STATISTICAL MODELS FOR MULTIHAZARDS

A thesis presented in partial fulfilment of the requirements for the degree of
Doctor of Philosophy
in
Statistics
at Massey University, Palmerston North,
New Zealand.

Gabriele Frigerio Porta

2020

Contents

Abstract **xi**

Acknowledgements **xii**

1 Introduction **1**
1.1 Motivation 1
1.2 Overview 4
1.2.1 Point process for hazard interactions 4
1.2.2 Survival analysis of landslide dams 7
1.2.3 Potentials 8

2 Methodology review **10**
2.1 Introduction 10
2.2 Point Processes 10
2.2.1 Definition and properties of temporal point processes 11
2.2.2 Spatio-temporal models 17
2.2.3 Discretization 19
2.3 Survival Analysis 22
2.4 Bayesian imputation of missing data 24

3 Literature review **26**
3.1 Scope of study 27
3.2 Single hazard assessments 29
3.2.1 Earthquake 29
3.2.2 Rainfall 30
3.2.3 Landslides 31
3.2.4 Landslide dams 31
3.3 Hazard triggering 32
3.3.1 Rainfall induced landslides 34
3.3.2 Earthquake induced landslides 38

3.4 Multi-hazard assessment attempts . . . 39
3.5 A new idea: hazard potential . . . 49

4 Earthquake/Rainfall triggering landslides 51
4.1 Introduction . . . 51
4.2 Data . . . 53
4.2.1 Earthquake data . . . 53
4.2.2 Rainfall data . . . 54
4.2.3 Landslide data . . . 55
4.3 Stochastic model(s) for landslide triggering . . . 58
4.3.1 Breaking down the triggering factors . . . 60
4.3.2 Three interaction models . . . 62
4.3.3 ZIP terms . . . 63
4.4 Results . . . 64
4.5 Discussion . . . 71
4.6 Appendix . . . 72
4.6.1 Estimation . . . 72
4.6.2 Landslide dating accuracy problem . . . 73
4.6.3 Log-likelihood of the ZIP model . . . 74

5 Landslide Dams 75
5.1 Introduction . . . 75
5.2 Landslide Dams . . . 76
5.3 Data . . . 79
5.4 Methods . . . 87
5.4.1 Covariate effects . . . 87
5.4.2 Imputation of missing values . . . 88
5.4.3 Bayesian framework and censoring . . . 91
5.4.4 Implementation . . . 91
5.5 Results . . . 92
5.6 Results with known lifetime class only . . . 94
5.7 Sicilian data . . . 94
5.8 Forecasting . . . 96
5.9 Discussion . . . 97
5.10 Conclusions . . . 99

6 A concept of natural hazard potentials 100
6.1 Definition of potential . . . 100
6.2 Potential for a single hazard . . . 107

6.2.1 Hawkes process . . . 107
6.2.2 Epidemic type aftershock-sequences model . . . 109
6.2.3 Stress release model . . . 110
6.3 Linking simple systems together - potential in a multi-triggering environment . . . 112
6.3.1 Earthquake/rainfall triggered landslides . . . 114
6.3.2 Linked stress release model . . . 121
6.3.3 Landslides leading to landslide dams . . . 121
6.4 Inhibition/threshold . . . 129
6.4.1 Debris flows . . . 132
6.5 A framework for multi-hazard simulation. . . . 134

7 Discussion and future research **137**

Bibliography **146**

List of Tables

3.2 list of all hazards divided into groups 42
3.3 Table of six hazards (main diagonal) and their interaction from De Pippo et al. (2008). In the top triangle of the matrix, each cell is at the intersection between a hazard on the left (being the cause) and a hazard at the bottom (being the effect). In the bottom triangle, it's the other way around. 47

4.1 Description of the datasets used. 59
4.2 Parameter estimates (normalised components) and resulting log-likelihood for each model. 64

5.1 Variables in the dataset. 80
5.2 Frequency of the events with known dam lifetime and condition. "Drained" and "Stabilized" refer to human intervention. Disappearance of the lake by aggradation is indicated as "alluviation". The counts in brackets are the events with exact damming and failure times. 81
5.3 Glossary of the variables used. 81
5.4 Spearman correlation matrix. Non-significant (at 5% level) correlations are in parentheses. 84
5.5 Estimates of the b coefficients for predictor imputation. 92
5.6 Estimates of the parameters expressing the effects of the dam measurements, catchment area, river slope and type of dam on its survival time. 93
5.7 Estimates of the b coefficients for predictor imputation. 94
5.8 Estimates of the parameters expressing the effects of the dam measurements, catchment area, river slope and type of dam on its survival time. 94
5.9 Estimates (main model) without using Sicilian data 95
5.10 Resutls with specific intercepts for Sicily. 96
5.11 Table showing the probability of survival at five and fifty years for a canonical dam. 97

List of Figures

1.1 Example of complex chain of hazards. The image is from Orchiston et al. (2018), modified to include rainfall, and to highlight the reduced chain of hazards that has been chosen for this thesis. Each arrow represent an interaction between events. The bold arrows highlight the chain of events that is used in this thesis. 5

2.1 Example of Poisson process. The first plot show the cumulated events over time, while the second plot shows the occurrences. 13
2.2 Example of Hawkes process: the line represents the conditional intensity of the process. At the bottom, the realization of the process, i.e. event occurrences. Each event produces a jump in the conditional intensity, represent by a spike in the line, followed by a temporal decay. A cluster of events produced a bigger jump and a longer decay. 16

3.1 Schematic diagram of the selected chain of hazards: earthquakes and rainfall can trigger landslides; landslides may fall on a river causing a landslide dam. The failure of a landslide dam can cause floods. 28
3.2 Rainfall thresholds obtained using the Bayesian (T_B, green line) and the Frequentist (light blue line is the 1% threshold, T_1; red line is the 5% threshold T_5) methods. Error bars on the rainfall mean intensity I show systematic error, assumed fixed and equal to 10%. Errors on the rainfall duration D were considered negligible and are not shown. Picture and original caption from Brunetti et al. (2010) 37
3.3 The figure shows the possible interaction between hazards, divided the type of interaction and the potential number of triggered events. For example, a volcanic eruption has the potential to trigger directly and increase the probability of a large number of landslide events. Reproduced from Gill and Malamud (2014) 43

3.4 The table shows the interaction between hazards classified into nine classes, depending on the chance of spatial overlapping and temporal likelihood. It is possible to see that earthquakes have high probabilities of both with other seismic events, while tsunami is very high in temporal likelihood with ground heave, but limited in spatial overlapping. 44

4.1 Earthquake distribution in centre-north Italy from 1981 to 2018. The L'Aquila sequence is visible in the bottom-right corner of the map. In Emilia-Romagna (shaded inset), earthquakes mostly follow the Apennines ridge but also some events occurred on foothills (between Rimini and Ravenna) and flat lands of Po Valley (all the top part of the region). The light blue concentration below Verona is the 2012 earthquake series. Municipality boundaries are shown in black. 55
4.2 Geographical distribution of daily rainfall average (mm) 1981-2018 in the Emilia-Romagna region. Municipality boundaries are shown in black. 56
4.3 Landslide locations 1981-2018 in the region of Emilia-Romagna. Municipality boundaries are shown in black. 57
4.4 Distribution of landslides by day of the month (A) and in relation to the accuracy code provided (B). The dataset includes an "accuracy code", which should give an idea of the precision of the date. A landslide with code 1 indicates a claimed daily precision of dating, while code 2 means a short period precision ($1 - 7$ days). Codes 3 to 7 indicate a monthly, bimonthly, seasonal, biannual and yearly precision. Code 8 stands for a multiple year precision, while 9 and 10 indicate the date on the document or an uncertain date. Finally, 11 stands for events with unknown accuracy. For the period $2016 - 2018$ we have no information about the accuracy, but plot C exhibits the same distribution as plot A. 58
4.5 Histograms of the normalised three components. 65
4.6 Estimated location susceptibilities for Model 2A (left) and 2D (right). . 66
4.7 Observed/expected landslides and normalised rainfall and earthquake components in Emilia-Romagna (Model 2B). 67
4.8 Observed/expected landslides and normalised rainfall and earthquake components in Emilia-Romagna (Model 2D). 68
4.9 Component values across all days and all municipalities (Model 2B). . . 70

5.1 Geographical distribution of landslide dams in our dataset. 79
5.2 Box plots for the variables by dam conditions. 82
5.3 Histogram plots for the log-transformed selected variables. 83

5.4 Cumulative distribution function plot of each variable (dotted line): the log-normal distribution (solid line) is a good fit. 84
5.5 Box plots for the variables split by lifetime class 85
5.6 Scatter plots of Dam length against Dam width. 86
5.8 Variable distributions of events in Sicily and in the rest of Italy. The Sicilian events tend to have smaller values. 86
5.7 Scatter plots of River slope against Dam length. 87
5.9 Concatenated conditional structure of the selected variables. The lines in black represents the conditional structure used. The grey lines reflects other postulated connections that were not significant in the regression models. 89
5.10 Distribution of median dam failure time against DBI index for each event in the presented dataset 98

6.1 Diagram of possible interactions, according to Gill and Malamud (2014). The full arrow (—) represents a "triggering and increased provability" interaction, the dashed arrow (- - -) represents an "triggering" interaction and the dotted arrow (·····) represents a "increased probability" interaction. Adapted from Gill and Malamud (2014), Figures 2 and 4 of the paper. FL = flood, GH = ground heave, SS = soil subsidence, GC = ground collapse, RS = regional subsidence, AV = avalanche, LA = landslide, VO = volcanic eruption, TS = tsunami, EQ = earthquake, WF = wildfire, ET(C) = extreme temperature (cold), ET(H) = extreme temperature (hot),, LN = lightning, SN = snowstorm,HA = hailstorm, TO = tornado, ST = storm, DR = draught 102
6.2 Example of a potential framework. The susceptibility is the baseline for the potential function $U(t)$, which is affected by slow and persistent fast processes and affect the likelihood of the hazard. Transient fast processes can affect the likelihood of the hazard. A direct triggering of the hazard occurrence is visible as a spike in the conditional intensity $\lambda(t)$, as it will be explained in Figure 6.3 106

6.3 Examples of three different earthquakes affecting the change on a landslide conditional intensity function (CIF) and potential function between t and $t + \Delta t$, as per (6.5). In all three cases, the earthquake increases instantaneously in time the conditional intensity, up to a value which can cause consequences (e.g. landslides). In case 6.3a, the potential is not affected, as after the event it is equal to the pre-earthquake level. In case 6.3b, the earthquake produces the effect of keeping the potential higher than before the seismic event for a while (weakening). In case 6.3c, the earthquakes lower the potential to less than the value pre-earthquake (healing), which suggests an inhibition of following consequences. 108
6.4 Example of conditional intensity and potential functions for stress release model. The drops in the function correspond to event occurrences. The potential function has been obtained from (6.15). Data from Harte (2010) 111
6.5 Example of interaction between hazards in a system. An external/internal process can affect the potential of a hazard (represented by the function $U(t)$) and consequently provoke a secondary hazard occurrence. The latter may then trigger a third hazard. The dashed line encompasses the hazard system. 113
6.6 Examples of plots describing the two possible failures of dam: immediate failure after the landslide fall (red line) and failure due to accumulation of water (blue line). This plot schematically represents the two functions (6.49) and (6.50, left y-axis) in relation to the amount of water in the dam (orange line, right y-axis). 126
6.7 Examples of the resulting functions shown in Figure 6.6. The red line represents the product in (6.52), while the blue line the sum of of (6.49) and (6.50). 127
6.8 Comparison between the global ID thresholds defined in this study and published global (worldwide) ID rainfall thresholds. 1 Caine (1980); 2 Innes (1983); 3 Clarizia et al. (1996); 4 Crosta and Frattini (2001); 5 Cannon and Gartner (2005); 6 threshold inferred from the entire set of ID rainfall data (this work); 7 thresholds inferred from the probability estimates of the rainfall conditions, for two different rainfall periods ($D < 48h$, and $D \geq 48h$) (this work). Dashed line shows $0.25mm\ h^{-1}$ rainfall intensity. Picture and original caption from Guzzetti et al. (2008) 130
6.9 Examples of threshold function influenced by two factors. On the two axes on the horizontal plane, there are two components producing a probability surface. 131

7.1 The first plot (by Ermini and Casagli (2003) shows Italian dams in terms of watershed area (A_b) and the ratio volume/height split by DBI index (UD = unstable dam, SD = stable dam). The second plot is built on the dataset used for our analysis in Chapter 5. 142
7.2 Distribution of log median dam failure time against DBI index (on a log scale) for each event in the presented dataset. The two lines indicate the two DBI thresholds indicated in Ermini and Casagli (2003). 142

Abstract

Natural hazards such as earthquakes, floods and landslides threaten communities in every part of the world. Exposure to such perils can be reduced by mitigation and forward planning. These procedures require the estimation of event likelihoods, a process which is well understood for single hazards. However, spatio-temporal interaction between natural hazards, through triggering or simple coincidence, is not uncommon (e.g. Alaska 1964, the Armero tragedy, the Kaikoura earthquake), and can lead to more severe consequences than the simple sum of two separate events. Hence single hazard assessments may underestimate, or incorrectly estimate, the real risk through a lack of interaction analysis. In the existing research literature, multi-hazards assessments are most commonly approached qualitatively or semi-quantitatively, evaluating hazards via an interaction matrix, without formal quantification of the risk. This thesis presents a quantitative framework, using point processes as the key tool, to evaluate the interaction of primary hazards in the occurrence of secondary (triggered) ones. The concept of the 'hazard potential' is developed, as a means of generalizing hazard interactions in space and time, allowing event outcomes to be simulated within a simple point process framework. Two particular examples of multiple hazard interactions are presented: rainfall and/or earthquake-induced landslides, and the survival of landslide dams. In the first case, point processes are used to model the triggering influence of multiple factors in a large real dataset collected from various sources. By discretizing space and time to match the data resolution, a daily-spatio-temporal hazard model to evaluate the relative and combined effects on landslide triggering due to earthquakes and rainfall is created. The case study on the Italian region of Emilia-Romagna suggests that the triggering effects are additive. In the second example, a Bayesian survival model is developed to forecast the time to failure of landslide dams, based on their characteristics and those of the potential reservoir. A case study on heterogeneous Italian events is presented, together with examples of potential results (forecasting) and possible generalizations of the model.

Acknowledgements

First, I would like to thank Sandro Nanni (ARPAE) for providing us the rainfall dataset, Marco Pizziolo (Emilia-Romagna Region) for the landslide data and Matteo Taroni (INGV) for the earthquake catalogue. This work was supported by the Resilience to Natures Challenges National Science Challenge, New Zealand.

I am incredibly grateful with my supervisors, Professor Mark Bebbington, Dr. Xun Xiao and Professor Geoff Jones. Their support, wisdom and words of encouragement have been very helpful throughout my PhD. Thanks for challenging me to push over my limits, and for teaching me so many valuable lessons. Thank you for your constant patience and your precious advice.

Thanks to the staff of the Department of Statistics at Massey University, Palmerston North, and to all my postgraduate friends, with whom I shared four great years.

Thanks to my family, supporting me from far far away. This thesis is dedicated to my grandmother Vittoria, she would have been so proud of me.

Finally, I would like to thank my wife Daniela for her support. May this thesis be of inspiration for her academic career. There is no limit to what you can achieve.

Chapter 1

Introduction

1.1 Motivation

Natural hazards are a significant threat to human society, beyond death and injury, as their consequences can disrupt the activities and the economy of the affected communities. As part of efforts to mitigate the effects produced by natural hazards, hazard assessments are carried out to evaluate where, when and how these events may occur. Historically, hazard assessments have been single-hazard analyses (Camassi and Stucchi, 1996; Brunetti et al., 2006; Zhuang et al., 2012), focusing on the occurrences of one hazard, without considering any others. In some cases, a triggering mechanism from another hazard (e.g. rainfall triggering landslides) has been involved in the analyses (Guzzetti et al., 1999; Dai et al., 2002; Casagli et al., 2003; Berti et al., 2012). However, it is known that hazards are not necessarily independent (De Pippo et al., 2008; Kappes et al., 2010). In fact, they can occur in the same region within the same time window (Gill and Malamud, 2014), potentially aggravating the consequences on the communities. That suggests the need for an all-encompassing analysis of hazards and their interactions, known in the literature as multi-hazard analysis. This is a relatively new topic that has received increased interest in recent years (Tilloy et al., 2019), as an effect of the rising awareness that the combined effect of interacting hazards potentially produces aggravated consequences on communities (Komendantova et al., 2014; Liu et al., 2015; UNDDR, 2019). Hence, several multi-hazard or resilience programs such as HAZUS (Schneider and Schauer, 2006), RiskScape (Reese et al., 2007), Resilience Challenge (National Science Challenges, 2019), MATRIX (Komendantova et al., 2014) have shed light on this topic.

Although there is a widely expressed need for a multi-hazard framework, the examples presented in the literature are limited in number and mostly focused on the qualitative description of interactions. The obvious missing aspect is a tool to estimate the likelihood or the consequences of hazard occurrences. Several aspects make the

building of such framework challenging, particularly from a quantitative point of view. Multi-hazard analysis has to deal with "elements of quite different kinds" (Hewitt and Burton, 1971), because hazards differ in characteristics, in frequency, in consequences: so their assessments differ too. Furthermore, their interactions can be equally diverse. Gill and Malamud (2014) classified four different types of triggering: direct triggering, increased probability, decreased probability, spatio-temporal overlapping. Kappes et al. (2010) and Heinimann (1998) hypothesised that the interaction between hazards is the result of disposition, of the area to the hazard, and triggering, from other hazards. The disposition is related to environmental elements. Therefore, the assessment of potential interactions requires a well-conceived multi-hazard framework (Marzocchi and Woo, 2009; Chen et al., 2010; Kappes et al., 2010, 2012a; Gill and Malamud, 2014; Esharti et al., 2015). This framework would need to include characteristics and interactions of hazards, with the final aim of more effective risk mitigation and resilience procedures (Mahendra et al., 2011).

Multi-hazard methodologies are in the early stages, especially in quantitative terms. Many qualitative or semi-quantitative examples (e.g. Gill and Malamud 2014; Komendantova et al. 2016) have appeared in the literature, but the lack of statistical models or analyses hinders the replicability of the proposed method to any group of interacting hazards, as well as to any area in the world. For example, the use of qualitative or semi-quantitative matrices for the evaluation of hazards (De Pippo et al., 2008; Kappes et al., 2012b) represents a step towards a multi-risk framework, as there is the acknowledgement and the attempt to evaluate interactions between hazards. Qualitative methods can only provide a description of the effects, which is a solid knowledge basis for quantitative models, but cannot provide any estimates. Semi-quantitative methods propose a categorization of quantitative effect (e.g. Kappes et al. (2012b) classify actual ranges of hazard intensities into three categories: low, moderate, high) or attempt the creation of indices based on ranked qualitative categories and occurrence frequencies (Gill and Malamud, 2014). Furthermore, these analyses are often based on a specific area (De Pippo et al., 2008), which may result in models that are too specific to be applied to other areas or hazards. With the aim of proposing a more reliable and widely-applicable framework, some authors have proposed quantitative multi-hazard methods based on generated data that do not reflect the complexity of natural hazards (Mignan et al., 2014). The simulation of data was justified by the lack of complete and consistent data for natural hazards of different natures. Hence, the challenging idea is to build a framework starting from real data, based on a specific hazard chain, but that can be generalised and used in other regions and for other hazard chains.

The reason for preferring a stochastic modelling approach is that it would allow us to calculate the likelihood of a certain event, i.e. the probability of occurrence. Therefore,

it is vital to be able to study the frequency and magnitude of a hazard, particularly in combination with the frequency and magnitude of triggering and/or interacting hazards. In this way, it is possible to observe the occurrences of hazards over time and evaluate more realistically the future occurrence probabilities of the triggered hazard.

The best way of achieving this is to build a fully quantitative and dynamical statistical model capable of evaluating the interactions among natural hazards. The existing literature proposes limited use of statistical methods (mostly copulas and extreme value distributions), to model hazard thresholds (Tilloy et al., 2019), such as the rainfall amount necessary to trigger a landslide (Berti et al., 2012). However, these cannot be considered as stochastic models, because they do not study the evolution of a hazard over time. More flexible methods, such as point processes (Vere-Jones, 1978; Isham and Westcott, 1979; Ogata, 1988), would be useful for this task because they can model the occurrence of events over time, while taking into consideration their magnitude or other characteristics. However, the shortage of data, which characterises natural hazard analyses on multi-hazard occurrences (Aksha et al., 2020), has heavily limited the research so far.

Point processes are stochastic models used to describe the occurrence of events in time or space. They have commonly been used in hazard analysis, particularly to model earthquake occurrences (Zhuang et al., 2004; Ogata and Zhuang, 2006; Lombardi and Marzocchi, 2010; Zhuang, 2011), rainfall and storms (Isham and Westcott, 1979; Rodriguez-Iturbe and Eagleson, 1987; Cowpertwait et al., 2007), eruptions (De la Cruz-Reyna, 1991; Bebbington, 2008; Wang et al., 2020) and wildfires (Xu and Schoenberg, 2011; Yang et al., 2015; Aragó et al., 2016). One of the main advantages of point processes is that they can be adapted to model various types of events, in relation to their interactions.

In particular, we need processes that can embrace the features typical of the possible interactions between hazards. Processes such as self-exciting ones (Hawkes, 1971; Ogata, 1988), self-correcting ones (Isham and Westcott, 1979), and their combinations (Schoenberg and Bolt, 2000) are able to capture the characteristics of the events in terms of frequency distribution, from short-term-high-frequency ones (clustering) to long-term-low-frequency ones (fast decay).

Features such as these can be adapted and extended to model not only the hazard occurrences, but the interaction between hazards. In fact, there is a connection with the classes of hazard interactions (direct triggering, increased/decreased probability, spatial/temporal overlapping) proposed by Gill and Malamud (2014). For example, the occurrence of one type of event in a short time period may increase the probability of occurrence of another hazard. Hence, these models can be useful methodologies for our framework.

To comprise all possible aspects related to the hazard interactions, more sophisticated models, such as marked point processes or processes with multiple layers of stochasticity (Rodriguez-Iturbe et al., 1987), can be used to include additional dimensions (e.g. magnitude, space) or combine more than one point process to express different features of a phenomenon, making the overall model a better representation of the data.

1.2 Overview

This thesis aims to explicitly examine the possibilities for placing multi-hazard analysis in a quantitative footing. As a first step, I will present a quantitative multi-triggering model for landslides triggered by earthquakes and rainfall. The second step will be to investigate the feasibility of extending the multi-hazard framework further into the proposed hazard chain. This will be done by studying the time to failure of landslide dams. Finally, the concept of potentials in multi-hazard analysis will be introduced, as a first step to a broad and comprehensive framework for multi-hazard analysis.

1.2.1 Point process for hazard interactions

The first part of the thesis is a review of the quantitative methodologies that can be used for a multi-hazard analysis in Chapter 2. Chapter 3 assesses the state of the art for hazard assessments, with a particular focus on a specific hazard chain, that will be described below.

To produce a first attempt at a quantitative framework for multi-hazard model, the scope of the study has been restricted to the chain of hazards constituted by earthquakes and rainfall triggering landslides and extended to landslide dams, as shown in Figure 1.1. Earthquakes and rainfall overlap in space, but are considered to be only coincidentally related, meaning that their coincidence is considered random (Gill and Malamud, 2014; Havenith et al., 2016). However, they can both trigger landslides (Shou et al., 2011; Zhou et al., 2015b; Iverson, 2000; Berti et al., 2012), so there is a need (Kappes et al., 2012b) for a statistical model that incorporates both features, in such a manner that the separate and joint triggering effects can be estimated.

As mentioned above, there is a lack of the type of data required for a quantitative multi-hazard analysis. More specifically, it is not easy to find time-homogeneous, time-stamped, geographically congruent data, ideally with magnitude data for at least two hazards. In some cases, data is available but with different levels of completeness and homogeneity. For example, rainfall datasets from different institutions (NIWA (2012), New Zealand; TRMM (2017), USA; Arpa-Piedmont (2018), Italy), and landslide datasets (ISPRA, 2019) differ substantially in completeness and homogeneity.

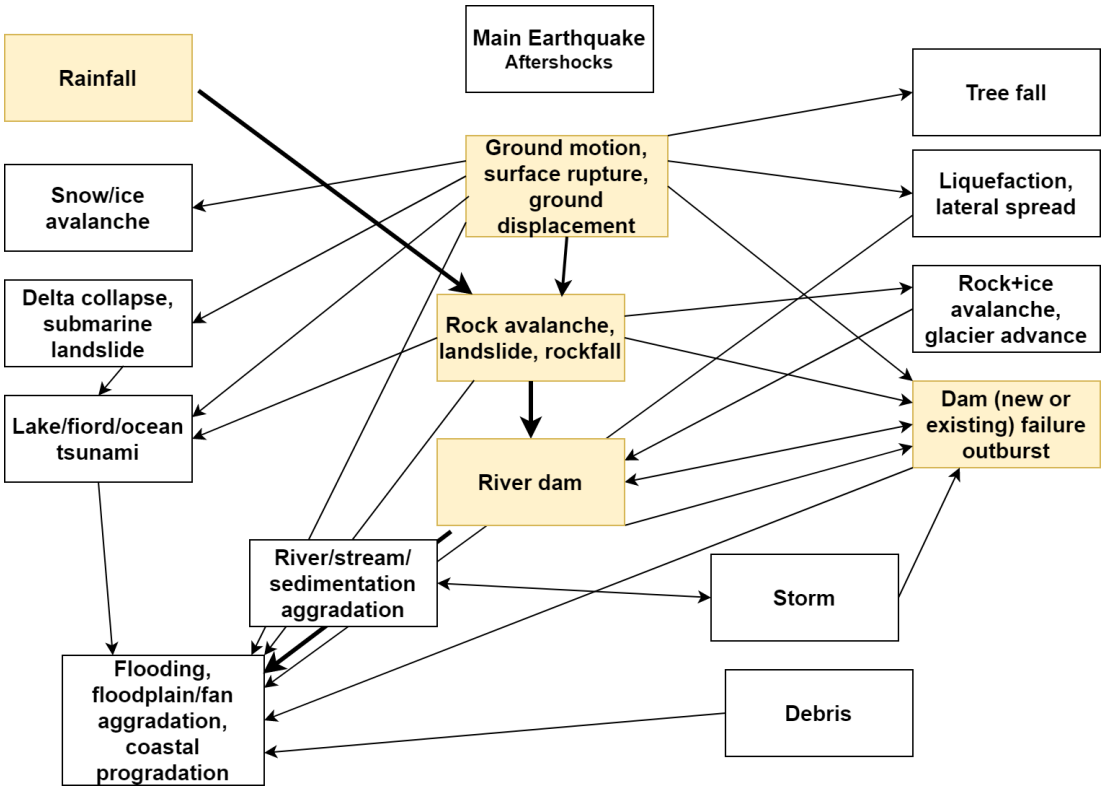


Figure 1.1: Example of complex chain of hazards. The image is from Orchiston et al. (2018), modified to include rainfall, and to highlight the reduced chain of hazards that has been chosen for this thesis. Each arrow represent an interaction between events. The bold arrows highlight the chain of events that is used in this thesis.

Hence, combining these datasets in one framework is challenging, and it is one of the challenges on quantitative multi-hazard modelling. Hazard occurrences may be also relatively rare for certain hazards, such as landslide dams (Costa and Schuster, 1988; Tacconi Stefanelli et al., 2015), due to the nature of the hazard. Furthermore, the difficulties in the collection process lead to recording errors and incompleteness, resulting in incorrect or missing data (Santi, 2018). Finally, due to these issues, datasets have different resolutions: for instance, earthquakes are typically recorded to the second, landslides to the day or to lesser precision. Therefore, not only it is difficult to obtain data, but there is often also the need for a substantial process of data cleaning and data imputation, to merge multiple datasets for a multi-triggering assessment. Chapter 4 includes considerable material on how this problem has been dealt with, in the case of three datasets for the Italian region of Emilia-Romagna.

In order to provide a first attempt at a quantitative multi-triggering model, Chapter 4 proposes a process to model the interaction between rainfall and earthquakes in the triggering of landslides. The region was chosen because of its pronounced seismic

activity: the 2012 earthquake sequence studied in Govoni et al. (2014) is an example of a long sequence with large earthquakes and aftershocks, in a region with diffuse seismicity (Chiarabba et al., 2005; Boccaletti et al., 2011). Furthermore, the region is particularly interesting for the presence of one of the largest landslide datasets available (ISPRA, 2019).

In the chosen hazard chain in Figure 1.1, landslides are commonly assessed via susceptibility maps, where the susceptibility is proportional to the spatial likelihood of a landslide, given a landslide has occurred. With appropriate normalization, it represents a spatial intensity. The only relevant interactions that have been studied so far are the triggering interactions earthquake-landslide and rainfall-landslide. These are usually studied via threshold analysis, e.g. intensity/duration (Berti et al., 2012), or susceptibility analysis, logistic regression Ayalew and Yamagishi (2005).

As mentioned earlier, point processes are useful methods to model the triggering effects of hazards over time. Nevertheless, landslides do not occur on the vast majority of days because, being predominantly triggered by heavy rainfall or earthquakes, not all days are suitable for landslide occurrence. A Zero-inflated Poisson (ZIP) model is proposed in Chapter 4, in order to deal with extremely sparse data over time, as it can model the excess of zeroes separately to the landslide process. Another challenge encountered was that landslides are recorded as discrete, often grouped data: events occur in remote areas and recorders are not able to recover the exact time of the event, but can only provide an approximation within bounds. Consequently, landslides with uncertain dates have been reallocated in proportion to the conditional probability of occurrence calculated from those recorded events considered to be precise.

The model is composed of terms specifically tailored to express the triggering effect on landslides: the earthquake triggering term, the rainfall triggering term, and the interaction term between earthquakes and rainfall in the triggering of landslides. In particular, the coseismic effect is based on existing empirical relationships (Utsu, 1970; Ogata, 1988; Wetzler et al., 2016) between main event magnitude and aftershock productivity as a proxy for ground shaking. Hence, the driving factors of the coseismic term are the magnitude of the earthquake and the distance between landslide and epicentre. In terms of rainfall, two terms expressing the average of the last two days of rainfall (short-term) and the weighted average of the last Δ days of rainfall (long-term) will express the effect of rainfall on landslides occurrences (Monsieurs et al., 2019; Rossi et al., 2010).

Although rainfall and earthquakes overlap in space and time, it is unclear how these factors interact in the triggering of landslides. The long-term rainfall effect is the main candidate to provide interaction with seismicity, as its prolonged effect is more likely to temporally overlap with seismic occurrences (Brain et al., 2017). Furthermore, the soil

saturation might have an effect on coseisimic landslides (Martino et al., 2020). Because of this knowledge gap, different formulations of the model has been assessed to view the best way to model the interaction.

1.2.2 Survival analysis of landslide dams

Chapter 5 moves further into the hazard chain shown in Figure 1.1. I apply a survival model to a landslide dam dataset in order to study the time-to-failure of these hazards. This kind of analysis is especially important for those hazards whose risk needs to be quickly assessed, as their situation can evolve rapidly. Insights will also be sought on how to deal with multivariate data with missing values in one or more variables.

Landslide dams are a common hazard which threatens downstream human settlement or infrastructure, as their collapse may result in a flash flood. The accumulated water compound the danger; therefore, the estimation of the time to failure becomes crucial for assessing proposed engineering procedures, and other risk mitigation techniques. The survival analysis is carried out by building a Bayesian model to predict the time to failure of landslide dams, based on the characteristics of landslide dams present in a dataset containing nearly 300 events occurred in Italy.

There are very few datasets for landslide dams (Costa and Schuster, 1988; Ermini and Casagli, 2003; Tacconi Stefanelli et al., 2015), and due to the nature of the event, they are characterised by a high percentage of missing data. The dataset by Tacconi Stefanelli et al. (2015) was chosen as one of the most complete in terms of variety of events: it includes dams that failed within hours or days, which are usually missing from other datasets. However, only 8 observations out of 300 were complete, including the time of damming and time of failure, which are vital in a time-to-event analysis. Together with the lack of time-to-failure data, there are also missing values within the covariates. Hence, Bayesian imputation was used to impute the missing values, via an analysis of the covariance structure of the variables.

The dimensions of landslide dams are considered as crucial factors for their survival, and they have been used to evaluate the dichotomy between failure and non-failure of landslide dams (Ermini and Casagli, 2003; Korup, 2004; Liao et al., 2018). By including these covariates in a survival model, it is possible to understand which ones in particular affect a landslide dam the most. This will allow for a quantitative assessment of the time-to-failure, which is an important improvement from the qualitative or semi-quantitative failure/non-failure assessments. The proposed model will lead to a significant advancement in the field, as it proves that robust quantitative results can be obtained even from very limited data. Furthermore, this approach will provide vital tools for the engineers and the stakeholders, taking decisions on a hazard that might produce severe consequences within a certain period of time.

1.2.3 Potentials

The large variety of hazards and interactions (De Pippo et al., 2008; Gill and Malamud, 2014) complicate multi-hazard analysis, due to different evaluation methods and hazard characteristics. Nevertheless, the identification of some patterns among the interaction of hazards might help finding a method that would transform this problem into an opportunity. To do so, in Chapter 6 I propose a conceptual framework for the occurrence of hazards, which considers the history of the hazards, their characteristics and their interactions. Starting from the concept of "potential" (Mignan et al., 2014), ideas from several authors (Kappes et al., 2010; Gill and Malamud, 2014) will be combined, particularly focusing on the notions of susceptibility and triggering (Heinimann, 1998; Kappes et al., 2010). Taking the example of landslides once again, the ground features, e.g. type of soil, affect the susceptibility of a land to landslides, but it is the combination of triggering hazard (rainfall and earthquakes) in space and time that actually increase or decrease the potential of a landslide.

Hence, in Chapter 6 I define the potential as a scalar-valued function encapsulating the history of the process up to a given time t into a single value. This would simplify the estimate of the current state of a system of natural hazards, producing an outcome that summarizes the combined effect of the disposition and triggering of the hazard over time. A key concept is the possibility to isolate and trace the different processes affecting the change of potential over time. For instance, in Chapter 4 the rainfall hazard is split into two components, expressing the short and long term effects of rainfall on landslide occurrence. Similarly, rainfall will be considered in the potential framework as the combination of two distinct processes affecting the potential of landslides. Examples of application of the potential concept on pre-existing models are presented. We will start with single-hazard models: self-exciting processes, such as the Hawkes process, are natural candidates for this framework because of how their conditional intensities are expressed. On the contrary, other models do not allow for the use of a potential function to summarize the history of the process. This is the case with Epidemic Type Aftershock Sequences, or ETAS (Ogata, 1988), because of the formulation of part of the conditional intensity. Multi-hazard cases will be investigated, with a focus on the different functions that can be chosen to model the different processes and their interactions. The cases of earthquake/rainfall triggered landslides, landslide dams and debris flow are explored. I will demonstrate the possibility of building modules of processes with concatenated marks to simplify the structure of the overall model and at the same time to mark the interaction among hazards. Also, I will show that in some cases processes can be approached from different angles, by introducing the concept of a masking function. This function shows or hides the potential of a hazard, while a second function delivers the effect of an continuous process. For instance, the

presence/absence of debris is a masking function for the potential of debris flow, given that the rainfall is the continuous process. It will be shown that it is possible to switch these two functions, using the rainfall as a masking function and a continuous process for the presence of debris. This provides a certain degree of freedom to adjust the model to other chain of hazards.

Similarly to Mignan et al. (2014), the framework will be assembled with a building block approach, which will allow for the combination of processes in order to express their interactions. The use of point processes as a basic structure for the framework will facilitate this process. Furthermore, the building block approach will also have an important benefit in the simulation of multi-hazard processes. Existing algorithms allow for single-hazard or limited multi-hazard examples, with drawbacks of lack of speed and the need of a large amount of process memory. Instead, the flexibility of this framework can simplify and increase the speed of the simulation by using a block for each hazard and taking into consideration that marks are inherited from a triggering hazard to a triggered hazard. The reduced process memory requirements will consequently facilitate forecasting and provide a solid base for more complex systems.

Finally, Chapter 7 discusses the results of the thesis and proposes suggestions for future research.

Chapter 2

Methodology review

2.1 Introduction

In aiming for a quantitative framework for hazard assessment, it is paramount to use appropriate statistical methods. Such methods should allow for the representation of event occurrences in time and/or space, describing events with respect to their history. A good level of flexibility is also essential so that these models can be combined, or enhanced with additional hazard features. A family of suitable models that hence will be used in this thesis is that of point processes. A point process is a "stochastic model that defines probabilistic rules for the occurrence of points in time and/or space" (Zhuang et al., 2012), and every new temporal or spatial point of the process is related to the history of previous points.

In particular, this chapter will review the temporal point process as a possible tool for the analysis of hazard interactions over time. Then, the time discretization of point processes will be investigated, as necessary means to deal with data characterised by different levels of temporal resolution (see Chapter 4).

Then, a review of survival analysis is presented, in order to have a methodology for time-to-event data, of particular interest when a hazard occurrence produces an impending threat. Furthermore, Bayesian imputation will be introduced, as a useful tool to extract as much information as possible from datasets naturally characterised by missing data.

2.2 Point Processes

The history of point processes can be traced back to Siméon-Denis Poisson in the mid-19th century, although their main development occurred in the 20th century as a result of the contribution of measure theory. A good portion of the terminology now used in point process, as well as the use of models that we associate with point processes

(e.g. exponential distribution), is derived from renewal theory. This latter studies the sequence of intervals between occurrences of an event, the concept of life tables (from actuarial statistics), and the enumeration of occurrences of events in specific time periods or space (Daley and Vere-Jones, 2003).

2.2.1 Definition and properties of temporal point processes

There are several approaches to define a point process, thanks to their applicability to many disciplines. One of them defines the point process as a counting process, which enumerates the occurrences of an event in a specific set, usually time or space. It is possible to imagine a temporal point process as points (events) on a line. In each interval on the line, there might be zero, one or more points. Each point occurs according to a rule, an intensity function, and to the previous history of the process.

A temporal point process is described by a counting process $N(t)$, the number of events occurring up to and including t, where $N(0) = 0$. In particular, in a sufficiently short time interval $(t, t+\Delta t]$, with Δt small, the probability of an event is $\lambda(t|H_t)\Delta t + \circ(\Delta t)$. The function $\lambda(t|H_t)$ is called the conditional intensity function and it expresses the expected rate of events at time t given the history of the process H_t up to time t:

$$\lambda(t|H_t) = \lim_{\Delta t \to 0} \frac{P[N(t+\Delta t) - N(t) > 0|H_t]}{\Delta t} \tag{2.1}$$

where $N(t+\Delta t) - N(t)$ is the number of events in a small interval $(t, t+\Delta t]$, also called differential. (2.1) can be explained as the instantaneous conditional probability of an event. In other words, it represents the conditional risk of an event occurrence in t given the previous realizations of the process in the interval from the time origin to t (Daley and Vere-Jones, 2003). The conditional intensity $\lambda(t|H_t)$ is then a rate function based on the process history, and because the probability of an event in any interval is non-negative, then the conditional intensity function needs to be non-negative everywhere as well.

If the conditional intensity $\lambda(t|H_t)$ is parameterized, then the parameters can be estimated by maximizing the log-likelihood:

$$\log L = \sum_{i=1}^{N(T)} \log[\lambda(t_i|H_{t_i})] - \int_0^T \lambda(t|H_t)dt \tag{2.2}$$

where the process is observed on the window $[0, T]$ and events occur at times $\{t_i\}$ where $0 \leq t_1 < t_2 < \ldots < t_n \leq T$.

A point process is defined as stationary if the structure of the process, its conditional intensity, is not altered by the translation of time by an arbitrary shift (Cox and Isham, 1980). In other words, $P[N(t+\Delta t) - N(t) > 0|H_t]$ depends on the length of Δt, and

not on the time instance t.

Renewal process and Poisson process

The simplest class of temporal point processes is the renewal process, a random point process with interarrival times X_i with $i = 1, \ldots, n$ defined as independent and identically distributed random variables. Defining the arrival times as

$$T_n = \sum_{i=1}^{n} X_i, \quad n \in \mathbb{N} \tag{2.3}$$

where T_n is the time for the n^{th} arrival for $n \in \mathbb{N}$. The sequence $\boldsymbol{T} = (T_0, T_1, \ldots)$ is the arrival time process, although $T_0 = 0$ is not necessarily considered an arrival. Then, the interarrival times can be recovered from the arrival times:

$$X_i = T_i - T_{i-1} \tag{2.4}$$

Hence, the process is a renewal process if the interarrival times $\{X_n : n \geq 1\}$ are non negative i.i.d. with distribution function

$$F(x) = P(X \leq x), \quad x \in [0, \infty) \tag{2.5}$$

It has to be noted that X_1 might have a different distribution from the other times possibly due to incomplete observation, as the time at which observation of the process began may not have been an event.

For the elementary renewal theorem (Feller, 1941; Smith, 1958), the rate of the process is equal to the reciprocal of the mean interarrival times, asymptotically. With

$$\lim_{t \to \infty} \frac{N(t)}{t} = \lambda \tag{2.6}$$

Because $t_{N(t)} \leq t < t_{N(t)+1}$ and $t_{N(t)} = X_1 + \ldots + X_{N(t)}$, dividing by $N(t)$ we obtain

$$\frac{1}{N(t)} \sum_{i=1}^{N(t)} X_i \leq \frac{t}{N(t)} \leq \frac{1}{N(t)} \sum_{i=1}^{N(t)+1} X_i \tag{2.7}$$

Because the extremes converge to $E(X)$ with $t \to \infty$ (for the law of large numbers), then also the element in the middle will converge too.

An important special case of renewal process is the Poisson process, which is used in many fields. A Poisson process is a temporal point process, with independent increments. In fact, given $0 \leq q < r \leq s < t$, $N(r) - N(q)$ and $N(t) - N(s)$ are independent Poisson random variables. The interarrival times $\{X_n : n \geq 1\}$ in a Poisson process are

i.i.d following an exponential distribution with probability density

$$f(x) = \lambda \exp(-\lambda x) \text{ for } x \geq 0 \tag{2.8}$$

The parameter λ is the rate of the process and it is constant over time. In other words, for any unit time interval, there is a constant expected number of occurrences. This can be compared with a general renewal process, whose rate is constant only asymptotically. Furthermore, the mean interarrival time is equal to the reciprocal of the rate: $1/\lambda = E(X)$. Consequently, the Poisson process is characterised by the memoryless property: the time to the next occurrence is not dependent on the time elapsed from the previous one. Thus, the expected number of arrivals is the same for two non-overlapping intervals. Given $t \geq 0$ and $s \geq 0$, the probability of event occurrence after $t + s$, given s, is equal to the probability of event occurrence after t: $P(X > t + s|X > s) = P(X > t)$. Therefore, the counting process $N(t)$ is a Poisson process of rate λ, independent of the history of the process.

$$\Pr\{N(t + \Delta t) - N(t) = 1|H_t\} = \lambda \Delta t + \circ(\Delta t) \tag{2.9}$$

Figure 2.1 shows an example of Poisson process: there is no pattern among the events.

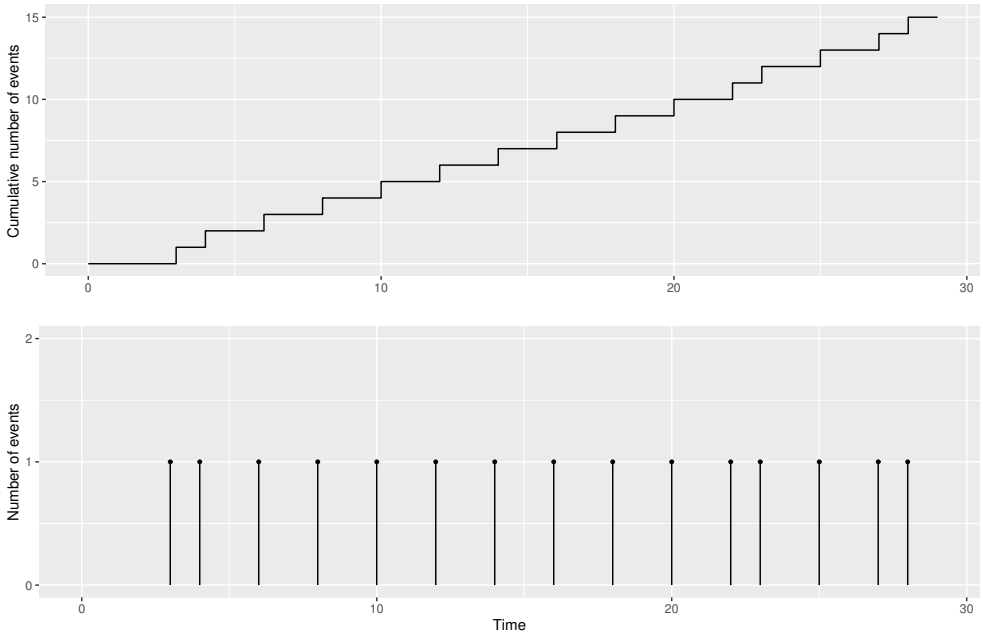


Figure 2.1: Example of Poisson process. The first plot show the cumulated events over time, while the second plot shows the occurrences.

Self-correcting processes

Self-correcting processes model the occurrence of events over time in an environment that is more general when compared to a renewal process. The key difference is the self-correcting process is dependent on earlier occurrences in such a way that allows for automatic correction of deviations from the mean. This can be achieved by specific models with a negative correlation between event occurrences so that the likelihood of an event is somewhat reduced by the occurrence of the earlier ones.

The first self-correcting model was proposed by Isham and Westcott (1979) as a model capable of producing ρt events over $(0, t]$ for a process with rate ρ. The self-correction seeks to have a difference between $N(t)$ and ρt close to zero. In other words, it corrects the deviations from the mean. To do so, the conditional intensity is defined as

$$\lambda(t|H_t) = \rho\phi(N(t) - \rho t) \tag{2.10}$$

where $\rho > 0$ and the function $\phi(\cdot)$ is chosen so that $0 \leq \phi(x) < \infty$ and $\phi(x) \geq c \; \forall x > 0$. In the same period, Vere-Jones (1978) introduced a substantially similar model, based on the elastic rebound theory (Knopoff, 1971).

Stress release model The Stress Release Model (SRM) (Vere-Jones, 1978; Vere-Jones and Deng, 1988), is used for phenomena characterised by characterised by loading and release (Reid, 1910). As parametrized by Vere-Jones (1978),

$$X(t) = X(0) + \rho t - S(t) \tag{2.11}$$

where $X(t)$ is the Benioff stress of the system at time t, $X(0)$ the initial level of stress, ρ the constant loading rate from external tectonic forces (which makes the stress increase linearly), and $S(t)$ the accumulated stress release from earthquakes within the region in: $S(t) = \sum_{t_i<t} S_i$ (Bebbington and Harte, 2003).

The conditional intensity is an exponential function of $X(t)$:

$$\lambda(t|H_t) = \exp[\mu + \nu X(t)] = \exp[\mu + \nu X(0) + \nu(\rho t - S(t))] \tag{2.12}$$

The function $\mu/\nu + X(0)$ can be interpreted as a parameter for the initial value of stress, which is unknown. The parameter ν can then be interpreted as the composition of strength and heterogeneity of the area, respectively. It is possible to re-parametrize (2.12) as shown by Ogata and Vere-Jones (1984):

$$\lambda(t|H_t) = \exp[\mu + \nu X(0) + \nu\rho t - \nu S(t)] = \exp[a + bt - cS(t)] \tag{2.13}$$

where the last step is an alternative parametrization used for numerical optimization

(Harte, 1999), where $\alpha + \beta X(0) = a$, $\beta\rho = b$ and $\beta = c$. The conditional intensity in (2.12) or (2.13) evaluates probability of an earthquake in relation to the stress level. The SRM accounts for the constant increase of stress over time, but also for the stress release (S_i) after an earthquake i. A large release of stress can make $X(t)$ become even negative. Ergo, during such recharge time, any subsequent event will be unlikely to occur. A variant was proposed by Bebbington et al. (2010), where the maximum magnitude was limited in order to keep $X(t)$ non-negative, and thus exclude the large events, except when the stress level was high. In this particular case, the model, studying the San Francisco Bay Region data, showed a reduced activity before a large event (Schwartz et al., 2014). For each event i with magnitude M_i, the energy released is computed as $S_i = 10^{0.75(M_i - M_0)}$, where a convenient lower magnitude threshold M_0 is subtracted from the magnitude of every earthquake, for computational reasons.

Self-exciting processes

On the opposite side, self-exciting point processes describe random sequences of events, where every event increases the likelihood of occurrence of future events. This increased likelihood leads to clustering of events, hence this type of stochastic models can represent the parent-offspring relationship between earthquake main shock and aftershocks (Hawkes and Oakes, 1974; Reinhart, 2017; Derek Tucker et al., 2019). In such a case, the covariance between interarrivals will be positive, so that the next event will be in a way attracted by the occurrence of the former one. These models have applications in a wide range of topics, from epidemiology (Meyer et al., 2012) to seismology (Musmeci and Vere-Jones, 1992), from criminology (Mohler et al., 2011) to finance (Errais et al., 2010). The concept behind the self-exciting process is that the seismicity can be seen as the sum of "background" earthquakes caused by tectonic loading and earthquakes "triggered" by other earthquakes. Earthquakes are complex phenomena to study, but the use of stochastic model, in the last thirty years (Kattamanchi et al., 2017), has helped the progression of this field.

Hawkes process The original self-exciting process is the Hawkes process (Hawkes, 1971), on which several self-exciting processes are based. The conditional intensity of the process is:

$$\lambda(t) = \mu + \int_{-\infty}^{t} h(t-s)N(ds) \tag{2.14}$$

where μ is the constant background rate and the function $h(\cdot)$ is the triggering function, which is often exponential, gathering the effects of the events occurring in a certain time window. Each event will increase the likelihood in relation to its magnitude, and this effect will decrease, e.g., exponentially at a given rate. This function will depend on

recent history or rather take into consideration a longer period, in relation to the form chosen for $h(\cdot)$. Figure 2.2 shows an example of Hawkes process.

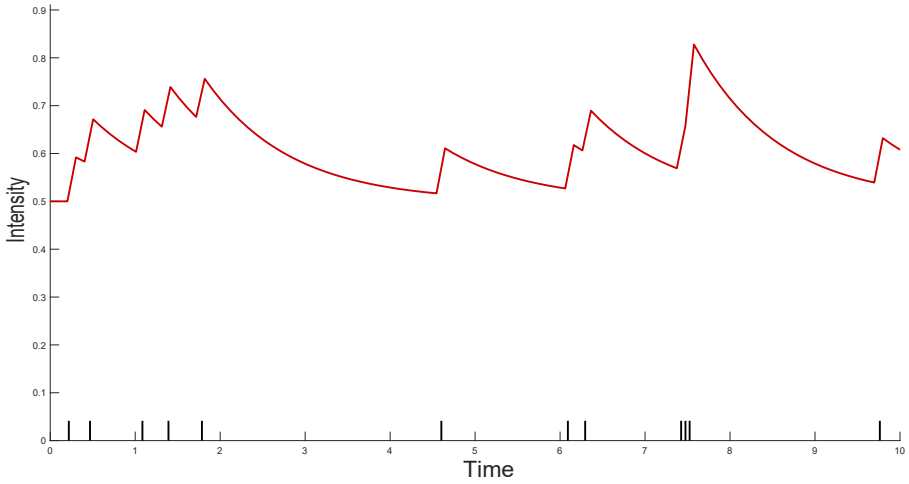


Figure 2.2: Example of Hawkes process: the line represents the conditional intensity of the process. At the bottom, the realization of the process, i.e. event occurrences. Each event produces a jump in the conditional intensity, represent by a spike in the line, followed by a temporal decay. A cluster of events produced a bigger jump and a longer decay.

ETAS model The ETAS model (Ogata, 1988, 1998a; Console et al., 2009) is formulated to incorporate empirical rules (such as Omori's law) to improve the modelling of earthquake sequences. The starting point of this model is that an earthquake event can trigger aftershocks in relation to its magnitude. With this approach, we consider that each event, large or small, can produce an aftershock cascade. The temporal conditional intensity of this model is:

$$\lambda(t) = \mu + \sum_{i:t_i<t} \frac{K\exp[\alpha(M_m - M_c)]}{(t - t_i + c)^p} \tag{2.15}$$

where μ is the background rate. The model includes the Omori decaying frequency law for aftershocks (Omori, 1894) $\lambda(t) = k/(t - t_i - c)^p$ in its modified version by Utsu and Ogata (1995) and the productivity law (Utsu, 1969) $N = K\exp[\alpha(M_m - M_c)]$: the number of aftershocks exponentially distributed and dependent on the mainshock magnitude. The multiplicative parameter K is called "aftershock productivity" and regulates the triggering intensity. It can be computed as the mean number of direct aftershocks per earthquake (Sornette and Werner, 2005). The parameter α controls how the productivity increases with magnitude.

Marked Point Processes

Before introducing spatial and spatio-temporal point processes, it is worthwhile to introduce a type of enhanced point processes, as they are also used to include spatial information in temporal point processes. In fact, marked point processes are useful tools for more complex models. It is possible to define a series of events as a marked point process (MPP) $\{(t_j,m_j) : t_j \in (0,T],m_j \in M\}$, where $\{t_j\}$ is a temporal point process and $\{m_j\}$ are the associated marks. The conditional intensity function is formed from two parts (Harte, 2010): the ground intensity function and the distribution of the marks:

$$\lambda(t,m|H_t) = \lambda(t|H_t) \times f(m|t,H_t) \tag{2.16}$$

The ground intensity function $\lambda(t|H_t)$ is the temporal point process describing the occurrence of events, while the density function of the mark distributions $f(m|t,H_t)$ describes the characteristics of the events, making the model more realistic.

For instance, it is possible to apply MMP to earthquakes, modelling the mark distribution to follow the Gutenberg-Richter frequency-magnitude law. This law expresses that the earthquakes N with a magnitude greater or equal to M in a region are distributed as a power law: $N = 10^{a-bM}$. It can be also used to model the mainshock-aftershock distribution (Utsu, 1969). The b parameter then regulates the frequency-magnitude of aftershocks (Rotondi and Varini, 2019). It has been observed that b is large after a strong earthquake, and it reduces only after a large aftershock (Wiemer et al., 2002). The magnitude of earthquakes was already present in the conditional intensity functions of the SRM and ETAS models, but in MPP they are generated by a dedicated function (additional to the intensity function), rather than being provided independently to the process. The overall conditional intensity $\lambda(t,m|H_t)$ of the marked point process expresses the infinitesimal expected rate of events occurring at time t with mark m. Therefore, a marked point process can be interpreted as an extended version of a point process, taking into account the distribution of marks.

2.2.2 Spatio-temporal models

So far, this chapter has covered the review of temporal models, but natural hazards occur and interact in space and time. Then, it essential to identify and use models that incorporate both spatial and temporal information in a multi-hazard framework (Kappes et al., 2012a; Gill and Malamud, 2014). In this section we will propose a review of space-time models. Before that, we need to briefly introduce spatial models, in order to have a basic understanding of both the models (temporal and spatial) forming spatio-temporal ones. Because the data of some natural hazards may not suit models with continuous dimensions, this section will also discuss the discretization of

at least one dimension.

Brief introduction to spatial point processes

Spatial point processes describe the distribution of events in space (Cressie, 1993; Diggle, 2003). The main difference from the temporal processes is that time is monotonically ordered, while space is not, and so they have to be treated differently. Spatial point processes are used in different environments, from medicine (Jing et al., 2011) to forestry (Mateu et al., 1998) and wildfire (Aragó et al., 2016).

In the case of a Poisson spatial process, with $N(B)$ being the number of points falling into a subset of B a 2-D space, the probability of an event occurring in a given area is:

$$\Pr\{N(B) = 1|\epsilon_B\} = \lambda|B| + \circ(|B|) \tag{2.17}$$

where ϵ_B represents all the points occurring outside B, $|B|$ is the measure of B (usually area), considered and λ is the constant rate of event occurrences per unit area. If λ is not constant but instead a function of position, the process is called a non-homogeneous Poisson process. One of the main issues with this process is that geographical unit cannot be considered independent. In the example of a wildfire, one unit is likely to be similar to the next ones in terms of vegetation and geomorphology. This issue can be dealt by using complex dependency structures, or using a kernel smoothing estimator.

