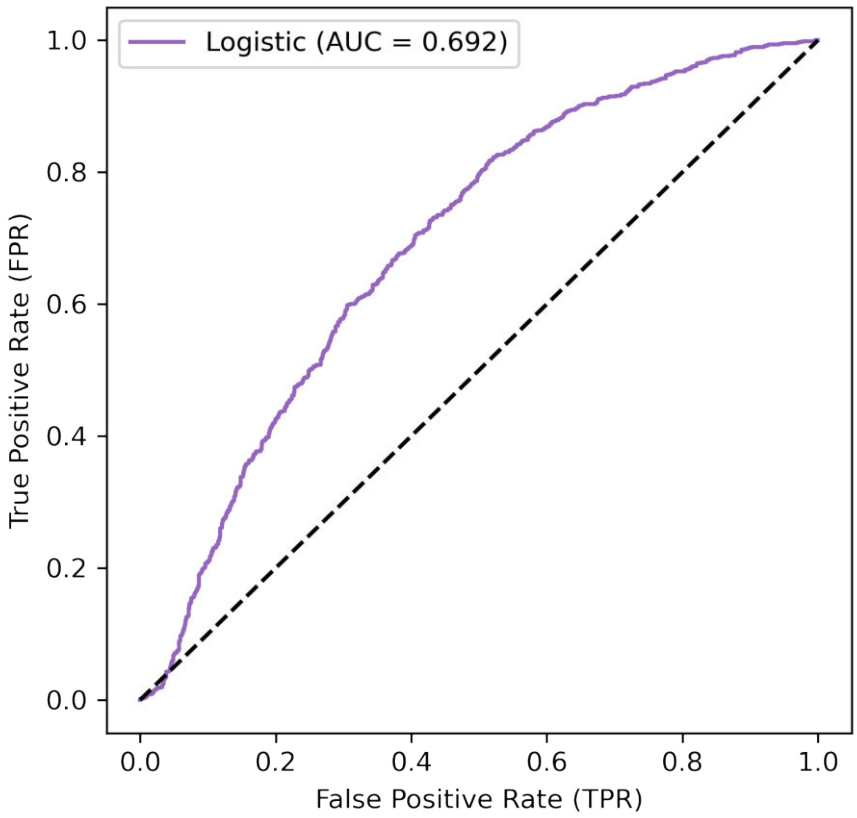

Figure 5.27: ROC curve for the application of the logistic model in Berguedà

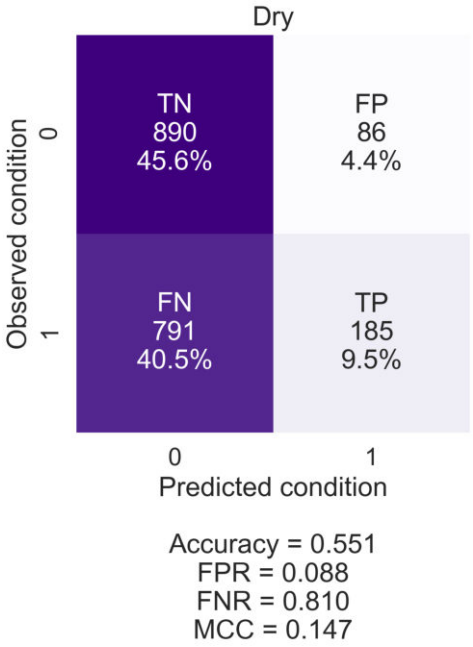

Figure 5.28: Confusion matrix for application of the logistic model in Berguedà

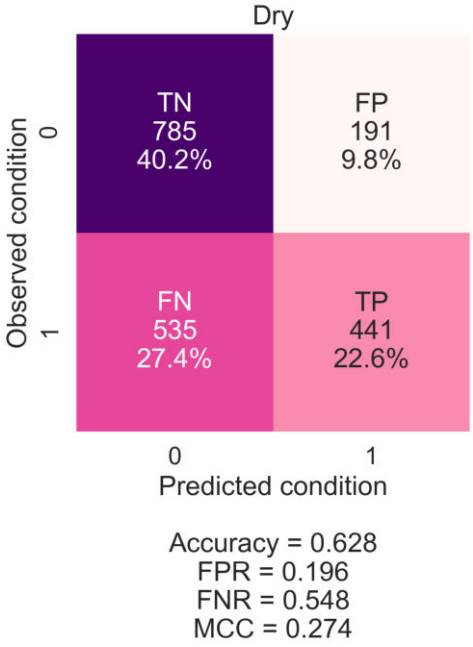

Figure 5.29: Confusion matrix for application of the decision tree in Berguedà

Chapter 6

Conclusions and Recommendations

6.1 Conclusions

The main goal of this research project was to develop a data-driven model to determine the susceptibility of a cell in a raster map regarding rainfall-induced shallow landslides. It can be said that this objective was achieved successfully mainly by two models based on the logistic regression for two different application cases. Measuring performance in terms of AUC-ROC, the logistic regression models scored at 0.806 when so-called "dry" explanatory variables were considered, and 0.800 when rainfall for a single landslide-inducing event was also considered as part of the explanatory variables.