However, spatial processes cannot represent the whole behaviour of natural hazards, as the temporal aspect is not considered. Spatio-temporal point processes have been used for modelling the occurrences of wildfire (Yang et al., 2015; Díaz-Avalos et al., 2016) and earthquakes (van Lieshout and Stein, 2012; Zhang and Huang, 2017). One of the most recognised of such models is the spatio-temporal application of the earlier presented ETAS models.

Spatio-temporal ETAS Spatio-temporal ETAS are well established models for modelling short-term seismicity (Console et al., 2009; Iwata, 2015; Guo et al., 2015), formulated by Ogata (1998a). In these cases, the background occurrence rate is assumed to be a function of spatial location and time (Ogata, 1998b). The model is characterized by a probability distribution of the time until the appearance of a child event, and a probability distribution of location (x,y) and magnitude m of a child event (an aftershock), dependent on the magnitude m_k and location (x_k,y_k) of its parent. The conditional intensity function is:

$$\lambda(t,x,y,m) = \lambda_0(x,y) + \sum_{k:t_k<t} h(m_k)g(t-t_k)f(x-x_k,y-y_k,m_k) \tag{2.18}$$

where $k(m_k)$ is the magnitude distribution of new events above a threshold magnitude. This function can be linked to the productivity law expressed in (2.15). The functions $g(t-t_k)$ and $f(x-x_k,y-y_k,m_k)$ are the probability density for the occurrence in time and space of the offspring events from an ancestor of magnitude m_k, at time t_k and location (x_k,y_k); $\lambda_0(x,y)$ is the background event in the given location.

2.2.3 Discretization

In the previous section it has been noted that spatial and spatio-temporal processes can be used to model natural hazards. Nevertheless, these models often cannot be entirely continuous because of the type of available data. Some hazards have naturally discretized data, which might hamper the application of point processes. This is the case of landslide data. As will be discussed in Chapter 4, landslides times have a daily precision at the most, and their localization may not be accurate enough to allow for continuous spatial data (such as geographical coordinates. Building a multi-hazard framework implies combining hazards with different data precision both in space and time. Hence, we need to find models which let us express events over possibly discretized time and space data. Because this issue is not infrequent with natural hazards, there is the need to find a way to treat data that are naturally discretized in a point process environment.

Space discretization

In Section 2.2.2, we have discussed point processes that treat both space and time in a continuous way. However, some natural hazards may have naturally discretized spatial data, to model which the point processes need to be modified. In the literature there are examples of point processes in which the discretization of space and/or time is used to adapt to the available data and carry out complex models. For example, Preisler et al. (2004) have used discretized voxels (three-dimensional pixels) to study the propagation of wildfire both in time and space. Wotton and Martell (2005) used discretized space to model lightning as cause of wildfires in Canadian forest. The use of pixels, voxels or spatio-temporal units allow spatial or spatio-temporal models to express the spread of hazards from one geographical and/or temporal unit to the next ones, hence in a non continuous way.

Often the reason for such discretization is the different data precision among natural hazards. The use of pixels or geographical units requires to take into consideration the possible dependence between them: a wildfire in a pixel increases the likelihood of wildfire in the neighbour pixels. This not often considered, as demonstrated but some landslide susceptibility analyses, where the factors are considered in function of the

presence or not presence of a landslide in a single pixel (Garcia-Rodriguez et al., 2008; Aristizábal et al., 2015; Feng et al., 2016).

In other cases, there is a problem of spatial non-homogeneity, which pertains to a non-equally partitioned space. A terrain can be subdivided into a grid of equally measured squares, but in the risk assessment of natural hazards it is common to work with areas of different sizes. This is the case if political units or seismic regions used instead of continuous space. Landslides might be spatially recorded as occurrences in a defined region, rather than with provided geographical coordinates. In such cases, the daily rate of landslides per town might be equal between two towns, but the effect of landslides will be higher in the smaller of the two. This is a case considered in Chapter 4, where the landslide data used only have the location name information. A simple example of discretization in point processes is that used by the linked stress release model (LSRM).

LSRM An example of a point process with discrete space is the linked stress release model, based on the previously mentioned SRM, proposed by Lu et al. (1999) and further investigated by Bebbington and Harte (2003). The model updates (2.11) to capture the propagation of stress in space (from one seismic region to another), as it is not realistic to consider that one region is independent to another one. In particular, Bebbington and Harte (2003) used a transfer parameter θ_{ij} to measures the proportion of stress transferred from region i to region j.

$$X_i(t) = X_i(0) + \rho_i t - \sum_j \theta_{ij} S^{(j)}(t) \tag{2.19}$$

where $S^{(j)}(t)$ is the accumulated stress release in region j up to time t; the coefficient θ_{ij} can be positive or negative, which results in discharge or excitation. Following (2.12), the conditional intensity function for the LSRM can use an exponential function for each region $\lambda_i(t|H_t) = \exp[\mu_i + \nu_i X_i(t)]$. Hence, for $i = 1,2,...$ we obtain:

$$\lambda_i(t|H_t) = \exp\left\{ \mu_i + \nu_i \left[X_i(0) + \rho_i t - \sum_j \theta_{ij} S^{(j)}(t) \right] \right\} \tag{2.20}$$

where for each region i μ_i and ν_i are parameters to be estimated similarly to (2.12), ρ_i the constant loading rate and θ_{ij} reflects the stress drop transferred from region i to region j.

Overall, the model evaluates the stress release in time (treated as continuous) between regions, and so with space treated in a discrete fashion. Therefore, this is an example of how the spatial discretization can be used to model natural hazards with point processes.

In Chapter 4 it will be shown how this type of space discretization can be applied to a nonhomogeneous Poisson process for landslide occurrences. For this study, the landslides were provided with discretized space information (the town, or municipality). The centroid of each municipality was used as the nominal location of corresponding landslides. In such cases, rather than having a temporal ground intensity function and a mark density functions, the time and the severity of triggering events is expressed through "components" (C_1,...), which define the effect of each triggering hazards:

$$\lambda(x,t) = \lambda(x) \cdot g(C_1(x,t),C_2(x,t),...,C_n(x,t)). \tag{2.21}$$

where $\lambda_0(x)$ is the baseline parameter for each location x and $g(\cdot)$ is a non-negative valued function. With this discretization, each municipality is treated as a single spatial element, as we only observed the number $N(x,t)$ of landslides at location x and time t.

Time discretization

Time in point processes is generally considered a type of continuous data. Nevertheless, the lack of data availability and different precision levels between datasets sometimes require a temporal discretization. In Chapter 4, a time discretization of a point process is proposed because of the naturally time-discretized landslide data, and the level of temporal detail provided in the rainfall dataset (daily precipitation).

A continuous time interval can be discretized into a partition of discrete intervals. For instance, a continuous time interval $(0,T)$ can be partitioned in days $\{1,2,...,T\}$. If $N(t)$ is a Poisson process with rate λ, we can redefine the rate as:

$$\mu_d = \int_{d-1}^{d} \lambda(t)dt \tag{2.22}$$

where $d \in \{1,2,...,T\}$. Hence, $\int_{d-1}^{d}\lambda(t)dt$ is an average of the ground intensity over $(d-1,d]$. Therefore, in the partition of the time interval we have one Poisson random variable $N(d,d+1) = N(d+1) - N(d)$ for each discrete time interval.

Taking again the case of landslides in Chapter 4 as an example, the process of landslides in the region of Emilia-Romagna is spatially discretized in municipalities (Italian political territories), and temporally discretized in days. A large input from earthquakes or rainfall can trigger more than one landslides within a municipality. Hence, if the Poisson process for landslide occurrences over time is $N(d,d+1)$, the conditional intensity of a Poisson process on a given day is μ_d, with a rate equal to the mean daily rate over day d.

2.3 Survival Analysis

Survival analysis is a branch of statistics that studies the lifespans of the individuals in a population. The focus is on time-to-event data, which is the time elapsed from a time origin to an endpoint. The endpoint may be the occurrence of a specific event. For instance, survival analysis can be applied to cardiology to evaluate the death or survival for patients with a cardiac disease (Ahmad et al., 2017). For an insurance company, survival analysis can be used to estimate the time before an accident (Brockett et al., 2008)

Assuming that T is a random variable of the time of an event, we can define the survival function $S(t)$. This function describes the distribution of the survival times, hence the probability that the failure will not happen before a specific time.

$$S(t) = P(T \geq t) \tag{2.23}$$

and it can also be defined as the integral of the density function of the survival times $f(t)$.

$$S(t) = \int_t^\infty f(u)du \tag{2.24}$$

The hazard function expresses the conditional failure rate at time t for the portion of population still not failed in t or the instantaneous failure rate at t. It is the ratio between the density function and the survival function:

$$h(t) = \lim_{\Delta t \to 0} \frac{P[t \leq T < t + \Delta t | T \geq t]}{\Delta t} = \frac{f(t)}{S(t)} \tag{2.25}$$

The hazard function can be linked back to the conditional intensity function in (2.1), as they both evaluate a conditional rate of occurrence, given the history, within a small interval.

$$\lambda(t|H_t) = \lim_{\Delta t \to 0} \frac{P[N(t + \Delta t) - N(t) = 1 | H_t]}{\Delta t} = \frac{f(t)}{S(t)} \tag{2.26}$$

The main difference is that while point process can observe more than one event, in survival analysis there is only one possible event, the failure. However, in a renewal process, each failure time is identically distributed, and so survival analysis techniques can be used to analyze series of events.

Together with Poisson processes, survival analysis is commonly used for time-related analysis (In and Lee, 2018; Hancock et al., 2014). Survival analysis is a statistical methodology commonly used to evaluate time-to-event data in many scientific areas. Although it is more common in clinical studies (Rossouw et al., 2002; Stupp et al., 2005), it has been used for hazard analysis as well. Some examples exist in geological

applications, where this method is used for time to failure in landsliding (Federico et al., 2012; Segalini et al., 2018). The variable of interest is then the time to occurrence of a landslide, a landslide dam, etc., which can provide helpful information, particularly in case of hazards whose magnitude is related to the time elapsed. For example, landslide dams can lead to floods, and the size of the flood is dependent on how much water have been accumulated behind the landslide dam. The amount of water depends on how much time has passed since the damming. The times to failure can be estimated with respect to covariates affecting the hazard. A landslide dam failure time will then be evaluated in relation to the covariates, e.g. the characteristics, that affect its stability: landslide dam dimensions, material, water accumulation. Hence, the survival analysis can also help understanding which covariates are critical in influencing the hazard failure.

An important concept of survival models is related to censoring. Survival analysis hypothetically studies subjects over time until their death/failure. However, studies are usually limited in time, so the subjects are studied from a starting time to a set end time. Those observations whose failure has not been observed by the end of the study are an example of (right) censored data. For instance, the time to failure for landslide dam is usually censored: short-term landslide dams, collapsing within hours or days, may be recorded only after their failure (Tacconi Stefanelli et al., 2015; Chen et al., 2014). In other words, the exact life time is not known, as the observation is recorded after observing the failure. Similarly, if engineering techniques are put in place to stabilize the landslide dam, we will not observe the failure, due to this external event.

Among the types of censoring, the most common one is right-censoring, which is the case of a non-failed subject (a landslide hasn't failed by the end of the study), either because the study reached a predetermined end of time (Type I censoring) or because a random event precludes subsequent observation of the event of interest (random censoring). In this case, the time of failure is higher than the censoring time $T > C$. In Chapter 5, we will see a particular case of Type I censoring, called observational censoring: some failures have not been observed yet, but there is no predetermined end of the study. Left-censoring is related to a subject whose failure has occurred before the beginning of the study. Hence, we only know that the time of failure is less than the time of the beginning of the study $T < t_0$. Interval censoring is when the event has occurred, but the only information available is that the event time is within a certain interval $(a,b]$.

The estimation of the parameters in (2.23) and (2.25) is usually done via maximizing the log-likelihood. In the case of all uncensored observations, the log-likelihood is the

sum of all log densities evaluated at the observed failure times:

$$l(\theta) = \log L(\theta) = \sum_{i=1}^{n} \log f(t_i,\theta) \tag{2.27}$$

If some observations are right-censored, then these will contribute to the log-likelihood with their survival function $S(c_i,\theta)$ (where c_i is the censoring time), as they have not failed yet. Hence the log-likelihood becomes:

$$l(\theta) = \log L(\theta) = \sum_{i=1}^{n} \{\delta_i \log[f(t_i,\theta)] + (1-\delta_i)\log[f(S(c_i,\theta)]\} \tag{2.28}$$

where δ_i is an index to distinguish censored and uncensored data: when $\delta_i = 1$ the observation is uncensored. The survival function $S(c_i,\theta)$ represents the contribution to likelihood from censored data c_i. Similarly, the log-likelihood for interval censored data will be:

$$l(\theta) = \sum_{i=1}^{n} \{\delta_i \log[f(t_i;\theta)] + (1-\delta_i)\log[S(a_i;\theta) - S(b_i;\theta)]\} \tag{2.29}$$

2.4 Bayesian imputation of missing data

Missing data are quite common in statistical analysis. Usually, computer programs such as R can deal with missing data by just ignoring them in the analyses (Rubin, 1976; Gelman and Hill, 2007). However, in some cases this approach is not viable. For example, if the data availability is limited, then discarding one observation because of one missing value becomes counterproductive, given then small population or sample.

As each data is vital for the analysis on small datasets, then it would be better to impute the missing value, which means to estimate what the missing value would have been if it were to be observed. There are several imputation methods, that range from simple ones, such as mean imputation, to probabilistic ones, such as random imputation, to model-based methods. In the case of mean imputation, imputing the value of missing observations can be done using the mean of all other existing values. For example, if there are missing data under the variable "height" of the landslide dam, then those will be imputed by calculating the mean of the observed heights of the other events.

Random imputation implies the random selection of a value from the range of observed data of the given variable. This selection is obtained sampling a random number from the observed ones (e.g. simple functions can be written in any statistical software) or predicting (e.g.) from a linear regression function, where a parametric model is used to impute the missing values. For example, it is possible to create a regression of the variable with missing values, as a function of all the other covariates.

Then, the prediction of the missing values can be obtained by using the known values of the other variables.

A further step in the methodology is represented by multiple imputation, where replacement values are drawn repeatedly from the estimated conditional distribution of the variable. A good approach is to model it as conditional on the other variables, similarly to the regression approach mentioned above. In Chapter 5, an example of Bayesian imputation is provided, where the distribution of each variable is modelled conditional on the other variables. This allows for the imputation of missing values while maintaining the existing correlation structure among the variables.

Chapter 3

Literature review

There has been a long debate on the definition of hazard, but it can be broadly defined as a phenomenon or a process which can affect human life, settlements and halt social and economic activities (UNDDR, 2016). This interpretation is not akin to the statistical definition of hazard, which refers to the likelihood of occurrence of a perilous event (Papathoma-Kohle et al. (2007) defines a hazard as "the probability of occurrence of a potentially damaging event within a specified time and given area"). For clarity, this thesis will use a non-statistical definition of hazard, to build a framework of models to evaluate the risk of hazards in a multi-hazard environment.

Marzocchi et al. (2012) define a hazard as "anything that can potentially generate adverse events and consequently create damage to the population or environment". Thus, a natural hazard is a natural event, linked to geographical processes (Bokwa, 2013). Many regions of the world are affected by multiple natural hazards (Liu et al., 2016; Kappes et al., 2012b,c). In each region these events depend on natural factors (Liu et al., 2016; Kappes et al., 2010) and occur within environmental systems. If a system can be affected by more than one natural hazard, there might be interaction among hazards, which can alter their occurrences and/or consequences. For example, the strong M_W7.8 Kaikoura earthquake in New Zealand on 14 November 2016 have contributed to the triggering of thousands of landslides (Massey et al., 2018). The subsequent rainfall and storms contributed as well. The earthquake might have weakened or predisposed - influenced (Gill and Malamud, 2014) - slopes to fail or be reactivated if a successive triggering event (rain) occurred (Dellow et al., 2017; Mason et al., 2018). Hence, the question arising is how much the interaction among hazards has the potential to affect the occurrence of other hazards in terms of when, how many and how large.

Hazards should not be evaluated in separated hazard assessments. Instead, it is crucial to consider the interactions among them (Lee and Rosowsky, 2006; Zuccaro et al., 2008; Kappes et al., 2010) within a single multi-hazard assessment. Furthermore,

hazards are history-dependent, which is particularly recognizable with earthquakes: large events are followed by aftershocks (Toda et al., 1998; Chiarabba et al., 2009), but at the same time catastrophic earthquakes may not occur for centuries (Stirling et al., 2012). Nevertheless, hazards depend on the history of triggering hazards as well. Landslides may or may not re-occur in the same area depending on their history, but also in relation to the history of, for instance, heavy antecedent rainfall. Hence we will define a hazard as is an event which occurs in time and space, in relation to the spatio-temporal and severity history of that hazard and those which can trigger it (Gill and Malamud, 2014).

Several authors (Marzocchi et al., 2009; Chen et al., 2010; Kappes et al., 2010, 2012a; Esharti et al., 2015; Gill and Malamud, 2014) have already pointed out the necessity for a multi-hazard framework so that risk mitigation and resilience procedures can be more effective (Mahendra et al., 2011). This awareness is also highlighted by the establishment of multi-hazard or resilience programs over the years, for example HAZUS (Schneider and Schauer, 2006), RiskScape (Reese et al., 2007), Resilience Challenge, MATRIX (Komendantova et al., 2014). Nevertheless, multi-hazard methodologies are not yet well established. The reason can be in the wide range of hazards, and consequently in the wide range of hazard assessments: the hazard description and characteristics, i.e. size, distribution in time and space, substantially differ from one hazard to another. For example, the clustering of earthquakes in time and their propagation in space is substantially different from the spatial and temporal distribution of rainfall and landslides. Consequently, models based on specific hazard features often cannot be used for other hazards. Hence, while a joint hazard analysis, treating the assessments as independent, would be possibly more convenient in terms of time and data consumption, it may not accurately depict the reality, unless a well-structured multi-hazard method is conceived (Kappes et al., 2012a). An example is provided by fragility curves, expressing the probability of damages by hazards per damage classes, which assume the absence of any disturbance from any hazard in the structure of any other hazard (Selva et al., 2013).

3.1 Scope of study

To build a multi-hazard framework, it is good practice to look at the state of the art of hazard assessments, whether the assessment is for a single hazard, for example the susceptibility analysis of an area to a hazard (Evans and Hungr 1993; Ayalew and Yamagishi 2005, see Section 3.2.3), a single hazard triggered by another hazard (Scott et al., 2005; Marc et al., 2016), or a multi-hazard approach (Thierry et al., 2008; Mahendra et al., 2011). Nevertheless, there is a considerable number of hazards,

and their differences ultimately reverberates on assessment methodologies. Hence, this thesis will focus on one specific chain of hazards, small enough to make the analysis feasible, and big enough to provide a good range of triggering interactions. This should provide an insight into the features of events and their interactions, providing a basis on which to build a generalised framework.

The selected chain is the one of landslides triggered by earthquakes and rainfall (Figure 3.1). The choice of landslides is due to their geographical extension, i.e. landslides occur in many countries. At the same time the modelling of earthquake and rainfall is advanced, compared to other hazards. Several authors have attempted the analysis of this geomorphic hazard chain after major earthquakes (Han et al., 2007; Zhou et al., 2015b; Fan et al., 2019). Relevant projects are underway to build resilience against coseismic hazards, such as Project AF8 in New Zealand (Orchiston et al., 2018) and rainfall-induced landslides, Landslide EVO Project in Nepal (Cieslik et al., 2019). A particular modelling extension of this chain is also considered: if a landslide falls

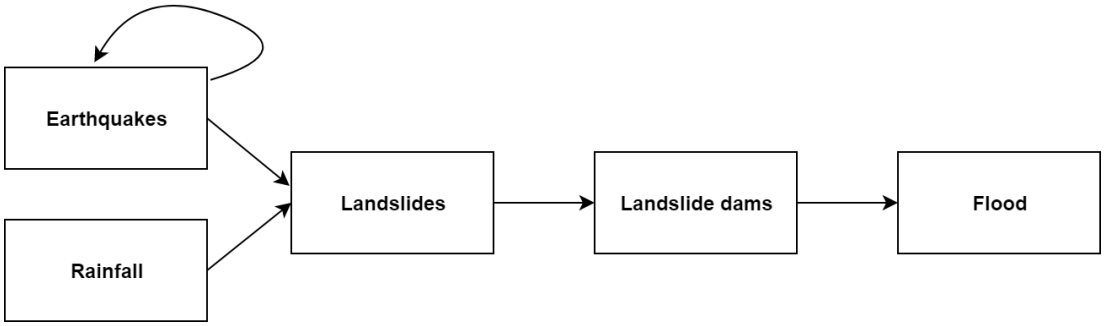


Figure 3.1: Schematic diagram of the selected chain of hazards: earthquakes and rainfall can trigger landslides; landslides may fall on a river causing a landslide dam. The failure of a landslide dam can cause floods.

on a river, there are chances for the formation of a natural dam and its consequent possible failure leading to a flood. Thus, the extended chain we will consider is the earthquake/rainfall→landslide→landslide dams, on which a preliminary multi-hazard framework will be built. Overall, the chain includes four different phenomena that will be used to guide a review of hazard assessments already existing in the literature. This review will provide a sound basis of knowledge, a starting point for the multi-hazard framework. In particular, as mentioned in Chapter 1, the main objective is to build quantitative models, so the purpose of this chapter is to present and review the state of the art on natural multi-hazard assessments, whether they are already quantitative or not, and how they can be improved or used for the framework.

3.2 Single hazard assessments

Understanding how single hazards are assessed is the first step for a multi-hazard methodology. The evaluation of magnitude and space-time distribution of a hazard will be used to formulate models so that the full extent of the hazard's characteristics is expressed.

3.2.1 Earthquake

Earthquakes are "sudden slip on a fault, and the resulting ground-shaking and radiated seismic energy caused by the slip"(USGS, 2019a). Earthquakes are recorded in catalogues together with their characteristics: time, hypocenter, size, and other technical information (Kagan, 1991). There are many well-detailed datasets, from local ones such as the Italian catalogue (Camassi and Stucchi, 1996), to the worldwide one (ISC, 2019; USGS, 2019a). With modern technology, seismic events are recorded to high temporal and spatial precision. However, catalogues are not time-homogeneous in their precision and completeness. This creates issues in dealing with long earthquake datasets, i.e. earthquake hazard, but is less of a problem when dealing with earthquakes triggering a secondary hazard, if the triggered hazard records, e.g. landslides, are not as long.

For the earthquake size (magnitude) there are several methodologies in use. The earthquake magnitude is a measure of the size (amplitude) of the seismic waves registered by seismographs. The most common measures used are the moment magnitude, the peak ground velocity (Joyner and Boore, 1981), peak ground acceleration (Boore et al., 1997) and Arias intensity (Arias, 1970). For instance, the moment magnitude is calculated from the seismic moment, i.e. the energy released, M_0 as $M_w = 2/3 \log M_0 - 10.7$ (Hanks and Kanamori, 1979). This measure is more reliable as an estimate of earthquake size compared to other previous measures, which were less efficient particularly in relation to the distance between the earthquake location and the seismograph registering the shake (Wells and Coppersmith, 1994). The Arias intensity determines the ground-shaking as the integral of the square of ground acceleration a of seismic waves over time: $I_A = \frac{\pi}{2g} \int_0^{T_d} a(t)^2 dt$.

As described in Chapter 2, point processes are widely used for the modelling of earthquake occurrences. Three features characterize the earthquake occurrences: time, space and magnitude. Several models for the temporal or spatial distribution of earthquakes have been proposed, where the occurrences are dependent on the temporal or spatial history of previous events, together with their magnitudes. This thesis focuses mainly on the continuous temporal occurrence of events: the spatial distribution is considered but discretized, as explained in Chapter 2 and Chapter 4. The analysis of seismic patters has developed substantially in the last thirty years (Zhuang et al., 2012),

from the use of Poisson processes for the number of earthquakes in a time interval, to the application of the theory of elastic rebound (Reid, 1910; Knopoff, 1971; Vere-Jones, 1978). These models are mostly for the modelling of mainshocks, but one major characteristics of earthquakes is that they tend to cluster. In fact, a mainshock can trigger or increase the probabilities of a sequence of aftershocks, which is a series of smaller events (in terms of magnitude), clustered right after the mainshock. For the aftershock modelling, it is common to use Omori's decaying frequency law of aftershocks (Omori, 1894; Ogata, 1988; Wetzler et al., 2016) $n(t) = \frac{k}{c+t}$ (the frequency of aftershock decreases hyperbolically over time), where t is the time from the mainshock, $k > 0$ and $c > 0$ are parameters related to productivity and time scale respectively. Alternatively, self-exciting models, such as the temporal Epidemic Type Aftershock Sequence (ETAS) model (Ogata, 1999; Helmstetter and Sornette, 2002), are used as described in Chapter 2.

3.2.2 Rainfall

Conventional models for rainfall data are time series (Mekis and Hogg, 1999; Brunetti et al., 2006; Park et al., 2019) or point processes (Rodriguez-Iturbe and Eagleson, 1987; Onof et al., 2000; Cowpertwait et al., 2007). As long-term rainfall time series are often inhomogeneous, due to the tools used (e.g. faulty or relocated rain gauge - when the political boundaries are changed), time series are usually homogenised (Auer et al., 2007) by interpolation, filtering and statistical testing, to improve data quality and to allow comparison and analysis over time. Although they might be presented as homogenised over time, long-term catalogues on a much finer location base (not at a regional level, but town by town) tend to be much less homogenised. As will be seen in Chapter 4: the rainfall dataset in use has thousands of missing values, some negative values, and a non constant indication of snow-type precipitation data over the time window. This required a time-consuming process of cleaning before being able to use the data.

Point processes are used to model and simulate precipitation (Rodriguez-Iturbe et al., 1984, 1987; Onof et al., 2000; Isham et al., 2005; Cowpertwait et al., 2007). The first model proposed (Rodriguez-Iturbe et al., 1984, 1987) is based on the concept that a storm (a rainfall event) is produced by pulses of overlapping rain cells. Each rain cell (the atomic component of the point process, with random duration and depth) does not have a physical interpretation but provides random pulses. Their magnitude and length define the intensity and duration of the storm. The intensity is set to be constant within the rain cell, although further models (Cowpertwait et al., 2007) overcome this limitation assuming that each rain cell is characterised by a Poisson process of rainfall pulses, which provides some fluctuation within the cell.

Precipitation data are gathered in databases (e.g. Arpa-Piedmont (2018)), together with information on the location of the rain gauges (the instrument that records the amount of precipitation), the amount (in millimetres), and the time. Typical measurements of the precipitation amount are cumulative rainfall values (daily, monthly, seasonal, etc.), as well as averages across wet days (rainfall events) (Haylock et al., 2006).

3.2.3 Landslides

A landslide is "the movement of a mass of rock, debris, or earth down a slope" (USGS, 2019b). Landslides are a common hazard in many terrains. Usually they are triggered by rainfall (Berti et al., 2012; Aristizábal et al., 2015; Peruccacci et al., 2017) or seismic activity (Lee, 2014a; Havenith et al., 2016; Robinson et al., 2016a). They are recorded in catalogues (e.g. Kirschbaum et al. 2010; Piacentini et al. 2018; ISPRA 2019), with information about the location, the time and the size. While the size of a landslide is often available and well-documented, the time and location are often missing, particularly for historical events (Steger et al., 2017). If a landslide occurs in a remote area, the time of the event can only be constrained between successive observations, which is the concept of interval-censored data that will be used in Chapter 5. Nowadays, the location can be specified with satellite or aerial methods, but for historical inventories the location can be as generic as the topographic name(Guzzetti et al., 2012). Accuracy and completeness are challenging to achieve, but there are examples of catalogues (e.g. ISPRA 2019) which combine different databases into a single one, providing a good step towards completeness, although it still lacks accuracy, due to the extent of the task and the numerous sources used. Landslide assessments are usually done via susceptibility mapping (Guzzetti et al., 1999; Ayalew and Yamagishi, 2005; Aleotti and Chowdhury, 1999; Dai et al., 2002), where GIS-based data and ground characteristics data are combined, usually through logistic regression, which attempts to find the best combination of factors to describe the susceptibility of areas to landslides. The derived susceptibility values are reported in a map, using a color-coding scheme in relation to the level of susceptibility.

3.2.4 Landslide dams

Landslide dams can be considered as a natural extension of the earthquake/rainfall-landslide chain. Landslide dams involve a mass of debris (landslide) falling on a river and damming it (Costa and Schuster, 1988; Ermini and Casagli, 2003; Korup, 2004; Zhou et al., 2015a; Tacconi Stefanelli et al., 2016; Liao et al., 2018). The subsequent accumulation of water behind such barriers creates a danger for the population downstream, increasing with the size of the lake, which may be released if the dam fails.

Many disastrous events have been observed over the years: the dams associated with the Wenchuan earthquake (Yin et al., 2009), rock avalanches in the Himalayas (Hewitt, 1998), local storms or typhoons (Tsou et al., 2011).

As for landslides, the number of existing catalogues for landslide dams is limited, but the few available ones (Costa and Schuster, 1988; Tacconi Stefanelli et al., 2015) are precious material for any landslide dam assessment. Generally, these catalogues have poor temporal information, with more detailed information in terms of location and dam characteristics, due to the scarcity of long-lasting events, and the rapidity at which the majority of them fail.

The stability of landslide dams is usually evaluated via indices, used to partition the landslide dams into groups by setting threshold values. An example is the Blockage Index (Casagli and Ermini, 1999), BI = $\log(V/C)$ the logarithm of the ratio between the volume of the landslide dam (V) and the catchment area (C). Ermini and Casagli (2003) further developed this approach by proposing the Dimensionless Blockage Index DBI = $\log(CH/V)$, where H is the dam height. Although these indices provide a picture, despite being static over time, of the dam conditions, they are subject to miss-classification, due to the limited number of results (e.g. failure/non-failure, or stable/unstable/uncertain). Furthermore, these indices are built on location-specific datasets, which are naturally biased to reflect the characteristics of events occurring within that region. The application of these indices may lead to miss-classification of events in other parts of the world. On the other hand, if the indices are built on a world-wide dataset, they will again represent medium to large events (Casagli and Ermini, 1999), which are only a part of landslide dams. Finally, these indices represent a snapshot of dams that have failed at the time of analysis, not considering the time elapsed since the dam was formed. This is a limitation, as these indices only provide a static picture of the current state of a landslide dam, rather than information on the evolution over time.

3.3 Hazard triggering

In a system of hazards, such as the chain in Figure 3.1, hazards are not independent entities, as they interact and occur with potential overlapping in space and time. Although it would be possible to ignore the interaction among hazards and analyse them as independent (Fleming et al., 2016), it is not a good idea, because there is a high risk of underestimating the overall effect of the hazards (Kappes et al., 2010). The first step towards a multi-hazard assessment is the analysis of hazard interactions, particularly the most direct mechanism: one hazard inducing one or more other subsequent events. The terminology and definition differ slightly from author to author. Delmonaco et al. (2006) refers to this phenomenon as domino effect or cascading failure which is a:

> "failure in a system of interconnected parts, where the service provided depends on the operation of a preceding part, and the failure of a preceding part can trigger the failure of successive parts".

Marzocchi et al. (2009) defines them as "coupled events" where "an adverse event triggers one or more sequential events (synergistic event)". Hence, a triggering model assessment will depend on the characteristics (frequency, magnitude, location) of the triggering hazard, as this information can provide evidence of the elements affecting the occurrence of the triggered hazard. A landslide assessment should be based on factors related to rainfall or earthquakes, a tsunami assessment on volcanic eruptions or earthquakes, periglacial debris flows are linked to ice melting, weather and temperature.

The analysis of the history of past events becomes then crucial to find patterns and relationships among factors to allow for the prediction of future occurrences (Yalcin, 2008). The temporal distribution of seismic events, i.e. the past release of tectonic stress (Reid, 1910; Ogata, 1988; Bebbington and Harte, 2001), is an example of history dependence.

In general, single hazard assessment methods can be divided between temporal or spatial, depending on whether the purpose is to evaluate the temporal or spatial likelihood of a hazard occurrence. These methods can be qualitative or quantitative, where the former are mostly based on expert knowledge, often expressing the risk of a hazard with categorical or at most ordinal scales (e.g. more likely, less likely), rather than being based on the analysis of data. The latter are based on on statistical and mathematical models, often informed by physical notions (e.g. the system of forces in the occurrence of a landslide), in order to estimate or predict hazards. Part of the analysis of a triggering mechanisms is focused on the intensity of the triggering events. Against the definition of hazard selected at the beginning of this chapter, some authors do not include specific processes in the definition of hazard (Kappes et al., 2010). An example is provided by heavy rainfall which is considered by Kappes et al. (2012b) as a natural process, that becomes a hazard within certain conditions. Nevertheless, the author herself states that heavy rainfall can cause floods and debris flows. The keyword is "heavy", which suggests a certain level of intensity. Retaining the definition of hazard proposed at the beginning of this chapter, it is possible to state that a primary hazard has the potential to trigger or influence the occurrence of a secondary one if certain conditions are met. Such conditions are related to the temporal and magnitude distribution of the triggering event.

Within the scope established in Section 3.1, it is possible to evaluate how the triggering mechanisms are apportioned between rainfall and landslides, and between earthquakes and landslides. Therefore, the following section presents a review of such triggering assessments.

3.3.1 Rainfall induced landslides

The literature provides many examples of rainfall-triggered landslides: whether it is seasonal precipitation or sudden storm, the accumulation of water in the ground can modify and weaken the structure of the soil until a portion of a slope detaches (Iverson, 2000). Therefore, several studies have focused on the rainfall-landslide mechanism, mostly following two paths. In the first case, studies have focused on the production of hazard maps to pinpoint those soil characteristics that would make an area more or less prone to landslides (Dahal et al., 2008; Yalcin, 2008), using GIS-data combined with geological and environmental information, and sometimes complex models such as neural networks (Pradhan and Lee, 2010). In the second case, authors have worked on the definition of a rainfall threshold in terms of intensity and duration (Berti et al., 2012; Aristizábal et al., 2015; Peruccacci et al., 2017) in order to provide useful information for early-warning systems.

Landslide hazard maps show the proneness of an area to landslides, evaluated as a function of geological or environmental factors. For example, the water saturation of the ground may lead to a landslide, as the ground is weakened. The amount of water in the ground is related to another event (e.g. precipitation, flood, snow-melting, etc.) which can be considered the trigger event for the landslide. Earlier literature proposed spatial assessment methods that are mostly qualitative (Humbert, 1977; Godefroy and Humbert, 1983; Zimmermann et al., 1986), as they are conditional to the opinions of experts and expressed in descriptive terms. Usually, such methods are based on landslide inventories: experts are used to evaluate the landslide risk of an area by identifying similarities with known events listed in the inventories. The elements used for the comparison range from slope characteristics (e.g. steepness, position) to other geomorphological features of the area (e.g. soil composition). For instance, Zimmermann et al. (1986) created a landslide hazard map of an area in Nepal, by a visual comparison of maps and photos of the field, and producing indices by filling checklists. These factors can be in quantitative analyses, provided that their selection occurs via statistical or mathematical tools (e.g. logistic regression). In some cases these approaches become semi-quantitative by using hierarchical methods (Barredo et al., 2000), a decision-making tool where factors are weighted in relation to what the experts believe are the essential elements discerning the different levels of risk. Nevertheless, the weighting of ordinal indices is once again questionable, as an opinion-based selection method is being preferred over a method based on data analysis. This also hampers the possibility of extending the applicability of such method to other areas.

In more recent literature, authors have taken advantage of GIS-data (Guzzetti et al., 1999; Pradhan and Lee, 2010; Yalcin, 2008; Ohlmacher and Davis, 2003) for a higher

level of data precision on vast areas. In some cases, weighted linear combination (Ayalew and Yamagishi, 2005) was used for variable selection, assigning weights (scores) to the factors using expert opinions. Another example is the weights-of-evidence model used by Dahal et al. (2008) in combination with GIS data, to obtain landslide susceptibility maps. The method is based on the evaluation of the predictive power of factors (e.g. soil type, aspect, relief, flow accumulation, etc.) on the occurrences of landslides. The proposed formula is a log odds ratio of occurrence of a landslide

$$W_i = \ln \frac{B|L}{B|\overline{L}} \tag{3.1}$$

where the weight of each factor (B) is calculated in relation to the presence or absence of a landslide (L) in a specific location. These weights are defined as positive or negative, expressing the presence or the absence of a factor for a specific event. For each event i the positive W_i^+ and negative weights W_i^- express the positive correlation between the presence/absence of the factor and the presence/absence of the landslide. In the end, the magnitude of the differences between the positive and negative weights provides a spatial association between the factor and the landslide. Overall, it is an estimation of the conditional probability of a landslide occurrence given a specific factor. The method is possibly over-complicated, and highlights that a different method, perhaps quantitative, may simplify and produce a more rigorous approach to hazard assessment.

All the above mentioned methods are decisional tools that are based on expert knowledge, even if combined with high-resolution data. However, the main issue with landslide hazard maps is that the proneness to the hazard is expressed for each map unit (e.g. a pixel). Each, say, pixel is then more or less susceptible to a hazard in relation to the level of some factors. Several papers, like the ones above cited, do not mention whether the map units are considered dependent. Indeed, they cannot be independent, as one small portion of land is very likely to be similar, and so at least related, to an adjacent one. Furthermore, the selection and description of the factors in some papers (Dahal et al., 2008) differ from others, making the comparison with other methods difficult. The selection methods of factors have to be rigorous and supported by quantitative analysis.

Other proposed methods used in landslide assessments range from principal components (Londoño-Linares, 2017; Santos et al., 2019) and logistic regression (Guzzetti et al., 1999; Dai et al., 2002; Yesilnacar and Topal, 2005; Ayalew and Yamagishi, 2005) to neural networks, often combined with spatial information, usually GIS-based. Although they tend towards a quantitative assessment, they still produce a classification of landslides based on the combination of the proposed factors. For example, Santos et al. (2019) have used what they call a “boosting technique”, discriminant analysis

applied on the principal components in a landslides dataset, claiming a better selection of classifiers, and an improved separation between stable and unstable slopes.

Logistic regression models the probability of presence or absence of an event (a landslides), against a set of regressors, which describe specific characteristics of the area (e.g. lithology, groundwater, slope angle). For instance, the odds ratio of the landslide probability occurrence can be modelled as

$$\frac{p}{1-p} = \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_k X_k) \tag{3.2}$$

Among the most common regressors affecting the stability of slopes there are the aspect (direction of the slope), the elevation, the lithology and the slope angle (Dai et al., 2002; Ayalew and Yamagishi, 2005; Yesilnacar and Topal, 2005). Some of these, such as lithology, are usually described qualitatively (Guzzetti et al., 1999), hence the inclusion of this type of data in a quantitative model might be challenging requiring the use of multiple intercepts.

The advantage provided by the logistic regression is that it assigns a probability, rather than a two point classification. This is an improvement compared to the mere distinction between stability and instability of a slope to establish whether a specific area or pixel is prone to landslides or not. A twofold classification is likely to be restrictive: highlighting one area as unstable does not imply when the landslide may occur, and under which circumstances.

However, the logistic regression is still lacking of important aspects such as time. Geomorphological characteristics are somewhat seen as static in time but the triggering of hazards has a temporal aspect that needs to be included. For example, the size of landslide dams represent a static snapshot of the hazard. Nevertheless, in Chapter 5 I will present a model which uses these to evaluate the time-to-failure of landslide dams. Therefore, it is vital to consider the temporal distribution of landslides in relation to the rainfall levels over time.

Several authors have attempted to evaluate rainfall and water runoff thresholds (Glade et al., 2000; Guzzetti et al., 2007; Rossi et al., 2017), a tool to evaluate the levels of rainfall above which landslide occurrence is likely, particularly useful with real-time rainfall analysis. Hence, the two crucial aspects for the triggering mechanism become the intensity and the duration of rainfall events. Evaluating their combination can provide information on the risk of landslides as a function of rainfall and time. Brunetti et al. (2010) focused on rainfall threshold curves in Italy, considering that the minimum level of rainfall to trigger a landslide can variate in relation to the intensity/-duration ratio of rainfall. These threshold curves follow a power law $I = \alpha D^{-\beta}$, where I is the intensity of the rainfall, D the duration, α the scaling constant and β is a shape parameter. Brunetti et al. (2010) tested both frequentist and Bayesian approaches to

fitting the best threshold curve to separate events based on the combined intensity/duration. Figure 3.2 shows rainfall events that resulted in landslides, with the thresholds for the frequentist (T_1) and Bayesian (T_B) performing similarly. Rossi et al. (2017)

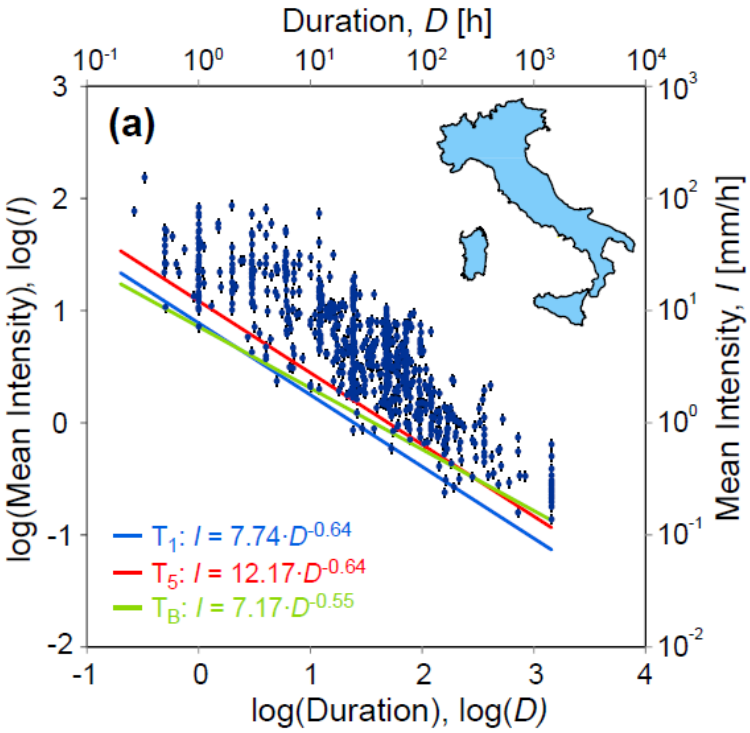


Figure 3.2: Rainfall thresholds obtained using the Bayesian (T_B, green line) and the Frequentist (light blue line is the 1% threshold, T_1; red line is the 5% threshold T_5) methods. Error bars on the rainfall mean intensity I show systematic error, assumed fixed and equal to 10%. Errors on the rainfall duration D were considered negligible and are not shown. Picture and original caption from Brunetti et al. (2010)

found similar results by applying a power-law to the intensity/duration threshold the the Italian region of Umbria. In an antecedent paper, Rossi et al. (2010) also studied the rainfall-triggered landslides phenomenon in Emilia-Romagna, our region of interest in Chapter 4. Rossi et al. (2010) took a slightly different approach to Rossi et al. (2017) focusing specifically on two statistics: the number of landslides in a day D_L and the number of landslides in an event S_{event} (which is a series of consecutive days with rainfall). Both these measures were once again linked with a function of the power-law family (a Zipf). Interestingly, it was found that the minimum amount of rainfall required to trigger landslides differs between short-term rainfall and long-antecedent rainfall. In the first case, the minimum amount varies in relation to the landslide intensity: considerably much more rainfall required for a high number of landslides within the same one-day rainfall. In the second case, the long-antecedent rainfall does needs little to actually trigger landslides, possibly due to ground saturation (Peruccacci et al.,

2012). However, the results by Rossi et al. (2010) did not consider any spatial diversity of the region (Emilia-Romagna, Italy), which is equivalent to considering the whole region as a single point in space. Nevertheless, they demonstrate that a rainfall-landslide assessment needs to be non-static over time in the evaluation of the triggering mechanism. In Chapter 4 we will explore how to link the likelihood of landslide occurrences to the changing rainfall levels over time, taking into consideration that short-term and long-term precipitations can produce different effects.

3.3.2 Earthquake induced landslides

Coseismic landslides are a major hazard, that occurs in correspondence with an earthquake. Many large earthquakes have been followed by extensive landsliding. This is the case of Chi-Chi in 1999 (Shou et al., 2011), Kashmir in 2005 (Owen et al., 2008), Wenchuan in 2008 (Zhou et al., 2015b), Kaikoura in 2016 (Dellow et al., 2017) and Molise 2018 (Martino et al., 2020).

Past studies have highlighted that earthquakes generate superficial ground movements which produces instability on slopes (USGS, 2019b). Keefer (1984) and then Rodriguez et al. (1999) and Keefer (2002) determined that the mechanism of coseismic triggering needs a minimum earthquake magnitude ($M_W 4$), that the number of coseismic landslides is often underestimated and it can reaches several thousands, spanning to hundreds of kilometers from the epicentre, in relation to the magnitude of the earthquake. Also, some studies suggest that the earthquakes might have a long-term effect on the potential occurrence of landslides for several years (Keefer, 1994; Wei et al., 2014).

Among the most important aspects affecting coseismic landslide occurrences, the most commonly used are: earthquake intensity (Arias intensity), fault location, altitude, climate and lithology (Keefer and Wilson, 1989; Syvitski and Schafer, 1996; Rodriguez et al., 1999; Keefer, 2002; Garcia-Rodriguez et al., 2008; Havenith et al., 2016). Susceptibility analysis, as for rainfall-landslide triggering, has been used to evaluate the factors affecting the susceptibility of landslide-prone areas are assessed in relation to earthquake triggering (Lee and Evangelista, 2006; Umar et al., 2014). Logistic regression has been used focusing on the geomorphological characteristics (e.g. slope gradient, elevation) and the features of the earthquakes (e.g. peak ground acceleration, distance from fault; Yin et al. 2009; Alfaro et al. 2012; Vessia et al. 2013), sometimes using physical models (Jibson et al., 2000). Li et al. (2013) produced coseismic landslides susceptibility maps for the Wenchuan 2008 and Lushan 2013 earthquakes, using a scale of colors, here with defined thresholds to have 5 color bands, rather than using a continuous scale. The results highlighted the importance of slope, lithology, elevation, distance from the fault.

Overall, the earthquake-landslide triggering mechanism is well known in the literature: with a minimum magnitude, an earthquake can trigger a landslide in the near proximity of a fault. Susceptibility analysis has been helpful to highlight the most important factors in the coseismic landslides triggering mechanism. Furthermore, it has been shown that large earthquakes can have a prolonged effect on the susceptibility of landslides (Yan et al., 2017; Li et al., 2020). Nevertheless, the literature still lacks of a solid temporal analysis between earthquakes and landslides, as all coseismic analysis are focused on the distribution of landslides immediately after a large earthquake has occurred. It is vital to evaluate the temporal distribution of earthquakes and landslides together, in order to capture the evolution of the process, and improve the knowledge of the triggering mechanism. In this view, Chapter 4 will propose a model in which earthquakes and landslides are treated as point processes, to incorporate the evolution of the triggering mechanism over time.

3.4 Multi-hazard assessment attempts

Hazard systems are usually quite complex and may cover different types of hazards, which can overlap in space and time. The interactions among hazards influence the overall hazard level, both in terms of size and frequency (Kappes et al., 2012b). Marzocchi et al. (2009) discussed the concept of "transfer", the possible amplification of the overall hazard due to interaction, because of their non independence. A debris flow resulting from the failure of a landslide dam might be of a higher magnitude than expected channel or slope debris flows. Hence, it is not possible to reduce a multi-hazard assessment to the sum of single, independent hazard assessments. Earthquake and rainfall can both trigger landslides, but they are only coincidentally related; they can occur randomly in the same area at the same time (Gill and Malamud, 2014; Havenith et al., 2016). However, their joint occurrence is likely to produce effects on the occurrence of landslides (Dellow et al., 2017). In the example of the Kaikoura earthquake, it is worth questioning whether the number of occurred landslides have been affected by the spatio-temporal overlap of the earthquake, the aftershock sequence and the rainfall, recalling also that large earthquakes may produce a prolonged effect on the landslide susceptibility of the affected area, as mentioned in Section 3.3.2. Nevertheless, the state of art of these methods is far from reliable quantitative approaches (Selva, 2013; Mignan et al., 2014; Kappes et al., 2012c; Gill and Malamud, 2014), and we will review the most important examples in this section.

Kappes et al. (2010) suggested that the first step for a multi-hazard assessment is the identification of the interactions between hazards, followed by the formulation of the links between hazard models, taking into consideration their characteristics and

differences. The identification of interactions has been the subject of important review works (Gill and Malamud, 2014), as we will see later on in this section. The second step is even more complicated, as the interactions are many and varying, and combining hazard, interactions and assessments that area substantially different it is not a simple process.

Hazard interactions For Kappes et al. (2010) an event is caused by the combination of two aspects: disposition and triggering event. The disposition represents a slow process changing the environmental settings over time, while the triggering event is a much faster process, able to modify the system quickly enough to lead to a possible immediate hazard occurrence An example is water filtering through the ground: after a certain time, depending on the ground composition), the leaking may have increased the ground water saturation enough to trigger a landslide. Otherwise, an intense day of rainfall can trigger the landslide in a much shorter period of time, with a much higher level of intensity needed from the primary hazard, the rainfall.

Liu et al. (2015) refers to cascading events (a directly produced occurrence of a secondary event), conjoint events (two phenomena occurring in the same area and time window), and dynamic hazards (the occurrence of one hazard affecting the chances of the occurrence of a secondary hazard).

Kumasaki et al. (2016) analysed cascading natural disasters in Japan through information on local and national newspapers, in order to have a wide review of all possible events. Interactions were classified as striking (primary disaster has significant further impact by imparting sufficient energy to move a significant mass or to propagate energy through media), undermining (a primary disaster lowers the resistance or weakens a system maintaining mass and the mass collapses) and compounding (a primary disaster lowers the resistance of a system and adds to the amount of mass affected)

While Kumasaki et al. (2016) provided a classification that seems more related to the effect on the communities, Liu et al. (2015) shows similarities with the classification seen in another paper (Gill and Malamud, 2014), a major review of the state of art of multi-hazard methodologies and a call for quantitative ones.

One of the most interesting classification was proposed by Gill and Malamud (2014). Starting from the distinction between primary and secondary hazards, they have presented a classification based on the assumption that in nature there are different resulting effects depending on the type of interaction. The proposed classification is the following:

- Triggering
- Increased hazard probability

- Decreased hazard probability
- Spatial/Temporal coincidence of hazards.

The first type is the most direct interaction, where one primary hazard triggers one or more secondary hazards. Every secondary event can also trigger other events itself, producing a chain of hazards. Hence, there is a direct activation of an event by another. Without the first event, there would be no triggering. The triggering can be caused also by the simultaneous occurrence of two or more events. Examples from the first class are volcanic eruptions triggering earthquakes, or earthquakes triggering landslides. The second and third classes reflect the mechanism of one hazard changing the environmental circumstances, hence altering the frequency or magnitude of another hazard. The primary hazard may or may not directly trigger a secondary natural hazard, leading however to its occurrence. Examples from the second and third classes are events related to rain and drought. Long rain-less periods may increase the chances for wildfire and landslides or debris flow. However, rainfall can directly trigger landslides or debris flow when certain levels of intensity and duration are reached. Finally, spatial/temporal coincidence of hazards occurs when two or more hazards (neither triggering the other) overlap in space and time, producing compounding losses potentially bigger than the sum of the effects of the hazards considered independently (Kappes et al., 2010; Mignan et al., 2014). For example, a volcano erupting during a typhoon may cause violent lahars and floods (Mount Pinatubo eruption, Umbal and Rodolfo (1996)).

Because the topic is new and potentially complex, it is important to review the proposed methods for multi-hazard analysis present in the literature. Even if such methods are qualitative or semi-quantitative they can provide precious information on how to build a multi-hazard framework. The final aim is to translate the interaction among hazards and how it affects the triggering mechanism into a quantitative statistical model, one that can be modified and applied to several chains of hazards and in different geographical areas, so that it is not limited to a case study or to specific events.

The most widely used approach is the use of risk matrices, a discrete approach which describes the hazard level coming from the combination of two hazards. It usually requires less data than other approaches (Kappes et al., 2012b), but its qualitative nature limits any further statistical analyses. In fact, hazard combinations are qualitatively described and perhaps ranked on the basis of expert judgement. Nevertheless, these risk matrices represent a milestone to build a quantitative hazard framework.

De Pippo et al. (2008) has proposed a matrix (Tab. 3.4) which helps with understanding the possible hazard interactions in a coastal hazard assessment. With the aim

of avoiding the underestimation of risks, the matrix is built by listing all the identified connections between hazards, in terms of inputs and outcomes.

For instance, landslides can accelerate cliff erosion and dam rivers. On the other hand, shoreline erosion and river flooding may increase the chance for landslides. This approach is totally qualitative, as it presents all possible interaction into a matrix-style representation, probably more helpful for decision-making at management level, or rather as a starting point for a quantitative risk assessment. In fact, this matrix states what should be included in a hazard assessment for this chain of hazards, but not how these interactions should be evaluated.

Another, perhaps more complete, version of a hazard matrix has been proposed by Gill and Malamud (2014). Their work is actually more than just a matrix. It is an extensive review of hazards and their possible connections and the presentation of some methods of representation. The hazards identified by the authors are listed in Table 3.2.

Hazard group	Hazard
Geophysical	Earthquake Tsunami Volcanic Eruption Landslide Snow avalanche
Hydrological	Flood Drought
Shallow Earth processes	Regional subsidence Ground collapse Soil (local) subsidence Ground heave
Atmospheric	Storm Tornado Hailstorm Snowstorm Lightning Extreme temperatures (Hot) Extreme temperatures (Cold)
Biophysical	Wildfire
Space/Celestial	Geomagnetic storms Impact events

Table 3.2: list of all hazards divided into groups

The proposed hazard matrix (Figure 3.3), shows the interaction between hazards, together with the potential of such interactions in terms of number of possible triggered

hazards. For instance, the impact of a meteorite can directly trigger a tsunami, and potentially a large number of earthquakes. Snowstorms can either trigger or increase the probability of a large number of avalanches. Droughts can increase the probability of wildfire.

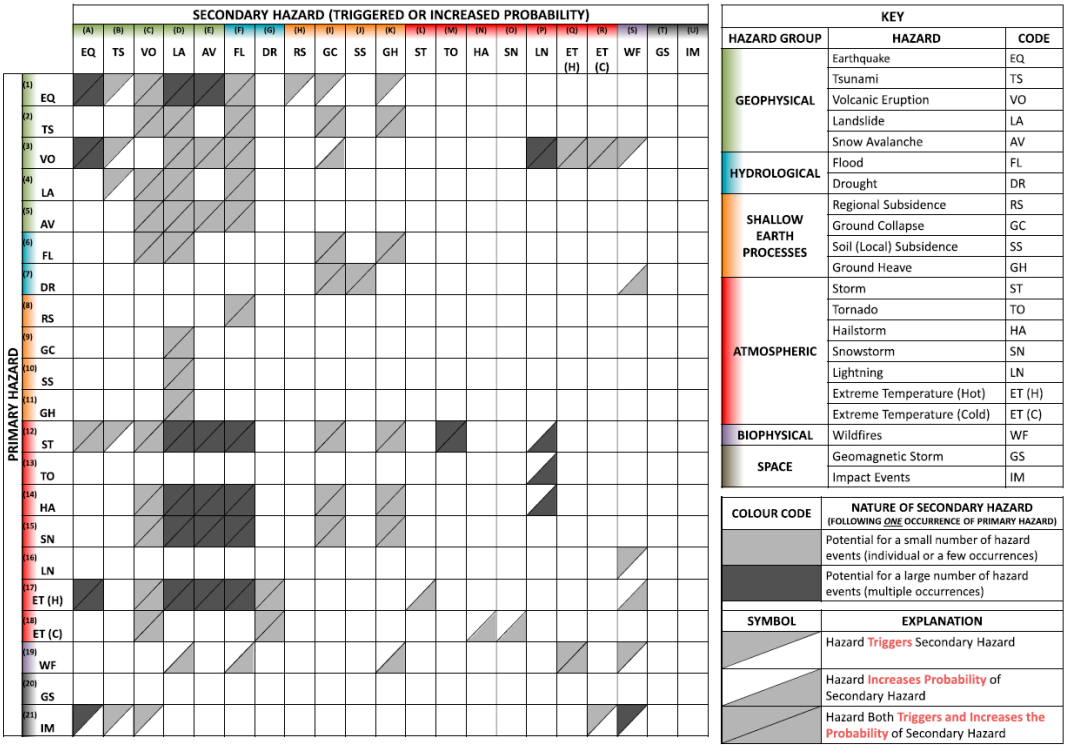


Figure 3.3: The figure shows the possible interaction between hazards, divided the type of interaction and the potential number of triggered events. For example, a volcanic eruption has the potential to trigger directly and increase the probability of a large number of landslide events. Reproduced from Gill and Malamud (2014)

Qualitative Multiple Hazard Assessment

The above matrices are a good starting point in terms of descriptive knowledge of causes and effects, but they also provide attempts of qualitative evaluations for a multi-hazard assessment.