Although the objective was achieved with the previously stated models, several other statistical and machine learning methods were tested to evaluate their performance at predicting landslide susceptibility. Considering the two cases established: "dry" and "2013 event", the SVC scored 0.813 and 0.772 respectively and the random forest scored 0.837 and 0.853 respectively. A single cost-complexity pruned classification tree was also evaluated and scoring 0.689 and 0.0735 in terms of overall model accuracy, but due to its simplicity, it was also discussed how useful this model would be in a decision-making process that involved stakeholders without a technical background.

Each of the discussed models gave insights into their inner workings and how data-driven models use explanatory variables to make decisions. In particular, the use and importance of different explanatory variables was found to be highly dependant on the model used, with some models like the logistic regression giving more importance to categorical variables and other models, such as those based on a random forest, focusing on continuous variables derived from the DEM and more related to terrain characteristics.

When comparing a data-driven model with a physically-based model, it was concluded that the approach of each model to determine landslide susceptibility was inherently different. FSLAM determines landslide susceptibility by determining which cells are stable or unstable based on their mechanical properties. This makes FS-

LAM much better at distinguishing with much more certainty cells that are stable or unstable. By the contrary, the logistic regression models take and approach in which landslide susceptibility is computed more directly base on a recorded inventory of landslides for a single event. This makes the approach more uncertain although it is able to achieve better overall performance. In the end, in regards to landslide susceptibility mapping, the uncertainty associated to data-driven models makes them less useful than the susceptibility maps as computed by a physically-based model. Even though setting up a physically-based model such as FSLAM is more time consuming, the time invested would be worth it if the aim of the model is to be used for mapping. Data-driven models can be more suitable if a quick preliminary assessment is required.

The use of outputs from the physically-based model to improve the performance of the data-driven model proved to be successful. The main issue found was that performance was increased at the cost of losing the insights provided by the outputs of the physically based model. Additionally, improvements in performance were very small (around 2%), which matches the current literature available. Such small increases of performance give rise to the question if is it really of worth to setup another model in order to achieve a slight increase in performance.

A test case for the use of data-driven models in regions within the Pyrenees trained by using data from only one particular area showed promise. Models trained using data from Val d'Aran provided adequate results when computing susceptibility in a different region within the Pyrenees. The results in terms of performance were mixed, more testing and data is required before saying with certainty if these types of model can be used successfully in the domain of the Pyrenees while only being trained using data from a single region.

6.2 Recommendations for further research

A number of aspects that were touched upon on this project can be explored further in a number of ways. In this section, some ideas and recommendations for future research are discussed.

Investigate the performance of a global model. During this research project, several attempts were made to propose a data-driven model for landslide susceptibility based mainly on freely available global data sets. These data-sets included topographic information from the Shuttle Radar Topography Mission (SRTM) (Farr & Kobrick, 2000), land cover data from the European Spatial Agency (ESA) (ESA, 2017) and soil texture classes by FutureWater (Simons et al., 2020). These models showed promise but there was a limitation in the number of inventories available to test their performance. The case for extrapolation within the Catalan Pyrenees was an attempt of preparing a model for a regional analysis but increasing the scope of data-driven models to a global scope should be explored further.

On the topic of landslide inventories, for this research project a single high-quality landslide inventory was used but it is limited to one single snowmelt-rainfall episode that occurred in 2013. If available, a multi-temporal database including the collected records of landslides for a particular region would prove to be richer and offer more insights that could be explored using data-driven models.

Rainfall was also only explored in a limited capacity in this work mainly due to the established framework which FSLAM uses and the landslide inventory related to a single episode. If rainfall is considered as the main triggering mechanism for landslides, an inventory that collects data for different time periods would permit further analysis of rainfall and how its interaction with the land cover and soil generates landslides. In addition, this analysis could serve as the starting point for an early warning system (EWS) based on a data-driven model using predictions for future rainfall.

Finally, the threshold set in order for a model to classify a cell as susceptible or not should be analysed further. In this work, the threshold was set always at 0.5 because it provided the best performance and is the typical value used in literature. Nonetheless, some authors have suggested that for mapping, the threshold should be set as the value that obtains the best TPR while at the same time maintaining a FPR of 10% (Goetz et al., 2015). This approach is interesting but the literature is limited in the consequences that this decision would have in real-life scenarios.

Bibliography

Baeza, C., & Corominas, J. (2001). Assessment of shallow landslide susceptibility by means of multivariate statistical techniques. *Earth Surface Processes and Landforms*, *26*(12), 1251–1263. https://doi.org/10.1002/esp.263

Bazan, J. G., & Szczuka, M. (2005). The Rough Set Exploration System [Series Title: Lecture Notes in Computer Science]. In D. Hutchison, T. Kanade, J. Kittler, J. M. Kleinberg, F. Mattern, J. C. Mitchell, M. Naor, O. Nierstrasz, C. Pandu Rangan, B. Steffen, M. Sudan, D. Terzopoulos, D. Tygar, M. Y. Vardi, G. Weikum, J. F. Peters, & A. Skowron (Eds.), *Transactions on Rough Sets III* (pp. 37–56). Series Title: Lecture Notes in Computer Science. Berlin, Heidelberg, Springer Berlin Heidelberg. https://doi.org/10.1007/11427834_2

CatalanNews. (2013). Severe floods in the north-western Catalan Pyrenees. Retrieved July 31, 2021, from http://www.catalannews.com/society-science/item/severe-floods-in-the-north-western-catalan-pyrenees

Cho, S. E. (2007). Effects of spatial variability of soil properties on slope stability. *Engineering Geology*, *92*(3-4), 97–109. https://doi.org/10.1016/j.enggeo.2007.03.006

Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V., & Böhner, J. (2015). System for Automated Geoscientific Analyses (SAGA) v. 2.1.4. *Geoscientific Model Development*, *8*(7), 1991–2007. https://doi.org/10.5194/gmd-8-1991-2015

Corominas, J., & Moya, J. (1999). Reconstructing recent landslide activity in relation to rainfall in the Llobregat River basin, Eastern Pyrenees, Spain. *Geomorphology*, *30*(1-2), 79–93. https://doi.org/10.1016/S0169-555X(99)00046-X

ESA. (2017). Land Cover CCI Product User Guide Version 2. Tech. Rep. Retrieved July 27, 2021, from http://maps.elie.ucl.ac.be/CCI/viewer/index.php

Farr, T. G., & Kobrick, M. (2000). Shuttle radar topography mission produces a wealth of data. *Eos, Transactions American Geophysical Union*, *81*(48), 583–585. https://doi.org/10.1029/EO081i048p00583

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, *27*(8), 861–874. https://doi.org/10.1016/j.patrec.2005.10.010

Fell, R., Corominas, J., Bonnard, C., Cascini, L., Leroi, E., & Savage, W. Z. (2008). Guidelines for landslide susceptibility, hazard and risk zoning for land use planning. *Engineering Geology*, *102*(3-4), 85–98. https://doi.org/10.1016/j.enggeo.2008.03.022

Froude, M. J., & Petley, D. N. (2018). Global fatal landslide occurrence from 2004 to 2016. *Natural Hazards and Earth System Sciences*, *18*(8), 2161–2181. https://doi.org/10.5194/nhess-18-2161-2018

Goetz, J., Brenning, A., Petschko, H., & Leopold, P. (2015). Evaluating machine learning and statistical prediction techniques for landslide susceptibility modeling. *Computers & Geosciences*, *81*, 1–11. https://doi.org/10.1016/j.cageo.2015.04.007

Guzzetti, F., Galli, M., Reichenbach, P., Ardizzone, F., & Cardinali, M. (2006). Landslide hazard assessment in the Collazzone area, Umbria, Central Italy. *Natural Hazards and Earth System Sciences*, *6*(1), 115–131. https://doi.org/10.5194/nhess-6-115-2006

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning*. New York, NY, Springer New York. https://doi.org/10.1007/978-0-387-84858-7

Highland, L., & Bobrowsky, P. (2008). *The landslide handbook - A guide to understanding landslides* (Circular No. 1325). U. S. Geological Survey. Reston, Virginia.

Hürlimann, M., Guo, Z., Puig-Polo, C., & Medina, V. (2021). Impacts of future climate and land cover changes on landslide susceptibility: Regional scale modelling in the Val d'Aran region (Pyrenees, Spain).