Gill and Malamud (2014) proposed a factor based indicator on an ordinal scale that may help to summarise the information enclosed in a matrix (Figure 3.3), using ranks. The values come from a literature review of hazards run by the authors:

$$OLF_T = \sum_{OLF=I}^{VI} (f_{OLF} \times OLF) \tag{3.3}$$

where the total score is the sum of the frequency f (number of times the hazard interaction is considered overlapping in the literature) multiplied by the overlap-likelihood factor, calculated as a rank. This Overlap-Likelihood Factor OLF represents the combination of spatial overlap (from 1 to 3) and temporal likelihood (from 1 to 3). These multiplications give values 1,2,3,4,6,9 that have been categorised into OLF values I to VI. For example, as from the Figure 3.4, snowstorms and volcanic eruptions are classified as non-likely to occur both at the same time (1) and in the same area (1), so $OLF = I$. Overall, this is a weighted average of ordinal variables, which does not assure

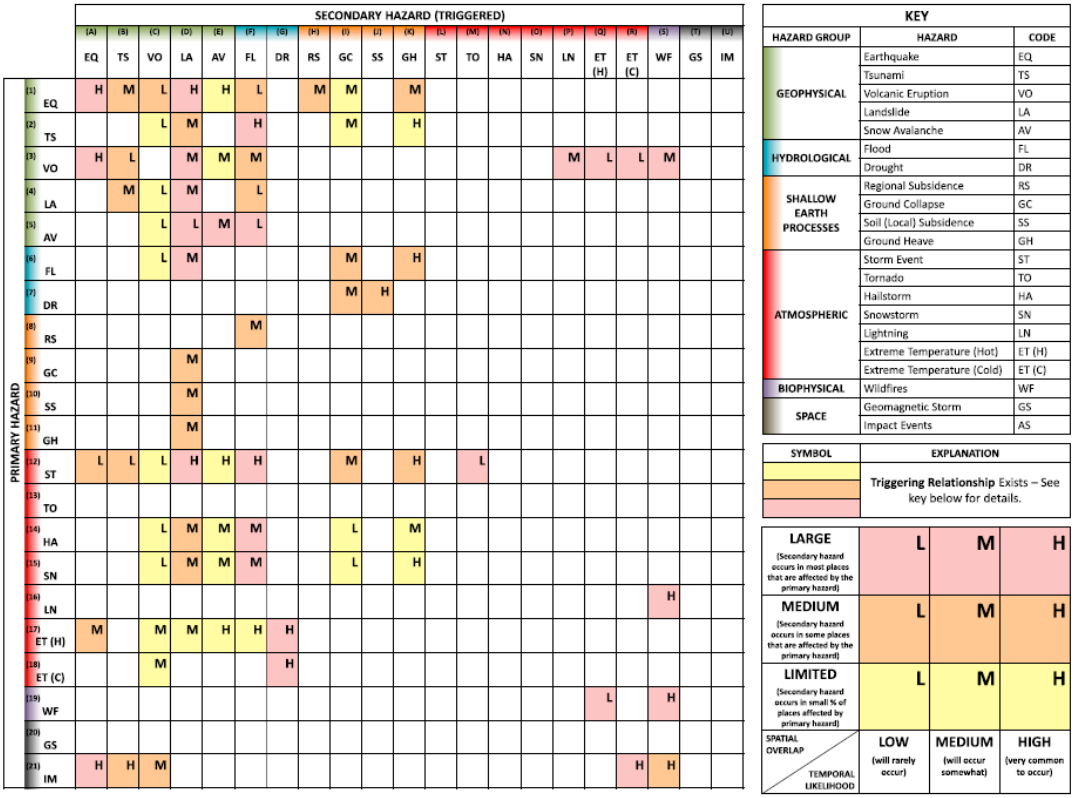


Figure 3.4: The table shows the interaction between hazards classified into nine classes, depending on the chance of spatial overlapping and temporal likelihood. It is possible to see that earthquakes have high probabilities of both with other seismic events, while tsunami is very high in temporal likelihood with ground heave, but limited in spatial overlapping.

the preservation of the ordering. Factors have been arbitrarily ordered, and the same can be said about the values assigned. Also, the information on hazards have been collected from a literature review of single hazard assessments: the evaluations may differ substantially, and the consequential artificial categorization may be meaningless. The classification is made by assigning numbers to classes from an ordinal scale, which is eventually qualitative. This association is not probabilistic or quantitative, and that

hinders the possibility of any comparison and to produce quantitative outcomes. Furthermore, it is not clear if the classification low to medium/limited to large is made assuming independence or not. There might be some further degree of interaction that is not captured by this ordinal classification.

Similarly, De Pippo et al. (2008) evaluated the interactions of the hazards affecting a given area by an interaction matrix in Table 3.4. Because every hazard is supposed to have different ranges of effects, not often comparable, the authors ranked the effects for each hazard and combined them into a weighted multi-hazard index H for the overall coastal hazard.

$$H = \sum_{n=1}^{6} Ip_n Xp_n \tag{3.4}$$

where Xp are ranks 0 to 6 expressing the percentage of influence of each parameter on the overall hazard, and Ip are a code of effects, ranked from 0 to 4 in relation to the severity of the effect (see Figure 3.4). The index H is a sort of cause/effect analysis, and is subjective, and somewhat arbitrary. The ranking of effects and parameter influence is not only fuzzy (ranks are arbitrarily assigned) but related to limited case studies which might not express the overall underlying process of interaction among hazards. The table in 3.4 shows the events and the possible effects. The result is still qualitative so hardly applicable to other events, and it is not possible to obtain quantitative and comparable results.

Another indicator, with similar issues, is the indicator-based RVI (relative vulnerability index), based on the PTVA (Papathoma Tsunami Vulnerability Approach; Papathoma et al. (2003)) and further developed by Kappes et al. (2012c), presented as a multi-hazard indicator for shallow landslides, large earthflows and debris flow triggered by intense rainfalls and summer storms. An intensity scale (low, medium, high) was used to classify the hazard by damage size, although Kappes et al. (2012c) conceded that the vulnerability to a hazard should not be limited only to the extent of the damages. A formula, similar to (3.4) was used, altough weights are now defined so that they sum to 1:

$$RVI = \sum_{i=1}^{m} w_m I_m S_m \tag{3.5}$$

where I_m are indicators of the vulnerability of an area and S_m are assigned scores of vulnerability. This equation bears the same issues of (3.4).

A different approach was proposed by Fleming et al. (2016), who combined risk curves (as the above cited Lee (2014b)) arising form different hazards, in order to calculate a common total risk. In particular, it is stated that independent risks can be combined using $P_i(L_j)$, the probability of exceedance of the j^{th} loss per annum for

independent i^{th} risk source:

$$P(L_j)_{tot} = 1 - \prod(1 - P_i(L_j)) \tag{3.6}$$

Each risk's impact is classified from insignificant to disastrous, and the probability of exceedance of a given loss is ranked from very unlikely (≤ 0.00001 per year) to very likely (up to 0.1 per year). With this information, a 5 by 5 matrix is built to provide a visual estimate of the relationship between probability of exceedance and impact of the overall risk. However, this equation is incorrect, unless at most one hazard can occur. There are anyway some possible changes that can be done to improve this method. As written before, we cannot consider the hazards independently, hence it would be useful here to consider a method to combine hazards together before plotting them. Furthermore, if the entire risk evaluation is quantitative, we may have a matrix plot with continuous changing colors instead of blocks with sets of colours based on qualitative ranks. That is the case of hazard maps made out of semi-quantitative models, which are presented in the next subsection.

The issue with these approaches is the non-replicability, as they are built on specific cases, with expert judgements on qualitative information. The application of any of these equations to other areas or hazard chains might require to restructure completely the equations. Another problem is that, assuming that they actually express the level of risk in a given area, they show a static picture of the multi-hazard interactions, as the time is not considered at all. As stated in Section 3.3, the effect of the triggering mechanism is definitively temporally related. Thus, the above equations represent a qualitative and vague idea of vulnerability of an area to more than one hazard, and not a multi-hazard assessment of the likelihood of triggered events.

Shoreline erosion	No interaction	A narrow steep beach without berms is open to wave attack	Coastal retreat contributes to the decrease in strength coastal zone, replacing	No interaction	Erosion affects wide stretches of the high-density urbanized coastal zone, replacing the dune ridge
Flooding can cause extensive coastline retreat close to the river mouth or inlet	**River flooding**	The concurrence of large waves breaking and flooding along the same coast increases destabilization	Breaches or overwash related to flooding induce landslides	No interaction	Flooding increase the effects with the covering of the surface due to urbanization
A large fetch and/or a wide coastal sector exposure to the prevailing wind detennines the highest rate of erosion	The contemporary occurrence of flooding and large waves breaking on the same coast increase destabilization along it.	**Surges**	Surges can affect a high cliff, both eroding the base (wave-cut notch) and scattering the marine spray along the slope	No interaction	Storm waves without adequate protection can easily reach buildings and roads close to the coast
The occurrence of landslides, associated to the quick removal of the talus, can accelerate the rate of cliff recession	Landslides and related phenomena can contribute to cut off or divert a flow in a water course	No interaction	**Landslides**	No interaction	Landslides occur on slopes subject to urbanization without thorough and/or regular planning, and therefore without a safety setting of the slope
No interaction	No interaction	Earthquakes and rapid falls into the sea of debris from volcanic eruptions can induce anomalous wave heights (tsunami)	Seismic shakes can induce landslides	**Seismicity and volcanism**	The influence of earthquakes and volcanic eruptions has risen in proportion with the number of inhabitants living and working in the coastal zone
Urbanized areas close to the coastline results in the rapid loss of its natural features; inadequate engineering structures induce erosion along the unprotected coast	Surface sealing by urbanization and the piping of natural flows increase erosion by flooding	Inadequate engineering structures make natural coastal defences less effective	Unauthorized building or confused urbanization have destabilized the natural balance	Engineering structure on the coast (i.e. confinement of lava flows)	**Man-made structures**

Table 3.3: Table of six hazards (main diagonal) and their interaction from De Pippo et al. (2008). In the top triangle of the matrix, each cell is at the intersection between a hazard on the left (being the cause) and a hazard at the bottom (being the effect). In the bottom triangle, it's the other way around.

Because this thesis will be mostly based on point processes, we can consider the implication of these categories on the construction of a conditional intensity for a triggered hazard. The direct triggering and the increased/decreased probability can be specified in a conditional probability as function of time and the history of the primary triggering of influencing hazard, as well as the history of the triggered hazard itself:

$$\lambda_2(t|H_t) = f(t|H_t^{(1)},H_t^{(2)}) \tag{3.7}$$

where $H_t^{(1)}$ and $H_t^{(2)}$ are the histories for the triggering and triggered hazard, respectively. The function $f(\cdot)$ will need to be specified in a manner that the conditional intensity will express an increased/decreased probability of occurrence of the secondary hazard in relation to the behaviour of the primary one over time, with the possibility of direct triggering given that the primary hazard will occur with specific features (e.g. in the case of coseismic landslide, the landslide will occur depending on the distance from the epicentre and the magnitude of the earthquake). For example, taking inspiration from (2.14) we may have a function that sums the contribution of each earthquake i up to time t:

$$\lambda_{LS}(t|H_{EQ}) = f(t,m,d) = \alpha + \beta \sum_{i:t_i<t} \exp\left[\frac{g(m_i)}{h(d_i)}\right] \tag{3.8}$$

where the $g(m_i)$ and $h(d_i)$ are some functions of magnitude and distance of earthquake i, while α and β are parameters.

In the case of spatio/temporal overlap, the triggered hazard will occur in relation to its history and the history of the two primary hazards. As we expect possible compounding effect, we can imagine the resulting conditional intensity of the triggered event as the sum of the two single conditional intensities of the primary hazards:

$$\lambda_3(t|H_t) = \lambda_1(t|H_t^{(1)},H_t^{(3)}) + \lambda_2(t|H_t^{(2)},H_t^{(3)}) \tag{3.9}$$

where $\lambda_1(\cdot)$ and $\lambda_2(\cdot)$ are the conditional intensities for the two primary events, both dependent on their respective history $H_t^{(1)}$ or $H_t^{(2)}$ and the history of the triggered hazard $H_t^{(3)}$. There are other possible combinations of these interaction categories. For instance, the two triggering hazards may interact in space and time in the triggering mechanism of the third hazard (e.g. earthquakes and rainfall occurring in the same area and time, affecting landslide events). In such case, we might prefer the product of the two triggering conditional intensities:

$$\lambda_3(t|H_t) = \lambda_1(t|H_t^{(1)},H_t^{(3)}) \times \lambda_2(t|H_t^{(2)},H_t^{(3)}) \tag{3.10}$$

In general, there may be the need to build the conditional intensity of the triggered

hazard as a function $g(\cdot)$ of the combined effects of the triggering hazards, so that the complexity of the system is well expressed:

$$\lambda_3(t|H_t) = g(t,f_1(H_t^{(1)}),f_2(H_t^{(2)})) \tag{3.11}$$

where $f_1(\cdot)$ and $f_2(\cdot)$ are the functions of the histories of the two triggering hazards, expressing, for instance, their magnitudes. In the case of earthquakes and rainfall:

$$\lambda_{LS}(t|H_{EQ},H_{RF}) = \alpha + \beta\exp[\gamma f_{RF}(p,t) + \delta g_{EQ}(m,d,t)] \tag{3.12}$$

where $f_{RF}(p,t)$ is some function of the precipitation p and time t, $g_{EQ}(m,d,t)$ is function of the earthquake as previously mentioned in (3.8), while α, β, γ and δ are parameters. An example of (3.11) is developed for the rainfall/earthquake $\rightarrow$ landslide system in Chapter 4.

3.5 A new idea: hazard potential

As mentioned in Section 3.4, Kappes et al. (2010) proposed that the mechanism leading to the occurrence of a hazard can be seen as the combination of two aspects: a slow and a fast one, namely called disposition and triggering. This theory is based on Heinimann (1998), where the slow process is defined as "the general setting which favors the specific process", while the triggering event is the process "which leads to the threshold crossing of a factor relevant for the hazard incidence". The underlying idea of this vision is that the interactions among hazards can be broken down into two separate processes running at different speed. Considering a rainfall-induced landslide: the lithology of the area is the basis of the hazard occurrence, because the characteristics of the soil (e.g. porosity) determine the susceptibility of the area to landslides. Then, the rainfall can be a short and intense event or a long period of moderate events, as well as a combination of both.

Considering the approach with point processes in use in this thesis, it is possible to imagine that the conditional intensity for the possible occurrence of a hazard, such a landslide, starts from a baseline value (depending on the area susceptibility), and then increases or decreases in relation to faster or slower processes. The main difference between these two processes will need to be expressed in the equations, with the faster process showing a predominant ability to speed up the process and suddenly push the conditional intensity to a level at which the occurrence of the hazard is very likely. On the other hand, the slow process will increase or decrease the conditional intensity by small amounts over time, possibly triggering a hazard only under specific long-term conditions (recall duration and intensity in Section 3.3.1. This is also reminiscent of the increasing/decreasing probability of a hazard mentioned in Gill and Malamud (2014).

This is a pivotal concept which has been expressed more or less clearly by the above mentioned authors, but that one author have particularly emphasized. Mignan et al. (2014) channeled this theory into the concept of "potential" of an interaction process, indicating the ability of a hazard to trigger a secondary one. Therefore, in this thesis there will be a constant reference to this concept, as the purpose is not only the creation of a working and replicable multi-hazard model, but most importantly to have an extensive and in depth framework in which such model can be built. Such "potential" framework, will then be based on the idea of a hidden layer governing the occurrence of a hazard, in relation to the factors that may affect it.

In their work, Mignan et al. (2014) built a probabilistic multi-risk framework based on Monte Carlo method, in which time series (representing different scenarios) of events are generated from a Poisson distribution. The interactions between hazard are computed through a so called "hazard correlation matrix" filled with conditional probabilities generated by the algorithm used. Although the overall work is a simulation and so not based on real data, it does highlight an important point. If there is the possibility to develop the concept of potential throughout a wide base of hazards, or in other words if it is possible to find a common approach to define conditional intensities using functions with specific features, there is the chance to make a step forward a feasible multi-risk framework. In fact, the complexity of combining different hazards, data, functions can be reduced with the use of conditional intensities built as blocks, connected in relation to the type of interactions. With this approach, a new algorithms may be developed for the simulation and evaluation of hazard occurrences with a substantially reduced amount of computer memory requirement. In Chapter 6, the concept of potential will be explored to produce examples of hazard interactions that can be expressed in such framework.

Chapter 4

A statistical model for earthquake and/or rainfall triggered landslides

4.1 Introduction

As discussed in Chapters 1 and 3, hazard analysis is mostly single-hazard focused, and examples of multi-hazard analyses are limited to susceptibility maps and models tailored to specific areas and events. In all these cases, the temporal aspect is suppressed, as the focus is mostly on the study of characteristics that make a certain area more or less susceptible to an event (Feng et al., 2016), or on the distribution of secondary events given the features (such as the magnitude) of the triggering hazard (Fan et al., 2012). However, the spatio-temporal overlapping of hazards suggests the need for multi-hazard analysis, making it possible to take into account the interaction among primary hazards in the triggering of secondary ones.

Landslides are one of the most common types of hazards, as they occur in a large number of countries, and their separate triggering by rainfall or earthquakes has been extensively studied. Recent projects have focused on the creation of large landslide databases, merging smaller existing ones (e.g. ISPRA 2019). The new availability of this data makes possible new analyses on landslides and their interaction with their triggering hazard. Therefore, in this chapter landslides will be used to investigate the feasibility of quantitative multi-hazard analysis able to capture the interaction among rainfall and earthquakes in the process of triggering landslides.

Even though there are data for rainfall, earthquakes and landslides, the lack of multi-hazard analyses is to be ascribed to the limited availability of quantitative data with specific space-time requirements, as explained in Chapter 1 (time-homogeneous,

time-stamped, congruent in time and space, ideally with magnitude expressed), as well as to the complexity of combining datasets with substantially different hazard temporal resolutions. Nevertheless, in Chapter 2, point processes have been presented as a viable option for multi-triggering modelling, thanks to their flexibility. In fact, it is possible to formulate the conditional intensity of a point process in order to include all the natural hazards occurring in a specific multi-triggering process. In other words, all the triggering hazards can be combined in the conditional intensity for a triggered hazard occurrence. The use of point processes, with suitable data preparation, allows for the development of a quantitative multi-triggering model. Therefore, in this chapter we will demonstrate that point processes can be used to model the multi-triggering effects of primary hazards on secondary ones, evaluating the interactions in the triggering process. In particular, we will be apportioning the relative and combined effects on landslide triggering given by earthquakes and rainfall.

Landslides are a common hazard in many terrains. Usually they are triggered by rainfall (Berti et al., 2012; Aristizábal et al., 2015; Peruccacci et al., 2017) or seismic activity (Lee, 2014a; Havenith et al., 2016; Robinson et al., 2016a). The risk associated to landslides can be quantified (Papathoma et al., 2015; Vega and Hidalgo, 2016), by probabilistic modelling (Lari et al., 2014a). Landslides are thus part of an important, and relatively well-documented, hazard chain (Han et al., 2007; Gill and Malamud, 2014), which also includes landslide dams. Earthquakes and rainfall are only coincidentally related; they can occur randomly in the same area at the same time (Gill and Malamud, 2014; Havenith et al., 2016). Because landslides can be triggered by either, there is a need (Kappes et al., 2012b) for a statistical model that incorporates both features, in such a manner that the separate and joint triggering effects can be estimated. To formulate a probabilistic model, we require a large database with landslide, earthquakes and rainfall well distributed throughout the spatial and temporal extents. While seismic and precipitation databases are commonly available, landslide ones are rarer and usually incomplete (Malamud et al., 2004; Guzzetti et al., 2012; Xu, 2015; Steger et al., 2016).

Previous work on landslide triggering has commonly looked at modelling the susceptibility (Kritikos et al., 2015; Aristizábal et al., 2015; Feng et al., 2016), i.e., the spatial distribution of events, in which the temporal dependence in triggering is suppressed. Some studies have focused on high-resolution models specifically tailored for a single area or short time period. Such models incorporate location-specific factors like slope, presence of watersheds and soil characteristics (Montrasio et al., 2012; Lee, 2014a; Aristizábal et al., 2016) driving the occurrence of landslides. It is difficult to extend this approach to a scale that supports robust statistical analysis of triggering causes due to data demands: small and localized datasets are much more refined, however the

collection of such data is time-consuming and expensive. Also, the null (landslides) data can dominate the problem without remedy. Nevertheless, thanks to recent efforts in landslide risk management, there exist some datasets that suit our need. One of the largest and most complete datasets is the Italian historical archive of landslides, collected by the IFFI project (Trigila et al., 2010). This has combined all the local and historical landslide archives, together with modern aerial photos. Additionally, Italy is prone to medium to high intensity earthquakes (Gasperini et al., 2013), and in many areas intense seasonal rainfall that can lead to flooding and landslides. Of all the Italian regions, Emilia-Romagna has the longest complete record of landslides, and an exploratory analysis of part of the landslide record has been performed by Rossi et al. (2010).

The rest of this chapter is organised as follows: Section 4.2 describes the datasets used, from the region of Emilia Romagna. Then, Section 4.3 introduces the model and its components. Finally, Section 4.4 presents the results obtained and the conclusions.

4.2 Data

I have chosen to base our analysis on the Italian region of Emilia Romagna because one of the largest and most complete datasets is the Italian historical archive of landslides, collected by the IFFI project (Trigila et al., 2010). This has combined all the local and historical landslide archives, together with modern aerial photos. Of all the Italian regions, Emilia-Romagna has the longest complete record of landslides, and an exploratory analysis of a lengthy part of the landslide record has been performed by Rossi et al. (2010). Additionally, Italy is prone to medium to high intensity earthquakes (Gasperini et al., 2013), and in many areas intense seasonal rainfall that can lead to flooding and landslides.

The region of Emilia-Romagna occupies a large area in Northern Italy, the southern boundary of which follows the Apennines range from north-west to south east. Half of the region consists of plains (part of the Po valley), while the remaining part is equally split between hills and mountains. The landslide prone areas are located on the Apennines, which represent a complex geological and tectonic setting (Martelloni et al., 2012), of a "fold-and-thrust post-collisional belt" formed by the subduction of the Adriatic plate with the European one (Bertolini et al., 2005). The deposits are mainly formed by sandstones and calcarenites (Vai and Martini, 2001).

4.2.1 Earthquake data

The earthquake dataset covers all seismic events of magnitude 3 or above which occurred in Italy from 1981 to 2018. The dataset is an updated version of the one compiled by

Gasperini et al. (2013) (up to 2015), provided by INGV, and extended until 2018 with the INGV online database. Note that this sets the temporal limits of our analysis. The data includes the location (latitude and longitude) of the epicentre, the moment magnitude, and time (to the nearest second). A large number of earthquakes could not possibly trigger landslides in Emilia-Romagna due to their distance (e.g. events in Southern Italy). Therefore, we have pruned the dataset using the distance from each epicentre to the centroid of the Emilia-Romagna region, retaining only the events within 400km (Khazai and Sitar, 2004). This subsetting method is much less severe than those in other studies (Marc et al., 2015), but our model will automatically discount earthquakes at too great a distance for their magnitude. With this threshold, I retained the L'Aquila sequence, containing many of the most recent and strongest events in the dataset. The number of events in the dataset is 8584, with a moment magnitude ranging between 3 and 6.5. The per annum rate of earthquakes in the triennium $2016 - 2018$ is substantially greater than that for the $1981 - 2015$ portion of the dataset as the number of events per year in Italy doubled (Italian National Institute of Geophysics and Volcanology, 2019). Figure 4.1 shows that most of the earthquakes have occurred along the Apennines, affecting in particular the province of Forlì-Cesena.

4.2.2 Rainfall data

The rainfall dataset (from ARPAE, the Emilia-Romagna environmental agency) is a compendium of daily precipitation from $1981-2018$. The data are from 441 rain gauges across 328 municipalities of the region, and each day/gauge observation is characterized by the amount of precipitation (mm), the geographical location (latitude and longitude) of the municipality and the type of precipitation: daily (from 00:00 to 24:00), cumulated (over a number of days) or snow (whether the precipitation is flagged as snow precipitation or not). As mentioned in Chapter 3, catalogs over a long period, such as this one, tend to have inhomogenous features. In fact, this dataset required a lot of cleaning before it could be used in the model.

We have redistributed cumulated values equally over their given time periods and further averaged values for municipalities with more than one operating gauge, and imputed missing values at a given municipality with that from the closest municipality value available. We thus have created the potential for a finer spatial analysis of landslide triggering than that of Rossi et al. (2010), who analyzed the region as a whole using only a single rain record. The resulting geographical distribution of rainfall over Emilia-Romagna is shown in Figure 4.2, highlighting the higher levels of precipitation along the Apennines. In particular, Figure 4.2 shows two clusters of higher average levels of precipitation, one in the south-east area of the region, but especially the one in the north-west. This mountainous area characterized by higher rainfall values is in

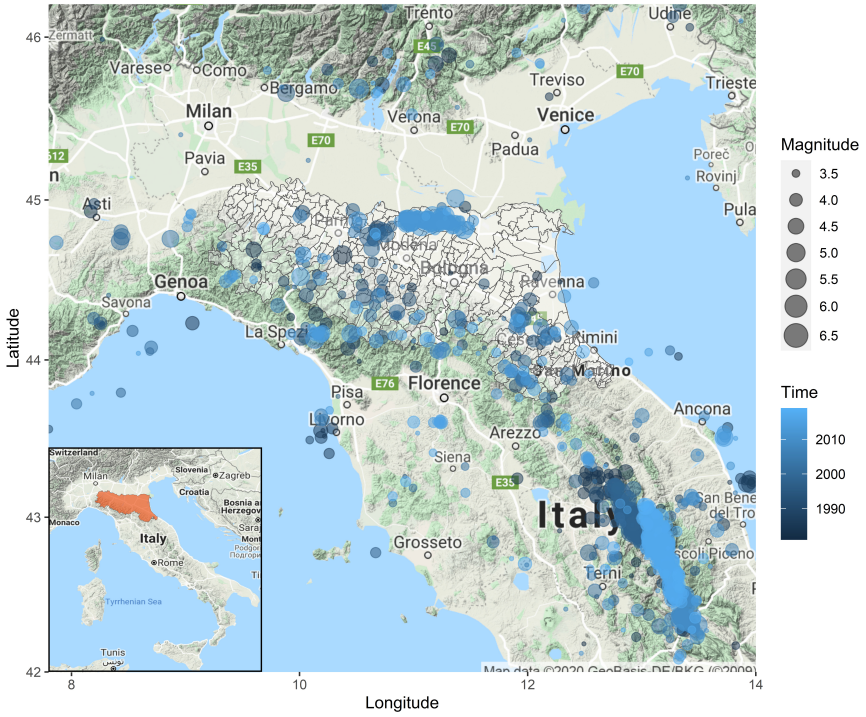


Figure 4.1: Earthquake distribution in centre-north Italy from 1981 to 2018. The L'Aquila sequence is visible in the bottom-right corner of the map. In Emilia-Romagna (shaded inset), earthquakes mostly follow the Apennines ridge but also some events occurred on foothills (between Rimini and Ravenna) and flat lands of Po Valley (all the top part of the region). The light blue concentration below Verona is the 2012 earthquake series. Municipality boundaries are shown in black.

the province of Piacenza.

4.2.3 Landslide data

The landslide data (ISPRA, 2019) for Emilia-Romagna contains 15118 landslides from prehistory to present. The data is heterogeneous, reflecting the multiple sources used to build the archive. The majority of landslides are reported with location (usually the name of the municipality), time (see paragraph below) and, when available, size. About 30% of records are incomplete.

As mentioned in Chapter 3, landslide data are often extracted from aerial photos or GIS data, incorporated with information on soil characteristics. This leads to datasets that are mainly focused on recording location and size of the event, but that rarely have temporal information with a level of precision and accuracy ideal for quantitative

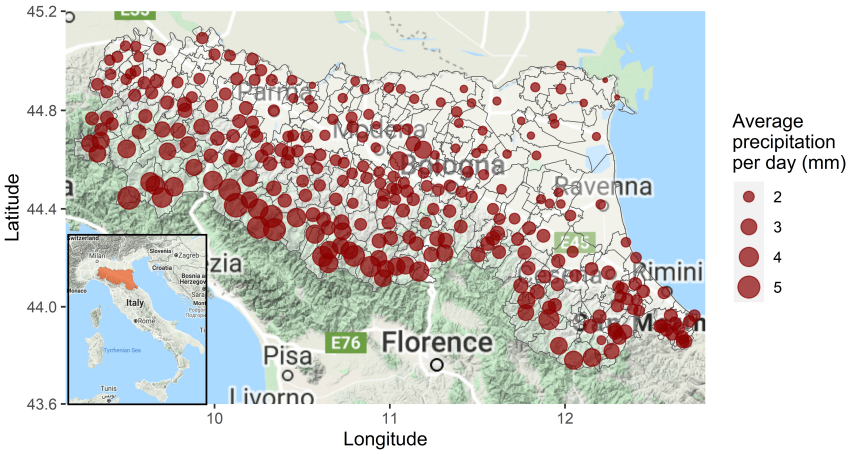


Figure 4.2: Geographical distribution of daily rainfall average (mm) 1981-2018 in the Emilia-Romagna region. Municipality boundaries are shown in black.

analysis. This is also attributable to the fact that landslides are often surveyed days or even weeks after a rainfall or earthquake event (Qi et al., 2010), which means that landslide clusters will typically be assigned to an arbitrary single date, even though they may have occurred days earlier or later. For instance, of the thousands of landslides occurred after the Kaikoura earthquakes (Mason et al., 2018), we don't know how many fell on that day, and how many the following days due to aftershocks or rainfall.

Rossi et al. (2010) discussed the completeness of a portion of this dataset at considerable length, arguing that it is complete enough for use from 1950 onward, the remaining time-inhomogeneities being ascribed to changes in triggering effects (primarily meteorological) and anthropogenic influences such as land use. The time window is further truncated to the period 1981-2018, to match the earthquake/rainfall catalog, leaving 7743 landslides. The main triggering factor of landslides in Emilia-Romagna is considered to be rainfall, while seismic-induced events are less frequent but still possible (Pizziolo et al., 2015; Piacentini et al., 2018; Troiani et al., 2017). Figure 4.3 shows the resulting geographical distribution of landslides, with the majority of landslides located in the mountainous area of the Apennines.

Landslide event time The major issue encountered in the landslide dataset is that of dating accuracy and precision. This is exemplified by the "first day problem" - the number of landslides recorded as occurring on the first day of a month is 2239 (Figure 4.4 A), 29% of the total amount.

This appears to be a consistent feature of the data, across years, seasons and municipalities. Figure 4.4 B shows the distribution of landslides in relation to the accuracy code, a variable included in the dataset that should express the dating precision of each

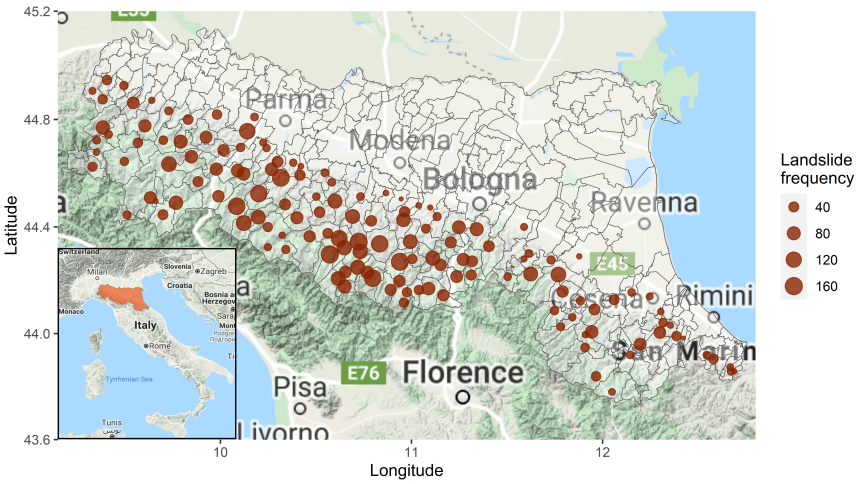


Figure 4.3: Landslide locations 1981-2018 in the region of Emilia-Romagna. Municipality boundaries are shown in black.

event. The figure indicates that the "first day" problem is spread across all accuracy classes, and hence these codes cannot be used to stochastically reallocate landslides over appropriate time intervals. Moreover, we can deduce that the accuracy codes on other days of the month are not reliable. From Figure 4.4A, there appears to be no clear pattern from day two onward, so it is assumed that days other than day one can be treated as precise to the day, and that day one events occurred on either day one (with a probability that will estimated), or on another day of the month, with some unknown distribution that I will likewise estimate. The rest of our solution to this problem is part of the model inference, which is covered in Sections 4.3 and 4.6.

Distances and geographical location While the earthquake data is specified to a high precision in space and time, the rainfall and landslide data are only geographically specified in terms of municipality name, and with (at best) daily precision in time. Hence, distances between earthquakes and potentially triggered landslides will be calculated from the earthquake epicentre to the municipality centroid. Our model will consider the number of landslides per day for each municipality, relative to the amount of precipitation within those municipalities and the distances to and magnitudes of seismic events. The metadata of the datasets used are summarized in Table 4.1.

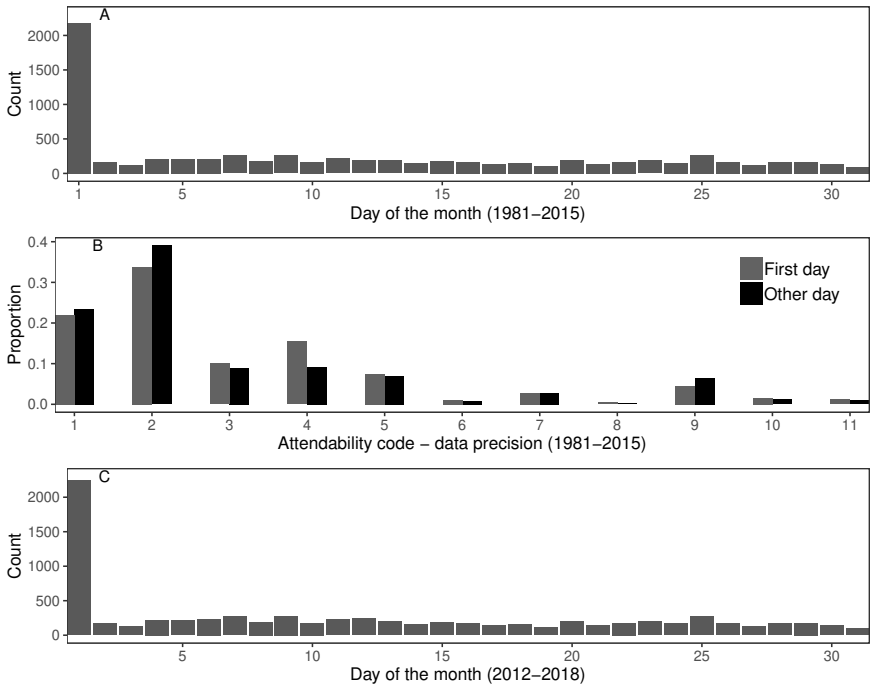


Figure 4.4: Distribution of landslides by day of the month (A) and in relation to the accuracy code provided (B). The dataset includes an "accuracy code", which should give an idea of the precision of the date. A landslide with code 1 indicates a claimed daily precision of dating, while code 2 means a short period precision ($1-7$ days). Codes 3 to 7 indicate a monthly, bimonthly, seasonal, biannual and yearly precision. Code 8 stands for a multiple year precision, while 9 and 10 indicate the date on the document or an uncertain date. Finally, 11 stands for events with unknown accuracy. For the period $2016-2018$ we have no information about the accuracy, but plot C exhibits the same distribution as plot A.

4.3 Stochastic model(s) for landslide triggering

As mentioned in Chapter 3 and in Section 4.1, the main triggering factors for landslides are rainfall and earthquake events. Additionally, many papers suggest a possible connection between these causes (Havenith et al., 2016; Marc et al., 2015). In this paper, we seek to quantify this influence by proposing a stochastic model that involves earthquakes, rainfall, landslides and their interactions in order to estimate the possibility of landslide events in relation to time and magnitude of seismicity and precipitation.

Because precipitation and earthquakes are two distinct types of triggering events, our model need to incorporate the effect of magnitude at distance of each of the triggering events and link their effects to landslide occurrences. As a suitable tool for this purpose (Chapter 2), we will consider a spatio-temporal point process over the space S (Emilia-Romagna) and the time horizon [1981,2018]. Particularly, we consider a non-homogeneous Poisson process denoted by a counting function $N(\mathcal{S},\mathcal{T})$ which counts

Table 4.1: Description of the datasets used.

Event Type	**Earthquakes**	**Rainfall**	**Landslides**
Time Window	1981-2018	1981-2018	1981-2018
Time Precision	Seconds	Days	Day at best
Location	Epicentre lat/long	Municipality centroid	Municipality centroid
Magnitude	Moment Magnitude	Millimeters per Day	Area or Volume
Features	$>$ 8000 events distance $<$400km magnitude $\geq$3	average 825.17mm per year	average 1.11 landslides per municipality/ year

the number of events occurring in $\mathcal{S} \subset S$ and $\mathcal{T} \subset$ [1981,2018]. It is commonly assumed that landslide occurrences follow a Poisson process (D'Odorico and Fagherazzi, 2003; Lari et al., 2014b), hence we will examine this assumption in our model. A Poisson process can be characterized by its intensity function (in continuous time and space) which describes the expected number of events. At the location $s \in S$ and time $\tau \in$ [1981,2018], the intensity function of landslides is a non-negative function denoted by $\lambda(s,\tau)$, where the probability of an event in a sufficiently small interval of space Δs and time $\Delta\tau$ is approximately $\lambda(s,\tau)\Delta s\Delta t$.

Discrete-time approximation As mentioned in Section 4.2, our datasets have different levels of spatial and temporal resolution, and that of rainfall and landslides is insufficient to fit a continuous (in time and space) model. This feature of the data implies a need to discretize time and space. Time is specified as days, which leads to a natural discretization of time in days (as mentioned in Chapter 1 and explained in Chapter 2). The centroid of each municipality was used as the nominal location of corresponding landslides. This leads to a discretization of space as $S = \cup_{x=1}^{X} S_x$ which is a disjoint union of spaces associated with municipalities over the index set of all municipalities $x \in \{1,...,X\}$. In other words, each municipality is treated in its entirety as a single spatial element. Therefore, we only observe the counts $N(S_x,[t,t+1))$ for landslides where $t \in \{1,...,T\}$ is the index set of different days. Basically, we count the landslide occurrence for each day t and each municipality x. For the sake of simplicity, we replace S_x by x and $[t,t+1)$ by t in the following analysis without ambiguity. Since we use a Poisson process with intensity $\lambda(s,\tau)$, the number of events $N(s,t)$ follows a Poisson distribution with mean

$$\mu(x,t) = \int_{s\in S_x}\int_{\tau\in[t,t+1)} \lambda(s,\tau)ds d\tau.$$

A ZIP model for landslides Because of the nature of the landslide series, the daily values are dominated by zeros (Witt et al., 2010), on 99.76% of the municipality-days. This suggests fitting a Poisson model is inappropriate, as the variance is nowhere near

the mean. The standard approach in such circumstances, which we adopt here, is to use a Zero-Inflated Poisson (ZIP) model, a random mixture of a Poisson variate and an atom at zero. If $N(x,t)$ is the number of landslides at location x and time t, the ZIP model augments the Poisson model by setting

$$\Pr(N(x,t)=n)=\begin{cases} q(x,t)+(1-q(x,t))\exp(-\mu(x,t)), & n=0 \\ (1-q(x,t))\exp(-\mu(x,t))(\mu(x,t))^n/n!, & n>0. \end{cases} \tag{4.1}$$

Briefly, a zero count can be produced either by the zero process or by the Poisson process, while a landslide occurrence will be produced only by the latter. Therefore, the probability of getting a zero count is $q(x,t)$, plus $1-q(x,t)$ times the probability that the Poisson distribution produces zero. The probability $q(x,t)$ is estimated with a logit model:

$$q(x,t)=\frac{\exp(-\nu_{x,t})}{1+\exp(-\nu_{x,t})} \tag{4.2}$$

where $\nu_{x,t}\in(-\infty,\infty)$ will defined as a linear function of the data.

4.3.1 Breaking down the triggering factors

The core of Eq. (4.1) is $\mu(x,t)$, a conditional mean function that links the occurrence of landslides with the possible triggering processes. We seek to parameterise it in terms of antecedent rainfall and earthquakes, modelling their temporal correlations and clustering, identified by Witt et al. (2010), as follows:

$$\mu(x,t)=\mu_0(x)\cdot g(C_1(x,t),C_2(x,t),...,C_n(x,t)). \tag{4.3}$$

where $\mu_0(x)$ is a baseline and $C_i(x,t)$ are *components* that capture the triggering effects of the primary events. The function $g(\cdot)$ is a link function, as is commonly used in generalised linear models. The purpose of the link function is simply to express the relationship between the components and the expected occurrence of landslides. The function $g(\cdot)$ and the components have now to be defined, based on our physical understanding of the triggering process. The components should increase with triggering propensity, and $g(\cdot)$ must be non-negative and monotonically increasing (Lawless, 1987; Daley and Vere-Jones, 2003).

One of our components will register the shaking effect of earthquakes, as they are one of the main triggers of landslides. Then, we want to differentiate between short and long-term rainfall, as it has been established that there are two distinct rainfall processes for triggering landslides (Rossi et al., 2010). In Eq.(4.3), we want to consider measures of short-term rainfall (denoted C_{RS}), long-term rainfall (C_{RL}) and seismic intensity (C_E) that increase with the triggering effect of the respective events. The

parameter μ_0 abstracts the susceptibility of municipality x to landslides, in terms of geography, lithology, soil structure and anthropogenic effects. This acknowledges the fact that landslides can be affected by local effects and isolates these to our search for a spatio-temporal relationship between earthquakes, rainfall and landslides. In the following paragraphs the components will be explained.

Seismic component An earthquake's ability to trigger a landslide is related to its magnitude within a certain period of time and within a certain distance from the epicentre (Robinson et al., 2016b; Parker et al., 2015; Kritikos et al., 2015; Marc et al., 2015). Kritikos et al. (2015) identified, via a fuzzy logic methodology, the most important factors in the triggering of coseismic landslides to be ground shaking intensity and distance (Chapter 3). Similar results were obtained by Parker et al. (2015, 2017) using logistic regression on datasets of various origins. As a proxy of ground shaking we will use the relationship (Utsu, 1970; Ogata, 1988; Wetzler et al., 2016) between main event magnitude and aftershock productivity, assuming that the forces that produce aftershocks are proportional to those that initiate landslides.

As it is not clear whether a landslide registered on day t has resulted from an earthquake on day t or day $t-1$, we aggregate the overall effect of seismic events occurred on both days. I thus propose a component

$$C_E(x,t) = \sum_{t-2<t_k\leq t} \frac{10^{1.5(m_k-3)}}{r_{x,k}^{\beta}} \tag{4.4}$$

where the k^{th} earthquake has magnitude m_k, at time t_k, a distance $r_{x,k}$ from location x. Spatial decay is modelled by a power law, with the distance being expressed in hundreds of kilometers (for numerical reasons). Following Meunier et al. (2007) and Zonno and Montaldo Falero (2009), we take $\beta = 1$, although a value of 1.8 (Travasarou et al., 2003) or even an exponential decay (Meunier et al., 2007) could be considered. The threshold of 3 in the magnitude simply reflects the cutoff in the catalogue.

Long-term rainfall component In order to define a tool for rainfall triggering landslides, a similar approach to the one by Monsieurs et al. (2019) is followed, using a measure of the antecedent rainfall rather than intensity-duration techniques (Chapter 3). Two rainfall components are considered in order to account for both the short and long-term effects on the triggering of landslides. While the short-term rainfall component will summarise the rainfall effect on the days t and $t-1$, for the long-term component we will use an exponential smoother (Montrasio et al., 2012) on a period of

Δ days prior to $t-1$ (i.e. days $t-\Delta-1$,...,$t-2$):

$$C_{RL}(x,t) = \frac{1}{\Delta}\sum_{\delta=1}^{\Delta}\omega^{\delta-1}P(x,t-\delta-1), \tag{4.5}$$

where $P(x,t)$ is the precipitation recorded at location x on day t. Some experimentation established that values of $\Delta = 150$, $\omega = 0.98$ produced the best fit to the data, although the fit was statistically similar for any $\Delta \in (120,180)$ days. The exponential smoother increases the effect of the days closer to the landslide day; with $\omega = 0.98$ it was found that day $t-152$ contributes approximately 5% as much as day $t-2$. The 150 days period, which may include multiple rain events (Palenzuela et al., 2016), agrees with the range of 42-400 days identified by Rossi et al. (2010), but is in excess of the approximately 30 days suggested by Guzzetti et al. (2012) and Berti et al. (2012). The sub-continental climate of Emilia-Romagna may drive the length of this influenced period as the precipitation is generally well-distributed during the year, with two peaks in spring and autumn (Nistor, 2016).

Short-term rainfall component Treating the long-term rainfall as in (4.5) allows us to use the simple average rainfall of the day of the landslide (t) and the day preceding ($t-1$) as a component expressing the mean intensity of the last two days.

$$C_{RS}(x,t) = \frac{P(x,t-1)+P(x,t)}{2} \tag{4.6}$$

As with earthquakes, this accounts for the inability to separate which day of rain may have triggered the landslide. The components (4.5) and (4.6) do not define a cumulated rainfall-duration threshold in the sense of Rossi et al. (2017). Instead we are using a "soft threshold", where events become more or less likely depending on their values, rather than possible/impossible. Effectively we are accounting for the uncertainty in the triggering conditions, driven by the fact that we include non-events (days without landslides) in our analysis. In other words, the model will not give dichotomous results, but rather the rate of daily landslides depending on the levels of the three components.

4.3.2 Three interaction models

I trialled three arrangements for the link function $g(\cdot)$ in (4.3), in order to test the interactions between the components.

Recalling Eq.(4.3), each model is a combination of the susceptibility term $\mu_0(x)$ and a function of the three components previously listed.All the components were normalised by dividing them by their grand mean across municipalities and time.

Model 1:

$$\mu(x,t) = \mu_0(x)\exp[\mu_1 C_{RS}(x,t) + \mu_2 C_{RL}(x,t) + \mu_3 C_E(x,t)] \tag{4.7}$$

treats the component effects as multiplicative.

Model 2:

$$\mu(x,t) = \mu_0(x)\{\exp[\mu_1 C_{RS}(x,t)] + \exp[\mu_2 C_{RL}(x,t)] + \exp[\mu_3 C_E(x,t)]\} \tag{4.8}$$

treats the effects as additive, while in

Model 3:

$$\mu(x,t) = \mu_0(x)\{\exp[\mu_1 C_{RS}(x,t) + \mu_2 C_{RL}(x,t)] + \exp[\mu_2 C_{RL}(x,t) + \mu_3 C_E(x,t)]\} \tag{4.9}$$

there are multiplicative effects between long-term rainfall and the other components, which are then added. This model represents long-term rainfall as a weakening factor, with the final impetus being provided by either intense precipitation or seismic shaking. The relative strengths of each component or interaction are measured by coefficient parameters μ_1 ,μ_2, μ_3.

4.3.3 ZIP terms

The ZIP form is compiled in four different forms (A to D), from the simplest ZIP form A, including only the intercept ν_0

$$A : \nu_{x,t} = \nu_0 \tag{4.10}$$

to models B and C, which are ZIP model forms that account for inflated zeros from the absence of the short or the long-term rainfall components, as suggested by the fact that rainfall is considered to be the dominant controlling mechanism for landslides in Emilia-Romagna:

$$B : \nu_{x,t} = \nu_0 + \nu_1 C_{RS}(x,t) \tag{4.11}$$

$$C : \nu_{x,t} = \nu_0 + \nu_2 C_{RL}(x,t) \tag{4.12}$$

and finally model D, which is the full model, allowing for inflated zeros to be influenced by all three components:

$$D : \nu_{x,t} = \nu_0 + \nu_1 C_{RS}(x,t) + \nu_2 C_{RL}(x,t) + \nu_3 C_E(x,t) \tag{4.13}$$

4.4 Results

The parameters were numerically optimized to maximize the likelihood. The susceptibility parameters $\mu_0(x)$, location based multipliers, were estimated as described in the Section 4.6. The model fitting was restricted to municipalities with at least ten recorded landslides in the period 1981 - 2018, to avoid the model fitting driven by individual landslides. The likelihood functions of the models are derived in the Appendix. The three components terms (4.4), (4.5) and (4.6) are used in the ZIP models (4.10)-(4.13) proposed in the previous section. Table 4.2 shows the estimated parameters and the value of the log-likelihood. Models are identified by a number denoting the form of $\mu(x,t)$ (Eq.4.7-4.9) and a letter identifying the ZIP model (Eq.4.10-4.13).

	Model 1		**Model 2**		**Model 3**	
	ZIP B	**ZIP D**	**ZIP B**	**ZIP D**	**ZIP B**	**ZIP D**
μ_0 **(mean)**	1.18E-03	3.95E-02	3.18E-03	3.93E-03	1.96E-03	1.53E-03
μ_0 **(SD)**	8.03E-04	4.09E-02	2.15E-03	2.68E-03	1.34E-03	1.04E-03
μ_1	0.23	0.23	0.18	0.18	0.21	0.15
μ_2	6.69E-08	0.16	0.06	0.16	0.02	0.15
μ_3	1.01E-05	1.02E-05	1.06E-06	1.44E-12	2.00E-04	6.00E-04
ν_0	7.64	8.05	7.55	7.92	7.58	11.9
ν_1	-22.7	-0.27	-2.24	-2.27	-22.9	-0.30
ν_2		-0.19		-0.45		-4.70E-75
ν_3		-3.20E-11		-1.00E-04		-0.01
Log-likelihood	-44710	-44670	-44393	-44337	-44587	-44792

Table 4.2: Parameter estimates (normalised components) and resulting log-likelihood for each model.

Of the ZIP forms presented in Eq.(4.10) - (4.13), only B and D have been retained as they were clearly superior to the others. Model 2 is preferred, as the log-likelihood is much the largest of the three, indicating that the earthquake and rainfall triggering effects on the number of landslides are best described as additive. The large differences in likelihood demonstrate that Models 1 and 3 are a poor reflection of reality in comparison with Model 2. Looking at the ZIP form, the difference in log-likelihood between Model 2B and Model 2D is not large, with a slight preference for 2D. Due to the complexity of the model, it is not clear if the improvement in log-likelihood from 2B to 2D is significant bearing in mind the two additional parameters (ν_2 and ν_3). These two extra parameters allow for different interpretations, particularly in terms of earthquakes, with one model (2B) including earthquakes as a term that mainly increases the number of landslides when at least one occurs, and the other (2D) having earthquakes affect the probability of there being any landslides at all.

The normalisation of the components (Figure 4.5) allows us to compare the importance of different components via the parameters with estimates in Table 4.2, while

the graphical representation in Figures 4.7 and 4.8 shows visually whether each model properly represents the data. Focusing on Models 2B and 2D, we first see that, regardless of which model is considered, the short-term rainfall parameter μ_1 is the largest contributor to landslide occurrences when only short-term rainfall is considered in the ZIP portion of the model. If the full ZIP parameterisation (4.13) is considered, the long-term rainfall parameter μ_2 increases in magnitude, but this is offset by the contribution from ν_2 in the ZIP portion. While the earthquake component parameter μ_3 is superficially low, we note that the values of the normalised seismic component can be orders of magnitude larger than the rainfall terms (Figure 4.5). Hence the seismic component is more variable, with a long tail, and the lower value of μ_3 means that the model is separating out the higher values of shaking. However, its effects can apparently be expressed through either μ_3 or ν_3, but not both. Turning to whether the models reflect the data, in Figure 4.7 we see the expected number of landslides consistently following the expected pattern from Model 2B. However, Model 2D shows a poor fit (Figure 4.8), where the expected landslides process is visually very different from the observed landslides one, being dominated by the 2016-2018 period which had slightly higher levels of rainfall overall. Hence Model 2D appears to be over-sensitive to the rainfall level. Model 2B shows a representation of landsliding which is more in line with other studies, which determined short-term rainfall to be the main driver of landsliding in Emilia-Romagna (Piacentini et al., 2018; Troiani et al., 2017).

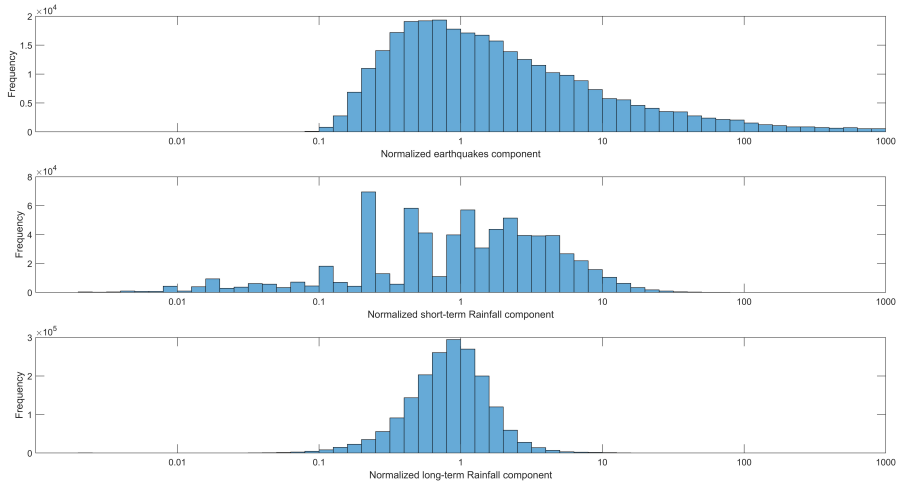


Figure 4.5: Histograms of the normalised three components.