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning* (Vol. 103). New York, NY, Springer New York. https://doi.org/10.1007/978-1-4614-7138-7

Matthews, B. (1975). Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochimica et Biophysica Acta (BBA) - Protein Structure*, *405*(2), 442–451. https://doi.org/10.1016/0005-2795(75)90109-9

Medina, V., Hürlimann, M., Guo, Z., Lloret, A., & Vaunat, J. (2021). Fast physically-based model for rainfall-induced landslide susceptibility assessment at regional scale. *CATENA*, *201*, 105213. https://doi.org/10.1016/j.catena.2021.105213

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Peng, L., Niu, R., Huang, B., Wu, X., Zhao, Y., & Ye, R. (2014). Landslide susceptibility mapping based on rough set theory and support vector machines: A case of the Three Gorges area, China. *Geomorphology*, *204*, 287–301. https://doi.org/10.1016/j.geomorph.2013.08.013

Reback, J., Jbrockmendel, McKinney, W., Van Den Bossche, J., Augspurger, T., Cloud, P., Hawkins, S., Gfyoung, Sinhrks, Roeschke, M., Klein, A., Petersen, T., Tratner, J., She, C., Ayd, W., Hoefler, P., Naveh, S., Garcia, M., Schendel, J., … Dong, K. (2021). Pandas-dev/pandas: Pandas 1.3.0. Zenodo. https://doi.org/10.5281/ZENODO.3509134

Reichenbach, P., Rossi, M., Malamud, B. D., Mihir, M., & Guzzetti, F. (2018). A review of statistically-based landslide susceptibility models. *Earth-Science Reviews*, *180*, 60–91. https://doi.org/10.1016/j.earscirev.2018.03.001

Shano, L., Raghuvanshi, T. K., & Meten, M. (2020). Landslide susceptibility evaluation and hazard zonation techniques – a review. *Geoenvironmental Disasters*, *7*(1), 18. https://doi.org/10.1186/s40677-020-00152-0

Shu, H., Hürlimann, M., Molowny-Horas, R., González, M., Pinyol, J., Abancó, C., & Ma, J. (2019). Relation between land cover and landslide susceptibility in Val d'Aran, Pyrenees (Spain): Historical aspects, present situation and forward prediction. *Science of The Total Environment*, *693*, 133557. https://doi.org/10.1016/j.scitotenv.2019.07.363

Simons, G., Koster, R., & Droogers, P. (2020). *HiHydroSoil v2.0 - A high resolution soil map of global hydraulic properties.* (tech. rep. Report 213). FutureWater.

Solomatine, D. P., & Ostfeld, A. (2008). Data-driven modelling: Some past experiences and new approaches. *Journal of Hydroinformatics*, *10*(1), 3–22. https://doi.org/10.2166/hydro.2008.015

Sun, J., Qin, S., Qiao, S., Chen, Y., Su, G., Cheng, Q., Zhang, Y., & Guo, X. (2021). Exploring the impact of introducing a physical model into statistical methods on the evaluation of regional scale debris flow susceptibility. *Natural Hazards*, *106*(1), 881–912. https://doi.org/10.1007/s11069-020-04498-4

Tarboton, D. G. (1997). A new method for the determination of flow directions and upslope areas in grid digital elevation models. *Water Resources Research*, *33*(2), 309–319. https://doi.org/10.1029/96WR03137

Thomas, K., Benjamin, R.-K., Fernando, P., Brian, G., Matthias, B., Jonathan, F., Kyle, K., Jessica, H., Jason, G., Sylvain, C., & al et, e. (2016). Jupyter Notebooks – a publishing format for reproducible computational workflows [Publisher: IOS Press]. *Stand Alone*, *0*(Positioning and Power in Academic Publishing: Players, Agents and Agendas), 87–90. https://doi.org/10.3233/978-1-61499-649-1-87

Van Rossum, G., & Drake, F. L. (2009). *Python 3 Reference Manual*. Scotts Valley, CA, CreateSpace.

Victoriano, A., García-Silvestre, M., Furdada, G., & Bordonau, J. (2016). Long-term entrenchment and consequences for present flood hazard in theGarona River (Val d'Aran, Central Pyrenees, Spain). *Natural Hazards and Earth System Sciences*, *16*(9), 2055–2070. https://doi.org/10.5194/nhess-16-2055-2016

Zevenbergen, L. W., & Thorne, C. R. (1987). Quantitative analysis of land surface topography. *Earth Surface Processes and Landforms*, *12*(1), 47–56. https://doi.org/10.1002/esp.3290120107

Zêzere, J., Pereira, S., Melo, R., Oliveira, S., & Garcia, R. (2017). Mapping landslide susceptibility using data-driven methods. *Science of The Total Environment*, *589*, 250–267. https://doi.org/10.1016/j.scitotenv.2017.02.188

Zhang, Y., Ge, T., Tian, W., & Liou, Y.-A. (2019). Debris Flow Susceptibility Mapping Using Machine-Learning Techniques in Shigatse Area, China. *Remote Sensing*, *11*(23), 2801. https://doi.org/10.3390/rs11232801