The location-specific susceptibilities $\mu_0(x)$ are shown in Figure 4.6 against the number of landslides per municipality: while μ_0 increases in general with the number of landslides, the triggeringeffects of the model are demonstrated in the variation around

a hypothetical straight line. Again, we see that Model 2D extracts less information from the triggering data. Figures 4.7 and 4.8 present three panels showing the compo-

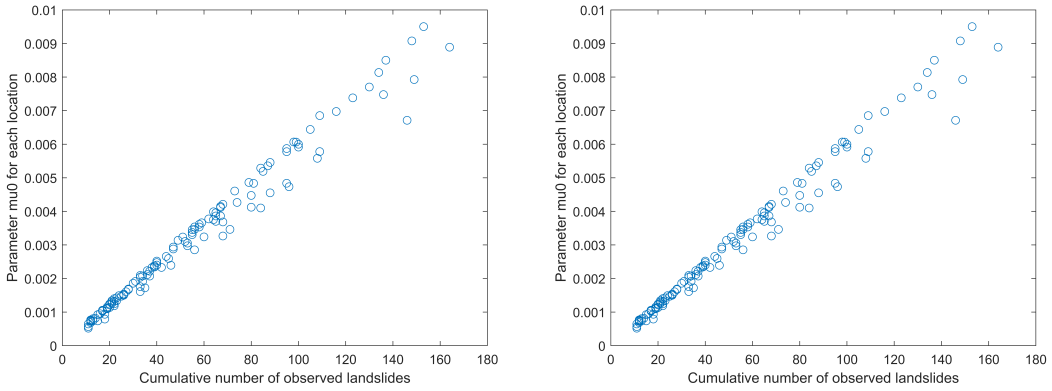


Figure 4.6: Estimated location susceptibilities for Model 2A (left) and 2D (right).

nents and the expected/observed landslides across all municipalities in the time window 1981-2018, respectively for Model 2B and 2D. The first plot displays short and long-term rainfall, the second one observed and expected landslides and the third one the earthquake component. As expected, looking at the first and second it is possible to see that the expected number of landslides has a temporal pattern which follows the rainfall one. At the end of the time window, due to the high peak of rainfall (the last three years of data show an increase in short-term rainfall), the expected landslides count is elevated. In the same period, there is a peak in the estimated earthquake effect which may have affected the triggered landslides. This 2016-2018 effect is seen more clearly in Figure 4.8, where the line expressing the expected landslides is higher compared to the one in Figure 4.7. and 4.8. Model 2B as illustrated in Figure 4.7 seems to provide a good representation of expected landslides against observed ones, remembering that many "day 1" landslides are obvious artifacts in the wrong temporal location. In contrast, Model 2D exhibits poor correlations between expected and observed landslide numbers, with the former being over sensitive to rainfall, and hence dominated by the higher levels of rainfall in 2016-2018.

A big spike in the expected number of landslides around 2012 reflects the anomaly of that year, where no landslides were recorded during or after the earthquake sequence mentioned in Section 4.2. It is interesting to notice that the correspondence in the number of observed and expected landslides around late 2008 to early 2009 includes considerable earthquake contributions, and some medium scale short- and long-term rainfall terms.

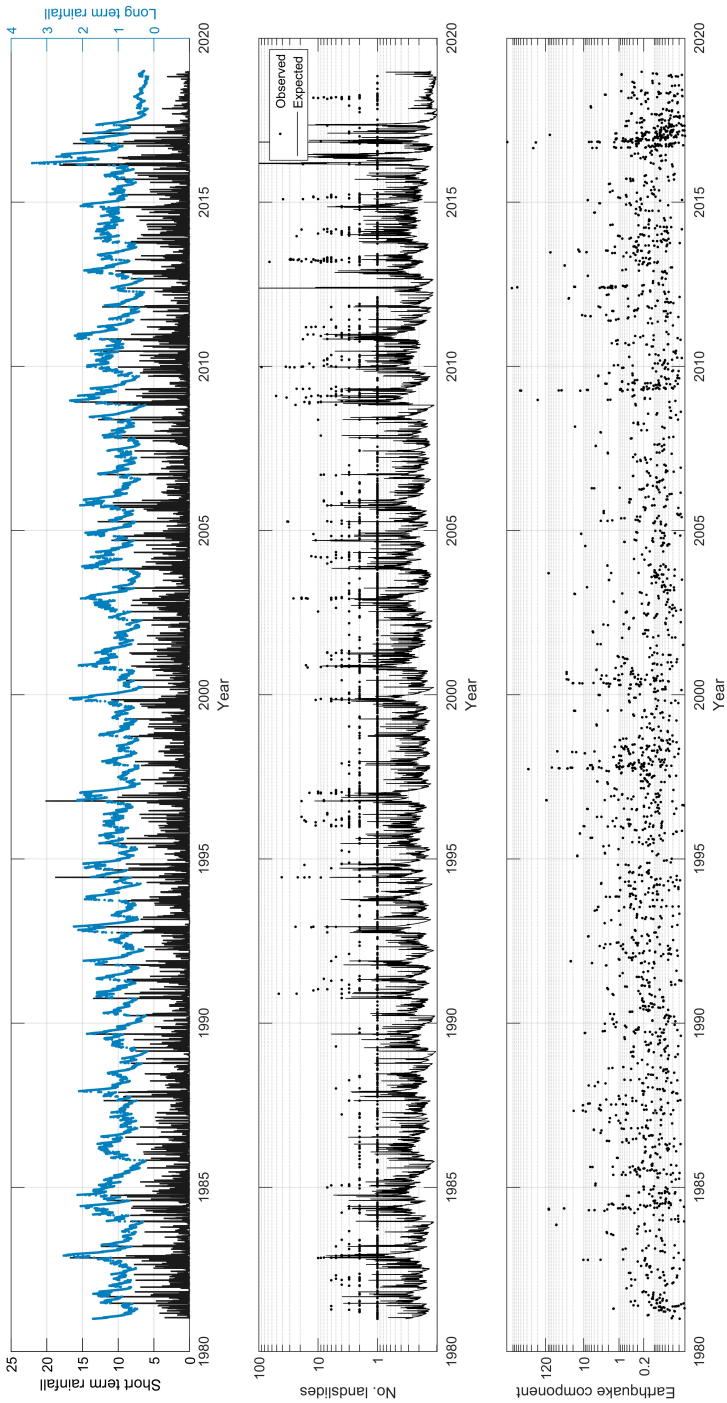


Figure 4.7: Observed/expected landslides and normalised rainfall and earthquake components in Emilia-Romagna (Model 2B).

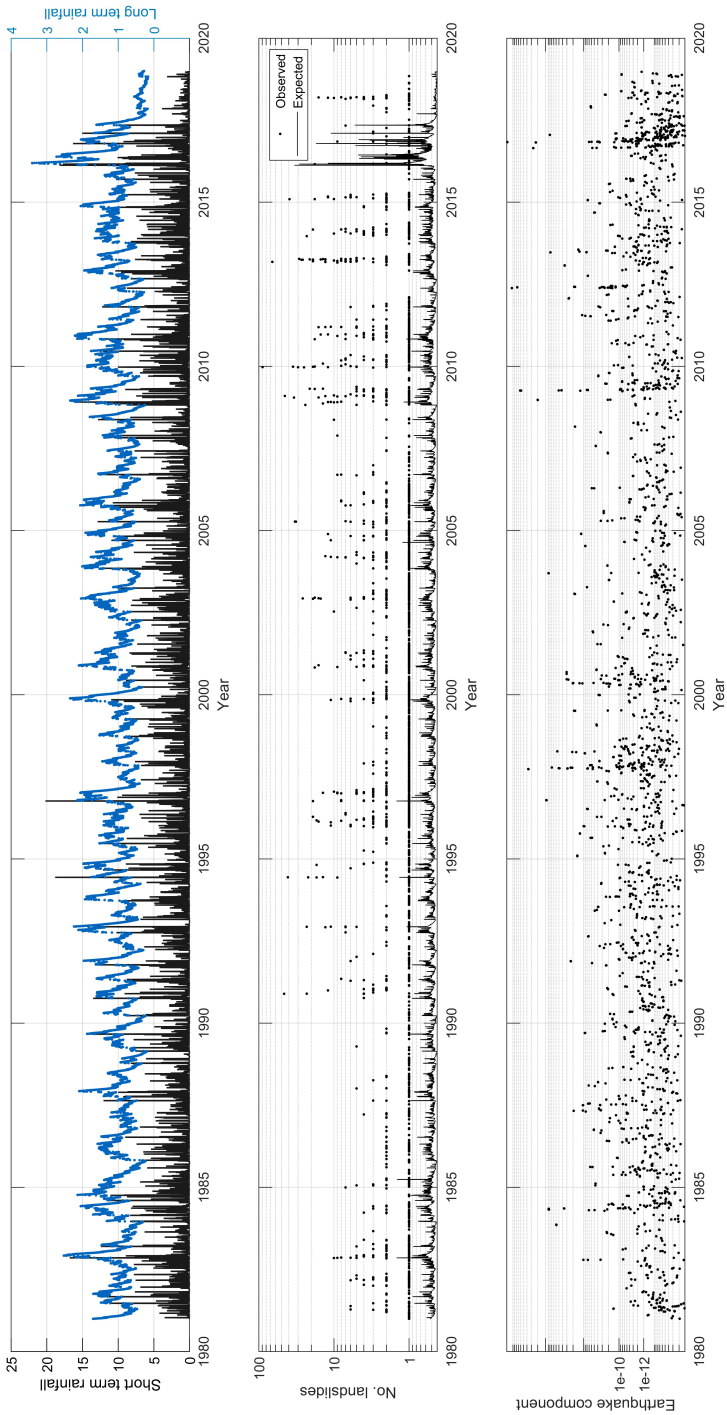


Figure 4.8: Observed/expected landslides and normalised rainfall and earthquake components in Emilia-Romagna (Model 2D).

The plots in Figure 4.9 shows the relationship between the three components (4.4), (4.5) and (4.6). The distributions of the components have been separated by days with landslides (black dots) and days without landslides (grey squares). In the first plot, the distributions of long-term against short term rainfall components exhibit interesting features: landslides tend to concentrate at medium to high values of long-term rainfall, as well as non-landslide values (see also the third plot, against the earthquake component). This indicates that high long-term precipitation values by themselves may not be a certain indication of landslide triggering. Recalling Table 4.2, the presence of a significant value of ν_2 term in Model 2D, confirms this finding: for high values of long-term precipitation, there are landslide occurrences but also many zero-landslide days (excess of zeros), resulting in a less than obvious link between landslides and long-term rainfall. This observation highlights the complexity of hazard interactions, mentioned in Chapter 3, that will be further discussed in Chapter 6.

The short-term rainfall values for days with landslides tend to be low, against either of the other two components, indicating that the short-term rainfall component by itself is not sufficient to trigger landslides, which is inconsistent with the estimates in Models 1B and 3B. When looking at each rainfall component against the earthquake one, it is possible to see a cluster of events at higher values of earthquakes component (in both the second and third plots) associated with low values of rainfall (short or long-term), suggesting the possible presence of coseismic landslide triggering expressed either as an amplifying term (Model 2B) or an initiating term (Model 2D), although this is somewhat difficult to identify in Figures 4.7 and 4.8.

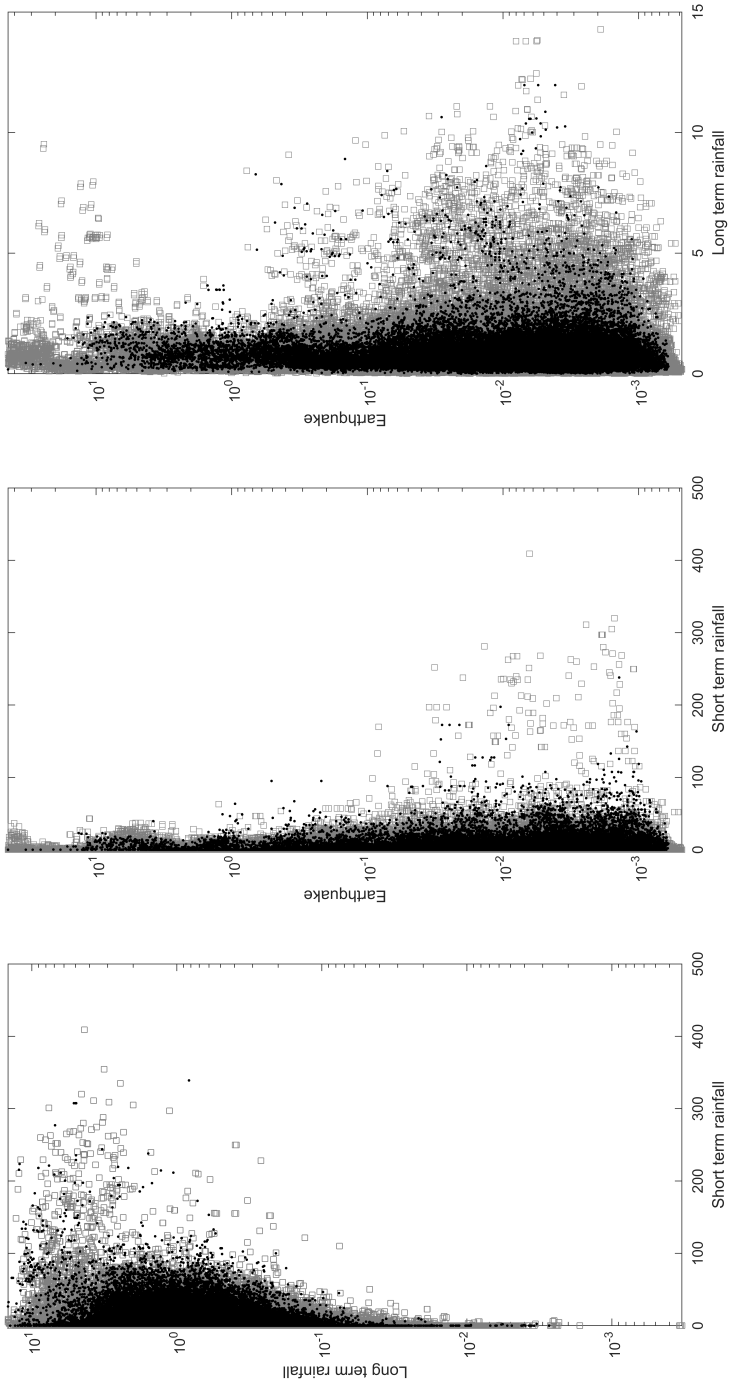


Figure 4.9: Component values across all days and all municipalities (Model 2B).

4.5 Discussion

I have demonstrated our proposition from Chapter 2: point processes can be used to model the triggering influence of multiple factors in a discrete approximation, with different configurations to examine possible hypotheses, and for a coarse resolution dataset. The formulation can be adapted to naturally discrete recorded data. Physical coefficients such as the weight $\{\mu_i\}$ for each component are treated as constants across space and time. Each location is assumed to have its own susceptibility to landslides, which acts as a multiplier. The temporal component is represented by a time-series of the triggering factors, the model structure remaining constant over time. The basis of the model is one that accommodates a spectrum of behavior from "increased probability" (Gill and Malamud, 2014), where the occurrence of an event increases the chances for the occurrence of a secondary event without directly triggering it, to almost direct triggering should the intensity rise quickly enough.

The available data for the landslide triggering problem in Emilia-Romagna is naturally at a daily precision. Hence the vast majority of location-days had no landslides. This over-abundance of zeros in the data required us to use a Zero-Inflated Poisson (ZIP) model (Chapter 2). This allowed us to treat the power-law decay in number of landslides per day (Rossi et al., 2010) as an aggregation across 139 municipalities of a few Poisson values and many zeros. I found that rainfall in particular exerted a strong effect on the likelihood of no landslides, agreeing with previous work by Rossi et al. (2010, 2017) and Peruccacci et al. (2017). With this foundation, the best triggering model has an additive form, where long-term and short-term (i.e., duration and intensity) rainfall, and coseismic triggering add together to raise the expected number of landslides. A multiplicative form was explicitly rejected by the data, as was a combined version where long-term rainfall was used as a multiplier for short-term rainfall and coseismic influence, and the terms added.

A possibility I did not examine, due to the already low level of coseismic landsliding in our dataset, is possibly transient triggering effects of earthquakes. It has been suggested that earthquakes can have a cumulative effect (even possibly a negative one) on landslide triggering (Brain et al., 2017), or that earthquakes and rainfall can interact in a complex manner over a period of years (Marc et al., 2015). This would require a new term in the model, where the cumulative effect of earthquakes is tracked (Bebbington and Harte, 2003). Considerable experimentation will be needed to identify characteristic time-windows and their dependence on data such as magnitude. A complicating fact is that the 2012 earthquake sequence represents a peculiar example of a seismic event without landslides. The reason may be due to a combination of factors, including the location of the epicentres of the two main shocks (about 50km away from the closest high ground) on a non-Apenninic fault, and the SE direction of seismic wave

propagation, which hence propagated unilaterally through the Po Valley towards the sea, rather than towards the mountains (Cesca et al., 2013; Pezzo et al., 2013; Vannoli et al., 2015).

At a finer level, if sufficient data on landslide location is available, the location susceptibility term could be parameterized in the usual manner (Parker et al., 2015), leading to a mapped intensity. However, this will require much more intensive development in the fitting process, as the data will be dichotomous (either a landslide occurs at that location and time, or not). Hence a spatial intensity will need to be fitted, possibly with a model for a size mark (Bebbington, 2015). The problem of whether a landslide inhibits (or encourages) a subsequent landslide at the same location would also need to be addressed.

Other avenues for future work include the possibility of including debris flows (Giannecchini et al., 2016) as a tertiary hazard, but with a complex triggering mechanism from rainfall, conditioned on the existence of previous landslides. The Melton ratio, measure of the area proneness to debris flows, of the catchment will then also have a role in the model (Welsh and Davies, 2011). A similar treatment could be accorded landslide dams (Tacconi Stefanelli et al., 2015).

4.6 Appendix

4.6.1 Estimation

Defining $N(x,t)$ as the actual number of landslides at location x on day t, we have that

$$\Pr(N(x,t) = n) = \frac{\exp[-\mu(x,t)][\mu(x,t)]^n}{n!}, \tag{4.14}$$

for $n = 0,1,2,....$

The log-likelihood for the process is therefore

$$\log L = \sum_{t=1}^{T}\sum_{x=1}^{X}\Big[- \mu(x,t) + N(x,t)\log(\mu(x,t)) - \log N(x,t)!\Big]. \tag{4.15}$$

Maximizing (4.15) is computationally expensive, due to different susceptibility terms $\mu_0(x)$ for every location. However, we can simplify this using a property of the point process Maximum Likelihood Estimate (MLE). Let us suppose that the conditional intensity for continuous time is written as $\mu(x,t) = \mu_0(x)h(x,t,\theta)$, where $h(x,t,\theta)$ is a function of components expressing the triggering mechanisms and of a vector θ of j parameters. The parameter $\mu_0(x)$ is a purely location based multiplier that expresses the susceptibility of a location to landslides. In static approaches an equivalent quantity is usually estimated via logistic regression (Garcia-Rodriguez et al., 2008; Minder et al.,

2009). For any value of θ, the log-likelihood at each location x is maximized by setting the conditional intensity (expected number of landslides) equal to the (observed) number of landslides in a given location across time

$$\mu_0(x)\sum_t h(x,t,\hat{\theta}) = \sum_t N(x,t),$$

where $\hat{\theta}$ is the MLE of θ. We now assume that this property is inherited by the ZIP model, thus equating the observed and expected numbers of landslides at that location, and hence

$$\mu_0(x)\sum_t h(x,t,\hat{\theta}) = \sum_t N(x,t). \tag{4.16}$$

This allows us to operate a two-step numerical optimization, where $\hat{\theta}$ is first updated, holding $\{\mu_0(x)\}$ fixed, and then the $\{\mu_0(x)\}$ are recalculated according to (4.16).

4.6.2 Landslide dating accuracy problem

As described in Section 4.2, we decided to consider all landslides that are reported on other than the first day of a month as reliable, and all landslides dated on the first day as potentially unreliable. Considering the intrinsic scarcity of landslide data (82% of days have no landslide events), it is not feasible to reduce our analysis to only the landslides not occurring after day one of each month, as we would not take account of a large portion of events and, more importantly, we would lose the continuity of their triggering effects. Moreover, the number of day one events differs sharply by month, and hence they still contain some information about triggering effects.

Let us define Y as the number of landslides at a given location in a specific month, with μ as their average daily rate of occurrence across the T days in the month. Then $Y = y_1 + y_2$, respectively the number of landslides recorded on the first day of a month and on the remaining days. Similarly, $T = t_1 + t_2$, split into the number of first days ($t_1 = 1$) and the number of remaining days t_2. Furthermore, let x_1 and x_2 denote the true (unobservable) number of landslides on the first day of a month and on the remaining days. Then if π is the unknown mis-specification rate at which landslides occurred from non-first days of a month but were recorded on the first day, we have $Y_1 = X_1 + \sum_{i=1}^{X_2} Z_i$, and $Y_2 = X_2 - \sum_{i=1}^{X_2} Z_i = \sum_{i=1}^{X_2}(1 - Z_i)$, where Z_i is a Bernoulli random variable with $\Pr(Z_i = 1) = \pi$. Taking expected values, we obtain $E[Y_1] = E[X_1] + \pi E[X_2] = \mu t_1 + \pi\mu t_2$ and $E[Y_2] = (1 - \pi)E[X_2] = (1 - \pi)\mu t_2$. We can now impute the missing data by replacing Y_1, Y_2 and μ with y_1, y_2 and $\hat{\mu} = \frac{y_1+y_2}{t_1+t_2}$, obtaining

$$\hat{\pi} = \frac{y_1/t_1 - y_2/t_2}{y_1/t_1 + y_2/t_1} \tag{4.17}$$

where we need $\frac{y_1}{t_1} > \frac{y_2}{t_2}$ (i.e., a noticeable excess of first day events) for a reasonable estimate. In order to obtain the probability that a landslide occurred on the first day of a month (event A), given that it has been recorded as such (event B), we can use Bayes Theorem to write

$$\widehat{\Pr(A|B)} = \frac{t_1}{t_1 + \pi t_2} = \frac{t_1/(t_1+t_2)}{y_1/(y_1+y_2)} \tag{4.18}$$

Thus we know $(1 - \widehat{\Pr(A|B)})y_1$ landslides need to be redistributed across other days of the month. We can do so by using the Expectation Maximization (EM) algorithm (Dempster et al., 1977), where they are allocated at each optimization step in proportion to the expected number of events. Note that we do not require landslides to be an integer at this point, as the log-likelihood calculations (4.15) or (4.19) do not require it (the factorial term in 4.15 is a constant, and thus does not feature in the maximization).

4.6.3 Log-likelihood of the ZIP model

The log-likelihood for the ZIP process is

$$\begin{aligned} \log L \;=\; & \sum_{t=1}^{T}\sum_{x=1}^{X} I_0(N(x,t))\log\{q(x,t) + [1-q(x,t)]\exp(-\mu(x,t))\} \\ & + \sum_{t=1}^{T}\sum_{x=1}^{X}\{[1 - I_0(N(x,t))][\log(1-q(x,t)) - \mu(x,t) + N(x,t)\log\mu(x,t) \\ & - \log(N(x,t)!)]\}, \end{aligned} \tag{4.19}$$

where $I_0(N(x,t))$ is 1 if no landslide occurred at location x on day t, 0 otherwise. The summation reveals that the model is fitted to the daily counts summed across municipalities (Rossi et al., 2010), but that each municipality contributes its own spatial triggering factors in rainfall, earthquake and susceptibility.

Data and code used can be found at the respective GitHub repository:
https://github.com/gfrigerioporta/eqrfls.git

Chapter 5

Landslide Dams

The material in this chapter forms the basis of a paper published in Landslides, (Frigerio Porta et al., 2020).

5.1 Introduction

In Chapter 4, I presented a model for the interactions in the triggering of landslides by rainfall and earthquakes. Going further into this hazard chain (see Figure 3.1), there is another consequential type of event whose analysis is similarly affected by the lack of sufficiently large and detailed datasets: landslide dams. A landslide dam occurs when a landslide falls on a river and blocks the water flow, causing water accumulation. The subsequent possible failure of the landslide dam will lead to a flash flood. This hazard threatens downstream human settlement or infrastructure. The danger is intensified by the amount of water accumulated, therefore estimation of the time to failure becomes crucial for assessing engineering risk mitigation procedures Chapter 3).

Modern engineering techniques have provided means to reduce the impact of a potential dam break, and in some cases the damage can be controlled (Wang et al., 2015). However, sometimes a failure occurs quickly and causes devastation and a high number of fatalities (Nibigira et al., 2018). Without any engineering mitigation, the presence of a dammed river creates a waiting game against nature, with high chances of devastating results (Inoue et al., 2013). The majority of landslide dam failures occur shortly after the formation (Dong et al., 2009; Korup, 2005), but the longer it takes the dam to fail, the more water is accumulated. Hence a quick assessment of the failure time of the dam becomes vital for determining if engineering options are necessary, or indeed feasible.

Other than by external events (e.g. earthquakes), the failure of a dam is mainly caused by overtopping or seepage (Awal et al., 2007). In overtopping, the water reaches the crest of a dam and spills over, eroding the structure, which eventually collapses

(Massey et al., 2013; Harrison et al., 2015). In case of seepage, the water filters through the mass of the dam (due to the porosity of the material) and weakens the structure until it collapses. Even if we exclude the effects of external events, the stability of a landslide dam is affected by multiple factors. Dams are subject to a certain amount of stress due to the pressure of the accumulated water. The type of material it is made of, the shape of the dam (McKillop and Clague, 2007), the size of the valley and of the resulting reservoir (Dong et al., 2009) are all crucial determinants of the durability of a landslide dam.

Previous studies (Casagli and Ermini, 1999; Korup, 2005; Deng et al., 2017) have focused on the dichotomy of failure/not failure using dam dimension indices, descriptive multivariate analysis and logistic regression (see Chapter 3). Although it is important to predict whether a dam will fail or not, it is perhaps more vital to know when it will fail. With a time to failure estimate, it is possible to put in place emergency procedures to secure the dam or to adopt other solutions such as safe release of the water (James and De Graff, 2012; Zhang et al., 2015; Chen et al., 2017) and implement evacuation plans. There is thus the need for a time-to-event model capable of forecasting the time of failure in relation to the dam characteristics. I therefore propose a Bayesian survival model to predict the time to failure of landslide dams, based on their characteristics and those of potential reservoir. A case study on heterogeneous Italian events is presented, where length and height of dams, and the catchment area behind them, are identified as the most important covariates controlling the time to failure.

The rest of the chapter is organized as follows. 5.2 discusses the landslide dam phenomenon. After that, we will introduce the dataset of landslide dams taken from the catalogue of Tacconi Stefanelli et al. (2016) in 5.3. 5.4 proposes a detailed description of our model, its inference and results. Finally, in 5.5 and 5.9 we discuss the results and conclude with possible generalizations of the model.

5.2 Landslide Dams

Our model will work with the dam characteristics only, in order to focus on the actual damming material, rather than the entire landslide. This also allows us to avoid the use of computed variables (e.g. volume). The dam characteristics consist of the material it is made of and its dimensions. As mentioned in Chapter 3, while the material is sometimes known prior to dam failure if geological studies in an area have been carried out, the dam dimensions play a more crucial role in the evaluation of time to failure (Dong et al., 2014). Several authors (Costa and Schuster, 1988; Casagli and Ermini, 1999; Tacconi Stefanelli et al., 2016) have also highlighted the importance of the dam dimensions in controlling dam failure. Therefore we will focus on the quantitative analysis of the dam dimensions, bearing in mind that our model is a first step towards

more complicated and tailored ones, that can incorporate other information such as the dam material.

A landslide dam is usually abstracted as a triangular prism, with a vertical surface spanning the valley on the upstream side. The length of a dam describes how large the blockage is in relation to the width of the valley, i.e. how likely the river is to be blocked, or to bypass the obstacle without forming a lake. The height of the dam, from the bottom of the valley to the crest of the structure, is a good indicator of how long it will take before the dam can be over-topped, and hence how long the material may have consolidated. The width of the dam is the along-valley dimension, and may indicate how robust the barrier should be, all other characteristics being equal.

Environmental (reservoir) characteristics are equally important as they give an idea of how much water can accumulate behind the dam (and hence the pressure on the dam) and of the inflow rate, which should indicate how fast the point of overtopping can be reached. A commonly used quantity for the latter is the catchment area, described as the area within which water flows (from rainfall and springs) towards the dam. Note that the treatment we are going to present here does not consider rainfall, which could be involved in decision making, but for which we lack data in this consideration of historical dams. Such factors contribute to the residual variation (standard error) in the model.

Any individual characteristic mentioned above cannot fully explain the durability of the dam, and they need to be combined into effective assessment tools. An example of these are geomorphological indices (Swanson et al., 1986; Ermini and Casagli, 2003), that have been developed to portray the condition of a dam using combinations of the dam and reservoir variables. These indices are usually calculated with only medium to large landslide dams, which lead to a biased representation of the landslide dams, as smaller ones, which are more likely to have failed by the present day, are not represented.

An example, previously mentioned in Section 3.2.4, is the Blockage Index (Casagli and Ermini, 1999), BI $= \log(V/C)$ the logarithm of the ratio between the volume of the landslide dam (V) and the catchment area (C). Ermini and Casagli (2003) further developed this approach by proposing the Dimensionless Blockage Index DBI $= \log(CH/V)$, where H is the dam height. These indices were used to partition the landslide dams into groups by setting threshold values. For example, Ermini and Casagli (2003) suggest that the dam is stable if DBI < 2.75 (H in m, V in m^3, C in km^2), unstable if DBI > 3.08 and uncertain otherwise.

Such calculations, however, only produce a static picture of the current state of the dam, without taking into consideration the temporal aspect of the hazard occurrence, such as the time since dam formation. The shortage of available data also hampers analysis, and restricts the use of these indices (Ermini and Casagli, 2003; Coico et al.,

2013) in other settings. Working on limited samples results in implicit biases given by the nature of the selected events. To combat this, Costa and Schuster (1988), as well as Ermini and Casagli (2003), compiled a worldwide databases of landslide dams, but the information about each dam is severely limited. Tacconi Stefanelli et al. (2016) instead attempted to avoid selection biases by assembling a more complete catalogue of Italian events. This dataset includes several types of events ranging from stable long-term dams with accompanying lake, to short-term failure events lasting just hours. This variety of events allows for more realistic analyses on the time to failure of dams. In fact short-term failure landslide dams are more frequent than long-lasting dams, and they are not free of risk. For example, Tsou et al. (2011) discussed the case of a typhoon-induced landslide that dammed the Chishan River in Taiwan and then failed a few hours later. Almost half of the events in the dataset of Tacconi Stefanelli et al. (2016) have failed after a short period of time with a lake still in the process of formation.

Furthermore, these short-term failures may be correlated with a categorical characteristic of the landslide. Costa and Schuster (1988) categorized landslides with the potential of damming a stream into six groups:

Type I: "..small landslides with respect to the width of the valley floor and do not reach opposite side of the valley";

Type II: "..larger and span the entire valley floor, occasionally depositing material high up on opposite valley sides";

Type III: "..fill the valley from side to side and move considerable distances up valley and down valley from the failure";

Type IV: "..form by the contemporaneous failure of material from both sides of a valley. The landslides can adjoin head-to-head in the middle of the valley, or they can juxtapose one another";

Type V: "..form when the same landslide has multiple lobes of debris that extend across a valley floor and form two or more landslide dams in the same reach of river";

Type VI: "..involve one or more failure surfaces that extend under the stream or river valley and emerge on the opposite valley side from the landslides. these dams typically involve slow basal sliding and slumping and form lakes by raising the elevation of the stream bed, changing the local gradient of the stream".

According to Costa and Schuster (1988), the most common landslide dams are type II and III (80% all together), followed by type I. The dataset summarized in the next section has a strong predominance of Type II (40%), followed by type I and III (26% and 24%).

5.3 Data

We have excluded 6 events from the dataset of Tacconi Stefanelli et al. (2016), due to missing information (lacking all the variables we need for our model). The remaining landslide dams include 295 events across Italy, Switzerland and San Marino, from the Alps (60% of events), through the central Apennines (10%) down to the mountains of Sicily (30%).

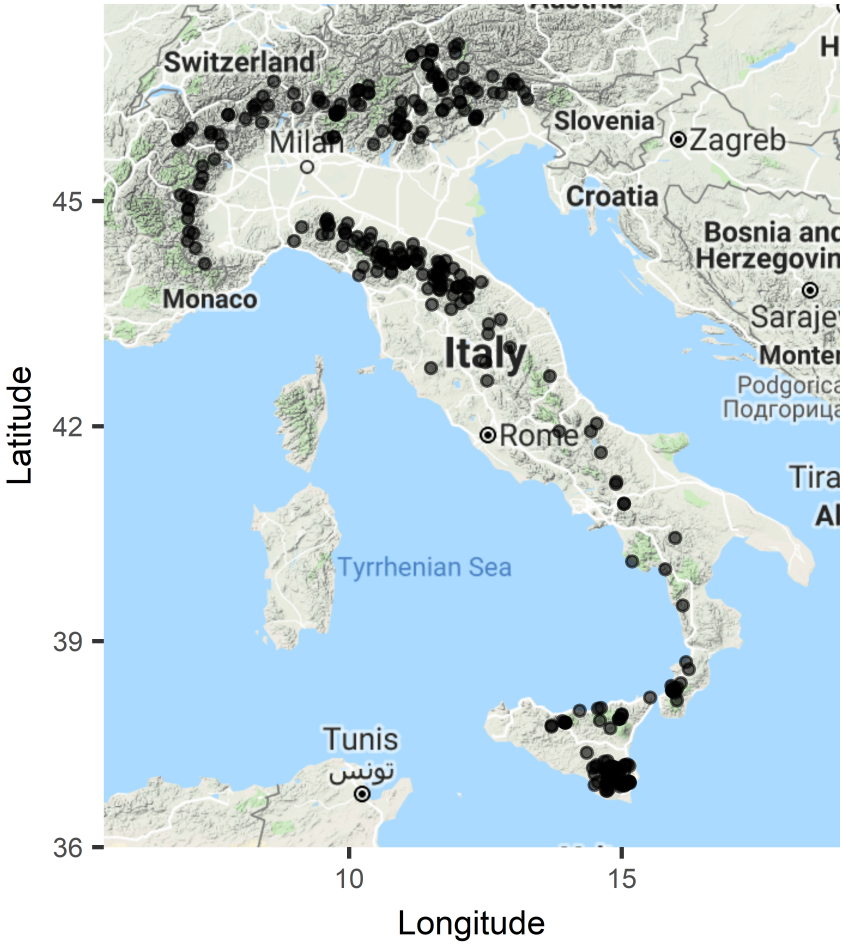


Figure 5.1: Geographical distribution of landslide dams in our dataset.

Each entry in the dataset consists of a single dam event, described by the quantitative variables presented in Table 5.1, and two categorical variables which are the landslide type (e.g. type I, etc.) from Costa and Schuster (1988), and the lifetime, ordered from "hours" to "millennia". The latter is, with a handful of exceptions, expressed as one of a number of ordinal categories. This is termed, in statistics, interval-censored data, in that we know (in our case we hypothesize) the lower and upper bound, but not the exact value.

Variables	% Missing	Group
Dam Width	40.73%	A
Dam Height	31.85%	A
Dam Length	31.07%	A
Dam Altitude	27.42%	A
Riverbed slope	22.98%	A
Catchment Area	1.57%	A
Dam Type	0%	A
Dam Surface	92.43%	B
Dam Volume	31.33%	B
Potential energy	97.13%	C
Lake Volume	82.25%	C
Lake Surface	67.89%	C
Lake Depth	79.37%	C
Max Flow capacity	76.50%	C
Lake Width	72.06%	C
Lake Length	69.19%	C
Landslide Surface	78.07%	D
Landslide Thickness	43.60%	D
Landslide Volume	13.05%	D

Table 5.1: Variables in the dataset.

As mentioned in Chapter 2, working with natural hazards such as landslide dams means dealing with incomplete or missing data. Table 5.1 summarizes the quantitative variables in the dataset with respect to completeness. The variables have been divided into four groups: Group A (variables we will use in our analysis), Group B (variables, such as volume, derived from those in Group A), Group C (variables with too many missing values to use, and no feasible way to impute them) and Group D (landslide variables). Bayesian imputation for Group C variables would not be robust, while using the variables in Group B would make the model over-specified. Therefore, it was decided to exclude these variables from our analysis and focus on the remaining ones in Group A. It is emphasized that only direct dam measurements (e.g. dam length) will be used, rather than computed ones (e.g. dam volume), in order to highlight the elementary/fundamental relationships among variables/dimensions driving the failure time of landslide dams. Hence the model we are going to build will be able to suggest the most important driving factors among direct measurements, and the best combination of them, if any. Furthermore, as noted above, we will use dam variables rather than landslide variables, as we are primarily interested in the potential mismatch between the size of the dam and the size of the valley (Costa and Schuster, 1988). If we did use compound variables such as dam volume or landslide volume, we would prefer the first one because it brings information about the landslide and, at the same time, the topography of the area (valley width). The landslide information could have a starring role, when combined with environmental variables, in forecasting the formation of landslide dams. This is, however, irrelevant to our aim of forecasting failure times

of dams that have occurred, and is left to future research.

Dam State	Lifetime Class							
	Hours	**Days**	**Months**	**Years**	**Centuries**	**Millennia**	**Missing**	**Total**
Alluviation			1	15	2	6	22	46
Drained		5	14	9			1	29
Existing			2	8	14	4	2	30
Stabilized					1			1
Not formed	2	3	1				92	98
Failed	21(2)	26(3)	10(2)	15(1)	6		13	84
Total	23	34	28	47	23	10	130	295

Table 5.2: Frequency of the events with known dam lifetime and condition. "Drained" and "Stabilized" refer to human intervention. Disappearance of the lake by aggradation is indicated as "alluviation". The counts in brackets are the events with exact damming and failure times.

Table 5.2 summarizes the lifetime and current status of the events in the dataset. In terms of time to failure, 51 out of 190 events (with known lifetime) have failed after a period of years, 100 within a year and 24 after just a few hours. Nine dams are still extant, with current lifetimes lasting from years to millennia. The "not-formed" class of landslide dams has many missing data. This was expected, as in such cases the dams disappear in a short time, making the data collection difficult. It is assumed that all these events would have a lifetime of hours.

Dam lifetime	Classification of the lifetime period, from days to millennia;
Dam length	Length of dam (m) from where the landslide has originated to as far as the dam goes towards the other side of the valley;
Dam height	Height of dam (m) from the bottom of the valley to the crest;
Dam width	Width of dam (m), indicating how far the dam goes upstream or downstream following the direction of the valley;
Dam altitude	The altitude of the dam crest above sea level (m);
Catchment area	The watershed above the dam (km^2);
River slope	Steepness (degrees) of the river bed
Dam type	Category of dam by Costa and Schuster (1988) representing the complexity of the landslide event and the dimensions and position of the dam in the valley.

Table 5.3: Glossary of the variables used.

The variables used are summarized in Table 5.3. Figure 5.2 shows the distribution of the variables by dam conditions. All variables but river slope are higher on average in the case of a formed dam, compared to the events in which dams did not form. If we look at the three dimensions of a dam (height, length, width), we can also see

these are higher in the case of piping, against overtopping. This is expected, as piping is a slower mechanism than overtopping, hence a larger dam is more likely to survive overtopping long enough to fail from piping. River slope appears to be the same on average among the lake conditions, suggesting that the formation of a lake is more related to the dimension of the dam, rather than the speed at which the water fills the valley.

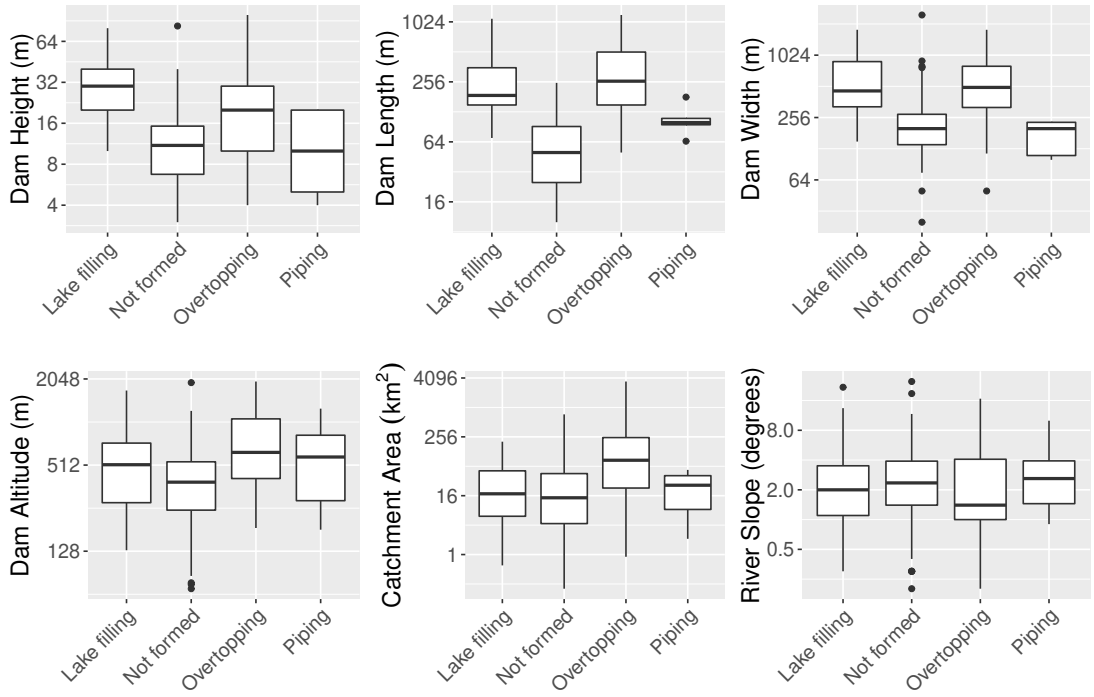


Figure 5.2: Box plots for the variables by dam conditions.

We will first examine the distribution of individual variables, and then consider their relationships. The histograms of the selected variables in Figure 5.3 exhibit highly right-skewed distributions. In order to simplify the analysis, all variables have been transformed by taking logarithms, after which all of the variables become approximately normally distributed, as can be seen in Figure 5.4. Hence from here on we will use the log-transformed variables and assume them to be normally distributed. The dam types defined by Costa and Schuster (1988) and the indices proposed by Casagli and Ermini (1999) and Ermini and Casagli (2003) highlight the importance of the relationship between variables when assessing the potential lifetime. A dam must be large enough to allow for quick water accumulation (or the river would just modify its path), high enough to not be easily overtopped by the accumulating water, and wide enough to resist the pressure of the lake that will form behind the wall. Therefore, a wider valley is less likely to be fully dammed, due to the larger volume landslide required to build

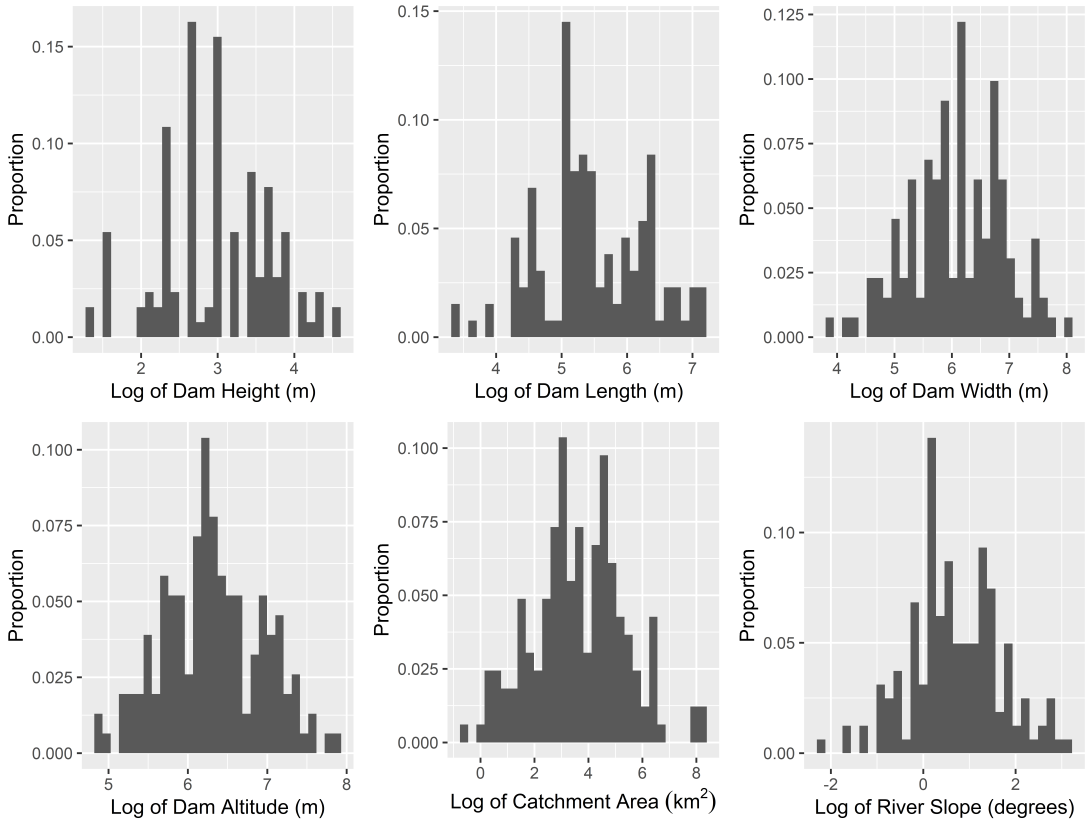


Figure 5.3: Histogram plots for the log-transformed selected variables.

a dam spanning the valley width. Narrower valleys are more prone to enduring events, as landslides are more likely to run over the entire section of the gorge. Such conditions are more common at higher altitudes, where valleys are narrower and deeper (Korup, 2004).

Table 5.4 shows the Spearman correlation coefficients between the variables. The highest correlation is observed between dam length and dam width followed by length with height, altitude and area, and height with width. Similar results have been found by Dong et al. (2009) in a smaller Japanese dataset. All the variables are significantly correlated except for altitude and catchment area, and all correlations are positive except those with river slope. Note that in such situations, correlation may be propagated between two variables that are not directly related by means of a third with which the first two are related. This will be used in the model specification.

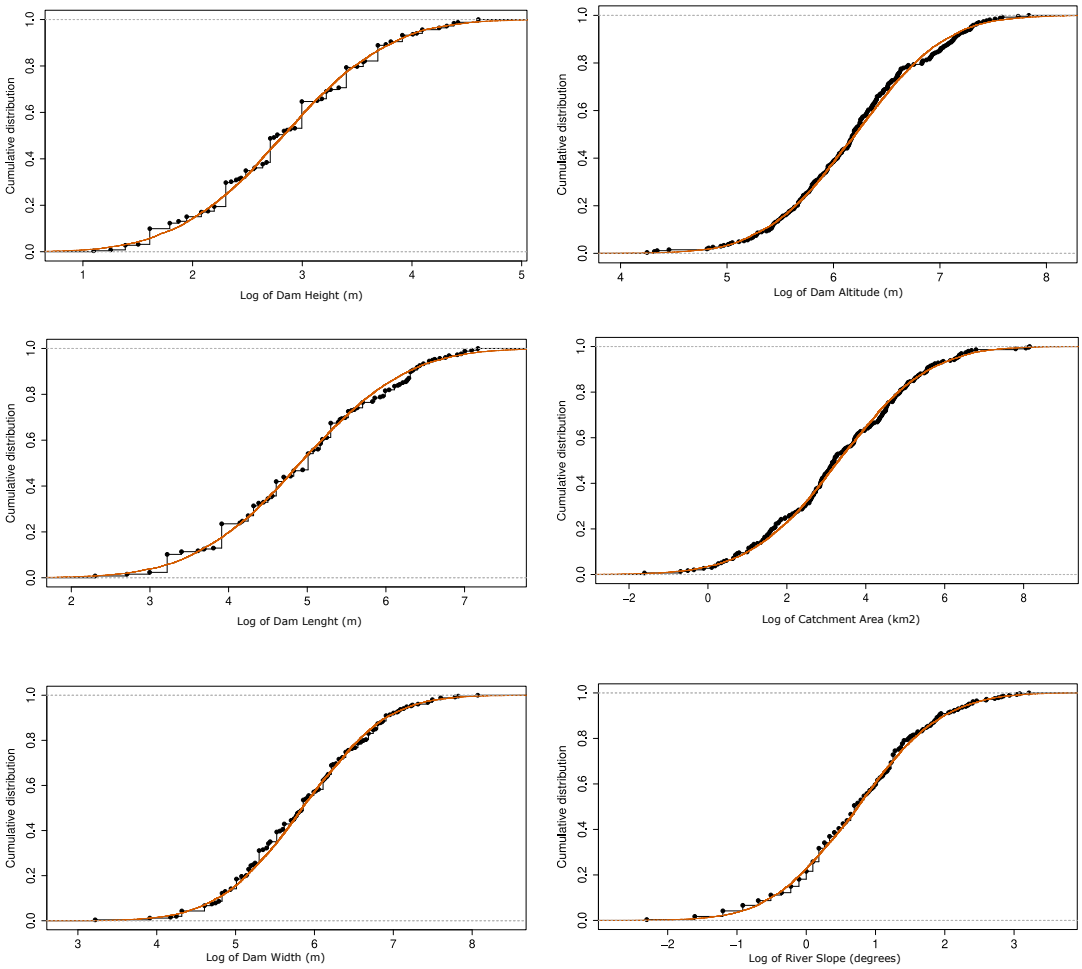


Figure 5.4: Cumulative distribution function plot of each variable (dotted line): the log-normal distribution (solid line) is a good fit.

Length	Height	Altitude	Width	Catchment area	River slope	
1	0.5421	0.4328	0.7066	0.4255	-0.2064	Length
	1	0.2343	0.5054	0.2132	-0.1658	Height
		1	0.3303	(0.0802)	0.1563	Altitude
			1	0.3893	-0.1656	Width
				1	-0.641	Catchment area
					1	River slope

Table 5.4: Spearman correlation matrix. Non-significant (at 5% level) correlations are in parentheses.

Figure 5.5 shows the distribution of the variables by dam lifetime class. Dams with greater dimensions (height, length, width) tend to survive longer. Similar separations are not observed in the other three variables.

Table 5.4 and Figure 5.5 suggest that we cannot consider only the marginal effect

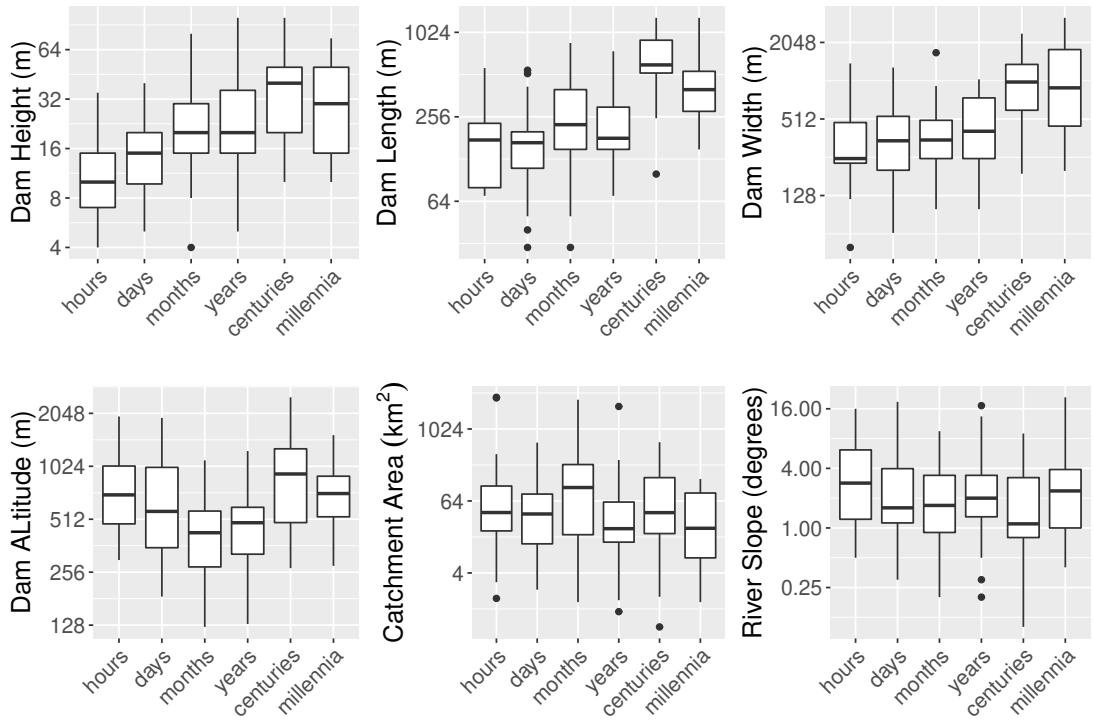


Figure 5.5: Box plots for the variables split by lifetime class

of each variable in our model; we need to consider the interaction between variables. Finding a suitable correlation structure between the variables will enable us to apply Bayesian imputation to the missing values, and hence to build a more robust model for the landslide dam failure time. Figure 5.6 shows the positive relationship between dam length and dam width, as a typical example of the positive correlation among the dam variables.

In contrast, Figure 5.7 shows the relationship between dam length and river slope, suggesting a relative lack of correlation. This occurs in every plot between river slope and the other variables, apart from catchment area. In that case we have a negative correlation, indicating that an increase in river slope implies a decrease in catchment area.

Importantly, Figures 5.6 and 5.7 contain data from events that occurred in Sicily, which appear dissimilar to the rest of the data. Sicily is mostly (86%) covered by hills and mountains (Barbera and Cullotta, 2012), and landslide dams are common. Nevertheless, the events in Sicily have, on average, smaller dam length and width, as well as catchment area, as can be seen in Figure 5.8. The fact that landslide dams in Sicily are smaller (and so more likely to fail quickly), together with the fact that these events in Figures 5.6 and 5.7 show different correlation patterns compared to the other events, suggests that we may need to consider this difference in the model formulation.

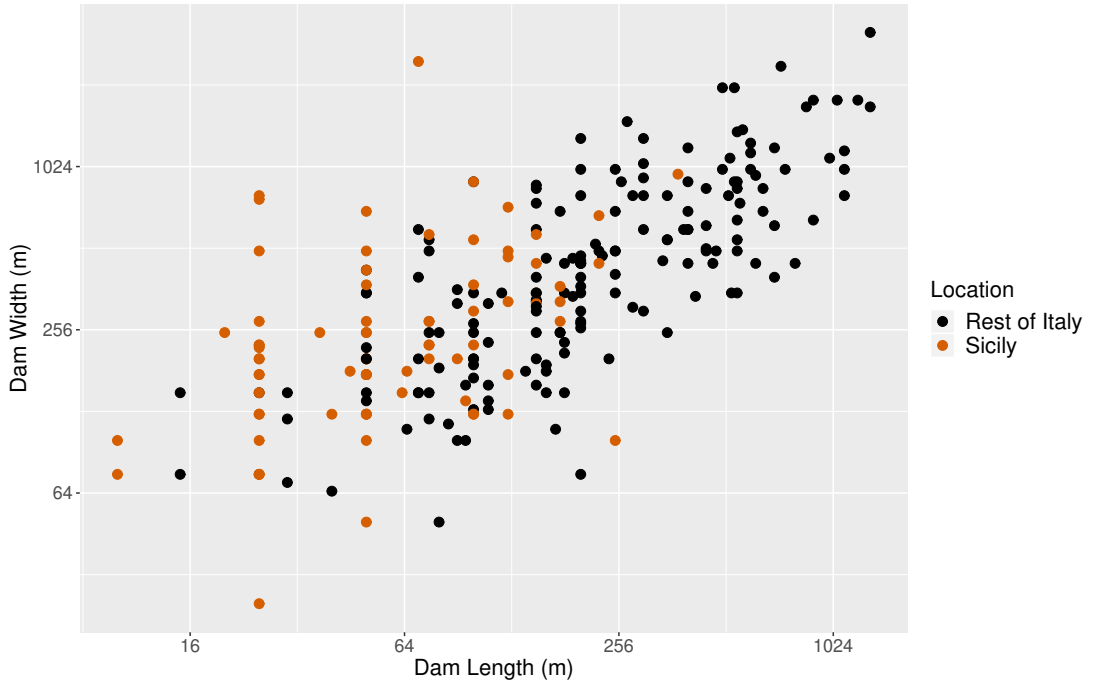


Figure 5.6: Scatter plots of Dam length against Dam width.

For the moment, these events will be retained in the dataset, with alternative models accounting for the differences presented in Section 5.7.

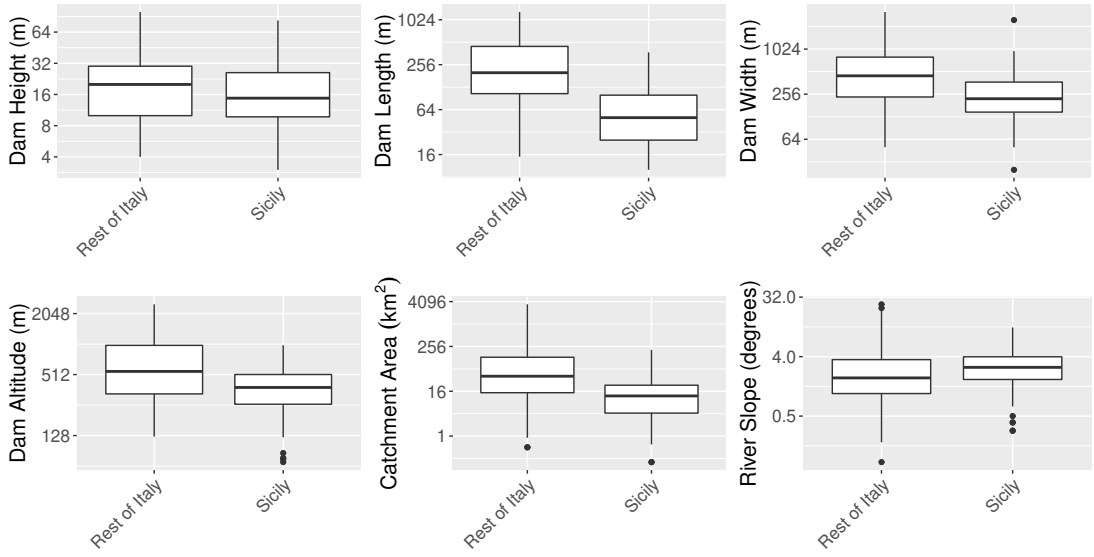


Figure 5.8: Variable distributions of events in Sicily and in the rest of Italy. The Sicilian events tend to have smaller values.

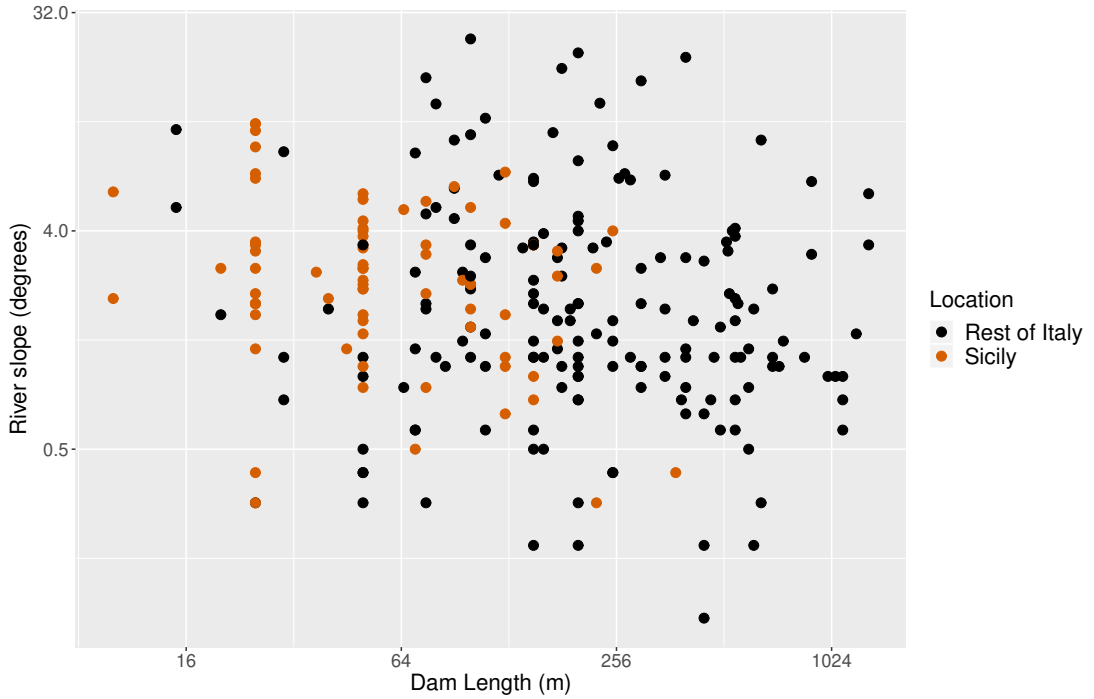


Figure 5.7: Scatter plots of River slope against Dam length.

5.4 Methods

As past studies have shown, certain specific characteristics of landslide dams can be used in creating discriminant functions to assess the stability of the dam (Casagli and Ermini, 1999; Ermini and Casagli, 2003). Other studies have estimated the probability of failure via logistic regression (Dong et al., 2011) on the dam dimensions, with good results in prediction for regionally localized datasets. However, together with the failure probability of a dam, it is crucial to know the time to failure, so that authorities can take decisions with the best information available. Survival analysis is a statistical methodology commonly used to evaluate time-to-event data in many scientific areas Chapter 2). It has been used, for example, in geological applications to evaluate the time to failure in landsliding (Federico et al., 2012; Segalini et al., 2018). Similarly to logistic regression, our survival model incorporate the variables describing the dam and the environment (discussed above) into the model as covariates, but instead of estimating the probability of failure, it estimates the survival time for each dam from its creation.

5.4.1 Covariate effects

The information obtained from the dataset can be combined into a model to predict the failure times of landslide dams. In particular, we are interested in understanding

how the covariates discussed above influence the failure time of a dam, denoted by a non-negative random variable T. For example, we might expect that a longer or higher dam is likely to survive longer than a shorter or narrower one, but to what extent is this related to the catchment area or the river slope?

A log-normal distribution is assumed for the failure times, i.e. log-transformed times are normally distributed as $\log T \sim N(\mu,\sigma^2)$, where μ is formulated as a function of the covariates to account for their effects on the expected dam failure time, and σ^2 the uncertainty (variance). These assumptions are, within the limitations of the data, consistent with them (see, e.g. , Ermini and Casagli (2003)[Figure 2]). The assumptions can be easily revisited should sufficient data be obtained to assess them. It is to be denoted the log-transformed dam length, width, height, altitude, catchment area and river slope by L_i,W_i,H_i,A_i,C_i,R_i for the i^{th} dam respectively, and each of the (log-transformed) variables is assumed to be normally distributed.

The model expresses the effect on the mean μ_i of the log-transformed survival times for the i^{th} dam as a linear combination of the covariates:

$$\mu_i = \beta_0 + \beta_H H_i + \beta_L L_i + \beta_W W_i + \beta_A A_i + \beta_C C_i + \beta_R R_i + \gamma_{D,i} \tag{5.1}$$

The parameter β_0 is the model baseline, and the other β are parameters representing the effects of the explanatory variables on the time to failure for the dams. The parameters γ_D represent the influence of the dam type D_i (Costa and Schuster, 1988) of the i^{th} event, assigning a different intercept for each category. Type II, the smallest type of dam reaching the other side of the valley, is used as a baseline, consequently γ_2 is set to zero to avoid over-parametrization. The use of Type I landslides as a baseline was explored, but those landslides do not reach the other side of the valley. As this type includes only dams that did not form or that failed within hours, this resulted in numerical instability. No events in the dataset are of dam type V, therefore we only have four of these effect terms.

5.4.2 Imputation of missing values

As mentioned above, each variable has a certain percentage of missing values, apart from dam type. In order to extract maximum value from the dataset, we will exploit the correlation structure between the variables to fill the gaps. We can do so by a method touched on in Chapter 2: Bayesian imputation (Little and Rubin, 1986), which involves the imputation of a missing value by looking at the distribution of that variable conditional on the others. The imputation is reiterated multiple times for each variable, in order to ensure the validity of the imputed values. It is important at this point to note that we have two completely different models here: the imputation model and the

survival model, previously explained. Once the first model has imputed the missing values, the second model estimates the relationships between the variables to establish which driving factors are affecting the time to failure of landslide dams.

Therefore the distribution of each variable has been modelled as conditional on others, creating a correlation structure among the variables. The relationship between the covariates is modelled using a sequence of conditional distributions in the structure illustrated in Figure 5.9, reflecting the correlation matrix in Table 5.4:

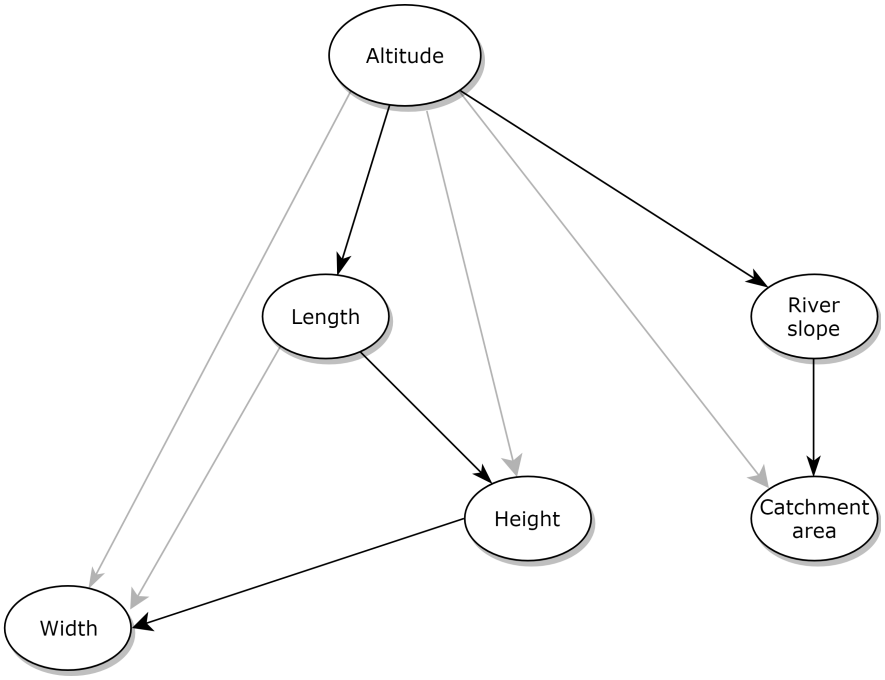


Figure 5.9: Concatenated conditional structure of the selected variables. The lines in black represents the conditional structure used. The grey lines reflects other postulated connections that were not significant in the regression models.

As can be seen in Figure 5.9, altitude is used as the starting point of a concatenation of conditional distributions. The length of a dam is given by the span of the landslide material across the valley, and the height by the amount of material accumulating on the floor of the valley. Assuming that the valley width is inversely correlated with altitude, for a certain landslide volume a dam is more likely to block the entire floor and to expand upwards in a narrow valley. On the contrary, in a wider valley, landslide volume being equal, a dam may be longer but lower. Hence, dam length is modelled conditional on altitude, and then height on altitude and length, and width on altitude, length, height and width. River slope being mildly correlated with the other variables in general, has been conditioned on altitude, as seems natural. Subsequently, catchment area is modelled conditional on river slope and altitude.

The altitude is used merely as a starting point for the conditional imputation of

missing values, taking advantage of the correlation structure among variables. There is no physical effect of the altitude on the other variables, or indeed the dam failure time. In particular, and as a check on the physical reality of the modelling, the survival model does not identify the altitude as a significant variable for the failure time of landslide dams (see Results Section)

The mean of each variable is conditional on other covariates. For example, the conditional mean of the dam height, given the dam length, is assessed by linear regression on the dam length, $\mu_{H|L} = b_1 + b_2 L$, where the b parameters are to be estimated. Frequentist linear regression using the `lm()` function (Chambers and Hastie, 1991) function in R was initially used to investigate the correlation structure, and provide starting values for the coefficients in the imputation models. As a first step, the hypothesized structure (Figure 5.9), assumes that the length is conditional on the altitude

$$L \sim N(b_L + b_L A, \sigma_L^2), \tag{5.2}$$

where σ_L^2 expresses the variability in the relationship. Subsequently, the height was modelled conditional on the length and altitude, but the latter was non-significant once the effect of height was accounted for. Hence the formulation is

$$H \sim N(a_H + b_H L, \sigma_H^2). \tag{5.3}$$

Finally, the width was modelled conditional to the height, length and altitude, with length being the only significant variable

$$W \sim N(a_W + b_W L, \sigma_W^2). \tag{5.4}$$

The river slope is conditional on altitude,

$$R \sim N(a_R + b_R L, \sigma_R^2), \tag{5.5}$$

and finally the catchment area is conditional on the river slope and altitude, both of which are significant,

$$C \sim N(a_C + b_{C,1} A + b_{C,2} R, \sigma_C^2). \tag{5.6}$$

The linear models suggest that some of the slopes become non-significant in the conditional structure (e.g. height as a predictor variable for width), because other factors (e.g. length) are co-linear, and explain the entirety of the relationship.

5.4.3 Bayesian framework and censoring

Given the amount of missing values across the entire dataset, we use a Bayesian survival model (Christensen et al., 2011), based on the conditional structure of variables described above. Ideally, the times of past events would be known, but in our dataset the exact failure time is known in only a few cases (8/295). Generally speaking, we can divide the observations up into three periods: events from Prehistory up to year 999, middle ages to nineteenth century (1000-1899), and post nineteenth century (1900 to today). For prehistoric events, we do not have a certain damming date, but we know if the dam (and the lake) still survives or not (in which case, we may know the failure date). For the middle age to nineteenth century events, we often do not have the damming date but we may have an approximate idea of the time of occurrence. For the third case, we usually have the damming date, but we are lacking the failure date for two main reasons. In the case of not-formed events, it is possible to suppose that the failure date is equal to the damming date or is a few days later, but there is no clear information about it. The second reason is due to the fact that from the twentieth century engineering applications have been put in place to control the dam. Therefore, the failure and inflow occur in an hours to days period, due to engineering solutions put in place. These cases are right-censored events, as these dams are known to have survived up to a certain time but may have failed subsequently if not for human intervention. In such cases, we do not have specific information on the time of the final extinction of the dam, only that it would have been beyond the date of stabilization or drainage.

With a Bayesian survival model we are able to work on data with limited available information, and we can use the dam lifetime information as an approximation of the survival time, incorporating it in our model to maximize the estimation power of our scheme. The Bayesian framework allows us to deal with missing data by generating them via MCMC methods from the conditional distribution of the respective variable, given the model and the other variables. This will preserve the correlation between variables by ensuring they follow the conditional structure deduced above.

5.4.4 Implementation

The model fitting was implemented in OpenBUGS, using MCMC methods to obtain samples of the posterior distributions (Lunn et al., 2012). The code is included in a GitHub repository, whose address is presented at the end of this chapter. Non-informative priors were used for all the parameters. For initial values, the b parameters (Eqs. (5.2)-(5.6)) were initialized with values estimated from their frequentist versions. The β parameters (Eq. 5.1) have been initialized as 0, expressing an initial state of no covariate effect on the failure time. The missing failure times were initialized by

apportioning specific starting values in relation to lifetime class. Parameter estimates were based on three chains of 10000 simulations each with different starting values for σ). Convergence was checked using the BGR diagnostic (Brooks and Gelman, 1998), based on which the first 1000 simulations were discarded as a burn-in period. Because of autocorrelation in the chains, iterations were thinned by a factor of 10 to obtain an approximately independent and identically distributed sample.

5.5 Results

Table 5.5 shows summary statistics for the posterior distributions of the a and b parameters.

Parameters	Mean	St.Dev	2.5%	Median	97.5%
a_L	0.839	0.573	-0.228	0.818	2.019
b_L	0.667	0.092	0.477	0.671	0.840
a_W	3.095	0.182	2.737	3.095	3.450
b_W	0.557	0.036	0.487	0.557	0.628
a_H	0.971	0.194	0.591	0.971	1.347
b_H	0.378	0.038	0.303	0.378	0.453
a_R	-0.521	0.577	-1.637	-0.524	0.589
b_R	0.201	0.093	0.022	0.201	0.380
a_C	1.250	0.813	-0.363	1.263	2.813
$b_{C,1}$	0.465	0.132	0.210	0.462	0.728
$b_{C,2}$	-1.104	0.084	-1.268	-1.104	-0.939

Table 5.5: Estimates of the b coefficients for predictor imputation.

In Bayesian analyses, one typically concentrates on the median, with the 95% 'credible interval' defined by the 2.5 and 97.5 percentiles serving to indicate whether an effect is "significant" (does not contain 0) or not. Only the intercepts b_1 and b_7, for dam length as a function of altitude and river slope as a function of altitude, are non-significant because the credible interval (2.5% to 97.5%) includes zero, indicating a possible null effect of the parameter. The point estimates (means, medians) of the slope coefficients are mostly positive, reflecting the positive correlations, but the effect of river slope on catchment area (b_{11}) is negative because a steeper slope indicates that the event has occurred at a higher altitude, where valleys tend to be narrower.

The parameter estimates of β and γ in Table 5.6 measure the effect of their respective covariates on the failure times.

Parameters	Mean	St.dev	2.5%	Median	97.5%
β_0	-3.149	5.790	-14.4	-3.192	8.149
β_1 (Height)	1.789	0.826	0.195	1.788	3.421
β_2 (Length)	5.202	0.715	3.831	5.189	6.633
β_3 (Width)	-1.179	0.953	-3.056	-1.17	0.6835
β_4 (Altitude)	0.829	0.909	-0.949	0.842	2.601
β_5 (Catchment Area)	-1.234	0.392	-2.010	-1.233	-0.4637
β_6 (River Slope)	0.535	0.526	-0.486	0.526	1.576
γ_1 (Type I Dam)	-5.979	1.406	-8.788	-5.965	-3.274
γ_3 (Type III Dam)	2.717	1.133	0.503	2.708	4.965
γ_4 (Type IV Dam)	-1.918	2.025	-5.937	-1.905	2.025
γ_5 (Type VI Dam)	1.080	1.979	-2.831	1.091	4.963

Table 5.6: Estimates of the parameters expressing the effects of the dam measurements, catchment area, river slope and type of dam on its survival time.

Each of the parameters from β_1 to β_6 in Table 5.6 expresses the marginal effect of the corresponding variable from Figure 5.9. The coefficients for the height and length are positive and significant, affirming that the there is a direct effect of these variables on the lifetime of a dam: the longer and higher the dam, the more likely it is to endure, given other covariates being equal. A longer dam is more likely to span across the entire valley width, having more chances of blocking the river completely. It also indicates that the lake may be wider, and hence fill more slowly, which as with a higher dam indicates that over-topping will take longer to occur. The credible intervals (the 2.5% to 97.5% values) for the altitude and river slope include zero, suggesting a non-significant residual effect, after accounting for the other variables, on the lifetime of the dam. Catchment area has a significant and negative correlation, suggesting an inverse effect on the lifetime of the dam. This is intuitively sensible for the catchment area, as water reaches the dam at a higher rate, making overtopping occur earlier.

With type II as a baseline, the estimated values for the γ parameters express how much the type of landslide affects the lifetime of a dam compared to an event with type II landslide. The parameter for type I (dams that do not span across the valley) is negative as expected, because this type indicates a temporary obstacle that the river can circumvent or wash away. The effect for type III is significantly positive, suggesting that when a landslide not only reaches the opposite side of the valley but also spans upstream and downstream, the resulting dam is more likely to survive longer than when the landslide only reaches the opposite side. Interestingly, the parameter for type IV dams (multiple failures from both sides of the valley) is negative and non-significant, while the one for type VI (more complex phenomena) is positive but still non-significant. These last two results may suggest that these types of events are so complex that they retain a certain intrinsic uncertainty in terms of failure time.

5.6 Results with known lifetime class only

Tables 5.7 and 5.8 show the results obtained from the 190 events with known lifetime class, therefore excluding those ones for which we have assumed a failure period of "hours".

Parameters	Mean	St.Dev	2.5%	Median	97.5%
a_L	0.819	0.572	-0.313	0.819	1.940
b_L	0.671	0.092	0.490	0.671	0.853
a_W	3.093	0.182	2.736	3.094	3.450
b_W	0.558	0.036	0.487	0.557	0.628
a_H	0.970	0.193	0.591	0.971	1.350
b_H	0.379	0.038	0.303	0.378	0.453
a_R	-0.563	0.572	-1.644	-0.575	0.581
b_R	0.207	0.092	0.024	0.209	0.382
a_C	1.240	0.804	-0.333	1.245	2.822
$b_{C,1}$	0.467	0.130	0.210	0.465	0.722
$b_{C,2}$	-1.104	0.084	-1.267	-1.103	-0.940

Table 5.7: Estimates of the b coefficients for predictor imputation.

Parameters	Mean	St.dev	2.5%	Median	97.5%
β_0	-3.303	5.666	-14.400	-3.333	7.603
β_1 (Height)	1.794	0.828	0.195	1.793	3.431
β_2 (Length)	5.225	0.714	3.856	5.211	6.655
β_3 (Width)	-1.171	0.947	-3.060	-1.161	0.657
β_4 (Altitude)	0.852	0.890	-0.879	0.858	2.599
β_5 (Catchment Area)	-1.247	0.394	-2.024	-1.243	-0.484
β_6 (River Slope)	0.532	0.527	-0.506	0.528	1.577
γ_1 (Type I Dam)	-6.014	1.388	-8.784	-6.005	-3.299
γ_3 (Type III Dam)	2.718	1.133	0.528	2.715	4.939
γ_4 (Type IV Dam)	-1.874	2.021	-5.837	-1.877	2.132
γ_5 (Type VI Dam)	1.120	1.982	-2.822	1.127	4.980

Table 5.8: Estimates of the parameters expressing the effects of the dam measurements, catchment area, river slope and type of dam on its survival time.

5.7 Sicilian data

Table 5.9 shows the results of the model outlined in the paper, but excluding the data from Sicily. We can see that the baseline β_0 has increased, suggesting that events outside Sicily last longer on average (recall that all dams in Sicily are non-formed ones). All other estimates differ little from the results when the Sicilian events are included.

	Mean	St.dev	2.5%	Median	97.5%
β_0	7.225	6.048	-4.781	7.253	19.1
β_1 **(Height)**	2.457	0.9362	0.633	2.455	4.303
β_2 **(Length)**	4.476	0.7807	2.947	4.47	6.017
β_3 **(Width)**	-0.07029	1.124	-2.272	-0.06569	2.125
β_4 **(Altitude)**	-0.5165	0.9426	-2.361	-0.5147	1.358
β_5 **(Catchment Area)**	-1.716	0.4429	-2.608	-1.706	-0.8715
β_6 **(River Slope)**	0.6104	0.551	-0.4688	0.6102	1.693
γ_1 **(Type I)**	-3.587	1.834	-7.234	-3.566	-0.02069
γ_3 **(Type III)**	1.718	1.155	-0.5077	1.715	3.987
γ_4 **(Type IV)**	-1.793	2.058	-5.843	-1.798	2.25
γ_5 **(Type VI)**	1.462	2.054	-2.594	1.474	5.453

Table 5.9: Estimates (main model) without using Sicilian data

A more elaborated model was created to consider dissimilarities among dams within and outside Sicily. This introduces some new parameters in the model that allow for separate baselines in Sicily ($\beta_{0,\text{Sicily}}$) and elsewhere, and adds additional effects for Sicily in the calculation of the conditional means for length, height and width. For example, the conditional mean for height on length becomes $\mu_{H|L} = b_1 + b_2 L + \phi_H$. The parameter ϕ_H differs from zero only in case of Sicilian events. Table 5.10 shows some interesting results. First of all, the overall intercept β_0 has become positive (as for the the results not considering Sicily), although it retains its variability and its confidence interval includes zero. The Sicilian baseline ($\beta_{0,\text{Sicily}}$) is negative and significant, suggesting a stronger tendency of failure of these events in the short term. The three different added effects for the width, length and height have different characters. That for length is negative, suggesting that these landslides are smaller compared to the non-Sicilian ones (many of these events are type I dams). That for height is small but positive, indicating that Sicilian dams are higher for their length than elsewhere, and that for the width is almost null and non-significant. In terms of dam types, type IV and VI have become similar to type II, and type III is negative but non-significant.

	Mean	St.dev	2.5%	Median	97.5%
β_0	9.619	5.866	-1.725	9.563	21.15
$\beta_{0,Sicily}$	-14.57	1.932	-18.54	-14.5	-10.97
β_1 **(Height)**	2.753	0.8692	1.056	2.751	4.483
β_2 **(Length)**	5.378	0.7277	3.985	5.366	6.838
β_3 **(Width)**	-1.184	0.9696	-3.099	-1.177	0.6996
β_4 **(Altitude)**	-0.5864	0.9129	-2.373	-0.5771	1.184
β_5 **(Catchment Area)**	-1.612	0.4011	-2.413	-1.609	-0.8414
β_6 **(River Slope)**	0.7403	0.5351	-0.3039	0.7358	1.803
γ_1 **(Type I Dam)**	-1.986	1.098	-4.155	-1.979	0.1557
γ_3 **(Type III Dam)**	-2.839	2.049	-6.864	-2.855	1.177
γ_4 **(Type IV Dam)**	-0.1886	2.019	-4.132	-0.1793	3.765
γ_5 **(Type VI Dam)**	0.02522	3.148	-6.134	0.03943	6.15
ϕ_H	0.52	0.09955	0.3241	0.5191	0.7159
ϕ_L	-1.062	0.1202	-1.299	-1.062	-0.828
ϕ_W	0.1857	0.09825	-0.00697	0.1862	0.3765

Table 5.10: Resutls with specific intercepts for Sicily.

5.8 Forecasting

As an example of the potential results we can obtain with this model, Table 5.11 shows the probability of survival for a canonical dam (type VI, $25m$ tall, $250m$ long, $500m$ wide, at an altitude of $585m$ above sea level with a catchment area of $76.9km^2$ and a river slope of $1.4°$). We can see that the probability of survival at 50 years is almost halved from that at 5 years. If we double the length or the height, then both probabilities increase. Contrarily, the survival probability decreases if we double catchment area. Doubling the width does not significantly affect the probabilities. If we switch to dam type II and IV we have higher probabilities both at 5 and 50 years, while types I and III result in inferior survival chances.

	Year	**Probability**	**St.dev**	**2.5%**	**Median**	**97.5%**
Original	5	0.433	0.219	0.065	0.422	0.861
	50	0.287	0.193	0.023	0.252	0.727
Length doubled	5	0.594	0.222	0.146	0.616	0.949
	50	0.440	0.225	0.063	0.430	0.875
Height doubled	5	0.542	0.224	0.116	0.551	0.926
	50	0.387	0.217	0.046	0.367	0.832
Width doubled	5	0.413	0.221	0.053	0.397	0.853
	50	0.271	0.192	0.018	0.232	0.716
C. Area doubled	5	0.344	0.211	0.032	0.315	0.796
	50	0.214	0.172	0.010	0.171	0.639
Type I Dam	5	0.422	0.052	0.322	0.421	0.527
	50	0.253	0.046	0.169	0.251	0.348
Type II Dam	5	0.553	0.063	0.428	0.554	0.674
	50	0.369	0.062	0.252	0.368	0.495
Type III Dam	5	0.292	0.116	0.107	0.282	0.531
	50	0.160	0.081	0.041	0.146	0.352
Type IV Dam	5	0.456	0.139	0.196	0.453	0.731
	50	0.290	0.122	0.090	0.278	0.560

Table 5.11: Table showing the probability of survival at five and fifty years for a canonical dam.

5.9 Discussion

The main message we obtain from these results is that there is a positive effect of length and height on the survival of dams, and a negative effect of catchment area. This can be summarized as $T \propto LH/C$, a ratio which is reminiscent of that (CH/V) used in the Dimensionless Blockage Index suggested by Casagli et al. (2003). Fig. 5.10 shows how the DBI index, applied to our dataset, fails to produce a clear distinction among events with different characteristics in terms of predicted median time to dam failure. Looking more closely, we see that $CH/V = 2CH/(LHW) = 2C/(LW)$), using the typical representation ($V = LWH/2$) of a dam as a triangular prism. So, in qualitative terms, our conclusion differs from that of Casagli et al. (2003) primarily in seeing the height rather than the width of the dam as important. However, our model is essentially quantitative, rather than qualitative, and provides a statistical distribution of failure time, rather than a "hard" prediction

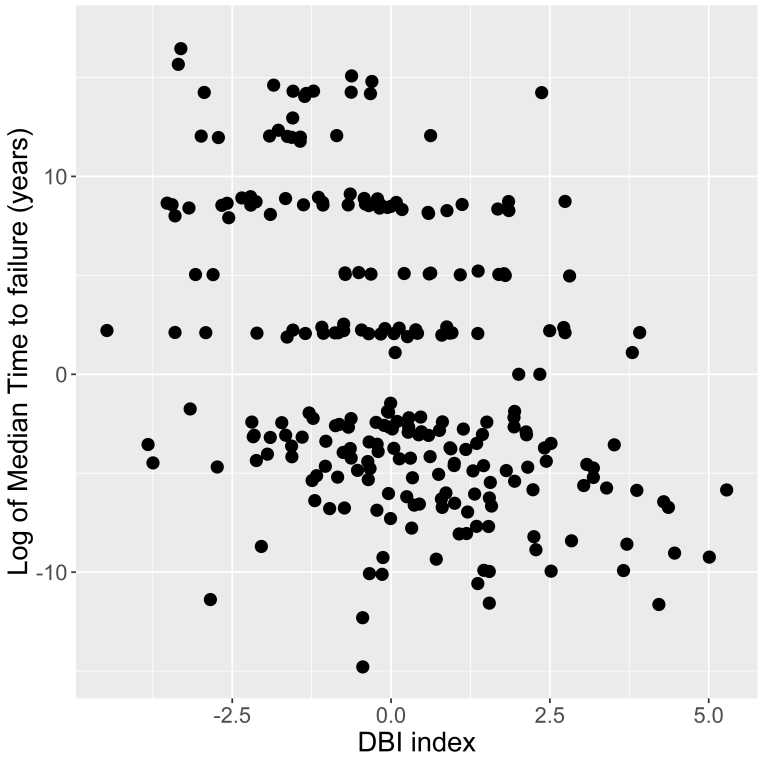


Figure 5.10: Distribution of median dam failure time against DBI index for each event in the presented dataset

Concerning the dam width, the correlation matrix presented in the data section suggests that width is correlated with height and length, therefore we would expect a positive slope in this model. Instead, the estimated value is negative and non-significant, as the credible interval includes zero.

The baseline parameter, β_0 has a high standard deviation, which suggests a high variability among the failure times of different events, regardless of the significant information obtained from other parameters such as length and height. This variability results from factors not included in our model due to lack of data, such as rainfall, material properties, local topography (spill ways, etc.).

As mentioned earlier, Sicilian events show some dissimilarities from dams in the rest of the dataset. These events typically occurred at lower altitudes, with a peculiar ratio between width and height or length. In fact, they often span upstream and downstream, but they are not tall or wide enough to create a resistant blockage for the local seasonal streams (landslides here are mostly rainfall triggered, therefore the stream flow is at its peak). If we run our model on all the events but the Sicilian ones, the parameter for width moves towards zero, confirming the non-significance of the width in the prediction of failure times.

To further consider such dissimilarities among dams' characteristics, I created an

extended version of the model, in which I introduce new additive parameters to allow for differences between Sicily and the other regions in the survival model (Eq. 5.1) and the imputation model (Eqs. (5.2)-(5.6)). The results, in Table 5.9, show that Sicilian events are effectively different, with wider but shorter (in terms of length) dams. These are mainly Type I dams, where the entire section of the side of a valley fails without fully damming the river. Because all Sicilian dams in the dataset have failed, the most predominant characteristics of these dams, width, is non-significant. The original model, that without Sicilian events, and the one modified with specific terms for the Sicilian events, show qualitatively similar results, confirming the robustness of our methodology.

These results suggest also that future work may explore the use of area-level random effects, rather than fixed intercepts, as we did for the Sicilian events. Such addition may further improve the estimation and forecasting ability of our model, towards a better understanding of events with area-related specific characteristics.

The model has the potential to be made more elaborate with more accurate and complete data. This could also include examining the most appropriate measurement for quantities such as river slope, which is currently recorded at the dam site, but obviously varies above the dam.

5.10 Conclusions

Landslide dams are a significant threat because the sudden release of water can have devastating consequences. I have presented a model that quantifies the time to failure of a newly formed dam, based on its characteristics and those of the potential reservoir. It demonstrates how a survival model for a natural hazard can be combined with a Bayesian imputation model to create a feasible tool for the analysis of time-to-event data where a large portion of the data is missing. The model can be used to probabilistically forecast the probable lifetime of a dam, and is robust, based on the results in fitting it to a heterogeneous national dataset. The dam height and length, and the catchment area above the dam are the most important variables controlling the time to failure, but the width of the dam does not significantly affect the lifetime after the length and height are accounted for. Using the Sicilian data, I have shown how the model can be generalised (maintaining its robustness) to include more specific geomorphological settings, if such information were to become available.

Data and code used can be found at the at the respective GitHub repository:
https://github.com/gfrigerioporta/Landslide-Dams.git

Chapter 6

A concept of natural hazard potentials

6.1 Definition of potential

In Chapter 3, we mentioned that there is a wide range of hazards and interactions (De Pippo et al., 2008; Gill and Malamud, 2014), which makes multi-hazard analysis particularly challenging. Despite that, hazard interactions show some patterns and similarities that might be helpful from a multi-hazard framework point of view. Figure 6.1 shows the large number of interactions that exist between a selection of natural hazards, but also how they can be grouped by the interaction mechanism. In fact, hazards can overlap in space and time, can be directly triggered, or influenced by other hazards, where the latter refers to the occurrence of a hazard increases or decreases the probability of occurrence of another hazard (Gill and Malamud, 2014).

This complexity is reflected in the hazard assessments: in Chapters 3 and 4, we mentioned that single-hazard assessments are usually based on the analysis of the susceptibility or disposition (Zimmermann et al., 1997) of a given area to a hazard. Part of susceptibility analysis is based on the study of local environmental factors, which affect the hazard likelihood but do not change much over time. For example, the characteristics of the soil can suggest how much a portion of land is prone to landslides. However, susceptibility analyses provide an image that is static in time, in relation to the characteristics of the environment that can affect the hazard occurrence, e.g. soil features for landslides. Instead, it is of great interest to assess how the interaction of factors and events can affect the process of the hazards and consequently, the probability of possible occurrence of hazards over time. The evaluation of future hazard occurrences in time is very arduous, particularly from a multi-hazard framework point of view, due to the complexity of hazard interactions and, consequently, the variety of their assessment methods.

To model the future occurrences of hazards in a multi-hazard framework, we can use the simplifying concept of "potential", defined as the ability of a hazard to trigger a secondary one (Mignan et al., 2014). In the dictionary, potential is defined as the ".. latent qualities or abilities that may be developed and lead to future success or usefulness.." and "..the possibility of something happening" (Oxford Dictionary, 2010). Therefore, we have here the opportunity to fill a knowledge gap by developing a conceptual framework to include the wide variety of hazards and their interactions into an all-embracing hazard assessment scheme. We will produce a potential framework for multi-hazard systems through a building block approach, where the different types of factors and interactions affecting the hazards can be modelled together. This will facilitate estimating the current state of a system at a certain point in time, in relation to all the inputs. Some existing models can already fit this framework, as will be demonstrated later on in the chapter, while others can be easily modified to do so. The framework will also serve as a base for simulation of natural hazard occurrences in different systems, where a low dimensional memory, or current state, facilitates multiple runs to estimate uncertainty.

In the literature, there are several examples of triggered hazard analysis (Yasuo et al., 2015; Brain et al., 2015) or susceptibility analysis (Yalcin, 2008; Monsieurs et al., 2019). However, a complete multi-hazard analysis with a quantification for both triggering mechanism and susceptibility has not yet been carried out, due to the difficulties of obtaining and working on multi-hazard triggering data, as described in previous chapters. Robinson et al. (2016b) studied coseismic landslides in New Zealand, mentioning both triggering and susceptibility aspects, but the earthquake triggering was not quantified. Ogata (1998a) and Ogata and Zhuang (2006) considered both aspects for earthquakes but not in a multi-hazard environment.

Hazards differ substantially in terms of their description and characteristics. Consequently, hazard assessments will differ as well, and one assessment may not be used for another hazard. Therefore, combining everything in one scheme is even more challenging, unless a well-structured multi-hazard method is conceived (Kappes et al., 2012a). In Chapter 4, we have presented a novel multi-hazard model in which differently assessed hazards are considered to trigger the occurrence of a secondary hazard. Nevertheless, there is an opportunity here to build a framework which would simplify multi-hazard analysis from a simulation and computational point of view. The analysis of the possible occurrence of hazards over time is also limited due to data availability. For example, earthquakes and rainfall over time have been analysed with specific models (Cowpertwait et al., 2007; Harte, 2013), while landslides are mostly assessed in term of the susceptibility of the ground, with no temporal aspect involved (Feng et al., 2016). Therefore, there is a need for a framework in which these two elements can be

combined to model the occurrence of hazards. Such a framework has to be adaptable to several types of hazards and locations and be able to measure the possibility of hazard occurrences over time.

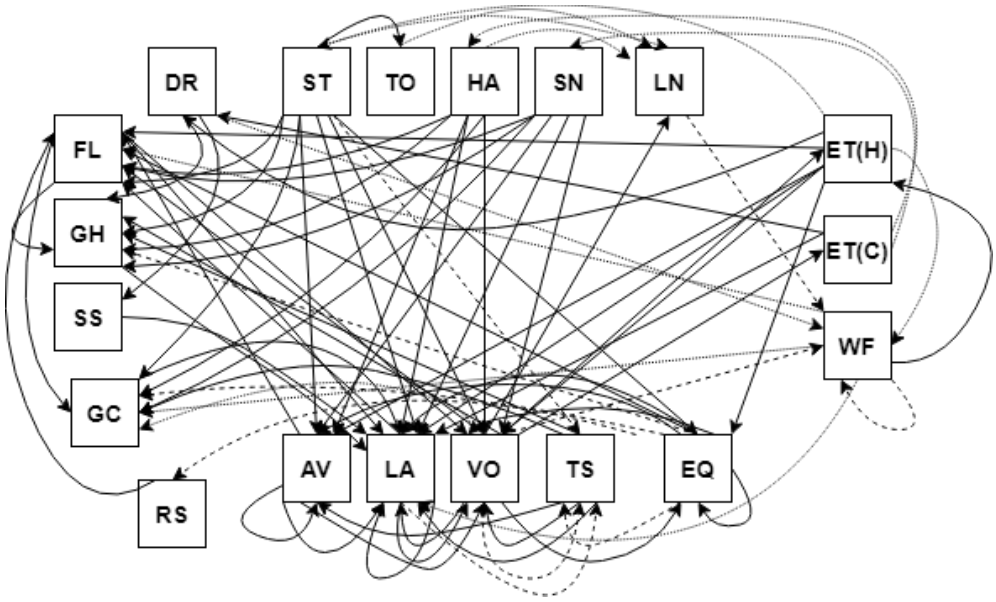


Figure 6.1: Diagram of possible interactions, according to Gill and Malamud (2014). The full arrow (——) represents a "triggering and increased provability" interaction, the dashed arrow (- - -) represents an "triggering" interaction and the dotted arrow (·····) represents a "increased probability" interaction. Adapted from Gill and Malamud (2014), Figures 2 and 4 of the paper. FL = flood, GH = ground heave, SS = soil subsidence, GC = ground collapse, RS = regional subsidence, AV = avalanche, LA = landslide, VO = volcanic eruption, TS = tsunami, EQ = earthquake, WF = wildfire, ET(C) = extreme temperature (cold), ET(H) = extreme temperature (hot),, LN = lightning, SN = snowstorm,HA = hailstorm, TO = tornado, ST = storm, DR = draught

A general concept to start with is the susceptibility, which is expressed via factors describing the environmental conditions that can increase or decrease the proneness of an area to a hazard. For example, the characteristics of the ground (type of soil, slope, water drainage) can be used to evaluate the susceptibility of an area to landslides. The fluctuation of the rainfall intensity/duration over time represents a process that can trigger the landslide, or at least that can speed up the process and lead to the occurrence. Because the susceptibility provides an instant snapshot of the environmental conditions, it can be used as a baseline of a hazard potential.

In addition to the basic idea of susceptibility, there may be a triggering mechanism that controls the hazard process. An event can be considered to be caused by the combination of disposition and triggering event (Heinimann, 1998; Kappes et al., 2010; Liu et al., 2016). Disposition is explained as "the general setting, which favours the specific process"; disposition can be further divided in basic disposition (constant factor) and variable disposition (related to seasonal or periodic changes). Hence, we can see the

basic disposition as another definition for susceptibility (and we will call it susceptibility from now on), while the variable disposition is a slow process affecting the hazard occurrence. The triggering event is the process "which leads to the threshold crossing of a factor relevant for the hazard incidence" Kappes et al. (2010). Building upon that, the suggestion by Kappes et al. (2010) is to consider the triggering mechanism as composed of two processes of different types: one reflecting a slow, long-term change and the second one a faster, short-term one. The slow process can be considered as slowly modifying the susceptibility of an area to the hazard. Conversely, the fast process is related to the occurrence of other hazards that may produce a sudden increase or decrease in the likelihood of a hazard.

Rainfall and/or seismically induced landslides can be used as an example to understand better the concepts expressed above. The occurrence of a landslide depends firstly on geomorphological characteristics (factors) of a specific area. The geomorphological factors can be considered the starting point of a landslide occurrence because the characteristics of the soil (e.g. porosity) determine the susceptibility of the area to landslides. Then, rainfall and earthquakes provide inputs that vary in intensity and duration over time. An earthquake produces an instantaneous effect (fast process) on the likelihood of the landslide, but it can also produce a long term effect on the potential of future events. In fact, strong earthquakes can weaken the soil and therefore increase the probability of future landslides (Cui et al., 2011), particularly in the case of subsequent rainfall events (Towhata et al., 2013). The rainfall is usually considered a long-term process that slowly (Polemio and Sdao, 1999) increases the likelihood of landslides (slow process), as the accumulation of water in the soil, slowly modified the stability of the slope. Nevertheless, short-term heavy rainstorms (fast process) can be intense enough to speed up the process by quickly saturating the soil moisture (Muntohar and Liao, 2010; Martha et al., 2014). In fact, the level of moisture changes the rainfall thresholds necessary for triggering a landslide (Baum and Godt, 2010; Ponziani et al., 2012). In particular, the presence of the short-term rainfall component in the model allows for an acceleration of the mechanism towards landslide occurrences (He et al., 2020). Hence, we need a framework in which models are able not only to accommodate direct triggering but also to include the concept of "increased probability" and the complexity of interactions between hazards.

If we combine the idea of Kappes et al. (2010) with those of Gill and Malamud (2014) and Mignan et al. (2014), we can build a conceptual framework for the occurrence of hazards, taking into consideration their interactions, triggering and disposition, under the single concept of "potential" of a hazard. The potential of a hazard should be a function of the hazard history, producing an outcome that summarizes the combined effect of the disposition and triggering of the hazard over time. The analysis of the

variation of potential over time becomes then a critical tool to evaluate possible future hazard occurrences. It is possible to construct a function affected by the slow and fast processes at difference paces over time. If we consider once again Gill and Malamud (2014), we can combine their types of interactions with the concepts expressed so far. In particular, the occurrence of certain hazards can increase or decrease the probability of the occurrence of other hazards. Hence, the probability of a hazard can go upward or downward depending on other hazards. As their relationship is driven by the hazard magnitude and frequency, consequently the increased/decreased probability is affected by these features as well. Going back to the example of rainfall: a long-term event slowly increases the probability of a landslide. If the soil moisture has not become saturated when the period of rainfall ends, then the likelihood of landslide will start dropping. In other words, a hazard occurrence depends initially on the susceptibility of the area (disposition), long-term changes (slow processes) and the occurrence and interaction of other hazards in the short term (fast processes). The combined analysis of these aspects constitute the hazard potential and will provide an estimate of the current state of the system. Therefore, this chapter will explore the concept of potential of the hazard and the possibility to quantitatively build such a multi-hazard framework.

We can assume that the potential of a hazard is dependent on time, as the factors that alter the process of hazard occurrence can change over time, e.g. the weakening of the soil due to long term rainfall.

Hazard occurrences are usually modelled as history-dependent (see Chapter 2). In a multi-hazard setting, the natural extension is to be multi-history dependent. The history of the process thus becomes high-dimensional. In order to fully exploit the framework, we want the potential to be low-dimensional, ideally a scalar, so that it simplifies the approach to multi-hazard assessment without overly compromising the inherent sophistication of combined natural hazards.

Definition. The potential of a specific hazard $U(t)$ is a **scalar**-valued function of the entire history of all the triggering factors in a multi-hazard system which control the intensity of the occurrence of this specific hazard. It increases or decreases in response to internal and external events.

Formally, we define $U(t)$ as a piece-wise continuous scalar function dependent on time, as we are interested in how the occurrences of point events and the modification of internal and external factors can alter the potential of a hazard over time. As a piece-wise function, $U(t)$ can jump in response to external point events. As a scalar, it summarizes the hazard in t as an overall effect of all processes affecting the system up to time t. Mathematically, we suppose $\lambda(t|H_t) = \lambda(t|U(t))$, which expresses that the potential synthesizes the history up to time t into one scalar value, $U(t)$ so that $U(t) : H_t \to R$. This allows us to evaluate the conditional intensity of a hazard at t

with the advantage of summarizing the history-dependence of the process into $\lambda(t|H_t)$ with just a scalar value $U(t)$. The intensity of the given hazard, λ, dependent on the time t and the history up to t, see equation (2.1), will then be defined as a function of the potential and consequently of time:

$$\lambda(t|H_t) = g(t,U(t);\theta) \tag{6.1}$$

This has important consequences, namely that the intensity can only depend on the potential and on any "external" process (i.e. not modelled through $U(\cdot)$) occurring at time t. We will explain this theme below. The function $g(t,U(t);\theta)$ should be a monotonic increasing function of $U(t)$, defined as $g : (0,\infty) \times (-\infty,\infty) \to [0,\infty)$ with θ as a vector of parameters. In particular, $g(\cdot)$ acts as a link function between the potential function and the conditional intensity function of the hazard. In some cases, $g(\cdot)$ can be restricted to a non-negative domain $(0,\infty)\times(0,\infty)$: this would happen if the potential is composed of elements that are strictly non-negative, meaning that there is only an exciting effect of the history on the conditional intensity and no inhibition effect (represented by negative values).

A simple candidate for $g(t,U(t);\theta)$ is the exponential function, a monotonically increasing function:

$$g(t,U(t);\theta) = \exp[\alpha + \beta U(t)], \text{ with } \alpha \in R, \beta > 0 \tag{6.2}$$

with $\theta = (\alpha,\beta)$. Another example is provided by the Heaviside function $H(\cdot)$, which allows $g(\cdot)$ to be written as a stepwise function.

$$g(t,U(t);\theta) = H(U(t)) \tag{6.3}$$

or, more generally:

$$g(t,U(t);\theta) = \alpha U(t)^{\beta} H(U(t)) \tag{6.4}$$

Because the occurrences of hazards over time modify the potential of a hazard, we are interested not only in the current state of the system at time t but also in the change of state over time. Hence, let us define $U(t+\Delta t) - U(t)$, the change in potential from t to $t + \Delta t$, with Δt small, as in (6.5).

$$U(t + \Delta t) = U(t) \quad + \quad \psi U(t)\Delta t \quad + \quad \sum_{j=1}^{J} \rho_j \Delta t \quad + \sum_{i:t_i \in [t,t+\Delta t)} \phi_i(t_i) \tag{6.5}$$

where $\rho_j \in (-\infty,\infty)$ for $j = 1,...,J$ are locally linear input/outputs (slow process), that can be called factors and $\phi_i(t_i)$ are the effects of jump events i (fast process) that

occur between t and $t + \Delta t$. The factor ψ allows for proportional decay (or inflation) in the potential. The third and fourth components are the exogenous inputs that can increase and decrease the potential. Hence, if ρ_j and ϕ_i are positive, then inhibition is not present and so $U(t) > 0$, where only excitation is possible.

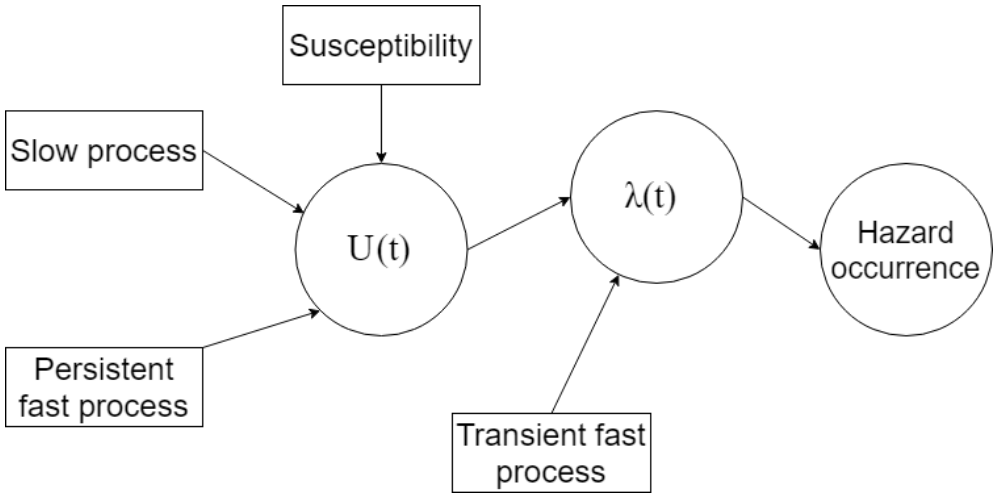


Figure 6.2: Example of a potential framework. The susceptibility is the baseline for the potential function $U(t)$, which is affected by slow and persistent fast processes and affect the likelihood of the hazard. Transient fast processes can affect the likelihood of the hazard. A direct triggering of the hazard occurrence is visible as a spike in the conditional intensity $\lambda(t)$, as it will be explained in Figure 6.3

The change in potential over time depends on different phenomena. As it is a function of the history, in our definition $U(t + \Delta t)$ depends both on $U(t)$ and the change in potential over $t + \Delta t$. Then, hazards occurring in the system may interact and add to this change with their specific processes. The slow process of input j has an effect on the potential equal to $\rho_j \Delta t$ in the time interval. Similarly, a persistent fast process i (which is a fast process endogenous to the system or, alternatively, events that can be modelled with point processes) affects the change in potential in the interval (see Figure 6.2) through the sum of $\phi_i(\cdot)$. These events are also referred to as "immigrants" (Hawkes and Oakes, 1974), as we will explain later with Figure 6.3. For example, if we split the rainfall process into two, following what we have achieved in Chapter 4, we can designate the long-term rainfall component as a slow process and the short-term rainfall component as a persistent fast process. This second process would be formulated as $\phi_i(t_i)$ in (6.5).

Another category of process able to produce a change is the one of transient processes. A transient process is represented by an exogenous hazard temporarily affecting a system (Pavel et al., 2018). These transient processes do not affect the history, and hence the potential, of the process, but directly the conditional intensity $\lambda(t)$ under

certain conditions, as shown in Figure 6.3. For example, an earthquake, considered external to a system, may trigger a landslide if the epicentre is close and the magnitude is large enough (see Chapter 4).

These occurrences enter the process at time t and produce a shock (the vertical spike in Figure 6.3) on the conditional intensity of the hazard $\lambda(t)$, which produces consequently an effect on the potential function for a certain period of time. An example is given in the Figures 6.3b and 6.3c, where an earthquake instantaneously increases the potential function $U(t)$ but then affects it for longer as $U(t)$ does not return to the value before the earthquake occurrence. Transient processes can be modelled with a separate point process or components, for example, in the fashion of the third component of (6.5) (the sum of ϕ_i), added to $U(t)$. (6.1) can be rewritten as (6.6), where μ represents a transient process, not defined as part of $U(t)$.

$$\lambda(t|H_t) = g(t,U(t);\theta) + \mu \mathbf{I}_{[s,s+\Delta s]}(t) \tag{6.6}$$

In the case of Figure 6.3b, the potential function after the immigrant earthquake remains higher than the previous level with $U(t + \Delta t) = U(t) + \alpha\mu$, where $\alpha > 0$ (Cui et al., 2011). However, different sequences of earthquakes (different combinations of frequency/magnitude) lead to different effects on landslide stability in the post-seismic period (Brain et al., 2017). Hence in Figure 6.3c, the effect is to reduce the potential level compared to the level before the earthquake: $U(t + \Delta t) = U(t) - \alpha\mu$. This can be linked to the "healing" effect: in other words, the reduction of susceptibility or likelihood of hazard over time (Marc et al., 2015).

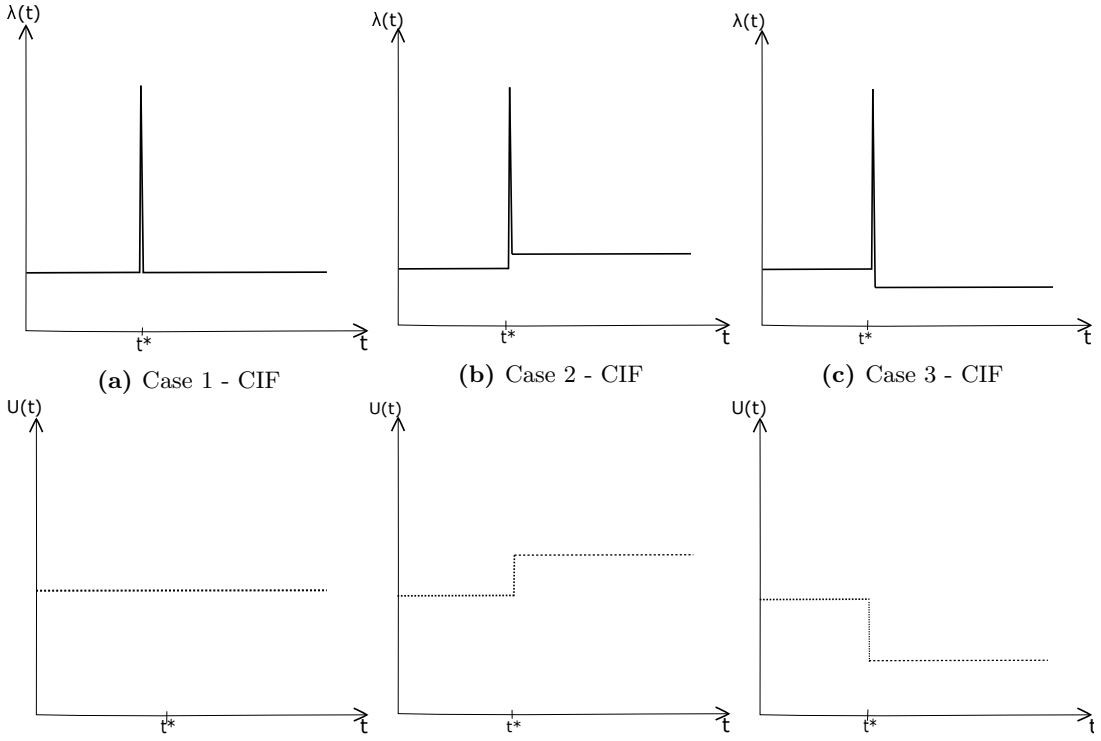
6.2 Potential for a single hazard

As a first step, we are going to investigate some examples of a potential function for a single hazard, hence without considering any interaction between triggering events.

6.2.1 Hawkes process

Self-exciting point processes (Ogata, 1999; Helmstetter and Sornette, 2002), reviewed in Chapter 2, are good candidates for the potential representation of a hazard. The main characteristic of these processes is that they allow for the modelling of events over time, with a focus on clustering events (Richter, 1958; Bak et al., 2002; Gu et al., 2013).

One of the earliest self-exciting processes is the Hawkes process (Hawkes, 1971). The Hawkes process can be used in the potential framework in the form without the



(a) Case 1 - CIF **(b)** Case 2 - CIF **(c)** Case 3 - CIF

(d) Case 1 - Potential function **(e)** Case 2 - Potential function **(f)** Case 3 - Potential function

Figure 6.3: Examples of three different earthquakes affecting the change on a landslide conditional intensity function (CIF) and potential function between t and $t + \Delta t$, as per (6.5). In all three cases, the earthquake increases instantaneously in time the conditional intensity, up to a value which can cause consequences (e.g. landslides). In case 6.3a, the potential is not affected, as after the event it is equal to the pre-earthquake level. In case 6.3b, the earthquake produces the effect of keeping the potential higher than before the seismic event for a while (weakening). In case 6.3c, the earthquakes lower the potential to less than the value pre-earthquake (healing), which suggests an inhibition of following consequences.

immigration term μ. However, we can retain μ provided that it is not a function of t:

$$\lambda(t) = \mu + \eta \sum_{i:t_i<t} \exp[\nu(t_i - t)] \tag{6.7}$$

where μ,η,ν are positive constants, and events occur at times $\{t_i\}$. In (6.7) μ is a parameter representing the background rate of the process, while the self-excitation is expressed by the sum of the exponential of $\nu(t_i - t)$ in relation to the time elapsed, where t_i are the time points prior to t. Hence, every event (say an earthquake) occurring at time t_i before t increases the likelihood of further events, by self-excitation. Figure 2.2 in Chapter 2 shows an example of Hawkes process: the occurrence of an event increases the conditional intensity of the process. Hence, further events are more likely to occur, further increasing the intensity.

For a self exciting process such as the Hawkes process, $U(t)$ needs to be strictly

positive. This will allow us to retain the self-excitation mechanism, as we want the potential function to react in the case of multiple events occurring within a short time period. As we have seen in (6.2), it is possible to use an exponential function for $g(\cdot)$, to link the conditional intensity function of the Hawkes process to its potential function. Hence, we can generalize (6.2) to provide a $g(t,U(t);\theta)$ function:

$$g(t,U(t);\theta) = \alpha + \beta U(t), \text{ with } \alpha \geq 0, \beta > 0, \tag{6.8}$$

with $\theta = (\alpha,\beta) = (\mu,\eta)$. The structure of the function mimics the one in (6.7), as desired. As mentioned earlier, we are interested in understanding the change in potential from t to $t + \Delta t$, so below we attempt to derive a form of (6.5) for (6.7). Hence, we start from the change from t to $t+\Delta t$ in the conditional intensity function, taking into consideration that in such time interval we have to consider the possible occurrence of further events, expressed by the second sum in the equation below:

$$\begin{aligned}
\lambda(t+\Delta t) &= \mu + \quad \eta \sum_{i:t_i<t} \exp[\nu(t_i - t - \Delta t)] \quad + \quad \eta \sum_{t\leq s_i<t+\Delta t} \exp[\nu(s_i - t - \Delta t)] \\
&= \mu + \quad \eta\exp(-\nu\Delta t) \sum_{i:t_i<t} \exp[\nu(t_i - t)] \quad + \quad \eta \sum_{t\leq s_i<t+\Delta t} \exp[\nu(s_i - t - \Delta t)] \\
&= \mu + \quad \eta\exp(-\nu\Delta t)U(t) \quad + \quad \eta \sum_{t\leq s_i<t+\Delta t} \exp[\nu(s_i - t - \Delta t)] \\
&= \mu + \quad \eta U(t+\Delta t)
\end{aligned} \tag{6.9}$$

where

$$\begin{aligned}
U(t+\Delta t) &= \exp(-\nu\Delta t)U(t) + \sum_{t\leq s_i<t+\Delta t} \exp[\nu(s_i - t - \Delta t)] \\
&= \psi + \sum_{i:t_i\in[t,t+\Delta t)} \phi_i(t_i)
\end{aligned} \tag{6.10}$$

Therefore, the change in potential from t to $t+\Delta t$ is given by an exponential effect $\exp(-\nu\Delta t)$ multiplied by the value of the potential function in t, plus the sum of the effects obtained from new events. Recalling (6.5), the first component can be seen as a constant ψ, providing the proportional decay and the sum of the effects from jump events ϕ_i. Therefore, it is proved that the Hawkes process, as defined in (6.7), can be expressed in terms of a scalar potential function (6.5).

6.2.2 Epidemic type aftershock-sequences model

Another example of self-exciting process is provided by the Epidemic Type Aftershock-Sequences (ETAS) processes Ogata (1988), a form of Hawkes process (Hawkes, 1971),

based on the idea that the intensity can be considered as the sum of "background" earthquakes caused by tectonic loading and "triggered" earthquakes. Furthermore, the ETAS models incorporate Omori's decaying frequency law for aftershocks (Omori, 1894), in the Utsu and Ogata (1995) modified version $\lambda(t) = k/(t - t_i - c)^{-p}$, and the productivity law (Utsu, 1969) $N = K\exp[\alpha(M_m - M_c)]$, by which the number of aftershocks is an exponential function of the mainshock magnitude. The temporal conditional intensity of this model is:

$$\lambda(t) = \mu + \sum_{i:t_i<t} \frac{k}{(t - t_i - c)^p} \exp[\alpha(m_i - m_c)] \tag{6.11}$$

where μ is the background rate (immigrants), and the second part expresses the self-exciting effect.

We want to know if we can derive (6.11) in the form of (6.5), to find the potential function for an ETAS model and to evaluate its change over time. Substituting (6.11) in (6.5) we get (6.12):

$$\lambda(t + \Delta t) = \mu + k \sum_{i:t_i<t} \frac{\exp[\alpha(m_i - m_c)]}{(t + \Delta t - t_i - c)^p} \frac{(t - t_i - c)^p}{(t - t_i - c)^p} + k \sum_{j:s_j \in [t,t+\Delta t)} \frac{\exp[\alpha(m_i - m_c)]}{(t + \Delta t - s_i - c)^p} \tag{6.12}$$

which is not a recursive function anymore.

The Hawkes process allowed for this procedure thanks to the exponential function used in the conditional intensity equation, which allowed for the separation of the Δt effect from the rest of the function. Here, instead, the power function does not allow for the extraction of the Δt effect without compromising the function, because we would need to know all the t_i times to calculate the effect coming from each i event.

This model cannot be written using (6.5). In fact, even assuming that the ETAS can be written without μ, it would not be possible to separate the Δt effect from the rest of the function. Nevertheless, there are alternative methods to ETAS that are able to characterise the mechanism mainshock-aftershock mimicking the modified Omori's law and avoiding the issue we have found in the ETAS model. For example, Borovkov and Bebbington (2003) proposed a two-nodes model for the stress release and production of aftershocks, which might be a candidate to represent ETAS-like models in a potential framework. A hierarchical Hawkes process was introduced by Wang et al. (2012) to model the earthquake cycle, including aftershock regimes.

6.2.3 Stress release model

Another example of $U(t)$ is provided by the stress release model (SRM) (Vere-Jones, 1978), mentioned in Chapter 3. The SRM is a process that describes how the tectonic

stress level of a system increases over time and is released in the case of earthquakes.

$$U(t) = X(t) = X(0) + \rho t - S(t) \tag{6.13}$$

where $X(t)$ is the Benioff stress of the system at time t, $X(0)$ is the initial level of stress, ρ is the constant loading rate from external tectonic forces (which makes the stress increase linearly) and $S(t)$ is the accumulated stress release from earthquakes within the region in (t): $S(t) = \sum_{i:t_i<t} S_i$. Evidently, a large release of stress (S_i large) can make $X(t)$ become very small, which means that any subsequent event within this recharge time will have a very low likelihood of occurrence.

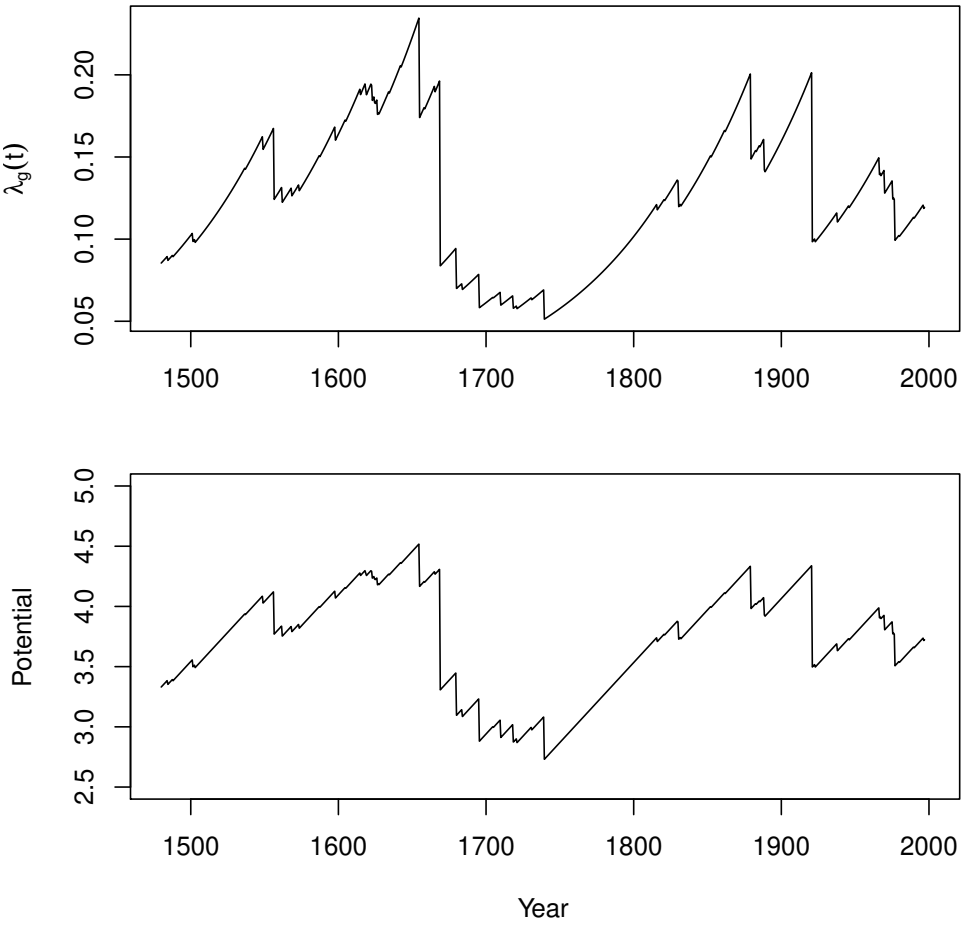


Figure 6.4: Example of conditional intensity and potential functions for stress release model. The drops in the function correspond to event occurrences. The potential function has been obtained from (6.15). Data from Harte (2010)

In a short time interval $[t,t+\Delta t)$, the function becomes:

$$\begin{aligned} U(t+\Delta t) &= U(t) + \rho\Delta t - [S(t+\Delta t) - S(t)] \\ &= U(t) + \rho\Delta t - \sum_{t_j \in [t,t+\Delta t)} S_j \end{aligned} \tag{6.14}$$

It is possible to link $\rho\Delta t$ in (6.14) to the effect by factors $\rho_j(\Delta t)$ in (6.5), as well as $S(t+\Delta t) - S(t) = \sum_{j:t_j \in (t,t+\Delta t]} S_j$ as the effect of events $\phi_i(t_i)$. Notably, here $\phi_i < 0$, which is the earlier mentioned case of inhibition. We can use (6.1) to write the conditional intensity corresponding to (6.13) using an exponential:

$$\lambda(t|H_t) = \exp[\alpha + \beta U(t)] = \exp[\alpha + \beta X(0) + \beta\rho t - \beta S(t)] = \exp[a + bt - cS(t)] \tag{6.15}$$

where the last step in (6.15) shows an alternative parametrization used for numerical optimization (Harte, 1999), where $\alpha + \beta X(0) = a$, $\beta\rho = b$ and $\beta = c$. The parameter a can then be interpreted as the initial value of stress of the system, while b represents the characteristics of the crust (Borovkov and Bebbington, 2003)

6.3 Linking simple systems together - potential in a multi-triggering environment

As mentioned earlier and in Chapter 3, hazards occur and interact in natural systems, therefore the models have to reflect such complexity. Triggering events do not produce the same level of effects in every system (e.g. different locations equate to different geological, morphological and climatological settings) and the occurrence of an event in one place may affect the potential of a hazard in another space or time differently. Furthermore, there might be interaction among the processes (slow, persistent fast or transient fast) behind each of the hazards we take into consideration. Figure 6.5 is a schematic representation of hazard interactions in a system: notice that external and internal processes affect $U_1(t)$ differently, as the internal process influences the recursive behaviour of the potential function (e.g. as in the SRM, where the system loads stress until the discharge). The dashed line depicts the boundary of one concept of a hazard system, where the internal process and the first hazard are included.

In Chapter 4, we demonstrated that it is possible to produce a quantitative model to evaluate the multi-triggering effect of the interaction between two primary hazards (earthquakes and rainfall) on the occurrences on secondary hazards. Each of the primary hazards, indexed with $p = 1,...,n$, can be described by a process with a conditional intensity, based on the history of that process. Applying (6.1), each history is expressed in terms of the potential function for that hazard:

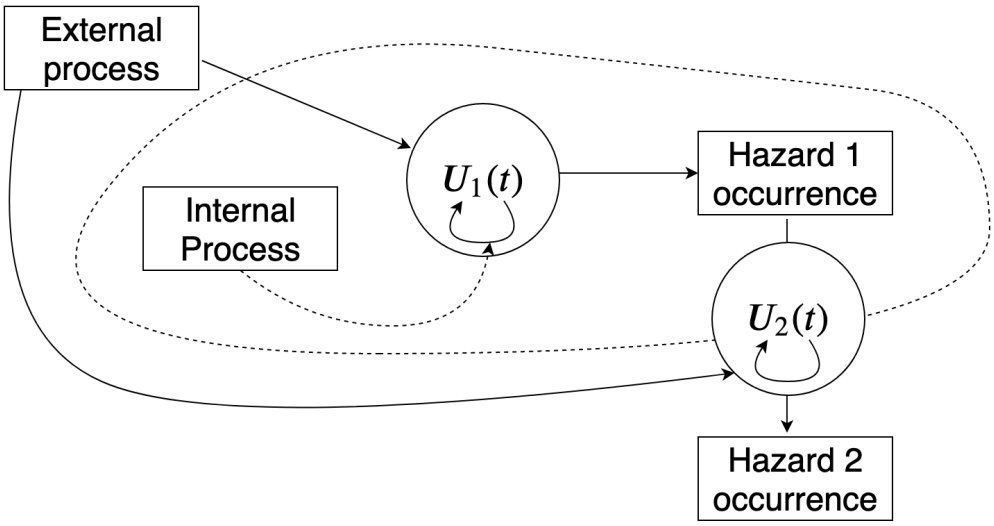


Figure 6.5: Example of interaction between hazards in a system. An external/internal process can affect the potential of a hazard (represented by the function $U(t)$) and consequently provoke a secondary hazard occurrence. The latter may then trigger a third hazard. The dashed line encompasses the hazard system.

$$\lambda_p(t|H_t^{(p)}) = \lambda_p(t|U^{(p)}(t)) \tag{6.16}$$

Secondary hazards can be triggered by multiple triggering events, which may interact. Therefore, we can consider that the conditional intensity of a secondary hazard is the result of the combination of the history of the secondary hazard and the histories of possible primary triggering hazards:

$$\lambda_s(t|H_t^{(s)},H_t^{(p_1)},...,H_t^{(p_n)}) = \lambda_s(t|U^{(s)}(t)) \tag{6.17}$$

where the potential function of the secondary hazard, following (6.1), can be expressed as function of the history of the secondary hazard and the history of all the primary hazards that can trigger it:

$$U^{(s)}(t) = h(H_t^{(s)},H_t^{(p_1)},...,H_t^{(p_n)}) \tag{6.18}$$

As seen earlier with ETAS (Section 6.2.2), there are many models that do not fit into the potential framework as they are. In particular, they do not allow for the simplification of the history to a scalar value and therefore do not enable a clear separation of the temporal effect on the hazard potential, as in (6.5). Nevertheless, models can be modified to meet the requirements of the potential framework, retaining their characteristics in terms of triggering mechanism. An example is the two-nodes model by Borovkov and Bebbington (2003). This issue becomes even more important

when combining models together. Here below, we present two examples of linked models to show how it is possible to build a link between simple models using the potential concept.

6.3.1 Earthquake/rainfall triggered landslides

As discussed in Chapter 3, landslides are usually triggered by rainfall (Berti et al., 2012; Aristizábal et al., 2015; Peruccacci et al., 2017) or seismic activity (Lee, 2014a; Havenith et al., 2016; Robinson et al., 2016a). Rainfall-triggered landslides occur based on the groundwater saturation, which is linked to the intensity/duration of the precipitation. Coseismic landslides are caused by a combination of the proximity to the epicentre and the amount of energy released.

Rainfall is characterised by fluctuation in terms of duration and intensity: there can be long-term events with limited intensity, as well as short-term heavy periods of rain. The amount of groundwater, which is the water present in the soil, affects the stability of slopes. Hence, the amount of rainfall water necessary to trigger a landslide depends on the amount of groundwater, and the rainfall intensity/duration is then crucial (Iverson, 2000).

Rodriguez-Iturbe et al. (1984), Rodriguez-Iturbe et al. (1987), Onof et al. (2000), Isham et al. (2005), Cowpertwait et al. (2007) and many others have used point processes to model rainfall over time. The atomic component of these processes is a rain cell, with random duration and depth. A rain cell produces pulses (which provide the intensity of the rain), and the overlapping of pulses constitutes a storm. Modelling the duration and the intensity of each rain cell with a specific point process results in the possibility of simulating different types of rainfall events, in terms of time and intensity. These models can be included in the potential framework if it is possible to find a function that adequately represents the history of the process as a scalar, as per (6.1). In some cases, they might be too complicated to be fit, for example, due to auto-correlation (Kaczmarska et al., 2014), but we can still include them in the framework as exogenous processes. Hence the very important ability to simulate them easily is allowed for. Several papers have proposed water runoff thresholds (Glade et al., 2000; Guzzetti et al., 2007; Rossi et al., 2017) to evaluate the required rainfall process to trigger landslides. Rossi et al. (2010) also highlighted the connection between landsliding and a preceding short rainfall period.

However, there is another aspect to be considered: the runoff thresholds might change over time (He et al., 2020) in relation to the amount of rain filter through the ground. Long-term and short-term rainfall, as well as the intensity of it, affect this process, as the natural drainage of the soil might take more time and so make the ground unstable for longer. This suggests that a temporal model for rainfall-triggered

landslides need to be sophisticated enough to consider all these features. The model we have proposed in Chapter 4 is capable of taking into consideration all aspects involved in the process.

Coseismic landslides are primarily linked to the superficial ground movements generated by earthquakes. The energy released produces instability on slopes (USGS, 2019b). The two key factors to consider are the distance from the epicentre and the intensity of the energy released (Kritikos et al., 2015; Parker et al., 2015; Havenith et al., 2016; Parker et al., 2017). Both factors are closely linked together: the larger the earthquake, the larger the area of landsliding (Keefer, 1984; Gorum et al., 2011), all other things being equal. Similarly to rainfall, there are many earthquake simulators using complex numerical simulations (Vere-Jones, 1978; Ogata, 1998a; Field et al., 2014), of varying degrees of complexity (and hence memory) that can be used to include earthquakes as exogenous processes. As mentioned earlier in Figure 6.3, the occurrence of an earthquake produces a spike in the potential function of a landslide. Such spikes represent an instantaneous release of energy which translates into an instantaneous increase in the likelihood of landslides, although the slope instability may persist, leading to post-seismic events (Robinson et al., 2016b; Parker et al., 2015; Kritikos et al., 2015; Marc et al., 2015). The post-seismic decay in potential is usually fast and can be described with a power-law (Travasarou et al., 2003) or an exponential (Meunier et al., 2007). The difference in the temporal distribution of the two hazards suggests that rainfall and earthquakes are quite different hazards: their simultaneous presence in a landslide model needs to be properly addressed.

In Chapter 4, the coseismic component was built using the relationship between main event magnitude and aftershock productivity (Utsu, 1970; Ogata, 1988; Wetzler et al., 2016) as a proxy of ground-shaking. The component includes both a ground-shaking measure and the distance from the epicentre. Their combination allows the component to mimic the sharp increases seen in Figure 6.3, in correspondence with strong earthquakes (Brain et al., 2015).

Provided that the two triggering mechanisms are differently modelled but both important in the landsliding process, the need for a statistical model that incorporates both hazards (Kappes et al., 2012b) is then substantial. Before our attempt (Chapter 4), some authors have tried to combine rainfall and earthquakes when assessing landslide risk (Vega and Hidalgo, 2016; Nguyen and Kim, 2019), attempting to establish a connection between seismicity and rainfall after a major earthquake (Shou et al., 2011; Fan et al., 2019). The study of the rainfall-induced landslides after the 1999 Chi-Chi (Taiwan) earthquake (Lin et al., 2006; Shou et al., 2011) has shown that the earthquake reduced the geomorphological stability of the area and two consequent rainfall events

(typhoons) in 2000 and 2001 have triggered more events than the earthquake itself (which caused over 9000 landslides) (Lin et al., 2006).

Shou et al. (2011) combined the empirical model by Uchiogi (1971) with an exponential decay to incorporate the effect of time from the Chi-Chi earthquake. This empirical model estimates the future landslide rate of a watershed (where another event has occurred) as a function of a future rainfall event. The landslide rate $Y = C_a/a = K(R - R_0)^a$ is defined as the landslide area (C_a) over watershed area (a) and assumed to be a function of rainfall, where K is a location-specific parameter, R the measure of one-day rainfall and R_0 is the critical rainfall level for landslides. To incorporate the temporal effect of the previous Chi-chi earthquake, this is multiplied by an exponential decay term $b_0 + b_1 e^{-bt}$, with $b_0 = b_1 = 1$ for the study area according to the author. Assuming $b > 0$, the landslide rate is then:

$$Y(t) = Y_0\left(1 + e^{-bt}\right) = K(R - R_0)^a\left(1 + e^{-bt}\right) \tag{6.19}$$

where Y_0 is the baseline, $(R - R_0)^a$ a function of rain and $\left(1 + e^{-bt}\right) \in (1,2)$. The function is not well parametrized for our needs: while in Chapter 4 the rainfall component is endogenous, here it is treated as an exogenous variable, as there can be no complex memory in the potential. Furthermore, the earthquake component should be linked at least to the magnitude of the earthquake and, in this example, the earthquake dependence is not general. Nevertheless, (6.19) can be seen as the multiplicative effect of a rainfall component and an earthquake component. We have explored this possibility in Chapter 4, with specific components for the Italian region of Emilia-Romagna.

However, if we modify slightly (6.19) to allow for updating, making R a function of time more clearly and generalize the magnitude in the earthquake component (taking inspiration from the productivity law), the equation can be seen as a conditional intensity. Hence, we have a function that has a rainfall function over time $K(R(t) - R_0)^a$, and the sum of all earthquake effects up to t considering the magnitude effect $\exp[\alpha(m_i - m_c)]$ and the time decay $\exp[-b(t - t_i)]$:

$$\lambda(t|H_t) = K(R(t) - R_0)^a\left\{1 + \sum_{i:t_i<t} \exp[\alpha(m_i - m_c)]\exp[-b(t - t_i)]\right\} \tag{6.20}$$

It is to be noticed that the resulting exponential term is a Hawkes process as per (6.7). We can update the potential to $(t + \Delta t)$ following (6.5), separating the effect of the seismic component over $(t,t + \Delta t)$. Also, we will use (6.8) as the $g(\cdot)$ function to link

the conditional intensity to the potential function.

$$\begin{aligned}
\lambda(t+\Delta t|H_t) &= K(R(t+\Delta t)-R_0)^a\Bigg\{1+\sum_{i:t_i<t}\exp[\alpha(m_i-m_c)]\exp[-b(t+\Delta t-t_i)] \\
&\quad+\sum_{t\le s_j<t+\Delta t}\exp[\alpha(m_j-m_c)]\exp[-b(t-s_j-\Delta t)]\Bigg\} \\
&= K(R(t+\Delta t)-R_0)^a\Bigg\{1+\exp[-b\Delta t]\sum_{i:t_i<t}\exp[\alpha(m_i-m_c)]\exp[-b(t-t_i)] \\
&\quad+\sum_{t\le s_j<t+\Delta t}\exp[\alpha(m_j-m_c)]\exp[-b(t-s_j-\Delta t)]\Bigg\} \\
&= K(R(t+\Delta t)-R_0)^a \\
&\quad\times\Bigg\{1+\exp[-b\Delta t]U(t)+\sum_{t\le s_j<t+\Delta t}\exp[\alpha(m_j-m_c)]\exp[-b(t-s_j-\Delta t)]\Bigg\} \\
&= K(R(t+\Delta t)-R_0)^a\{1+U(t+\Delta t)\}
\end{aligned} \tag{6.21}$$

where

$$U(t+\Delta t)=\exp[-b\Delta t]U(t)+\sum_{t\le s_j<t+\Delta t}\exp[\alpha(m_j-m_c)]\exp[-b(t-s_j-\Delta t)] \tag{6.22}$$

With Δt small we can approximate $\exp(-b\Delta t)\approx 1-b\Delta t$, which allows us to obtain $U(t+\Delta t)=U(t)-b\Delta tU(t)+\sum_{t\le s_j<t+\Delta t}\{\exp[\alpha(m_j-m_c)]\exp[-b(t-s_j-\Delta t)]\}$. This last result can be linked to (6.5) to obtain the following:

$$U(t+\Delta t)=\psi U(t)\Delta t+\sum_{i:t_i\in[t,t+\Delta t)}\phi_i(t_i) \tag{6.23}$$

where $\psi=-b$.

As mentioned earlier, the model by Shou et al. (2011) was not sufficiently parametrized to support multiple earthquake and rainfall triggering effects over time on landslide occurrences, and our proposed update (6.20) still includes the rainfall component as endogenous. The literature proposes limited examples on earthquake/rainfall triggered landslides from a temporal point of view (see Chapter 4), in Chapter 4 we have proposed a point process specifically built to evaluate the effects of both rainfall and earthquakes on landslide occurrences, and that can be modified to be included in the potential framework. Compared to (6.21), this model is more complex in terms of interactions of triggering hazards. In the Chapter 4 case, the earthquake component is the exogenous one as it has no memory within the system, while the rainfall is endogenous, as it

contributes in the long-term to the triggering, and so needs to form the potential. To better address the characteristics of rainfall, the phenomenon has been split into two separate components, in order to consider both the long-term effect of seasonal rainfall (slow process) and the high-intensity-short-term events (fast process). Furthermore, all the three components are functions of time and can be easily adapted to be recursive and so be used in the potential framework. The conditional intensity function from Chapter 4 is:

$$\lambda(x,t) = \lambda_0(x)\exp[\lambda_1 C_{RS}(x,t) + \lambda_2 C_{RL}(x,t) + \lambda_3 C_E(x,t)] \tag{6.24}$$

One obvious complication arises in formulating this model as a potential. As expressed in Chapter 4, the model works on discrete time, to reflect the naturally discretized data of landslides and the available rainfall data. Therefore, the potential function can be only linked to (6.5), only if we discretize the latter. Hence, we can re-write (6.5) as:

$$U(t+1) = U(t) \quad + \quad \psi U(t) \quad + \quad \sum_{j=1}^{J} \rho_j \quad + \sum_{i=1}^{I} \phi_i(t+1) \tag{6.25}$$

where all the decay/inflation, linear and exogenous inputs are considered only at time $t+1$.

We now need to consider whether this obeys the updating rule (6.25), and is thus within our potential framework. We will first examine each of the components in turn.

While the earthquake component is considered exogenous, the rainfall components can be adapted to (6.25), by writing (6.24) as $\lambda(t|H(t)) = g(t,U(t)) = g_1(t,U(t)) \times g_2(C_E)$, notationally suppressing the location x for simplicity.

The short-term rainfall component is the arithmetic mean of the last two days of rainfall.

$$C_{RS}(x,t) = \frac{P(x,t-1) + P(x,t)}{2} \tag{6.26}$$

While this is reasonably simple, it does require keeping two days of rain "in memory". Instead, we can define $C_S(x,t)$ as the linear combination of itself at $t-1$ and the rainfall value at t. Let

$$C_S(x,t) = 0.5[C_{RS}(x,t-1) + P(x,t)] \times I_{(0,\infty)}[P(x,t)] \tag{6.27}$$

This way, (6.27) is a recursive function of short term rainfall, where $C_S(x,t-1)$ is updated by taking the average between its value in $t-1$ and the amount of precipitation in t. The average is multiplied by an indicator function which makes $C_S(x,t) = 0$ if there is no precipitation in t. This way, the history of the component can be evaluated as required to be part of this framework, and the indicator function keeps the component

a short-term one, as per the original (6.26).

The long-term rainfall component is

$$C_{RL}(x,t) = \frac{1}{\Delta}\sum_{\delta=1}^{\Delta}\omega^{\delta-1}P(x,t-\delta-1) \tag{6.28}$$

where $\Delta = 150$. Hence 150 days of rainfall need to be kept in memory. Instead, we can use an alternative definition:

$$C_L(x,t) = \omega C_{RL}(x,t-1) + P(x,t) \tag{6.29}$$

which makes $C_L(x,t)$ an exponential smoother, and hence an updatable scalar. The new long-term component has become a true exponential smoother, expressing the change in the long-term rainfall trend. It is to be noted that these two new components work for "perfect" data, i.e. when every aspect of the data is known, and their accuracy is ideal.

The new model is subject to the same data quality issues as the old, except that additional approximation might be needed to initialize (6.29) in the case of missing rainfall data. Hence, we can rewrite (6.24) as follow:

$$\lambda(x,t) = \lambda_0(x)\exp[\lambda_1 C_S(x,t) + \lambda_2 C_L(x,t) + \lambda_3 C_E(x,t)] \tag{6.30}$$

Therefore, we can suppose that the link function is the generalization of (6.2):

$$g(t,U(t);\theta) = \alpha\exp[\beta U(t)], \text{ with } \alpha > 0,\beta > 0 \tag{6.31}$$

Thus, we can get the potential function for (6.24) as follow:

$$\alpha\exp[\beta U(t)] = \lambda_0(x)\exp[\lambda_1 C_S(x,t) + \lambda_2 C_L(x,t) + \lambda_3 C_E(x,t)] \tag{6.32}$$

We can define $\alpha = \lambda_0\exp[\lambda_3 C_E(x,t)]$ (as the earthquake component is exogenous), so that the equation becomes:

$$\exp[\beta U(t)] = \exp[\lambda_1 C_S(x,t) + \lambda_2 C_L(x,t)] \tag{6.33}$$

Applying a log transformation we obtain:

$$\beta U(t) = \lambda_1 C_S(x,t) + \lambda_2 C_L(x,t) \tag{6.34}$$

And dividing both terms by β we get:

$$U(t) = b_1 C_S(x,t) + b_2 C_L(x,t) \tag{6.35}$$

where $b_1 = \lambda_1/\beta$ and $b_2 = \lambda_2/\beta$. Substituting (6.27) and (6.29) into (6.35) we have:

$$\begin{aligned} U(t) = b_1 \frac{1}{2}[C_{RS}(x,t-1) + P(x,t)] \times I_{(0,\infty)}[P(x,t)] \\ + b_2[\omega C_{RL}(x,t-1) + P(x,t)] \end{aligned} \tag{6.36}$$

Grouping the elements of (6.36) by time of calculation t and $t+1$, we get

$$\begin{aligned} U(t) = \left\{ b_1 \frac{C_{RS}(x,t-1)}{2} \times I_{(0,\infty)}[P(x,t)] + b_2 \omega C_{RL}(x,t-1) \right\} \\ + \left\{ b_1 \frac{1}{2} P(x,t) + b_2 P(x,t) \right\} \end{aligned} \tag{6.37}$$

It is to be noticed that the indicator function in (6.36) is not required in the last term anymore, as rainfall is non-negative, and there is no contribution here if $p(x,t)$ is zero. We can now identify $U(t-1)$ as the first term in (6.37), hence

$$\begin{aligned} U(t) &= U(t-1) + \left\{ \frac{b_1}{2} P(x,t) + b_2 P(x,t) \right\} \\ &= U(t-1) + \left\{ \frac{b_1 + 2b_2}{2} P(x,t) \right\} \end{aligned} \tag{6.38}$$

Therefore, we can shift it by one temporal lag so that (6.38) can be compared to (6.25):

$$U(t+1) = U(t) + \left\{ \frac{b_1 + 2b_2}{2} P(x,t+1) \right\} = U(t) + bP(x,t+1) \tag{6.39}$$

where $b = (b_1 + 2b_2)/2$ for convenience and $P(x,t+1)$ provides a rainfall input to the potential in $t+1$, which contributes to both the rainfall components above presented. Therefore, recalling (6.25), we can write:

$$U(t+1) = U(t) \quad + \quad \phi(t+1) \tag{6.40}$$

In this model, the potential is the rainfall and the earthquakes exogenous while in (6.20) we have the opposite situation. Note also that this new model with (6.27) and (6.29) is a different model than the one proposed in Chapter 4 and it would need to be re-fitted, with possibly different conclusions. Nevertheless, it will still be able to embrace all the concepts expressed earlier on in this chapter, from direct triggering to increased probability and complex interaction among hazards over time. Hence, it might mimic the above mentioned soil moisture concept, in the case of data paucity, which is not uncommon with natural hazards. Nevertheless, as shown in Chapter 4, the large amount of missing data is a major characteristic of this type of analysis. The components proposed in (6.27) and (6.29) may need further work to be applied to missing data.

6.3.2 Linked stress release model

A limitation of the SRM is that (6.13) evaluates the accumulation and release of stress within the same region. Nevertheless, the occurrence of an earthquake in one region may affect a nearby region due to the propagation of seismic waves (considering space/time decay). Since this paper aims to introduce a framework for hazard potentials, as an example of a linked system, we will look into the Linked Stress release model (LSRM) proposed by Lu et al. (1999) and further investigated by Bebbington and Harte (2003). This modified version of the Stress release model takes into account the stress transfer between adjacent regions by using a link between systems. If (6.13) is valid for any given region, (6.41) represents the stress accumulation/release for region i:

$$U_i(t) = U_i(0) + \rho_i t - \sum_j \theta_{ij} S^{(j)}(t) \tag{6.41}$$

where $S^{(j)}(t)$ is the accumulated stress release in region j up to time t, and the coefficient θ_{ij} measures the fixed proportion of stress transferred from region j to region i.

Assuming that the conditional intensity for each region is exponential (as per (6.1)) and that each region has a different set of parameters $\theta_i = (\alpha_i,\beta_i)$, the conditional intensity function for region i, following (6.2) and (6.15):

$$\begin{aligned} \lambda_i(t) = \exp\{\alpha_i + \beta_i U_i(t)\} &= \exp\left\{ \alpha_i + \beta_i \left[\rho_i t - \sum_j \theta_{ij} S^{(j)}(t) \right] \right\} \\ &= \exp\left\{ a_i + b_i t - \sum_j c_{ij} S^{(j)}(t) \right\} \end{aligned} \tag{6.42}$$

6.3.3 Landslides leading to landslide dams

Landslide dams occur when a landslide falls into a river. Depending on the dam's and valley's characteristics, the dam may survive a period of time that spans from minutes (Dong et al., 2009; Korup, 2005) to centuries or more (Tacconi Stefanelli et al., 2016). It is possible to summarize the process of a landslide dam failure in three pivotal steps that may or may not lead to a dam failure: the occurrence of a landslide, the formation of the dam and the possible failure of the dam (Costa and Schuster, 1988). If a landslide has fallen on a river, it may or may not form a dam. The morphology of the landslide (in terms of material, layout and size) establishes if the dam may be washed away within a very short time period, or if it consolidates, allowing for water accumulation (Massey et al., 2013; Harrison et al., 2015). In the second case, there is a chance for flash flooding in the case of dam failure (see Chapter 5). Finally, the third phase is the

failure of the dam. This mechanism depends on the morphology of the dam (McKillop and Clague, 2007) and the water-flow (Dong et al., 2009), responsible for the water accumulation. The more cubic meters of waters there are behind the landslide dam, the higher is the risk of flooding. Therefore, we can consider these three phases one at a time, to build a complete conditional intensity function for this complex process.

A key concept is that there is a link between the size of a landslide and the possible formation of a landslide dam (Tacconi Stefanelli et al., 2020). The characteristics of a landslide dam are inherited, at least partially, from the landslide features and location. These features need to reflect the characteristics of the phenomenon so that it is possible to establish whether the landslide will fail or not, and how long it will take to be overtopped or washed away. Therefore, we could use the dam dimensions and the valley size (Costa and Schuster, 1988; Casagli and Ermini, 1999; Tacconi Stefanelli et al., 2016). The dam dimensions are directly related to the characteristics of the landslide originating the dam. Hence, the larger the landslide, the more likely it is that there is sufficient material to build a high or large dam blocking the river, all other things being equal. In addition to the dam dimensions, the catchment area (CA) at the landslide location is also important to evaluate the rate at which water accumulates behind the dam. Thus, it provides a measure of how much pressure of water the dam might need to sustain over time.

The first phase is the occurrence of a landslide on a river. We have already seen a conditional intensity for a landslide occurrence (6.30), as a function of the triggering processes produced by rainfall and earthquakes. This is the mechanism that initiates the landslide dam chain (e.g. an earthquake triggering a landslide that falls into a river). To model this phenomenon within our framework and particularly within the landslide dam chain, an ideal candidate is the marked point process. As explained in Chapter 2, the MPP combines a ground intensity function and a mark distribution (Harte, 2010). In this case, a possible conditional intensity for a landslide occurrence that might produce a landslide dam is given by:

$$\lambda_{LS}(t,H,L,W|H(t)) = g_{LS}(t,U_{LS}(t);\theta) \times \psi(m_{LS}|U_{LS}(t),x) I_{(L,W,H)}(f(m_{LS},x)) \quad (6.43)$$

The ground intensity function $g_{LS}(t,U_{LS}(t);\theta)$ is the conditional intensity describing the occurrence of a landslide. This function can be expressed as (e.g.) in (6.30). The $\psi(m_{LS}|U(t),x)$ function is the probability density for the mark m_{LS}, characteristics of the landslide size given the location x. For our analysis, we can reduce the concept of marks to solely the landslide volume $V_{LS}(x)$ for a landslide at location x, although other aspects could possibly be included. We have conceptualized this function as conditional on the potential, in order to allow for the possibility of a link between landslide occurrence probability and landslide size. The mark density is multiplied by

an indicator function $I_{(L,W,H)}(f(m_{LS},x)$ expressing that the three dimensions L,H and W will have to be the particular values that a landslide of volume m_{LS} would make in that location; also, the function $f(\cdot)$ maps the landslide volume and location into the the dimensions of the landslide dam in that location.

The landslide volume will be affected by the magnitude of the triggering event. We wish to stress, once again, that earthquakes are treated as exogenous events in Chapter 4 but as endogenous in (6.20). The earthquake magnitude M has been linked to the total landslide volume with $\log(V_{LS})(x) \sim aM - b$ (Gutenberg-Richter law, Keefer (1994); Malamud et al. (2004)), and the frequency of landslides per given volume is assumed to follow a negative power law $f(V_{LS}) \sim V_{LS}^{-c}$. A similar relationship has been observed between the number of landslides and earthquake magnitude (Keefer, 2002; Li et al., 2011; Alvioli et al., 2014). In terms of rainfall, it is not clear if there is a similar effect between rainfall magnitude and landslide volume. The rainfall-triggered landslides phenomenon is quite complex, as the size of landslides might be related to a combination of intensity and duration (hence the potential), with some papers suggesting that these events are more likely with longer rainfall periods (Chen et al., 2017). Rainfall-landslides related studies have suggested that the Gutenberg-Richter law applies to landslides frequency/volume distributions (Whitehouse and Griffiths, 1983; Dai and Lee, 2001; Gao et al., 2018), and a power-law would suit the frequency-area distribution (Hovius et al., 1997). Hence we can leave open the possibility of there being a similar dependence on the triggering process in the rainfall case, which we will assume can be expressed through the potential function. Therefore, assuming for simplicity $m_{LS} = V_{LS}(x)$, we can define the landslide marks as a power-law function, and the cumulative distribution would be:

$$\Psi(m_{LS}|U_{LS}(t),x) = 1 - k[U_{LS}V_{LS}(x)]^b \tag{6.44}$$

where $b < 0$ is the power-law exponent and k is a scaling factor which needs to be $k = U_{LS}^{-b}$ in order to allow (6.44) to integrate to 1, where $V_{LS} > 0$. The mark density is then the derivative of the cumulative distribution:

$$\psi(m_{LS}|U_{LS}(t),x) = \frac{d}{dV_{LS}(X)}\Psi(m_{LS}|U_{LS}(t),x) = -bkU(t)[U(t)V_{LS}(x)]^{b-1} \tag{6.45}$$

Alternatively, if $U(t) > 0$ for all t, as for (6.8), we can have $U(t)$ as part of the exponent. In this case, $k = 1$:

$$\Psi(m_{LS}|U_{LS}(t),x) = 1 - k[V_{LS}(x)]^{bU(t)} \tag{6.46}$$

The mark density is then:

$$\psi(m_{LS}|U_{LS}(t),x) = \frac{d}{dV_{LS}(X)}\Psi(m_{LS}|U_{LS}(t),x) = -kbU(t)V_{LS}(x)^{bU(t)-1} \tag{6.47}$$

Assuming that the landslide has fallen on a river, we move to the second phase, the formation of a landslide dam. Hence, we need a probability that the landslide will form a landslide dam, i.e. that the potential landslide dam actually blocks the river.

The dimensions of the dam are used to assess the stability of a dam via indices (Ermini and Casagli, 2003). Although these indices might be useful to distinguish landslide in terms of failure/non-failure, they are based on dams failed at a given time. Instead, we are interested in the failure at the time of dam formation. However, the dams types, as defined by Costa and Schuster (1988), offer greater insights. A type I event is defined as non-formed landslide dam, as the landslide does not reach the other side of the valley, which is when the length of the dam is not large enough compared to the width of the valley, so the landslide dam will not form as it is not able to retain water. Hence, a landslide dam that is not a type I is a dam that has formed. Types II to VI represent landslides that reach the other side of the valley and are differentiated by how much material is accumulated, e.g. depositing material higher up in the valley, or multiple landslides joined together. So the key question is whether the landslide extends across the entire valley. Hence, we can use the length of the landslide dam and the valley width as the crucial variable to assess the probability of formation. If the length is larger than the valley width, then the landslide dam formation may occur:

$$P(LSD|H,L,W,x) = \begin{cases} 1 & \text{if } L > W_V(x) + \epsilon(H,W) \\ 0 & \text{if } L < W_V(x) + \epsilon(H,W) \end{cases} \tag{6.48}$$

where $W_V(x)$ is the valley width at location x and ϵ is a positive, but decreasing, function of H and W, which represents whether the "tip" of the landslide seals off the far side of the valley.

The final phase is the failure of the dam. The dam failure is related to the dimension of the dam, but also to the accumulation of water behind it. Water inflow was not considered in Chapter 5, due to data availability, but it becomes necessary in this framework, as we want to model the dam failure over time with a point process. The water inflow can be expressed as a function of time, affecting the failure of the dam, as it can provide an estimate of how long it will take for the water to reach the crest of the dam, which is typically the point at which the dam will fail through overtopping if it is not sufficiently consolidated. It is important to remember that the dam failure can occur in two different ways, by overtopping and by seepage (Awal et al., 2007). In the case of seepage, the water filters through the landslide dam from the upstream face to

the downstream face; this is a fast failure, as the dam is weakened by water saturation until the collapse. In the case of overtopping, it takes more time for the dam to fail, as the water needs to accumulate. Once the water reaches the top of the dam, it erodes the material from the downstream face until the dam collapses.

We need to consider how the two types of failures can be modelled separately, before combining them into one conditional intensity for the failure of the landslide dam. One aspect that is in common between the two failure types is that they are both functions of the water inflow. Consequently, if $r(t)$ is the inflow rate at time t, we can define $R(t)$ as the cumulative inflow since the dam formation: $R(t) = \int_0^t r(u)du$. In particular, we can use the catchment area CA and write $r(t) \approx CA \times r_0(t)$, where $r_0(t) =$ rainfall rate across the catchment. This approach is advantageous as there are several existing point process models that can be used for rainfall (Rodriguez-Iturbe et al., 1987; Cowpertwait et al., 2007; Kaczmarska et al., 2014). The cumulative inflow rate is important, as it provides an insight of the time to failure. In fact, we can model the failure by seepage as a negative exponential of the cumulative inflow rate. Figure 6.6 shows schematically the two possible functions of these failures of a dam. The red line highlights the case of seepage that we shall call $\lambda_{D,S}(\cdot)$: because this is the fastest of the two failures: the function is at its maximum after the dam formation, and then it falls sharply. Therefore, we can suppose:

$$\lambda_{D,S}(t,R(t);\theta) = a_0 \exp\{b[V_{l,max} - R(t)]\} \tag{6.49}$$

where a_0 and b are parameters. Using the updated link function presented in (6.31), the potential function in (6.49) is thus equal to the cumulative inflow: $U_{D,S}(t) = V_{l,max} - R(t)$. The blue line in Figure 6.6 shows how the second type of failure can be modelled. The conditional intensity for overtopping, that we shall call $\lambda_{D,OV}(\cdot)$, can be regarded as a Gaussian-like function, increasing as the cumulative inflow (orange line) approaches the maximum lake volume ($R(t) = V_{l,max}$). Afterwards, the risk of failure reduces, suggesting a stabilization of the dam. Hence we can write:

$$\lambda_{D,OV}(t,R(t);\theta) = a_1 \exp\left[-\frac{(R(t) - V_{l,max})^2}{2c^2}\right] \tag{6.50}$$

where a_1 and c are parameters. In this case, the potential function is equal to the square difference between cumulative inflow and maximum water volume in the lake: $U_{D,OV}(t) = (R(t) - V_{l,max})^2$ Because there are two mechanisms involved but a single event, the failure of the landslide dam, we need a single scalar potential as a function of the two potentials of the conditional intensities presented in (6.49) and (6.50). Without considering the mark function, we can assume that the conditional intensity function of the potential failure of a landslide dam can be defined as $\lambda(t|H(t)) = g_D(t,U_i(t);\theta)$,

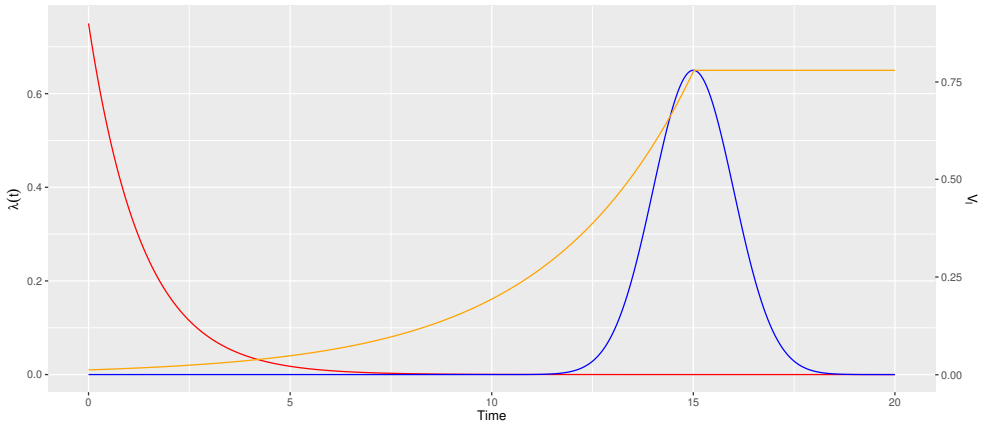


Figure 6.6: Examples of plots describing the two possible failures of dam: immediate failure after the landslide fall (red line) and failure due to accumulation of water (blue line). This plot schematically represents the two functions (6.49) and (6.50, left y-axis) in relation to the amount of water in the dam (orange line, right y-axis).

where $g(\cdot)$ is a single link function combining the two failure mechanisms. Hence, we need to identify which type of function is the best option. We can assume, from (6.49) and (6.50), that the link function is can be written as an exponential, so that the overall conditional intensity function, using the form of (6.1), is $\lambda_D(t) = \exp[U_D(t)]$ and so it can be seen as the product of the two marginal conditional intensities:

$$\lambda_D(t,U_D(t);\theta) = \lambda_1(t) \times \lambda_2(t) = a\exp[b_1U_1(t) + b_2U_2(t)] \tag{6.51}$$

Although the summation of the two conditional intensities (competing risks model) would be more natural, it does not fit the potential framework, as there is not then a single scalar potential. In fact, with a summation we would have obtained

$$\lambda(\cdot) = \lambda_1(\cdot) \times \lambda_2(\cdot) = a_1\exp(b_1U_1(\cdot)) + a_2\exp(b_2U_2(\cdot)) \tag{6.52}$$

which does not provide a single value. Given the degrees of freedom provided by the constants a, b_1 and b_2, it is possible for the overall failure intensity to be close enough in the product form (6.51) to the competing risks model, as shown in Figure 6.7.

Finally, we can combine all the equations we have proposed so far to formulate the conditional intensity function of a flood triggered by a landslide dam failure, that we shall name $\lambda_F(t|H(t))$. This function is built as the product of the three elements expressed above for two main reasons: first, as mentioned earlier, this method allows for the transfer of marks from one function to another; secondly, this transfer through the concatenation of potentials can simplify the simulation of the system. In fact, this method exploits the modularization of a simulation by splitting it into concatenated smaller ones and reducing the memory requirements, as all transfers between modules

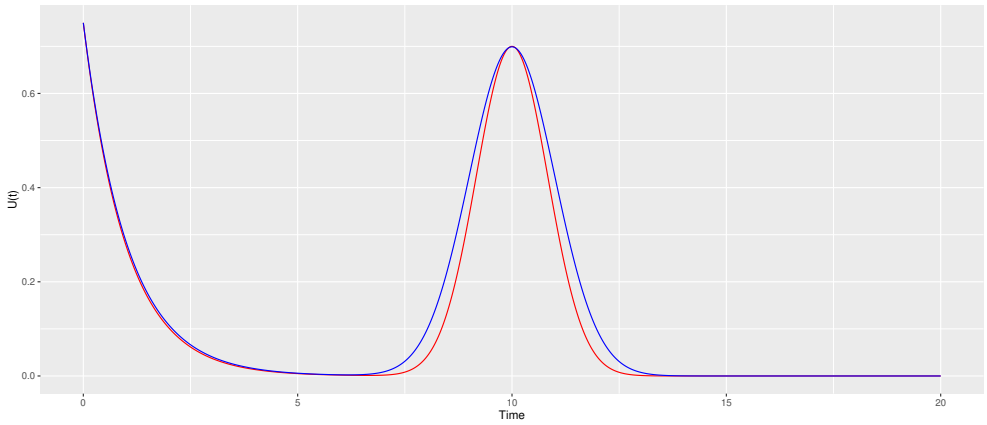


Figure 6.7: Examples of the resulting functions shown in Figure 6.6. The red line represents the product in (6.52), while the blue line the sum of of (6.49) and (6.50).

are performed by way of potentials. Because of this modularization, the conditional intensity of a flood triggered by a landslide dam failure is composed of several modules. Conceptually, the first one is the conditional intensity of the landslide occurrence $\lambda_{LS}(\cdot)$ at location x, for which we can use (6.43), followed, in order, by the probability distribution of the landslide dam formation (6.48), the cumulative water inflow $R(t)$, and the conditional intensity of the landslide dam failure (6.51). Furthermore, while the failure is assumed to occur at t, the rainfall, the landslide and the dam formation are assumed to occur at $s < t$.

To have a flood of volume m_F at time t, the landslide dam has to fail at t and at the same time the flood volume must be equal to the volume of water in the dam: $m_F = \min\{m_D, R(t-s)\}$, where $s < t$ is the time of the landslide and the formation of the landslide dam, with the maximum lake volume being at least equal to the volume of the flood: $m_D \geq m_F$. To build the flood conditional intensity, we can simplify for a moment the parametrization of $\lambda_{LS}(\cdot)$, into $\lambda_{ls}(\cdot)$, using the maximum lake volume rather than the landslide volume (recall that the lake volume is dependent on the landslide volume thanks to the cascading parametrization adopted above in (6.43) and (6.51)).

$$\begin{aligned}\lambda_F(t, m_F | U_D(t)) = \lambda_D(t, U_D(t); \theta) \int_{m_F}^{\infty} \int_{-\infty}^{t} P(LSD|H, L, W, x) \\ \times \lambda_{ls}(s, v) I_{\{m_F\}}(\min\{m_D, R(t-s)\}) \; ds \; dv\end{aligned} \tag{6.53}$$

This allows us to consider the time lapse $t - s$ between landslide occurrence/landslide dam formation and dam failure, particularly in terms of flood volume produced by a dam at location x, with certain dimensions and with the water accumulated between s and t. For the latter we have used an indicator function expressing that the minimum

of the lake volume and the inflow (i.e. the volume of water in the lake at failure) must be equal to m_F. It is to be noted that the conditional intensity of the dam failure $\lambda_D(\cdot)$ does not have a mark distribution associated with it, as the flood volume is deterministic. However, it can be obtained from the potential, given the knowledge of the maximum lake volume.

Next, we can replace $\lambda_{ls}(\cdot)$ with (6.48) and (6.43):

$$\lambda_{ls}(s,v) = \lambda_{LS}(s,H,L,W|U_{LS}(s);\theta) \times I_{\{h(v,x)\}}(H) \tag{6.54}$$

where $\lambda_{LS}(\cdot)$ is now characterised by the landslide volume. The maximum lake volume comes into play with the indicator function $I_{\{h(u,x)\}}(H)$, which requires that the height of the dam, H, at location x, be such as to produce a maximum lake volume of v. This provides the link between landslide, landslide dam, lake volume and flood. We can then combine (6.53) with $\lambda_{LS}(\cdot)$:

$$\begin{aligned}\lambda_F(t,m_F|U_F(t)) = \lambda_D(t,U_D(t);\theta) \int_{m_F}^{\infty}\int_{-\infty}^{t} & P(LSD|H,L,W,x) \\ & \times \lambda_{LS}(s,H,L,W|U_{LS}(s);\theta) \times I_{\{h(v,x)\}}(H) \\ & \times I_{\{m_F\}}(\min\{m_D,R(t-s)\})\ ds\ dv\end{aligned} \tag{6.55}$$

Going further, we can replace the conditional intensity function of the dam failure with the multiplication of the two mechanisms, and the conditional intensity function of the landslide dam with (6.43):

$$\begin{aligned}\lambda_F(t,m_F|U_F(t)) = & [\lambda_{D,S}(t,U_D(t);\theta) \times \lambda_{D,OV}(t,U_D(t);\theta)] \\ & \times \int_{m_F}^{\infty}\int_{-\infty}^{t} P(LSD|H,L,W,x) \times [g_{LS}(s,U_{LS}(s);\theta) \\ & \times \psi(m_{LS}|U_{LS}(s),x) I_{(L,W,H)}(f(m_{LS},x))] \\ & \times I_{\{h(v,x)\}}(H) \times I_{\{m_F\}}(\min\{m_D,R(t-s)\})\ ds\ dv\end{aligned} \tag{6.56}$$

Hence, we have obtained a flood conditional intensity function that takes into consideration all the aspects affecting the event, from the occurrence of the landslide (considering time elapsed, location and volume), the formation of the landslide dam (with its dimensions inherited from the landslide volume) to the landslide dam failure, taking into consideration the two different failure mechanisms. Overall, we have demonstrated that the conditional intensity function of a hazard such as a flood can be written with cascading modules expressing the potential of antecedent events, using their conditional intensities and mark densities.

6.4 Inhibition/threshold

The occurrence of a hazard depends, among other things, on the characteristics describing the other hazards involved in the triggering processes. For example, in Section 6.3.1 we have discussed how the characteristics of earthquakes (magnitude, distance) and rainfall (intensity, duration), may affect the occurrence of landslides, which can even cluster, depending on the marks of the primary hazards. Similarly, in Section 6.3.3, we have seen how the features of a landslide can be transferred to and alter the durability of a landslide dam. While in some cases the effects of marks can be modelled via continuous functions, as presented in the previous section, there are situations in which effects of marks need to be expressed differently. This is the case for qualitative marks, such as the geomorphologic features of an area, which are often expressed in categories. Thresholds can be used to overcome this problem. The underlying idea is that if a certain characteristic is present or absent, that will drastically change the conditional intensity of a hazard, although the potential may be unaffected. As mentioned at the beginning of this chapter, an event occurs if its disposition, frequency or magnitude for Kappes et al. (2012b), is altered by another one (Kappes et al., 2010). Gill and Malamud (2014) described this interaction as "increased probability": the characteristics of the secondary hazards can be moved towards a threshold (or tipping point, minimum value), which can be evaluated via analysis of the temporal likelihood. For example, strong earthquakes such as the $9.2Mw$ one in Alaska, 1964 produced regional subsidence, which increased the probability of future flooding. Hence, the phenomenon of subsidence produces a boosting effect on flooding potential. In fact, the modification of the ground produced by a powerful earthquake might affect a watercourse, by blocking the water or opening new paths. This could result in a drastic increase of the likelihood of a flood in the proximity of the river, even if the potential of a flood pre-earthquake had not suggested that, e.g. not enough rain in the preceding period to trigger a flood.

Thresholds have been used to evaluate the occurrence of hazard in relation to the properties of primary ones, particularly for the analysis of rainfall-triggered landslides (Keefer, 1984; Brunetti et al., 2010). Among others (Guzzetti et al., 2007; Badoux et al., 2009; Baum and Godt, 2010), Cannon et al. (2008) have used empirical intensity/duration rainfall thresholds to explain the occurrence of debris flows. The triggering mechanism of this phenomenon can be synthesised into two aspects: the presence of loose material (Iverson et al., 1997; Tang et al., 2009) and the occurrence of a hazard entraining the material, usually rainfall (Peng et al., 2015; Ferreira et al., 2016; Giannecchini et al., 2016). As the main purpose of these papers was to build a warning system for the occurrence of debris flow, the presence of loose material was considered as an existing condition, and the research focused on the study of rainfall thresholds, to

evaluate whether or not the debris flow would be triggered. Guzzetti et al. (2007) have shown different examples of intensity/duration thresholds from different studies on a dataset of debris flows in Figure 6.8. In our wider scope of the potential framework, we can consider the rainfall as a quantitative process expressed through a potential as in (6.40). Hence, the aspect that can be represented via threshold is the presence of debris. That would allow the threshold to have the function of either inhibiting or allowing for the occurrence of a hazard.

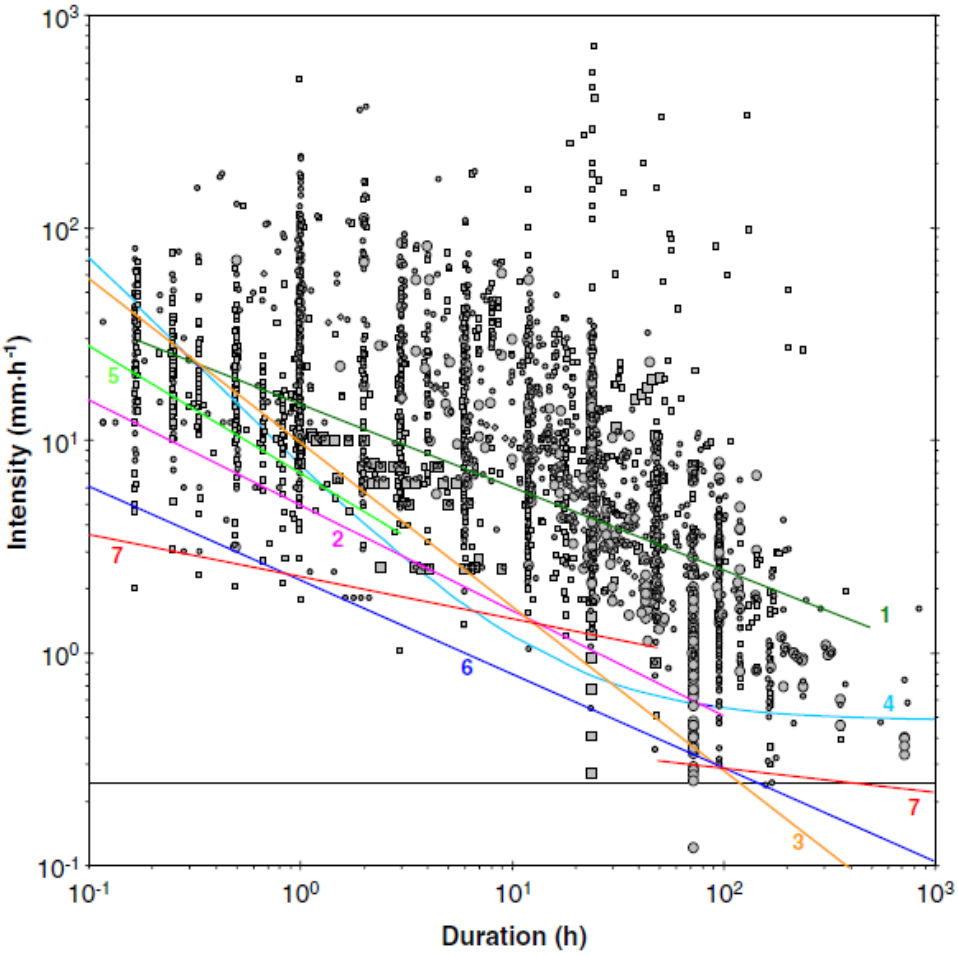


Figure 6.8: Comparison between the global ID thresholds defined in this study and published global (worldwide) ID rainfall thresholds. 1 Caine (1980); 2 Innes (1983); 3 Clarizia et al. (1996); 4 Crosta and Frattini (2001); 5 Cannon and Gartner (2005); 6 threshold inferred from the entire set of ID rainfall data (this work); 7 thresholds inferred from the probability estimates of the rainfall conditions, for two different rainfall periods ($D < 48h$, and $D \geq 48h$) (this work). Dashed line shows $0.25mm\ h^{-1}$ rainfall intensity. Picture and original caption from Guzzetti et al. (2008)

To pursue this idea, we can define, $W(t)$ as the Inhibition/Excitation function ("I/E function" from now on), which allows or inhibits the triggering of a hazard by

the process modelled using the potential function. Because there are different types of thresholds, we will need to define different types of functions. One function can be defined as "hard threshold": these are often used with natural hazards to evaluate the minimum magnitude values of hazards which can trigger secondary ones. A common example is the one of rainfall threshold as early warning systems for hazards such as landslides (Begueria and Vicente-Serrano, 2006; Tiranti and Rabuffetti, 2010). Similarly, the above mentioned Casagli and Ermini (1999) used a hard threshold to separate stable and at-risk landslide dams using an index based on the dam dimensions.

However, depending on the characteristic of the hazard and because events in nature may not have a clear cut between occurrence and non-occurrence, the hard threshold might be of a draconian nature that does not go well with all situations, particularly if the processes involved may produce different results at different levels of the processes expressed in the threshold. In this case, the use of "soft thresholds" can be a valid alternative. Figure 6.9 shows an example of soft threshold, where the likelihood of a hazard changes in relation to the progression of specific factors. In Chapter 4, we have presented two rainfall components that can serve as soft threshold for landslides. In fact, high levels of these components, which reflect high levels of short- and long-term rainfall, increase the likelihood of landslides, rather than suggesting a sharp landslide/no-landslide boundary. In Figure 6.9, we can imagine the two horizontal axes on the bottom plane as the two rainfall components, and the surface represents the probability of the hazard: at increasing levels of the components, the surface function shows a smooth transition through different levels of potential, rather than shifting abruptly from one line to another as per hard threshold curves, which may not have enough flexibility. It is then possible to see the I/E function as a function, which

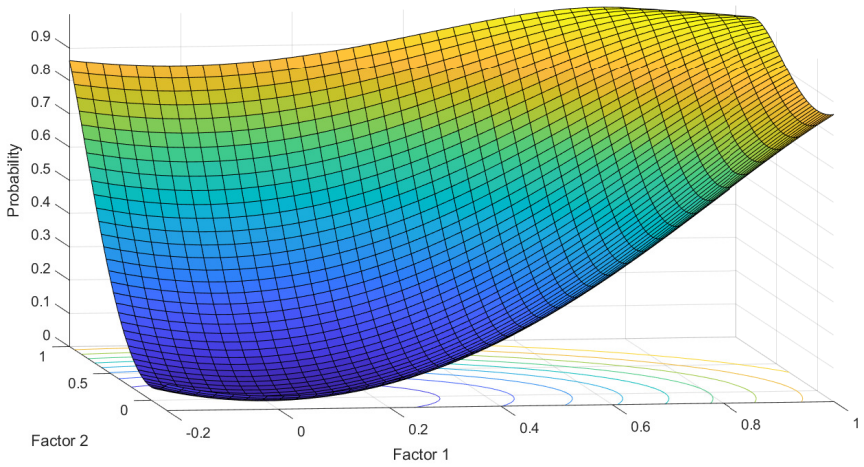


Figure 6.9: Examples of threshold function influenced by two factors. On the two axes on the horizontal plane, there are two components producing a probability surface.

"masks" it or "reveals" $U(t)$, inhibiting or allowing for hazard occurrence. In other words, the event can only occur if a certain condition is satisfied, conditions that are condensed in $W(t)$. A possible way to define $W(t)$ is presented in (6.57), where $W(t)$ is either 0 or 1 whether the condition for the occurrence of the external event is satisfied or not.

$$W(t) = \begin{cases} 1 & \text{condition satisfied} \\ 0 & \text{otherwise} \end{cases} \tag{6.57}$$

In the case of a soft threshold, the support of $W(t)$ can be [0,1], instead of {0,1} as in (6.57).

6.4.1 Debris flows

As mentioned earlier, debris flows provide a good example of the use of thresholds in the potential framework. Debris-flow occurs when intense precipitation mobilises loose sediments in a watershed or channel (Iverson, 2000). There are two main drivers in this hazard: water and loose material. A strong rainfall event, possibly a short-term-high-intensity one, acts as a vehicle, moving the loose material present on a slope (Peng et al., 2015). Furthermore, the topography of the area plays also an important role, as the presence of a watershed with steep slopes or stream channel with active erosion increases the chances of debris flow. Loose sediments may be the result of a recent landslide (e.g. portion of ground that is detached but that has not reached the bottom of the valley), a volcanic eruption (ashes), or wildfire (i.e. the reduced moisture in the ground increases the instability of it)(Welsh and Davies, 2011).

The case of debris flow can be expressed by two components. One is the presence of loose material (debris), and one is the mechanism that triggers the movement of the material (e.g. rainfall). While the second can be modelled with a point process, the presence of debris (consequence of a primary event such as a volcanic eruption or a landslide) can be defined using (6.57). Hence, when $W(t) = 1$, the debris flow potential is revealed, and the event can occur. With $W(t) = 0$, the event cannot occur, as there is no material to move. We assume $W(t)$ to be a function of a vector of parameters representing the satisfaction of certain conditions for the occurrence of the external event at a given time: $W(t) = f(\xi(t))$.

Therefore, the $W(t)$ function can be implemented in the potential framework as a new method to address the change of potential in relation to other events, particularly concerning those phenomena that do not produce any potential unless they occur, such as for the debris flow example. For instance, we can extend the concept of (6.1) by implementing $W(t)$:

$$g(t,U(t);\theta) = g(t,U(t);\theta) \times W_D(t) \tag{6.58}$$

Hence, in (6.58) $W(t)$ activates the link function $g(t,U(t);\theta)$ and ultimately the potential of a hazard.

In the literature (Melton, 1965; Jackson Jr. et al., 1987), a way to evaluate whether an area is prone to debris flow is to evaluate the steepness through the Melton ratio:

$$R = \frac{H_b}{\sqrt{A_b}} \tag{6.59}$$

where A_b is the catchment area in km^2 and H_b is the basin relief in km (the difference between the highest elevation and the lowest elevation in the catchment). The ratio is built so that the narrower and steeper is the channel, the more likely it is to observe debris-flows. In fact, a common subdivision of the values is the following:

- if $R \leq 0.3$ conventional fluvial processes occur;
- if $0.3 < R < 0.6$ debris floods occur;
- if $R \geq 0.6$ debris flows occur.

Debris floods are phenomena in between a flood and a debris flow. In fact, while the latter can have peak discharges up to 40 times greater than floods, debris floods discharges go up to twice those of flood.

Because debris floods and debris flows are similar in terms of mechanism (both need debris and rainfall), we can summarize the information from the Melton ratio into a new index, that we shall call $P(R)$ the probability density of debris flows occurrence. This will be a continuous function, monotonic in R. For example, one possibility is:

$$P(R) = \frac{\Phi\left[\frac{R-a}{b}\right] - \Phi\left(-\frac{a}{b}\right)}{1 - \Phi\left(-\frac{a}{b}\right)} \tag{6.60}$$

where Φ is the standard normal distribution function, and a,b are constants chosen so that $P(R = 0.3) = \epsilon$ and $P(R = 0.6) = 1 - \epsilon$ for some suitable (small) value of ϵ.

One option to model the potential of debris flow is to use a threshold function for the debris flow, as without debris the phenomenon would not occur. Debris flows get mobilised by a large amount of water fallen in a short period of time, which can be hours to days (Bacchini and Zannoni, 2003; Pan et al., 2018). Therefore, given how we have developed our rainfall components, it is safer to include both in this example. Once we have settled the rainfall and Melton Ratio contribution, we can look at the threshold function for the debris. The presence of debris can be written using an I/E function for the presence of debris: $W(t)$ is equal to 1 only if there are debris in the

area in analysis. Hence a step function will serve our need:

$$W_D(t) = \begin{cases} 1 & \text{debris in } t \\ 0 & \text{no debris in } t \end{cases} \tag{6.61}$$

Therefore, following (6.58), the link function between potential and conditional intensity function for the occurrence of debris flow $g_{DF}(\cdot)$ will be composed of a rainfall-related $g_{RF}(\cdot)$ function and the I/E function for the presence of debris.

$$g_{DF}(t,U(t);\theta) = P(R) \times g_{RF}(t,U_{RF}(t);\theta) \times W_{DF}(t) \tag{6.62}$$

However, we can also consider the converse formulation, with the presence of debris being modelled by the ground intensity function, based on the potential of debris flows $U_D(t)$, and the rainfall as an I/E function. Hence, (6.58) will become:

$$g(t,U_{DF}(t);\theta) = g_{DF}(t,U_{DF}(t),P(R);\theta) \times W_{RF}(t) \tag{6.63}$$

where $g_{DF}(t,U_D(t);\theta)$ is a link function for the amount of debris in t, which also takes into account the probability distribution of the Melton Ratio $P(R)$, to evaluate what kind of event may occur. Then, the $W_{RF}(t)$ function will need to assess the amount of debris in comparison to the amount of rainfall: if the two volumes are roughly equivalent, then debris floods would occur; if the amount of debris is much bigger than the rainfall, there is a higher probability of debris flows. In case of the amount of debris being smaller than the amount of rainfall, it is then the ordinal fluvial process. Hence, we can define D as the ratio between debris flow volume and rainfall volume, and update (6.60):

$$W_{RF}(t) = \frac{\Phi\left[\frac{D-\alpha}{\beta}\right] - \Phi\left(-\frac{\alpha}{\beta}\right)}{1 - \Phi\left(-\frac{\alpha}{\beta}\right)} \tag{6.64}$$

where Φ is the standard normal distribution function, and α,β are constants chosen so that $P(R = 0.3) = \eta$ and $P(R = 0.6) = 1 - \eta$ for some suitable value of $\eta < 0.5$.

6.5 A framework for multi-hazard simulation.

At the beginning of this chapter, we have mentioned that the complexity of interactions among natural hazards was substantially a call for a framework capable of unifying multi-hazard assessment under one model structure. We have proposed an initial framework that can accommodate both single-hazard and multi-hazard assessments. Using point process models as a basic structure, we have shown that pre-existing models can often be included, or adapted to our framework. This is the case of the

Hawkes process for earthquakes, of point processes for rainfall, and of our model for earthquake/rainfall triggering landslides from Chapter 4. Furthermore, the landslide dams section suggests that we can expand a process to take into account new hazards. The last model proposed for debris flows also shows that in some cases we also have the choice of modelling events from different perspectives.

An important aspect of this framework is that the flexibility of this all-encompassing framework can play a very important role in terms of simulation. There are packages that allow for simulation of quantitative models on the different platforms, such as *PtProcess* (Harte, 2010), which allows for simulation and fitting of marked point processes. Nevertheless, other hazards are modelled via physical models (CAESAR flooding model, Coulthard et al. (2002)) or via hazard maps (landslides, Dahal et al. (2008); Yalcin (2008)). Hence, with our framework there is an opportunity to combine hazards with different characteristics into one single framework to simulate interacting hazards.

As shown earlier in this chapter, point processes allow for a flexible framework where interacting hazards can be combined to evaluate their occurrences. The mentioned landslide dams chain is a natural example: concatenated marked point processes can evaluate the occurrence of landslide dam failures and consequently the size of floods. The package *PtProcess* (Harte, 2010), for instance, would allow writing a ground intensity function and mark distribution to build a specific marked point process, which can then be combined with another one to simulate a more complex phenomenon such as the one of landslide dams. This is a novel approach that is able to overcome the complexity of natural hazards, as it allows to split a potentially vast simulation algorithm in smaller modules, easier to handle and faster at simulating. Hence, we can imagine a package where we can define the conditional intensity function $\lambda(t|H_t)$ or the link function $g(\cdot)$, to evaluate, fit and simulate the potential function of a hazard. The functions would need to be specified in terms of characteristics of the environment such as the Melton ratio, landslide susceptibility characteristics, et cetera and to other processes affecting the triggering of the hazard (such as earthquakes for the landslide model).

In this view, the framework could facilitate the development and integration of the assessment for a natural hazard processes within our multi-hazard environment, using a building block approach of point process models. In fact, hazard estimates are generally obtained by simulation and require many simulations to quantify the uncertainty. The number of simulations required will rise steeply with the complexity of the system, while at the same time, the simulations become slower. A big part of this is the history, particularly in spatial environments. Hence the potential framework, with its lessened memory requirements, will facilitate forecasting.

A similar "brick-by-brick" concept was proposed by Mignan et al. (2014), where an initial approach to a potential framework was proposed. The authors produced Monte Carlo simulated time series, each one representing a different scenario of natural and man-made events. Each hazard as simulated from a Poisson distribution with specific long-term parameters, and events were combined with conditional probabilities. Our method proposes to go beyond this kind of approach, as each hazard would have a tailored conditional intensity and the modularization approach with cascade-like blocks would not only reduces time and memory requirements, but also provide flexibility and allow for realistic quantification of hazard interactions.

Chapter 7

Discussion and future research

This thesis contributes to the development of multi-hazard assessments, by providing examples of quantitative models for hazard interactions, and developing valuable concepts for spatio-temporal analysis of hazard occurrences. The statistical methodologies used have proved to be useful tools to produce substantially solid results and to overcome the challenging issues arising from the lack of data, typical of natural hazards. This final chapter closes the thesis by reviewing the achievements obtained in the previous chapters and proposing suggestions for future research.

The goals of this thesis, as defined in Chapter 1, can be summarised as follows:

- Formulate a quantitative multi-triggering model to evaluate the interaction among primary hazards in the triggering of secondary ones.
- Extend the hazard chain with a further hazard, using different techniques to investigate the feasibility of different types of analysis in relation to the type of hazard.
- Develop the concept of "hazard potential" in a quantitative framework, as an underlying process governing the occurrences of events.

Given these goals, I present here below a brief review of the work completed in the previous chapters, together with a discussion on possible future steps for these topics. Because of the variety of natural hazards and their interactions, the thesis is based on the earthquake/rainfall-landslide-landslide dams hazard chain, described in Chapter 3, large enough to possess a certain level of variety of hazards and their interactions.

Chapter 4 provided a first attempt at modelling a multi-triggering point process, developed to evaluate the landslide occurrences triggered by the interactions between earthquake and rainfall. Both earthquakes and rainfall are known to trigger landslides (Shou et al., 2011; Zhou et al., 2015b; Iverson, 2000; Berti et al., 2012), but their

concurrence is considered to be coincidental, although they spatially overlap. Nevertheless, there have been examples of potentially augmented landsliding events due to strict overlapping or temporal proximity (within a time window) of the two hazards, e.g. Kaikoura, Orchiston et al. (2018).

Landslides data are naturally discrete ones, and the Emilia-Romagna dataset I have used was not different, with temporal information recorded at a daily precision: from 1981 to 2018 (13789 days) there have been 7743 landslides over 328 municipalities; thus the count of landslides per day/location is zero for the majority of the day-municipality "cells" in the dataset (99.75%). Therefore, I have used a nonhomogeneous Poisson process then discretized it into a zero-inflated Poisson regression. This process was used to model the occurrences of landslides triggered by rainfall and earthquake, also taking into consideration the interaction among the triggering events. This method is able to model the excess of zeros naturally resulting from the triggering mechanisms of earthquake and rainfall, which do not produce landslides for the majority of the days.

The landslide process is conceived to embrace a range of interaction types, as in Gill and Malamud (2014), from the "increased probability" to almost direct triggering should the intensity rise quickly enough. That is achieved by formulating a Poisson process with components which expresses the triggering of rainfall, the triggering of earthquakes and the combined effect of the two primary hazards. Such components have been designed using proxies (Utsu, 1970; Ogata, 1988; Rossi et al., 2010; Wetzler et al., 2016; Monsieurs et al., 2019) which transfer the characteristics of the triggering effects (intensity, distance) into the landslide process. Hence, the model structure remains constant over time, with the baseline of the model, $\mu_0(x)$, representing the susceptibility of each location x, which acts as a multiplier, and the strength of each component $\{\mu_i\}$ (with i = 1,2,3) is treated as constant across space and time. The temporal component of the model is provided by the time-series of the triggering factors. Three different models have been tested to find the best fitting combination of the three components, and the preferred model (comparing their log-likelihoods) was the one with additive triggering effects (long-term and short-term rainfall, and coseismic triggering). Although the multiplicative form was rejected, other hazard chains or regions might show different results.

The model highlighted a strong effect of long-term rainfall on the likelihood of no landslides, for Emilia-Romagna, agreeing with previous work by (e.g.) Rossi et al. (2010, 2017) and Peruccacci et al. (2017). Two additive models are flagged as the best ones, with very close log-likelihood. Although these treat the earthquake triggering differently, due to the complexity of the models, it is not clear which one is the best. Considering the landslide occurrences across all municipalities individually is a novel approach compared to previous works, as it has allowed us to consider the region

not as a single point in space (Rossi et al., 2010), but as a set of multiple locations. Hence, each location contributes differently to the triggering mechanism, due to the specific geomorphic features of the municipality and the triggering effect received by the municipality from earthquakes and rainfall, which cannot be equal across the whole region.

Hence, Chapter 4 demonstrates that point processes are useful tools in a multi-triggering framework, particularly to model the triggering influence of multiple factors in a discrete approximation, as the formulation is adapted to naturally discrete recorded data.

It is to be noted that although the proposed model has been built on a specific hazard chain, it can be adjusted to include other hazards of the same chain, as well as to other hazard chains. The components expressing the triggering effects on landslides are built to model rainfall and earthquakes triggering. Hence, it is possible to tune model and components to different hazards. For instance, volcanic eruptions and earthquakes might be included in a similar model to evaluate their triggering effects on tsunami (Geist, 2014; Schmidt et al., 2011; Garza-Giron et al., 2018). In fact there is a known correlation between earthquakes and eruption, with the former able to trigger eruptions (Marzocchi, 2002; Bebbington and Marzocchi, 2011), and the latter able to trigger earthquakes. That is because the model has been built with the purpose of creating a tool in a multi-hazard framework, keeping in mind the variety of hazards and interactions among them (see Chapter 3), rather than building a model for that hazard chain only.

Two remarks can be made on the seismic triggering. First, I did not include any transient triggering effects of earthquakes: it has been suggested that there might be a (positive or negative) accumulative effect on landslide triggering (Brain et al., 2017), which might allow for more complex interactions with rainfall (Marc et al., 2015). In such a case, the model would need to include a new term, tracking the cumulative effect of earthquakes (see SRM in Sections 2.2.1 and 6.2.3). In order to examine this, it would be needed to find more suitable data, as the homogeneous period of the Emilia-Romagna landslide catalogue coincided with an unusually quiet period in the earthquake catalogue in the vicinity of Emilia-Romagna. The weak seismic shaking during that period of the catalogue have consequently produced limited evidence on earthquake triggering. Nevertheless, the analysis of the time series of the three components shows that landslides have occurred in correspondence to medium to strong earthquakes (with epicentres in or near the mountainous region) at times corresponding low rainfall. The two major earthquakes that occurred during the study period in Emilia-Romagna, on the contrary, were located in the flat areas of the Po valley, with characteristics that tend not to produce any landslides.

Some suggestions can be provided on this issue. A transient triggering component is needed to fill the gap and capture all the aspects of the coseismic triggering. Once again, this would require a specific set of data not available for our analysis. Furthermore, it is suggested that future work would test this model in other areas with more evidences of coseismic landsliding, so that the model can be tested further and improved. Nevertheless, this approach would find issues with inadequate available data, as coseismic landslide inventories for specific regions usually have no time dimensions (Tang et al., 2016; Tian et al., 2019). Hence, this problem will need to be addressed.

A second remark that can be made is about the location parameter $\mu_0(x)$. With more refined data, particularly in terms of locations, the landslide susceptibility can be spatially re-parametrized (Parker et al., 2015), leading to a mapped intensity. In order to calculate spatial intensity, a model for a size mark (e.g. Bebbington 2015) might be required, but that would also need geomorphological data with a certain level of definition, such as the one needed for our models (time-homogeneous, time-stamped, congruent in time and space, ideally with magnitude expressed). This would need to be done within each municipality to achieve better spatial definition.

Finally, one more aspect that can be addressed as an example of future work is that in Chapter 4 I have not considered if the probability of new landslides is either increased or decreased after an occurrence. In other words, there is here the possibility to investigate if there is inhibition or excitation (clustering) of landslides in one location. The model would need to be modified to address this purpose, in relation to the characteristics of the area, and the short and long term triggering effects. The type of data I have used do not allow for such analysis. In general, landslides are temporally recorded in relation to the moment they have been discovered, which produces low-definition temporal data.

Chapter 5 explored a natural extension of the landslide hazard chain, the landslide dam. Because of their nature, landslide dams are a type of hazard that needs to be assessed rapidly, as their failure can result in flash floods. The danger is compounded by the amount of water accumulated, hence the estimation of the time to failure becomes crucial for planning risk mitigation procedures.

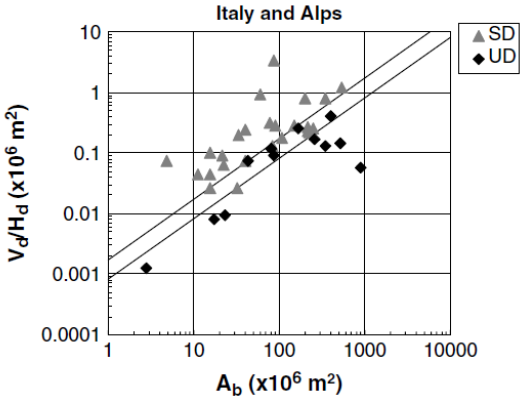
The lack of data has always limited the studies on this hazard, especially in terms of quantitative analysis. Chapter 5 aimed to overcome some of the limitations of landslide dams data by proposing a Bayesian model to predict the time to failure of landslide dams, based on imputing missing dam and reservoir measurements via an analysis of their covariate structure.

Survival analysis techniques were used to model the failure time of landslide dams in relation to their characteristics, improving the current state of the art of landslide dams analyses, which are mostly based on indices (Casagli and Ermini, 1999; Ermini

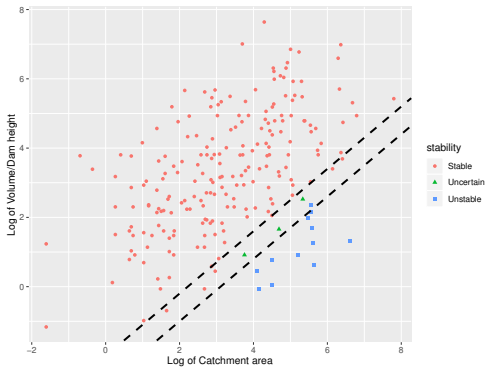
and Casagli, 2003; Korup, 2004) or logistic regression on the probability of failure (Dong et al., 2011). These methods produce a snapshot of a landslide dam conditions, which is an expression of whether the landslide dam will fall or not. Instead, a survival model allows the estimation of the failure times, hence quantifying the temporal aspect of the hazard. This is extremely important, especially in the process of decision making, if the water is filling the valley and engineering procedures might need to be put in place to avoid flash flood and disastrous consequences.

The biggest challenge in the dams analysis was the scale of missing data, an inherited feature of landslide dam datasets (Costa and Schuster, 1988; Ermini and Casagli, 2003). The dataset I have worked with, by Tacconi Stefanelli et al. (2015), has the unique peculiarity of including the type of landslide dams that Costa and Schuster (1988) classified mostly under Type I: those landslides that do not form a dam or form a dam that fails quickly. The nature of these events makes the data collection difficult. Hence, I have embraced the presence of these events in our model, together with the other landslide dam types, and used the concept of censored data to overcome the challenge of dealing with only 3% of data with exact time to failure. Concerning the missing values of the covariates, Bayesian imputation has been used, benefiting from the correlation structure among the variables. Linear regression was used to evaluate which of the landslide dams and valleys characteristics are the most important ones in determining the time-to-failure of the event.

Chapter 5 shows that the model can be used to probabilistically forecast the probable lifetime of a dam in a robust manner, based on the results in fitting it to a heterogeneous national dataset. The results obtained highlight that dams with longer length and higher height tend to survive longer while dams with larger catchment area survive for shorter times. This recalls the conclusions from other studies (Ermini and Casagli, 2003), with the important difference that our results have been obtained via quantitative analysis, which provides a statistical distribution of failure time, rather than a "hard" prediction. In fact, Figure 7.1 shows a comparison between a plot by Ermini and Casagli (2003) and one obtained from our data. The plot on the left visualizes how the DBI index (see Sections 3.2.4 and 5.2) divided landslide dams, looking at the relationship between the ratio volume/height and the catchment area. In the plot on the right, I have tried to recreate the same plot, using a computed version of the landslide dam volume (see Chapter 5). Ermini and Casagli (2003) have used the DBI to divide the dams into three groups, although only two are shown in the plot: stable, uncertain and unstable. Actually, the real meaning of "stable" should be "has not failed yet" and for "unstable" should be "has failed", with "uncertain" providing an in-between group of not yet failed dams. I have applied the same thresholds on our plot, and we can see that the DBI thresholds, obtained by Ermini and Casagli (2003),



(a) Original plot by Ermini and Casagli (2003)



(b) Plot from our analysis

Figure 7.1: The first plot (by Ermini and Casagli (2003) shows Italian dams in terms of watershed area (A_b) and the ratio volume/height split by DBI index (UD = unstable dam, SD = stable dam). The second plot is built on the dataset used for our analysis in Chapter 5.

are not correct for our dataset, as the majority of the dams are labelled as stable, while the majority of the dams in the dataset have disappeared within minutes or hours. This suggests that studying only large events introduces a bias in the analysis. Figure 7.2 shows the DBI index against median survival times. Again, the lines represent the thresholds suggested by Ermini and Casagli (2003): the DBI is not able to differentiate between landslide dams when the survival times are taken into consideration.

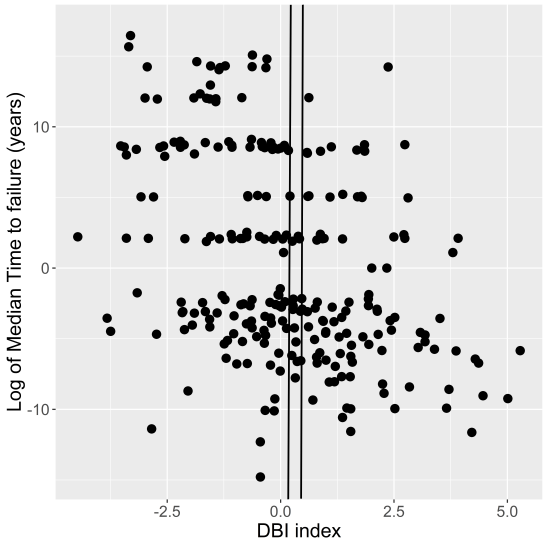


Figure 7.2: Distribution of log median dam failure time against DBI index (on a log scale) for each event in the presented dataset. The two lines indicate the two DBI thresholds indicated in Ermini and Casagli (2003).

The results of Chapter 5 also indicated that the dam width is not significant with a

negative coefficient. Other studies hypothesise that dam width should be as important as dam height (Ermini and Casagli, 2003; Liao et al., 2018). However, in Chapter 5 it was observed that the dam width has a medium to strong correlation with length and height. Given the significance of length and height, the effect produced by width is most likely absorbed by these two variables. As expressed by Costa and Schuster (1988), length is a crucial dimension, as it can be used as a proxy of dam formation/resistance (if the dam length is shorter than the valley width, it is unlikely that a lake will be formed). In this view, the dam width becomes secondary to other variables. It is also possible that our model is emphasizing the concept of dam length in comparison to valley width, but is important to remind that this is the first time that landslide dams with very short survival times have been included in a study. There might also be a river slope effect, produced by the fact that in a greater slope the width is less resistive (Costa and Schuster, 1988).

The addition of location-based intercepts suggests that geomorphic differences between regions may be quite distinctive. This is supported by the fact that the baseline parameter, β_0, has a high variability among survival times. The Sicilian events are a clear example of the need for regional intercepts. These events are characterised by lower altitudes and a different distribution of height-width ratio. Hence, further research may focus on incorporating geomorphic characteristics into the model, which are usually qualitative data. Another way to explore this aspect is using location-level random effects, rather than fixed intercepts, which should help to improve the estimation and forecasting ability of our model, towards a better understanding of events with location-related specific characteristics.

In addition, the model has the potential to be made more elaborate, with more accurate and complete data. This could also include examining the most appropriate measurement for quantities such as river slope, which is currently recorded at the dam site, but obviously varies above the dam.

Finally, Chapter 6 proposes an innovative approach to multi-hazard analysis using a framework based on the concept of potentials. The proposed idea is that the potential can synthesize and store the whole history of a process into a scalar value. Hence, it is expression of all the internal processes affecting the occurrences of a hazard, and their changes over time. This model relies on the use of monotonic increasing link functions of the potential $U(t)$, which are able to translate the history of the process H_t into a scalar value in $\lambda(t|H_t)$. Furthermore, I have categorized the components contributing to the change of potential over time, expressing slow and fast processes, decay or inflation and exogenous inputs.

I started by investigating how pre-existing single-hazard models can be adapted to this framework. Self-exciting processes (Hawkes process) can easily be incorporated

because the conditional intensity uses an exponential function, which can be updated to express the change of potential from t to $t + \Delta t$. In particular, the Hawkes process change of potential is characterised by a constant (proportional decay) and a sum of jump effects provided by events occurring within a small time interval. On the contrary, ETAS models cannot be used in this framework because of the use of power law decay functions, which do not allow $U(t)$ to be updated from t to $t + 1$.

Then, I moved to the core of the problem: the expression of multi-hazard processes in the potential framework. The crucial point is to be able to formulate the different type of hazard interactions and all the processes concurring in the potential of a hazard. With a building-block approach, it is possible to include each process separately, so that the conditional intensity is able to characterize the behaviour of each process in terms of time and magnitude. Then, all these blocks can be combined in relation to their interactions, taking advantage of the range of point processes explored in previous chapters

The adaptation of the multi-hazard model presented in Chapter 4 is a clear example of this concept. The two rainfall components (long-term and short-term) have been adapted to fit the definition of potential, highlighting the change in potential over time. Furthermore, this example is important because it shows that the potential concept extends to discrete time. This result further suggests the flexibility of this method, which is particularly important given the different precision between hazard data, as explored in previous chapters.

I have also mapped the extended hazard chain explored in Chapter 5 to the potential framework. This was a more challenging exercise, as it was crucial to consider all the interactions from the starting point of the chain (the occurrence of a landslide) to the size of the resulting event (the flood due to landslide dam failure). Hence, embracing the building-block type of model construction seen in Chapter 4, I have built the conditional intensity of a flood due to landslide dam failure as a composite of the single conditional intensities and probability distributions that lead to the flood: landslide occurrence, dam formation and dam failure. This particular chain of hazards also allowed us to explore the use of marks to model how the interaction between events could affect the size of a triggered hazard. In fact, the marks of the landslide are passed to the landslide dam in the first place, and then to the flood.The landslide dam dimensions are dependent on the size of the landslide originating the dam and the formation of the dam is once again determined by the size of the landslide (Costa and Schuster, 1988). Furthermore, it is possible to formulate the dam failure and flood size by taking into account the size of the lake as function of the location of the landslide and the size of the landslide dam and the inflow between formation and failure or filling of the dam. Another important aspect considered is that there is more than one type of failure

mechanism for a landslide dam, two in particular: seepage and overtopping. Hence, the conditional intensity expressing this event needs to include both. Although the sum of the two conditional intensities following a competing risk model would be more obvious, it would not fit the potential framework. Nevertheless, it has been shown that it is possible to approximate the same resulting conditional intensity with the use of a multiplication between the two failure mechanisms. In Chapter 6, I have also investigated the use of thresholds to model the potential of specific type of events. The example provided is the one of debris flow, as the occurrence of such event is strictly dependent on the presence/absence of debris. This type of scenario can be addressed as an extension of the concept of "increased probability" defined by Gill and Malamud (2014), assuming that the presence of debris would produce a non-zero probability of an occurrence of debris flow. Hence, I have defined a function $W(t)$ that activates the potential of debris flow, in relation to the presence of debris. It has also been shown, that this approach is flexible as it allows a certain level of degree to adjust the model as needed, by switching the role of hazard phenomena in the model. In fact while in one case the presence of debris is modelled with a threshold function and the rainfall with an intensity function, it is also possible to formulate the debris flow with an intensity function and the rainfall with a threshold function.

One of the major upsides provided by the potential framework is to accelerate multi-hazard simulation. The complexity of hazards and their interactions usually demands high memory requirements. I have used point processes as a base for our model, as they allow for multiple solutions that can be adapted to several hazard chains and reflect the hazard characteristics in terms of temporal occurrences and size. This allows for modularization of a complex conditional intensity (e.g. with several parameters within a single function) into concatenated block functions. For example, the earthquake/rainfall-landslides model proposed in Chapter 4 and revisited in Chapter 6. With this modularized approach it is possible to use marked point process: each block function can carry a mark distribution function, which can be linked to the following functions, so that marks are interconnected between hazards. For instance, the landslide volume is reflected on the landslide dam size and on the flood size. This would increase the simulation speed and lessen the memory requirements. This result, together with those obtained in the previous chapters, highlights that quantitative models of multi-hazard analyses can not only be carried out, but also that they are flexible to include and combine hazards with different characteristics. The examples in Chapters 4 and 5 provide tools that can be built upon in new case studies on other hazard chains and datasets. The possibility of improving computer simulations via the use of point processes and the modularization approach made feasible by the potential framework opens the possibility of multi-hazard estimation by point process simulation.

Bibliography

Ahmad, T., Munir, A., Bhatti, S., Aftab, M., and Raza, M. (2017), "Survival analysis of heart failure patients: A case study," *Plos One*, 12(7).

Aksha, S., Resler, L., Juran, L., and Carstensen, L. (2020), "A geospatial analysis of multi-hazard risk in Dharan, Nepal," *Geomatics, Natural Hazards and Risk*, 11(1), 88–111.

Aleotti, P. and Chowdhury, R. (1999), "Landslide hazard assessment: summary review and new perspectives," *Bulletin of Engineering Geology and the Environment*, 58(1), 21–44.

Alfaro, P., Delgado, J., García-Tortosa, F., Lenti, L., López, J., López-Casado, C., and Martino, S. (2012), "Widespread landslides induced by the Mw 5.1 earthquake of 11 May 2011 in Lorca, SE Spain," *Engineering Geology*, 137-138, 40 – 52.

Alvioli, M., Guzzetti, F., and Rossi, M. (2014), "Scaling properties of rainfall induced landslides predicted by a physically based model," *Geomorphology*, 213, 38–47.

Aragó, P., Juan, P., Díaz-Avalos, C., and Salvador, P. (2016), "Spatial point process modeling applied to the assessment of risk factors associated with forest wildfires incidence in Castellón, Spain," *European Journal of Forest Research*, 135(3), 451–464.

Arias, A. (1970), "Measure of earthquake intensity," in *Seismic Design for Nuclear Power Plants*, Hansen, Robert J. (ed.). Cambridge, Massachusetts Institute of Technology Press, pp. 438–483.

Aristizábal, E., García, E., and Martínez, C. (2015), "Susceptibility assessment of shallow landslides triggered by rainfall in tropical basins and mountainous terrains," *Natural Hazards*, 78(1), 621–634.

Aristizábal, E., Velez, J., Martínez, H., and Jaboyedoff, M. (2016), "SHIA Landslide: a

distributed conceptual and physically based model to forecast the temporal and spatial occurrence of shallow landslides triggered by rainfall in tropical and mountainous basins," *Landslides*, 13(3), 497–517.

Arpa-Piedmont (2018), "Landslide catalogue," `http://webgis.arpa.piemonte.it/Web22/sifrap/ii_livelli/002-00122-00.pdf`.

Auer, I., Böhm, R., Jurkovic, A., Lipa, W., Orlik, A., Potzmann, R., Schöner, W., Ungersböck, M., Matulla, C., Briffa, K., Jones, P., Efthymiadis, D., Brunetti, M., Nanni, T., Maugeri, M., Mercalli, L., Mestre, O., Moisselin, J.-M., Begert, M., Müller-Westermeier, G., Kveton, V., Bochnicek, O., Stastny, P., Lapin, M., Szalai, S., Szentimrey, T., Cegnar, T., Dolinar, M., Gajic-Capka, M., Zaninovic, K., Majstorovic, Z., and Nieplova, E. (2007), "HISTALP-historical instrumental climatological surface time series of the Greater Alpine Region," *International Journal of Climatology*, 27(1), 17–46.

Awal, R., Nakagawa, H., Baba, Y., Sharma, R., and Ito, N. (2007), "Study on Landslide Dam Failure by Sliding," *Annuals of the Disaster Prevention Research Institute, Kyoto University*, 50B, 653–659.

Ayalew, L. and Yamagishi, H. (2005), "The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda-Yahiko Mountains, Central Japan," *Geomorphology*, 65(1), 15 – 31.

Bacchini, M. and Zannoni, A. (2003), "Relations between rainfall and triggering of debris-flow: case study of Cancia (Dolomites, Northeastern Italy)." *Natural Hazards and Earth System Sciences*, 3(1/2), 71 – 79.

Badoux, A., Graf, C., Rhyner, J., Kuntner, R., and McArdell, B. W. (2009), "A debris-flow alarm system for the Alpine Illgraben catchment: design and performance," *Natural Hazards*, 49(3), 517–539.

Bak, P., Christensen, K., Danon, L., and Scanlon, T. (2002), "Unified scaling law for earthquakes," *Physical Review Letters*, 88, 178501.

Barbera, G. and Cullotta, S. (2012), "An Inventory Approach to the Assessment of Main Traditional Landscapes in Sicily (Central Mediterranean Basin)," *Landscape Research*, 37, 1–31.

Barredo, J., Benavides, A., Hervás, J., and van Westen, C. (2000), "Comparing heuristic landslide hazard assessment techniques using GIS in the Tirajana basin, Gran Canaria Island, Spain," *International Journal of Applied Earth Observation and Geoinformation*, 2(1), 9 – 23.

Baum, R. and Godt, J. (2010), "Early warning of rainfall-induced shallow landslides and debris flows in the USA," *Landslides*, 7(3), 259–272.

Bebbington, M. (2008), "Incorporating the eruptive history in a stochastic model for volcanic eruptions," *Journal of Volcanology and Geothermal Research*, 175(3), 325–333.

Bebbington, M. and Harte, D. (2001), "On the statistics of the Linked Stress Release Model," *Journal of Applied Probability*, 38(1), 176–187.

— (2003), "The linked stress release model for spatio-temporal seismicity: formulations, procedures and applications," *Geophysical Journal International*, 154(3), 925–946.

Bebbington, M., Harte, D., and Jaumé, S. (2010), "Repeated intermittent earthquake cycles in the San Francisco Bay Region," *Pure and Applied Geophysics*, 167(6), 801–818.

Bebbington, M. S. (2015), "Spatio-volumetric hazard estimation in the Auckland volcanic field," *Bulletin of Volcanology*, 77(5), 39.

Bebbington, M. S. and Marzocchi, W. (2011), "Stochastic models for earthquake triggering of volcanic eruptions." *Journal of Geophysical Research: Solid Earth*, 116.

Begueria, S. and Vicente-Serrano, S. M. (2006), "Mapping the Hazard of Extreme Rainfall by Peaks over Threshold Extreme Value Analysis and Spatial Regression Techniques," *Journal of Applied Meteorology and Climatology*, 45(1), 108–124.

Berti, M., Martina, M., Franceschini, S., Pignone, S., Simoni, A., and Pizziolo, M. (2012), "Probabilistic rainfall thresholds for landslide occurrence using a Bayesian approach," *Journal of Geophysical Research: Earth Surface*, 117(F4), 1–20.

Bertolini, G., Guida, M., and Pizziolo, M. (2005), "Landslides in Emilia-Romagna region (Italy): Strategies for hazard assessment and risk management," *Landslides*, 2(4), 302–312.

Boccaletti, M., Corti, G., and Martelli, L. (2011), "Recent and active tectonics of the external zone of the Northern Apennines (Italy)," *International Journal of Earth Sciences*, 100(6), 1331–1348.

Bokwa, A. (2013), "Natural hazard," *Encyclopedia of Earth Sciences Series*, 711–718.

Boore, D., Joyner, W., and Fumal, T. (1997), "Equations for estimating horizontal response spectra and peak acceleration from western North American earthquakes: a summary of recent work," *Seismological Research Letters*, 68(1), 128–153.

Borovkov, K. and Bebbington, M. S. (2003), "A Stochastic Two-node Stress Transfer Model Reproducing Omori's Law," *Pure and Applied Geophysics*, 160, 1429–1445.

Brain, M. J., Rosser, N. J., Sutton, J., Snelling, K., Tunstall, N., and Petley, D. N. (2015), "The effects of normal and shear stress wave phasing on coseismic landslide displacement," *Journal of Geophysical Research: Earth Surface*, 120(6), 1009–1022.

Brain, M. J., Rosser, N. J., and Tunstall, N. (2017), "The control of earthquake sequences on hillslope stability," *Geophysical Research Letters*, 44(2), 865–872.

Brockett, P., Golden, L., Guillen, M., Nielsen, J., Parner, J., and Perez-Marin, A. (2008), "Survival Analysis of a Household Portfolio of Insurance Policies: How Much Time Do You Have to Stop Total Customer Defection?" *The Journal of Risk and Insurance*, 75(3), 713–737.

Brooks, S. P. and Gelman, A. (1998), "General methods for monitoring convergence of iterative simulations," *Journal of Computational Graphical Statistics*, 7(4), 434–455.

Brunetti, M., Maugeri, M., Monti, F., and Nanni, T. (2006), "Temperature and precipitation variability in Italy in the last two centuries from homogenised instrumental time series," *International Journal of Climatology*, 26(3), 345–381.

Brunetti, M., Peruccacci, S., Rossi, M., Luciani, S., Valigi, D., and Guzzetti, F. (2010), "Rainfall thresholds for the possible occurrence of landslides in Italy," *Natural Hazards and Earth System Science*, 10(3), 447–458.

Caine, N. (1980), "The Rainfall Intensity: Duration Control of Shallow Landslides and Debris Flows," *Geografiska Annaler. Series A, Physical Geography*, 62(1/2), 23–27.

Camassi, R. and Stucchi, M. (1996), "NT4.1 un catalogo parametrico di terremoti di area italiana al di sopra della soglia del danno," *CNR-GNDT*, 1–86.

Cannon, S. and Gartner, J. (2005), "Wildfire-related debris flow from a hazards perspective," in *Debris-flow Hazards and Related Phenomena*, Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 363–385.

Cannon, S., Gartner, J., Wilson, R., Bowers, J., and Laber, J. (2008), "Storm rainfall conditions for floods and debris flows from recently burned areas in southwestern Colorado and southern California," *Geomorphology*, 96(3-4), 250–269.

Casagli, N. and Ermini, L. (1999), "Geomorphic analysis of landslide dams in the Northern Apennine," *Transactions of the Japanese Geomorphological Union*, 20(3), 219–249.

Casagli, N., Ermini, L., and Rosati, G. (2003), "Determining grain size distribution of the material composing landslide dams in the Northern Apennines: sampling and processing methods," *Engineering Geology*, 69(1), 83 – 97.

Cesca, S., Braun, T., Maccaferri, F., Passarelli, L., Rivalta, E., and Dahm, T. (2013), "Source modelling of the M5-6 Emilia-Romagna, Italy, earthquakes (2012 May 20-29)," *Geophysical Journal International*, 193(3), 1658–1672.

Chambers, J. and Hastie, T. (1991), *Statistical models in S*, Chapman and Hall/CRC.

Chen, K.-T., Kuo, Y.-S., and Shieh, C.-L. (2014), "Rapid geometry analysis for earthquake-induced and rainfall-induced landslide dams in Taiwan," *Journal of Mountain Science*, 11, 360–370.

Chen, X., Cui, P., You, Y., Cheng, Z., Khan, A., Ye, C., and Zhang, S. (2017), "Dam-break risk analysis of the Attabad landslide dam in Pakistan and emergency countermeasures," *Landslides*, 14(2), 675–683.

Chen, Y.-M., Fan, K., and Chen, L.-C. (2010), "Requirements and functional analysis of a multi-hazard disaster-risk analysis system," *Human and Ecological Risk Assessment: An International Journal*, 16(2), 413–428.

Chiarabba, C., Amato, A., Anselmi, M., Baccheschi, P., Bianchi, I., Cattaneo, M., Cecere, G., Chiaraluce, L., Ciaccio, M., De Gori, P., De Luca, G., Di Bona, M., Di Stefano, R., Faenza, L., Govoni, A., Improta, L., Lucente, F., Marchetti, A., Margheriti, L., Mele, F., Michelini, A., Monachesi, G., Moretti, M., Pastori, M., Piana Agostinetti, N., Piccinini, D., Roselli, P., Seccia, D., and Valoroso, L. (2009), "The 2009 L'Aquila (central Italy) Mw6.3 earthquake: main shock and aftershocks," *Geophysical Research Letters*, 36(18).

Chiarabba, C., Jovane, L., and DiStefano, R. (2005), "A new view of Italian seismicity using 20 years of instrumental recordings," *Tectonophysics*, 395(3), 251 – 268.

Christensen, R., Johnson, W., Branscum, A., and Hanson, T. (2011), *Bayesian Ideas and Data Analysis*, CRC Press.

Cieslik, K., Shakya, P., Uprety, M., Dewulf, A., Russell, C., Clark, J., Dhital, M. R., and Dhakal, A. (2019), "Building resilience to chronic landslide hazard through citizen science," *Frontiers in Earth Science*, 7, 278.

Clarizia, M., Gullà, G., and Sorbino, G. (1996), "Sui meccanismi di innesco dei soil slip," *International Conference Prevention of Hydrogeological Hazards: The Role of Scientific Research*, 1, 585–597.

Coico, P., Calcaterra, D., De Pippo, T., and Guida, D. (2013), "A preliminary perspective on landslide dams of Campania Region, Italy," *Landslide Science Practice*, 6, 83–90.

Console, R., Murru, M., and Falcone, G. (2009), "Probability gains of an epidemic-type aftershock sequence model in retrospective forecasting of M $\geq$ 5 earthquakes in Italy," *Journal of Seismology*, 14, 9–26.

Costa, J. E. and Schuster, R. L. (1988), "Formation and failure of natural dams." *Bulletin of the Geological Society of America*, 100(7), 1054–1068.

Coulthard, T., Macklin, M., and Kirkby, M. (2002), "A cellular model of Holocene upland river basin and alluvial fan evolution," *Earth Surface Processes and Landforms*, 27, 269 – 288.

Cowpertwait, P., Isham, V., and Onof, C. (2007), "Point process models of rainfall: developments for fine-scale structure," *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 463(2086), 2569–2587.

Cox, D. R. and Isham, V. (1980), *Point Processes*, Chapman and Hall/CRC, Routledge.

Cressie, N. A. C. (1993), *Statistics for Spatial Data.*, Wiley.

Crosta, G. and Frattini, P. (2001), "Rainfall thresholds for triggering soil slips and debris flow," *Proceedings of the 2nd EGS Plinius Conference on Mediterranean Storms*, 463–487.

Cui, P., Chen, X.-Q., Zhu, Y.-Y., Su, F.-H., Wei, F.-Q., Han, Y.-S., Liu, H.-J., and Zhuang, J.-Q. (2011), "The Wenchuan Earthquake (May 12, 2008), Sichuan Province, China, and resulting geohazards," *Natural Hazards*, 56(1), 19–36.

Dahal, R., Hasegawa, S., Nonomura, A., Yamanaka, M., Masuda, T., and Nishino, K. (2008), "GIS-based weights-of-evidence modelling of rainfall-induced landslides in small catchments for landslide susceptibility mapping," *Environmental Geology*, 54 (2), 311–324.

Dai, F. and Lee, C. (2001), "Frequency-volume relation and prediction of rainfall-induced landslides," *Engineering Geology*, 59(3-4), 253–266.

Dai, F., Lee, C., and Ngai, Y. (2002), "Landslide risk assessment and management: an overview," *Engineering Geology*, 64(1), 65 – 87.

Daley, D. J. and Vere-Jones, D. (2003), *An introduction to the theory of point processes. Vol. I*, Probability and its Applications (New York), New York: Springer-Verlag.

De la Cruz-Reyna, S. (1991), "Poisson-distributed patterns of explosive eruptive activity," *Bulletin of Volcanology*, 54(1), 57–67.

De Pippo, T., Donadio, C., Pennetta, M., Petrosino, C., Terlizzi, F., and Valente, A. (2008), "Costal Hazard assessment and mapping in Northern Campania," *Geomorphology*, 97(3-4), 451–466.

Dellow, S., Massey, C., Cox, S., Archibald, G., Begg, J., Bruce, Z., Carey, J., Davidson, J., Pasqua, F., Glassey, P., Hill, M.and Jones, K., Lyndsell, B., Lukovic, B., McColl, S., Rattenbury, M., Read, S., Rosser, B., Singeisen, C., and Little, M. (2017), "Landslides caused by the Mw7.8 Kaikōura earthquake and the immediate response," *Bulletin of the New Zealand Society for Earthquake Engineering*, 50, 106–116.

Delmonaco, G., Margottini, C., and Spizzichino, D. (2006), "New methodology for multi-risk assessment and the harmonisation of different natural risk maps," Tech. rep., European Commission - Applied multi-risk mapping of natural hazards for impact assessment.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977), "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society, Series B*, 39(1), 1–38.

Deng, D., Li, L., and Zhao, L. (2017), "Limit equilibrium method (LEM) of slope stability and calculation of comprehensive factor of safety with double strength-reduction technique," *Journal of Mountain Science*, 14(11), 2311–2324.

Derek Tucker, J., Shand, L., and Lewis, J. (2019), "Handling missing data in self-exciting point process models," *Spatial Statistics*, 29, 160 – 176.

Diggle, P. (2003), *Statistical Analysis of Spatial Point Patterns*, Arnold.

D'Odorico, P. and Fagherazzi, S. (2003), "A probabilistic model of rainfall-triggered shallow landslides in hollows: A long-term analysis," *Water Resources Research*, 39 (9), 1–14.

Dong, J., Lai, P., Chang, C., Yang, S., Yeh, K., Liao, J., and Pan, Y. (2014), "Deriving landslide dam geometry from remote sensing images for the rapid assessment of critical parameters related to dam-breach hazards," *Landslides*, 11(1), 93–105.

Dong, J.-J., Tung, Y.-H., Chen, C.-C., Liao, J.-J., and Pan, Y.-W. (2009), "Discriminant analysis of the geomorphic characteristics and stability of landslide dams," *Geomorphology*, 110(3-4), 162–171.

— (2011), "Logistic regression model for predicting the failure probability of a landslide dam," *Engineering Geology*, 117(1-2), 52–61.

Díaz-Avalos, C., Juan, P., and Serra-Saurina, L. (2016), "Modeling fire size of wildfires in Castellon (Spain), using spatiotemporal marked point processes," *Forest Ecology and Management*, 381, 360 – 369.

Ermini, L. and Casagli, N. (2003), "Prediction of the behaviour of landslide dams using a geomorphological dimensionless index," *Earth Surface Processes and Landforms*, 28(1), 31–47.

Errais, E., Giesecke, K., and Goldberg, L. (2010), "Affine point processes and portfolio credit risk," *SIAM Journal on Financial Mathematics*, 1(1), 642–665.

Esharti, L., Mahmoudzadeh, A., and Taghvaei, M. (2015), "Multi hazards risk assessment, a new methodology," *International Journal of Health System and Disaster Management*, 3(2), 79–88.

Evans, S. and Hungr, O. (1993), "The assessment of rockfall hazard at the base of talus slopes," *Canadian Geotechnical Journal*, 30(4), 620–636.

Fan, X., Scaringi, G., Domènech, G., Yang, F., Guo, X., Dai, L., He, C., Xu, Q., and Huang, R. (2019), "Two multi-temporal datasets that track the enhanced landsliding after the 2008 Wenchuan earthquake," *Earth System Science Data*, 11(1), 35–55.

Fan, X., van Westen, C., Xu, Q., Gorum, T., and Dai, F. (2012), "Analysis of landslide dams induced by the 2008 Wenchuan earthquake," *Journal of Asian Earth Sciences*, 57, 25 – 37.

Federico, A., Popescu, M., Elia, G., Fidelibus, C., Internò, G., and Murianni, A. (2012), "Prediction of time to slope failure: a general framework," *Environmental Earth Sciences*, 66(1), 245–256.

Feller, W. (1941), "On the integral equation of renewal theory," *The Annals of Mathematical Statistics*, 12(3), 243–267.

Feng, H., Yu, J., Zheng, J., Xiaoming, T., and Peng, C. (2016), "Evaluation of different models in rainfall-triggered landslide susceptibility mapping: a case study in Chunan, Southeast China," *Environmental Earth Sciences*, 75(21), 1399.

Ferreira, C., Rossini-Penteado, D., Brollo, M., Picanco, J., Da Silva, M., and Guimaraes, B. (2016), "Debris flow hazard and susceptibility zonation in small watersheds in Itaoca municipality, Sao Paulo state, Brazil," *Landslides and Engineered Slopes. Experience, Theory and Practice*, 2, 893–900.

Field, E., Arrowsmith, R., Biasi, G., Bird, P., Dawson, T., Felzer, K., Jackson, D., Johnson, K., Jordan, T., Madden, C., Michael, A., Milner, K., Page, M., Parsons, T., Powers, P., Shaw, B., Thatcher, W., Weldon II, R., and Zeng, Y. (2014), "Uniform California Earthquake Rupture Forecast, version 3 (UCERF3) -The time-independent model," *Bulletin of the Seismological Society of America*, 104(3), 1122–1180.

Fleming, K., Parolai, S., Garcia-Aristizábal, A., Tyagunov, S., Vorogushyn, S., Kreibich, H., and H., M. (2016), "Harmonizing and comparing single-type natural hazard risk estimations," *Annals of Geophysics*, 59(2), 1–9.

Frigerio Porta, G., Bebbington, M., Xiao, X., and Jones, G. (2020), "Bayesian lifetime analysis for landslide dams," *Landslides*, 17(8), 1835–1848.

Gao, L., Zhang, L. M., and Cheung, R. W. M. (2018), "Relationships between natural terrain landslide magnitudes and triggering rainfall based on a large landslide inventory in Hong Kong." *Landslides*, 15(4), 727.

Garcia-Rodriguez, M. J., Malpica, J., Benito, B., and Diaz, M. (2008), "Susceptibility assessment of earthquake-triggered landslides in El Salvador using logistic regression," *Geomorphology*, 95, 172 – 191.

Garza-Giron, R., Brodsky, E., and Prejean, S. (2018), "Mainshock-Aftershock Clustering in Volcanic Regions," *Geophysical Research Letters*, 45(3), 1370–1378.

Gasperini, P., Lolli, B., and V., G. (2013), "Empirical Calibration of Local Magnitude Data Sets Versus Moment Magnitude in Italy," *Bulletin of the Seismological Society of America*, 103(4), 22 – 27.

Geist, E. (2014), "Explanation of temporal clustering of Tsunami sources using the epidemic-type aftershock sequence model," *Bulletin of the Seismological Society of America*, 104(4), 2091–2103.

Gelman, A. and Hill, J. (2007), *Data Analysis Using Regression and Multilevel/Hierarchical Models*, Cambridge University Press.

Giannecchini, R., Galanti, Y., D'Amato-Avanzi, G., and Barsanti, M. (2016), "Probabilistic rainfall thresholds for triggering debris flows in a human-modified landscape," *Geomorphology*, 257, 94–107.

Gill, J. and Malamud, B. (2014), "Reviewing and visualizing the interactions of natural hazards," *Reviews of Geophysics*, 52(4), 680–722.

Glade, T., Crozier, M., and Smith, P. (2000), "Applying probability determination to refine landslide-triggering rainfall thresholds using an empirical 'Antecedent daily rainfall model'," *Pure and Applied Geophysics*, 157(6-8), 1059–1079.

Godefroy, P. and Humbert, M. (1983), "Cartography of natural hazards due to earth movements and seismic activity. Application of the 'plans pertaining to exposure to risks', PER, in France. [La cartographie des risques naturels lies aux movements de terrain et aux seismes. Application en France a l'elaboration des Plans d'Exposition aux Risques (PER).]," *Hydrogeologie - Geologie de l'Ingenieur*, 2, 69–90.

Gorum, T., Fan, X., van Westen, C. J., Huang, R. Q., Xu, Q., Tang, C., and Wang, G. (2011), "Distribution pattern of earthquake-induced landslides triggered by the 12 May 2008 Wenchuan earthquake," *Geomorphology*, 133(3-4), 152–167.

Govoni, A., Marchetti, A., De Gori, P., Di Bona, M., Lucente, F., Improta, L., Chiarabba, C., Nardi, A., Margheriti, L., Agostinetti, N., Di Giovambattista, R., Latorre, D., Anselmi, M., Ciaccio, M., Moretti, M., Castellano, C., and Piccinini, D. (2014), "The 2012 Emilia seismic sequence (Northern Italy): Imaging the thrust fault system by accurate aftershock location," *Tectonophysics*, 622, 44–55.

Gu, C., Schumann, A. Y., Baiesi, M., and Davidsen, J. (2013), "Triggering cascades and statistical properties of aftershocks," *Journal of Geophysical Research: Solid Earth*, 118(8), 4278–4295.

Guo, Y., Zhuang, J., and Zhou, S. (2015), "An improved space-time ETAS model for inverting the rupture geometry from seismicity triggering," *Journal of Geophysical Research: Solid Earth*, 120(5), 3309–3323.

Guzzetti, F., Carrara, A., Cardinali, M., and Reichenbach, P. (1999), "Landslide hazard evaluation: a review of current techniques and their application in a multi-scale study, Central Italy," *Geomorphology*, 31(1), 181 – 216.

Guzzetti, F., Mondini, A., Cardinali, M., Fiorucci, F., Santangelo, M., and Chang, K.-T. (2012), "Landslide inventory maps: new tools for an old problem," *Earth-Science Reviews*, 112(1), 42 – 66.

Guzzetti, F., Peruccacci, S., Rossi, M., and Stark, C. P. (2007), "Rainfall thresholds for the initiation of landslides in central and Southern Europe," *Meteorology and Atmospheric Physics*, 98(3), 239–267.

— (2008), "The rainfall intensity–duration control of shallow landslides and debris flows: an update." *Landslides*, 5(1), 3.

Han, J., Wu, S., and Wang, H. (2007), "Preliminary study on geological hazard chains," *Earth Science Frontiers*, 14(6), 11–23.

Hancock, M., Maher, C., da Cunha Menezes Costa, L., and Williams, C. (2014), "A guide to survival analysis for manual therapy clinicians and researchers," *Manual Therapy*, 19(6), 511 – 516.

Hanks, T. and Kanamori, H. (1979), "A moment magnitude scale," *Journal of Geophysical Research B: Solid Earth*, 84(B5), 2348–2350.

Harrison, L., Dunning, S., Woodward, J., and Davies, T. (2015), "Post-rock-avalanche dam outburst flood sedimentation in Ram Creek, Southern Alps, New Zealand," *Geomorphology*, 241, 135–144.

Harte, D. (1999), *Documentation for the Statistical Seismology Library. Research Report 98/10, revised edition*, School of Mathematical and Computing Sciences, Victoria University of Wellington, New Zealand.

— (2010), "PtProcess: An R package for modelling marked point processes indexed by time," *Journal of Statistical Software*, 35(8), 1–32.

— (2013), "Bias in fitting the ETAS model: A case study based on New Zealand seismicity," *Geophysical Journal International*, 192(1), 390–412.

Havenith, H., Torgoev, A., Braun, A., Schlogel, R., and Micu, M. (2016), "A new classification of earthquake-induced landslide event sized based on seismotectonic, topographic, climatic and geologic factors," *Geoenvironmental Disasters*, 3(6), 1–24.

Hawkes, A. (1971), "Spectra of some self-exciting and mutually exciting point processes," *Biometrika*, 58(1), 83–90.

Hawkes, A. and Oakes, D. (1974), "A Cluster Process Representation of a Self-Exciting Process," *Journal of Applied Probability*, 11(3), 493–503.

Haylock, M. R., Peterson, T. C., Alves, L. M., Ambrizzi, T., Anunciação, Y. M. T., Baez, J., Barros, V. R., Berlato, M. A., Bidegain, M., Coronel, G., Corradi, V., Garcia, V. J., Grimm, A. M., Karoly, D., Marengo, J. A., Marino, M. B., Moncunill, D. F., Nechet, D., Quintana, J., and Rebello, E. (2006), "Trends in total and extreme South American rainfall in 1960–2000 and links with sea surface temperature." *Journal of Climate*, 19(8), 1490 – 1512.

He, S., Wang, J., and Liu, S. (2020), "Rainfall event-duration thresholds for landslide occurrences in China," *Water*, 12(2), 1–17.

Heinimann, H. R. (1998), *Methoden zur analyse und bewertung von naturgefahren: eine risikoorientierte betrachtungsweise*, Umwelt-Materialien Naturgefahren, Bern: BUWAL.

Helmstetter, A. and Sornette, D. (2002), "Diffusion of epicenters of earthquake aftershocks, Omori's law, and generalized continuous-time random walk models," *Physical Review E*, 66(6), 24.

Hewitt, K. (1998), "Catastrophic landslides and their effects on the Upper Indus streams, Karakoram Himalaya, Northern Pakistan," *Geomorphology*, 26(1-3), 47–80.

Hewitt, K. and Burton, Ian, j. a. (1971), *The hazardousness of a place : a regional ecology of damaging events*, University of Toronto Press.

Hovius, N., Stark, C. P., and Allen, P. A. (1997), "Sediment flux from a mountain belt derived by landslide mapping," *Geology*, 25(3), 231–234.

Humbert, M. (1977), "ZERMOS Cartography. Modes of mapping the zones subject to hazards associated with soil and subsoil movements. [La Cartographie ZERMOS. Modalites d'etablissement des cartes des zones exposees a des risques lies aux mouvements du sol et du sous-sol.]," *Bulletin du Bureau de Recherches Geologiques et Minieres Section 3*, (1-2), 5–8.

In, J. and Lee, D. (2018), "Survival analysis: Part I - Analysis of time-to-event," *Korean Journal of Anesthesiology*, 71(3), 182–191.

Innes, J. L. (1983), "Debris flows," *Progress in Physical Geography*, 7(4), 469–501.

Inoue, K., Mori, T., and Mizuyama, T. (2013), "Three large historical landslide dams and outburst disasters in the North Fossa Magna Area, Central Japan," *International Journal of Erosion Control Engineering*, 5(2), 145–154.

ISC (2019), "International Seismological Centre," `http://www.isc.ac.uk/`.

Isham, V., Cox, D., Rodríguez-Iturbe, I., Porporato, A., and Manfreda, S. (2005), "Representation of space-time variability of soil moisture," *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 461(2064), 4035–4055.

Isham, V. and Westcott, M. (1979), "A self-correcting point process," *Stochastic Processes and their Applications*, 8(3), 335 – 347.

ISPRA (2019), "Inventory of landslide phenomena in Italy," `http://www.isprambiente.gov.it/en`.

Italian National Institute of Geophysics and Volcanology (2019), "INGV 2018 earthquakes report," `https://ingvterremoti.wordpress.com/2019/01/24/speciale-2018-un-anno-di-terremoti/`.

Iverson, R., Reid, M., and LaHusen, R. (1997), "Debris-flow mobilization from landslides," *Annual Review of Earth and Planetary Sciences*, 25, 85–138.

Iverson, R. M. (2000), "Landslide triggering by rain infiltration," *Water Resources Research*, 36(7), 1897–1910.

Iwata, T. (2015), "A variety of aftershock decays in the rate- and state-friction model due to the effect of secondary aftershocks: implications derived from an analysis of real aftershock sequences," *Pure and Applied Geophysics*, 173.

Jackson Jr., L., Kostaschuk, R., and MacDonald, G. (1987), *Identification of debris flow hazard on alluvial fans in the Canadian Rocky Mountains*, vol. 7 of *GSA Reviews in Engineering Geology*, Geological Survey of Canada, Terrain Sciences Division: Geological Society of America.

James, A. and De Graff, J. V. (2012), "The draining of Matthieu landslide-dam lake, Dominica, West Indies," *Landslides*, 9(4), 529–537.

Jibson, R., Harp, E., and Michael, J. (2000), "A method for producing digital probabilistic seismic landslide hazard maps," *Engineering Geology*, 58(3), 271 – 289.

Jing, H., Yang, Y., and Nishikawa, R. (2011), "Detection of clustered microcalcifications using spatial point process modeling," *Physics in Medicine and Biology*, 56(1), 1–17.

Joyner, W. B. and Boore, D. M. (1981), "Peak horizontal acceleration and velocity from strong-motion records including records from the 1979 Imperial valley, California, earthquake," *Bulletin of the Seismological Society of America*, 71(6), 2011–2038.

Kaczmarska, J., Isham, V., and Onof, O. (2014), "Point process models for fine-resolution rainfall," *Hydrological Sciences Journal*, 59(11), 1972–1991.

Kagan, Y. Y. (1991), "Likelihood analysis of earthquake catalogues," *Geophysical Journal International*, 106(1), 135.

Kappes, M., Gruber, K., Frigerio, S., Bell, R., Keiler, M., and Glade, T. (2012a), "The MultiRISK platform: The technical concept and application of a regional-scale multihazard exposure analysis tool," *Geomorphology*, 151-152, 139 – 155.

Kappes, M., Keiler, M., and Glade, T. (2010), "From single- to Multi - Hazard Risk Analyses: a concept addressing emerging challenges," in *Mountain risks International Conference*, pp. 351–356.

Kappes, M., Keiler, M., von Elverfeldt, K., and Glade, T. (2012b), "Challenges of analyzing multi-hazard risk: a review," *Journal of the International Society for the Prevention and Mitigation of Natural Hazards*, 64(2), 1925–1958.

Kappes, M., Papathoma-Kohle, M., and Keiler, M. (2012c), "Assessing physical vulnerability for multi-hazards using an indicator-based methodology," *Applied Geography*, 32(2), 577–590.

Kattamanchi, S., Krishna, R. T., and Ramesh, D. (2017), "Non-stationary ETAS to model earthquake occurrences affected by episodic aseismic transients." *Earth, Planets and Space*, 69(1), 1 – 14.

Keefer, D. (1994), "The importance of earthquake-induced landslides to long-term slope erosion and slope-failure hazards in seismically active regions," *Geomorphology*, 10 (1-4), 265–284.

— (2002), "Investigating landslides caused by earthquakes - a historical review," *Surveys in Geophysics*, 23(6), 473–510.

Keefer, D. and Wilson, R. (1989), "Predicting earthquake-induced landslides, with emphasis on arid and semi-arid environments," *Landslides in a semi-arid environment*, 2(PART 1), 118–149.

Keefer, D. K. (1984), "Landslides caused by earthquakes." *Bulletin of the Geological Society of America*, 95(4), 406–421.

Khazai, B. and Sitar, N. (2004), "Evaluation of factors controlling earthquake-induced landslides caused by Chi-Chi earthquake and comparison with the Northridge and Loma Prieta events," *Engineering Geology*, 79–95.

Kirschbaum, D., Adler, R., Hong, Y., Hill, S., and Lerner-Lam, A. (2010), "A global landslide catalog for hazard applications: method, results, and limitations," *Natural Hazards*, 52(3), 561–575.

Knopoff, L. (1971), "A stochastic model for the occurrence of main-sequence earthquakes," *Reviews of Geophysics*, 9(1), 175–188.

Komendantova, N., Mrzyglocki, R., Mignan, A., Khazai, B., Wenzel, F., Patt, A., and Fleming, K. (2014), "New multi-hazard and multi-risk assessment methods for Europe. Risk governance and the communication process from science to policy: evaluating perceptions of stakeholders from practice in multi-hazard and multi-risk decision support models: MATRIX results II and reference report," doi:10.1016/j.ijdrr.2013.12.006.

Komendantova, N., Scolobig, A., Garcia-Aristizábal, A., Monfort, D., and Fleming, K. (2016), "Multi-risk approach and urban resilience," *International Journal of Disaster Resilience in the Built Environment*, 7(2), 114–132.

Korup, O. (2004), "Geomorphometric characteristics of New Zealand landslide dams," *Engineering Geology*, 73(1-2), 13–35.

— (2005), "Distribution of landslides in southwest New Zealand," *Landslides*, 2(1), 43–51.

Kritikos, T., Robinson, T., and Davies, T. (2015), "Regional coseismic landslide hazard assessment without historical landslide inventories: A new approach," *Journal of Geophysical Research F: Earth Surface*, 120(4), 711–729.

Kumasaki, M., King, M., Arai, M., and Yang, L. (2016), "Anatomy of cascading natural disasters in Japan: main modes and linkages," *Journal of the International Society for the Prevention and Mitigation of Natural Hazards*, 80(3), 1425–1441.

Lari, S., Frattini, P., and Crosta, G. (2014a), "A probabilistic approach for landslide hazard analysis," *Engineering Geology*, 182, 3 – 14.

— (2014b), "A probabilistic approach for landslide hazard analysis," *Engineering Geology*, 182, 3–14.

Lawless, J. (1987), "Regression methods for Poisson process data," *Journal of the American Statistical Association*, 82(399), 808–815.

Lee, C. (2014a), "Multi-stage statistical landslide hazard analysis: earthquake-induced landslides," *Landslide Science for a Safer Geoenvironment*, 3, 205–211.

— (2014b), "Statistical seismic landslide hazard analysis: An example from Taiwan," *Engineering Geology*, 182(PB), 201–212.

Lee, K. and Rosowsky, D. (2006), "Fragility analysis of woodframe buildings considering combined snow and earthquake loading," *Structural Safety*, 28(3), 289–303.

Lee, S. and Evangelista, D. (2006), "Earthquake-induced landslide-susceptibility mapping using an artificial neural network," *Natural Hazards and Earth System Science*, 6.

Li, C., Ma, T., Zhu, X., and Li, W. (2011), "The power–law relationship between landslide occurrence and rainfall level," *Geomorphology*, 130(3), 221 – 229.

Li, N., Tang, C., Yang, T., and Chen, M. (2020), "Analysing post-earthquake landslide susceptibility using multi-temporal landslide inventories — a case study in Miansi Town of China," *Journal of Mountain Science*, 17(2), 358–372.

Li, W.-L., Huang, R.-Q., Xu, Q., and Tang, C. (2013), "Rapid susceptibility mapping of co-seismic landslides triggered by the 2013 Lushan Earthquake using the regression

model developed for the 2008 Wenchuan Earthquake," *Journal of Mountain Science*, 10(5), 699–715.

Liao, H.-M., Yang, X.-G., Xu, F.-G., Xu, H., and Zhou, J.-W. (2018), "A fuzzy comprehensive method for the risk assessment of a landslide-dammed lake," *Environmental Earth Sciences*, 77, 750.

Lin, C.-W., Liu, S.-H., Lee, S.-Y., and Liu, C.-C. (2006), "Impacts of the Chi-Chi earthquake on subsequent rainfall-induced landslides in central Taiwan," *Engineering Geology*, 86(2), 87 – 101.

Little, R. J. A. and Rubin, D. B. (1986), *Statistical Analysis with Missing Data*, New York, NY, USA: John Wiley & Sons, Inc.

Liu, B., Siu, Y., and Mitchell, G. (2016), "Hazard interaction analysis for multi-hazard risk assessment: a systematic classification based on hazard-forming environment," *Natural Hazards and Earth System Sciences*, 16(2), 629–642.

Liu, Z., Nadim, F., Garcia-Aristizábal, A., Mignan, A., Fleming, K., and Luna, B. (2015), "A three-level framework for multi-risk assessment," *Georisk*, 9(2), 59–74.

Lombardi, A. and Marzocchi, W. (2010), "The ETAS model for daily forecasting of italian seismicity in the CSEP experiment," *Annals of Geophysics*, 53(3), 155–164.

Londoño-Linares, J. (2017), "Càlculo de susceptibilidad a deslizamientos mediante análisis discriminante. Aplicación a escala regional," *DYNA*, 84(201), 278–289.

Lu, C., Harte, D., and Bebbington, M. S. (1999), "A linked stress release model for historical Japanese earthquakes: coupling among major seismic regions," *Earth, Planets and Space*, 51(9), 907–916.

Lunn, D., Jackson, C., Best, N., Thomas, A., and Spiegelhalter, D. (2012), *The Bugs Book: A Practical Introduction to Bayesian Analysis*, CRC Press.

Mahendra, R., Mohanty, P., Bisoyi, H., Srinivasa Kumar, T., and Nayak, S. (2011), "Assessment and management of coastal multi-hazard vulnerability along the Cuddalore–Villupuram, east coast of India using geospatial techniques," *Ocean & Coastal Management*, 54(4), 302 – 311.

Malamud, B. D., Turcotte, D. L., Guzzetti, F., and Reichenbach, P. (2004), "Landslide inventories and their statistical properties," *Earth Surface Processes and Landforms*, 29(6), 687–711.

Marc, O., Hovius, N., and Meunier, P. (2016), "The mass balance of earthquakes and earthquake sequences," *Geophysical Research Letters*, 43(8), 3708–3716.

Marc, O., Hovius, N., Meunier, P., Uchida, T., and Hayashi, S. (2015), "Transient changes of landslide rates after earthquakes," *Geology*, 43(10), 883–886.

Martelloni, G., Segoni, S., Fanti, R., and Catani, F. (2012), "Rainfall thresholds for the forecasting of landslide occurrence at regional scale," *Landslides*, 9(4), 485–495.

Martha, T., Roy, P., Govindharaj, K., Kumar, K., Diwakar, P., and Dadhwal, V. (2014), "Landslides triggered by the June 2013 extreme rainfall event in parts of Uttarakhand state, India," *Landslides*, 12(1), 135–146.

Martino, S., Antonielli, B., Bozzano, F., Caprari, P., Discenza, M. E., Esposito, C., Fiorucci, M., Iannucci, R., Marmoni, G., and Schilirò, L. (2020), "Landslides triggered after the 16 August 2018 Mw 5.1 Molise earthquake (Italy) by a combination of intense rainfalls and seismic shaking," *Landslides*, 17, 1177–1190.

Marzocchi, W. (2002), "Modeling the stress variations induced by great earthquakes on the largest eruptions of the 20th century," *Journal of Geophysical Research*, 107 (B11), 1–8.

Marzocchi, W., Garcia-Aristizábal, A., Gasparini, P., Mastellone, M., and Di Ruocco, A. (2012), "Basic principles of multi-risk assessment: a case study in Italy," *Journal of the International Society for the Prevention and Mitigation of Natural Hazards*, 62(2), 551–573.

Marzocchi, W., Mastellone, M., Di Ruocco, A., Novelli, P., Romeo, E., and Gasperini, P. (2009), "Principles of multi-risk assessment," *European Commission - Natural risk assessment*, `http://publications.europa.eu/resource/cellar/22eb788f-5d0a-496a-92d4-4759b0b57fde.0001.03/DOC_2`.

Marzocchi, W. and Woo, G. (2009), "Principles of volcano risk metrics: theory and the case study of Mt. Vesuvius and Campi Flegrei," *Journal of Geophysical Research: Solid Earth*, 114(B3), 1–12.

Mason, D., Justice, R., and Saul, G. (2018), "Kaikōura earthquake slope hazards – risk mitigation and network resilience," *NZ Geomechanics News*, (96).

Massey, C., McSaveney, M., and Davies, T. (2013), "Evolution of an overflow channel across the Young River landslide dam, New Zealand," in *Landslide Science and Practice: Risk Assessment, Management and Mitigation*, eds. Margottini, C., Canuti, P., and Sassa, K., Springer, pp. 43–49.

Massey, C., Townsend, D., Rathje, E., Allstadt, K. E., Lukovic, B., Kaneko, Y., Bradley, B., Wartman, J., Jibson, R. W., Petley, D. N., Horspool, N., Hamling,

I., Carey, J., Cox, S., Davidson, J., Dellow, S., Godt, J. W., Holden, C., Jones, K., Kaiser, A., Little, M., Lyndsell, B., McColl, S., Morgenstern, R., Rengers, F. K., Rhoades, D., Rosser, B., Strong, D., Singeisen, C., and Villeneuve, M. (2018), "Landslides triggered by the 14 November 2016 Mw 7.8 Kaikōura earthquake, New Zealand," *Bulletin of the Seismological Society of America*, 108(3B), 1630–1648.

Mateu, J., Uso, J., and Montes, F. (1998), "The spatial pattern of a forest ecosystem," *Ecological Modelling*, 108(1), 163 – 174.

McKillop, R. and Clague, J. (2007), "A procedure for making objective preliminary assessments of outburst flood hazard from moraine-dammed lakes in southwestern British Columbia," *Natural Hazards*, 41(1), 131–157.

Mekis, E. and Hogg, W. (1999), "Rehabilitation and analysis of Canadian daily precipitation time series," *Atmosphere-Ocean*, 37(1), 53–85.

Melton, M. A. (1965), "The Geomorphic and Paleoclimatic Significance of Alluvial Deposits in Southern Arizona," *The Journal of Geology*, 73(1), 1–38.

Meunier, P., Hovius, N., and Haines, A. J. (2007), "Regional patterns of earthquake-triggered landslides and their relation to ground motion," *Geophysical Research Letters*, 34(20), 1–5.

Meyer, S., Elias, J., and Höhle, M. (2012), "A Space-Time Conditional Intensity Model for Invasive Meningococcal Disease Occurrence," *Biometrics*, 68(2), 607–616.

Mignan, A., Wiemer, S., and Giardini, D. (2014), "The quantification of low-probability-high-consequences events: part I. A generic multi-risk approach," *Journal of the International Society for the Prevention and Mitigation of Natural Hazards*, 73(3), 1999–2022.

Minder, J. R., Roe, G. H., and Montgomery, D. R. (2009), "Spatial patterns of rainfall and shallow landslide susceptibility," *Water Resources Research*, 45(4), 1–11.

Mohler, G., Short, M., Brantingham, P., Schoenberg, F., and Tita, G. (2011), "Self-exciting point process modeling of crime," *Journal of the American Statistical Association*, 106(493), 100–108.

Monsieurs, E., Dewitte, O., and Demoulin, A. (2019), "A susceptibility-based rainfall threshold approach for landslide occurrence," *Natural Hazards and Earth System Sciences*, 19(4), 775–789.

Montrasio, L., Valentino, R., and Luca Losi, G. (2012), "Shallow landslides triggered by rainfalls: Modeling of some case histories in the Reggiano Apennine (Emilia-Romagna Region, Northern Italy)," *Natural Hazards*, 60, 1231–1254.

Muntohar, A. and Liao, H.-J. (2010), "Rainfall infiltration: Infinite slope model for landslides triggering by rainstorm," *Natural Hazards*, 54(3), 967–984.

Musmeci, F. and Vere-Jones, D. (1992), "A space-time clustering model for historical earthquakes," *Annals of the Institute of Statistical Mathematics*, 44(1), 1–11.

National Science Challenges (2019), "Resilience to Nature's Challenge," `http://https://resiliencechallenge.nz/`.

Nguyen, V. and Kim, Y.-T. (2019), "Rainfall-earthquake-induced landslide hazard prediction by Monte Carlo simulation: a case study of Mt. Umyeon in Korea," *KSCE Journal of Civil Engineering*, 24.

Nibigira, L., Havenith, H.-B., Archambeau, P., and Dewals, B. (2018), "Formation, breaching and flood consequences of a landslide dam near Bujumbura, Burundi," *Natural Hazards and Earth System Sciences*, 18(7), 1867–1890.

Nistor, M. M. (2016), "Spatial distribution of climate indices in the Emilia-Romagna region," *Meteorological Applications*, 23(2), 304–313.

NIWA (2012), "NIWA 1981-2010 mean monthly rainfall data," `https://niwa.co.nz/education-and-training/schools/resources/climate/meanrain`.

Ogata, Y. (1988), "Statistical models for earthquake occurrences and residual analysis for point processes," *Journal of the American Statistical Association*, 83(401), 9–27.

— (1998a), "Space-time point-process models for earthquake occurrences," *Annals of the Institute of Statistical Mathematics*, 50(2), 379–402.

— (1998b), "Space-time point-process models for earthquake occurrences," *Annals of the Institute of Statistical Mathematics*, 50(2), 379–402.

— (1999), "Seismicity analysis through point-process modeling: a review," *Pure and Applied Geophysics*, 155(2), 471–507.

Ogata, Y. and Vere-Jones, D. (1984), "Inference for earthquake models: a self-correcting model," *Stochastic Processes and their Applications*, 17(2), 337 – 347.

Ogata, Y. and Zhuang, J. (2006), "Space-time ETAS models and an improved extension," *Tectonophysics*, 413(1-2), 13–23.

Ohlmacher, G. and Davis, J. (2003), "Using multiple logistic regression and GIS technology to predict landslide hazard in northeast Kansas, USA," *Engineering Geology*, 69(3-4), 331–343.

Omori, F. (1894), "On the aftershocks of earthquakes," *Journal of the College of Science*, 7, 111–200.

Onof, C., Chandler, R., Kakou, A., Northrop, P., Wheater, H., and Isham, V. (2000), "Rainfall modelling using poisson-cluster processes: a review of developments," *Stochastic Environmental Research and Risk Assessment*, 14(6), 384–411.

Orchiston, C., Mitchell, J., Wilson, T., Langridge, R., Davies, T., Bradley, B., Johnston, D., Davies, A., Becker, J., and McKay, A. (2018), "Project AF8: developing a coordinated, multi-agency response plan for a future great Alpine Fault earthquake," *New Zealand Journal of Geology and Geophysics*, 61(3), 389–402.

Owen, L., Kamp, U., Khattak, G., Harp, E., Keefer, D., and Bauer, M. (2008), "Landslides triggered by the 8 October 2005 Kashmir earthquake," *Geomorphology*, 94(1), 1 – 9.

Oxford Dictionary (2010), *Oxford Dictionary of English - Angus Stevenson*, Oxford University Press.

Palenzuela, J., Jiménez-Perálvarez, J., Chacón, J., and Irigaray, C. (2016), "Assessing critical rainfall thresholds for landslide triggering by generating additional information from a reduced database: an approach with examples from the Betic Cordillera (Spain)," *Natural Hazards*, 84(1), 185–212.

Pan, H.-L., Jiang, Y.-J., Wang, J., and Ou, G.-Q. (2018), "Rainfall threshold calculation for debris flow early warning in areas with scarcity of data," *Natural Hazards and Earth System Sciences*, 18(5), 1395–1409.

Papathoma, M., Dominey-Howes, D., Zong, Y., and Smith, D. (2003), "Assessing tsunami vulnerability, an example from Herakleio, Crete," *Natural Hazards and Earth System Sciences*, 3(5), 377–389.

— (2015), "Loss estimation for landslides in mountain areas - An integrated toolbox for vulnerability assessment and damage documentation," *Environmental Modelling & Software*, 63, 156–169.

Papathoma-Kohle, M., Neuhäauser, B., Ratzinger, K., Wenzel, H., and Dominey-Howes, D. (2007), "Elements at risks as a framework for assessing for the vulnerability of communities to landslides," *Natural Hazards and Earth System Sciences*, 7 (6), 765–779.

Park, J., Onof, C., and Kim, D. (2019), "A hybrid stochastic rainfall model that reproduces some important rainfall characteristics at hourly to yearly timescales," *Hydrology and Earth System Sciences*, 23(2), 989–1014.

Parker, R., Hancox, G., Petley, D., Massey, C., Densmore, A., and Rosser, N. (2015), "Spatial distributions of earthquake-induced landslides and hillslope preconditioning in the Northwest South Island, New Zealand," *Earth Surface Dynamics*, 3(4), 501–525.

Parker, R. N., Rosser, N. J., and Hales, T. C. (2017), "Spatial prediction of earthquake-induced landslide probability," *Natural Hazards and Earth System Sciences Discussions*, 2017, 1–29.

Pavel, T., Hasan, S., Halim, N., and Mozumder, P. (2018), "Natural hazards and internal migration: The role of transient versus permanent shocks," GLO Discussion Paper Series 255, Global Labor Organization (GLO), `https://ideas.repec.org/p/zbw/glodps/255.html`.

Peng, J., Fan, Z., Wu, D., Zhuang, J., Dai, F., Chen, W., and Zhao, C. (2015), "Heavy rainfall triggered loess–mudstone landslide and subsequent debris flow in Tianshui, China," *Engineering Geology*, 186, 79 – 90.

Peruccacci, S., Brunetti, M., Luciani, S., Vennari, C., and Guzzetti, F. (2012), "Lithological and seasonal control on rainfall thresholds for the possible initiation of landslides in central Italy," *Geomorphology*, 139-140, 79–90.

Peruccacci, S., Brunetti, M. T., Gariano, S. L., Melillo, M., Rossi, M., and Guzzetti, F. (2017), "Rainfall thresholds for possible landslide occurrence in Italy," *Geomorphology*, 290(Supplement C), 39–57.

Pezzo, G., Boncori, J., Tolomei, C., Salvi, S., Atzori, S., Antonioli, A., Trasatti, E., Novali, F., Serpelloni, E., Candela, L., and Giuliani, R. (2013), "Coseismic deformation and source modeling of the May 2012 Emilia (Northern Italy) earthquakes," *Seismological Research Letters*, 84(4), 645–655.

Piacentini, D., Troiani, F., Daniele, G., and Pizziolo, M. (2018), "Historical geospatial database for landslide analysis: the Catalogue of Landslide Occurrences in the Emilia-Romagna Region (CLOCkER)," *Landslides*, 15(4), 811–822.

Pizziolo, M., Bernardi, M., Daniele, G., Generali, M., and Piacentini, D. (2015), "Landslide occurrences during the high-intensity rainfall event of March–April 2013 in the Emilia-Romagna Region (North Apennines, Italy)," in *Engineering Geology for Society and Territory-Volume 5*, Springer, pp. 777–780.

Polemio, M. and Sdao, F. (1999), "The role of rainfall in the landslide hazard: the case of the Avigliano urban area (Southern Apennines, Italy)," *Engineering Geology*, 53 (3), 297 – 309.

Ponziani, F., Pandolfo, C., Stelluti, M., Berni, N., Brocca, L., and Moramarco, T. (2012), "Assessment of rainfall thresholds and soil moisture modeling for operational hydrogeological risk prevention in the Umbria region (central Italy)," *Landslides*, 9 (2), 229–237.

Pradhan, B. and Lee, S. (2010), "Regional landslide susceptibility analysis using back-propagation neural network model at Cameron Highland, Malaysia," *Landslides*, 7 (1), 13–30.

Preisler, H., Brillinger, D., Burgan, R., and Benoit, J. (2004), "Probability based models for estimation of wildfire risk," *International Journal of Wildland Fire*, 13(2), 133–142.

Qi, S., Xu, Q., Lan, H., Zhang, B., and Liu, J. (2010), "Spatial distribution analysis of landslides triggered by 2008.5.12 Wenchuan Earthquake, China," *Engineering Geology*, 116(1-2), 95–108.

Reese, S., King, A., Bell, R., Schmidt, J., Pringle, R., and Henderson, R. (2007), "Regional RiskScape: a multi-hazard loss modelling tool," *International Congress on Modelling and Simulation*, 74–80.

Reid, H. F. (1910), "The mechanics of the earthquake - The California earthquake of April 18, 1906, report of the state earthquake investigation commission," *Report of the State Earthquake Investigation Commission*, 2, 16–28, `http://activetectonics.asu.edu/ActiveFaultingSeminar/Papers/Reid_1910.pdf`.

Reinhart, A. (2017), "A Review of Self-Exciting Spatio-Temporal Point Processes and Their Applications." 33, 1–30.

Richter, C. F. (1958), *Elementary seismology.*, W. H. Freeman.

Robinson, T., Davies, T., Wilson, T., and Orchiston, C. (2016b), "Coseismic landsliding estimates for an Alpine Fault earthquake and the consequences for erosion of the Southern Alps, New Zealand," *Geomorphology*, 263(15), 271–86.

Robinson, T., Davies, T., Wilson, T., Orchiston, C., and Barth, N. (2016a), "Evaluation of coseismic landslide hazard on the proposed Haast-Hollyford Highway, South Island, New Zealand," *Georisk*, 10(2), 146–163.

Rodriguez, C., Bommer, J., and Chandler, R. (1999), "Earthquake-induced landslides: 1980–1997," *Soil Dynamics and Earthquake Engineering*, 18(5), 325 – 346.

Rodriguez-Iturbe, I., Cox, D., and Isham, V. (1987), "Some models for rainfall based on stochastic Point processes," *Proceedings of The Royal Society of London, Series A: Mathematical and Physical Sciences*, 410(1839), 269–288.

Rodriguez-Iturbe, I. and Eagleson, P. (1987), "Mathematical models of rainstorm events in space and time," *Water Resources Research*, 23(1), 181–190.

Rodriguez-Iturbe, I., Gupta, V. K., and Waymire, E. (1984), "Scale considerations in the modeling of temporal rainfall," *Water Resources Research*, 20(11), 1611–1619.

Rossi, M., Luciani, S., Valigi, D., Kirschbaum, D., Brunetti, M., Peruccacci, S., and Guzzetti, F. (2017), "Statistical approaches for the definition of landslide rainfall thresholds and their uncertainty using rain gauge and satellite data," *Geomorphology*, 285, 16 – 27.

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

Rossouw, J., Anderson, G., Prentice, R., LaCroix, A., Kooperberg, C., Stefanick, M., Jackson, R., Beresford, S., Howard, B., Johnson, K., Kotchen, J., and Ockene, J. (2002), "Risks and benefits of estrogen plus progestin in healthy postmenopausal women: Principal results from the women's health initiative randomized controlled trial," *Journal of the American Medical Association*, 288(3), 321–333.

Rotondi, R. and Varini, E. (2019), "Failure models driven by a self-correcting point process in earthquake occurrence modeling," *Stochastic Environmental Research and Risk Assessment*, 33(3), 709–724.

Rubin, D. (1976), "Inference and missing data," *Biometrika*, 63(3), 581–592.

Santi, P. (2018), "Landslide Analysis with Incomplete Data: Examples from Colorado and Wyoming," in *Rocky Mountain Geo-Conference 2018*, American Society of Civil Engineers, pp. 15–25.

Santos, A., Lana, M., Cabral, I., Pereira, T., Zaré, M., de Fátima Santos da Silva, D., and Santos, T. (2019), "Evaluation of rock slope stability conditions through discriminant analysis," *Geotechnical and Geological Engineering*.

Schmidt, J., Matcham, I., Reese, S., King, A., Bell, R., Henderson, R., Smart, G., Cousins, J., Smith, W., and Heron, D. (2011), "Quantitative multi-risk analysis for natural hazards: A framework for multi-risk modelling," *Natural Hazards*, 58(3), 1169–1192.

Schneider, P. and Schauer, B. (2006), "HAZUS-its development and its future," *Natural Hazards Review*, 7.

Schoenberg, F. and Bolt, B. (2000), "Short-term exciting, long-term correcting models for earthquake catalogs," *Bulletin of the Seismological Society of America*, 90(4), 849–858.

Schwartz, D., Lienkaemper, J., Hecker, S., Kelson, K., Fumal, T., Baldwin, J., Seitz, G., and Niemi, T. (2014), "The earthquake cycle in the San Francisco Bay Region: A.D. 1600-2012," *Bulletin of the Seismological Society of America*, 104(3), 1299–1328.

Scott, K., Vallance, J., Kerle, N., Macías, J., Strauch, W., and Devoli, G. (2005), "Catastrophic precipitation-triggered lahar at Casita volcano, Nicaragua: Occurrence, bulking and transformation," *Earth Surface Processes and Landforms*, 30(1), 59–79.

Segalini, A., Valletta, A., and Carri, A. (2018), "Landslide time-of-failure forecast and alert threshold assessment: A generalized criterion," *Engineering Geology*, 245, 72 – 80.

Selva, J. (2013), "Long-term multi-risk assessment: statistical treatment of interaction among risks," *Journal of the International Society for the Prevention and Mitigation of Natural Hazards*, 67(2), 701–722.

Selva, J., Argyroudis, S., and Pitilakis, K. (2013), "Impact on loss/risk assessments of inter-model variability in vulnerability analysis," *Natural Hazards*, 63(2), 723–746.

Shou, K., Hong, C., Wu, C., Hsu, H., Fei, L., Lee, J., and Wei, C. (2011), "Spatial and temporal analysis of landslides in Central Taiwan after 1999 Chi-Chi earthquake," *Engineering Geology*, 123(1), 122 – 128.

Smith, W. (1958), "Renewal theory and its ramifications," *Journal of the Royal Statistical Society. Series B (Methodological)*, 20(2), 243–302.

Sornette, D. and Werner, M. (2005), "Apparent clustering and apparent background earthquakes biased by undetected seismicity," *Journal of Geophysical Research: Solid Earth*, 110(B9), 1–13.

Steger, S., Brenning, A., Bell, R., and Glade, T. (2017), "The influence of systematically incomplete shallow landslide inventories on statistical susceptibility models and suggestions for improvements," *Landslides*, 14(5), 1767–1781.

Steger, S., Brenning, A., Rainer, B., Petschko, H., and Glade, T. (2016), "Exploring discrepancies between quantitative validation results and the geomorphic plausibility of statistical landslide susceptibility maps," *Geomorphology*, 262, 8 – 23.

Stirling, M., McVerry, G., Gerstenberger, M., Litchfield, N., Van Dissen, R., Berryman, K., Barnes, P., Wallace, L., Villamor, P., Langridge, R., Lamarche, G., Nodder, S., Reyners, M., Bradley, B., Rhoades, D., Smith, W., Nicol, A., Pettinga, J., Clark, K., and Jacobs, K. (2012), "National seismic hazard model for New Zealand: 2010 update," *Bulletin of the Seismological Society of America*, 102(4), 1514–1542.

Stupp, R., Mason, W., Van Den Bent, M., Weller, M., Fisher, B., Taphoorn, M., Belanger, K., Brandes, A., Marosi, C., Bogdahn, U., Curschmann, J., Janzer, R., Ludwin, S., Gorlia, T., Allgeier, A., Lacombe, D., Cairncross, J., Eisenhauer, E., and Mirimanoff, R. (2005), "Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma," *New England Journal of Medicine*, 352(10), 987–996.

Swanson, F., Oyagi, N., and Tominaga, M. (1986), "Landslide dams in Japan," in *Landslide dams: process risk and mitigation*, ed. Schuster, R. L., American Society of Civil Engineers, New York, vol. 3, pp. 131–145.

Syvitski, J. and Schafer, C. (1996), "Evidence for an earthquake-triggered basin collapse in Saguenay Fjord, Canada," *Sedimentary Geology*, 104(1-4), 127–153.

Tacconi Stefanelli, C., Casagli, N., and Catani, F. (2020), "Landslide damming hazard susceptibility maps: a new GIS-based procedure for risk management." *Landslides*, 1.

Tacconi Stefanelli, C., Catani, F., and Casagli, N. (2015), "Geomorphological investigations on landslide dams," *Geoenvironmental Disasters*, 2(1), 21.

Tacconi Stefanelli, C., Segoni, S., Casagli, N., and Catani, F. (2016), "Geomorphic indexing of landslide dams evolution," *Engineering Geology*, 208, 1–10.

Tacconi Stefanelli, C., Segoni, S., Casagli, N., and Catani, F. (2016), "Geomorphic indexing of landslide dams evolution," *Engineering Geology*, 208, 1 – 10.

Tacconi Stefanelli, C., Segoni, S., Casagli, N., and Catani, F. (2016), "Geomorphological analysis for landslide dams," in *Landslides and engineered slopes. Experience, theory and practice*, eds. Aversa, S., Cascini, L., Picarelli, L., and Scavia, C., CRC Press, vol. 3, pp. 1883–1887.

Tang, C., Westen, C., Tanyas, H., and Jetten, V. (2016), "Analysing post-earthquake landslide activity using multi-temporal landslide inventories near the epicentral area of the 2008 Wenchuan earthquake," *Natural Hazards and Earth System Sciences*, 16, 2641–2655.

Tang, C., Zhu, J., Li, W.-l., and Liang, J. (2009), "Rainfall-triggered debris flows following the Wenchuan earthquake," *Bulletin of Engineering Geology and the Environment*, 68, 187–194.

Thierry, P., Stieltjes, L., Kouokam, E., Nguéya, P., and Salley, P. M. (2008), "Multi-hazard risk mapping and assessment on an active volcano: the GRINP project at Mount Cameroon," *Natural Hazards*, 45(3), 429–456.

Tian, Y., Xu, C., Ma, S., Xu, X., Wang, S., and Zhang, H. (2019), "Inventory and Spatial Distribution of Landslides Triggered by the 8th August 2017 M W 6.5 Jiuzhaigou Earthquake, China," *Journal of Earth Science*, 30(1), 206–217.

Tilloy, A., Malamud, B. D., Winter, H., and Joly-Laugel, A. (2019), "A review of quantification methodologies for multi-hazard interrelationships," *Earth-Science Reviews*, 196, 102881.

Tiranti, D. and Rabuffetti, D. (2010), "Estimation of rainfall thresholds triggering shallow landslides for an operational warning system implementation," *Landslides*, 7 (4), 471–481.

Toda, S., Stein, R., Reasenberg, P., Dieterich, J., and Yoshida, A. (1998), "Stress transferred by the 1995 Mw = 6.9 Kobe, Japan, shock: Effect on aftershocks and future earthquake probabilities," *Journal of Geophysical Research: Solid Earth*, 103 (10), 24543–24565.

Towhata, I., Yamada, S., Toyota, H., and Qureshi, M. (2013), "Long term effects of strong earthquake shaking on slope instability; Lessons from recent seismic events," in *14th European Conference on Earthquake Engineering 2010*, pp. 1–11.

Travasarou, T., Bray, J. D., and Abrahamson, N. A. (2003), "Empirical attenuation relationship for Arias Intensity," *Earthquake Engineering & Structural Dynamics*, 32 (7), 1133–1155.

Trigila, A., Iadanza, C., and Spizzichino, D. (2010), "Quality assessment of the Italian Landslide Inventory using GIS processing," *Landslides*, 7(4), 455–470.

TRMM (2017), "GPM TMI on TRMM unpacked data L1A 1.5 hours 13 km V05," `https://disc.gsfc.nasa.gov/datasets/GPM_1ATMI_05/summary`.

Troiani, F., Piacentini, D., Della Seta, M., and Galve, J. (2017), "Stream length-gradient hotspot and cluster analysis (SL-HCA) to fine-tune the detection and interpretation of knickzones on longitudinal profiles," *CATENA*, 156, 30–41.

Tsou, C.-Y., Feng, Z.-Y., and Chigira, M. (2011), "Catastrophic landslide induced by Typhoon Morakot, Shiaolin, Taiwan," *Geomorphology*, 127(3-4), 166–178.

Uchiogi, T. (1971), "Landslides due to one continual rainfall," *Journal of the Japan Society of Erosion Control Engineering*, 23(4), 21–34.

Umar, Z., Pradhan, B., Ahmad, A., Jebur, M., and Tehrany, M. (2014), "Earthquake induced landslide susceptibility mapping using an integrated ensemble frequency ratio and logistic regression models in West Sumatera Province, Indonesia," *CATENA*, 118, 124 – 135.

Umbal, J. and Rodolfo, K. (1996), "The 1991 lahars of Southwestern Mount Pinatubo and evolution of the lahar-dammed Mapanuepe lake," in *Fire and Mud: Eruptions and Lahars of Mount Pinatubo, Philippines*, University of Washington Press, pp. 951–970.

UNDDR (2016), "Report of the open-ended intergovernmental expert working group on indicators and terminology relating to disaster risk reduction," `https://www.preventionweb.net/files/50683_oiewgreportenglish.pdf`.

— (2019), "Global assessment report on disaster risk reduction," `https://www.undrr.org/publication/global-assessment-report-disaster-risk-reduction-2019`.

USGS (2019a), "U.S. Geological Survey - Earthquake Hazards," `https://www.usgs.gov/natural-hazards/earthquake-hazards`.

— (2019b), "U.S. Geological Survey - Landslide hazards," `https://www.usgs.gov/natural-hazards/landslide-hazards`.

Utsu, T. (1969), "Some problems of the distribution of earthquakes in time (1)," *Geophysical Bulletin, Hokkaido University*, 22, 73–93.

— (1970), "Aftershocks and earthquake statistics (2) : further investigation of aftershocks and other earthquake sequences based on a new classification of earthquake sequences," *Journal of the Faculty of Science, Hokkaido University, Series 7, Geophysics*, 4, 197–266.

Utsu, T. and Ogata, Y. Matsu'ura, S. (1995), "The centenary of the Omori formula for a decay law of aftershock activity," *Journal of Physics of the Earth*, 43(1), 1–33.

Vai, G. B. and Martini, I. P. (2001), *Anatomy of an Orogen: the Apennines and Adjacent Mediterranean Basins*, Springer.

van Lieshout, M. N. M. and Stein, A. (2012), "Earthquake modelling at the country level using aggregated spatio-temporal point processes," *Mathematical Geosciences*, 44(3), 309–326.

Vannoli, P., Burrato, P., and Valensise, G. (2015), "The Seismotectonics of the Po Plain (Northern Italy): Tectonic Diversity in a Blind Faulting Domain," *Pure and Applied Geophysics*, 172(5), 1105–1142.

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

Vere-Jones, D. (1978), "Earthquke prediction - a statistician's view," *Journal of Physics of the Earth*, 26(2), 129–146.

Vere-Jones, D. and Deng, Y. (1988), "A point process analysis of historical earthquakes from North China," *Earthquake Research in China*, 2(2), 165–181.

Vessia, G., Parise, M., and Tromba, G. (2013), "A strategy to address the task of seismic micro-zoning in landslide-prone areas," *Advances in Geosciences*, 35, 23–35.

Wang, B., Zhang, T., Zhou, Q., Wu, C., Chen, Y.-L., and Wu, P. (2015), "A case study of the Tangjiashan landslide dam-break," *Journal of Hydrodynamics*, 27(2), 223 – 233.

Wang, T., Bebbington, M., and Harte, D. (2012), "Markov-modulated Hawkes process with stepwise decay," *Annals of the Institute of Statistical Mathematics*, 64, 521–544.

Wang, T., Schofield, M., Bebbington, M., and Kiyosugi, K. (2020), "Bayesian modelling of marked point processes with incomplete records: volcanic eruptions," *Journal of the Royal Statistical Society. Series C: Applied Statistics*, 69(1), 109–130.

Wei, X.-L., Chen, N.-S., Cheng, Q.-G., He, N., Deng, M.-F., and Tanoli, J. (2014), "Long-term activity of earthquake-induced landslides: A case study from Qionghai Lake Basin, Southwest of China," *Journal of Mountain Science*, 11(3), 607–624.

Wells, D. L. and Coppersmith, K. J. (1994), "New empirical relationships among magnitude, rupture length, rupture width, rupture area, and surface displacement," *Bulletin of the Seismological Society of America*, 84(4), 974–1002.

Welsh, A. and Davies, T. (2011), "Identification of alluvial fans susceptible to debris-flow hazards," *Landslides*, 8(2), 183–194.

Wetzler, N., Brodsky, E. E., and Lay, T. (2016), "Regional and stress drop effects on aftershock productivity of large megathrust earthquakes," *Geophysical Research Letters*, 43(23), 12,012–12,020.

Whitehouse, I. and Griffiths, G. (1983), "Frequency and hazard of large rock avalanches in the central Southern Alps, New Zealand." *Geology*, 11(6), 331–334.

Wiemer, S., Gerstenberger, M., and Hauksson, E. (2002), "Properties of the aftershock sequence of the 1999 Mw 7.1 Hector Mine earthquake: Implications for aftershock hazard," *Bulletin of the Seismological Society of America*, 92(4), 1227–1240.

Witt, A., Malamud, B. D., Rossi, M., Guzzetti, F., and Peruccacci, S. (2010), "Temporal correlations and clustering of landslides," *Earth Surface Processes and Landforms*, 35(10), 1138–1156.

Wotton, B. and Martell, D. (2005), "A lightning fire occurrence model for Ontario," *Canadian Journal of Forest Research*, 35(6), 1389–1401.

Xu, C. (2015), "Preparation of earthquake-triggered landslide inventory maps using remote sensing and GIS technologies: Principles and case studies," *Geoscience Frontiers*, 6(6), 825 – 836.

Xu, H. and Schoenberg, F. (2011), "Point process modeling of wildfire hazard in Los Angeles county, California," *The Annals of Applied Statistics*, 5(2), 684–704.

Yalcin, A. (2008), "GIS-based landslide susceptibility mapping using analytical hierarchy process and bivariate statistics in Ardesen (Turkey): Comparisons of results and confirmations," *CATENA*, 72(1), 1–12.

Yan, Y., Cui, P., Chen, S.-C., Chen, X.-Q., Chen, H.-Y., and Chien, Y.-L. (2017), "Characteristics and interpretation of the seismic signal of a field-scale landslide dam failure experiment," *Journal of Mountain Science*, 14(2), 219–236.

Yang, J., Weisberg, P., Dilts, T., Loudermilk, L., Scheller, R., Stanton, A., and Skinner, C. (2015), "Predicting wildfire occurrence distribution with spatial point process models and its uncertainty assessment: a case study in the Lake Tahoe Basin, USA," *International Journal of Wildland Fire*, 24, 380–390.

Yasuo, M., Fukashi, M., and Setsuya, N. (2015), "The October 16, 2013 rainfall-induced landslides and associated lahars at Izu Oshima Volcano, Japan," *Journal of Volcanology and Geothermal Research*, 302, 242 – 256.

Yesilnacar, E. and Topal, T. (2005), "Landslide susceptibility mapping: A comparison of logistic regression and neural networks methods in a medium scale study, Hendek region (Turkey)," *Engineering Geology*, 79(3), 251 – 266.

Yin, Y., Wang, F., and Sun, P. (2009), "Landslide hazards triggered by the 2008 Wenchuan earthquake, Sichuan, China," *Landslides*, 6(2), 139–151.

Zhang, S., Xie, X., Wei, F., Chernomorets, S., Petrakov, D., Pavlova, I., and Tellez, R. D. (2015), "A seismically triggered landslide dam in Honshiyan, Yunnan, China: from emergency management to hydropower potential," *Landslides*, 12(6), 1147–1157.

Zhang, T. and Huang, Y.-N. (2017), "Gradient angle-based analysis for spatiotemporal point processes." *Electronic Journal of Statistics*, 11(2), 4424–4451.

Zhou, G., Cui, P., Zhu, X., Tang, J., Chen, H., and Sun, Q. (2015a), "A preliminary study of the failure Mechanisms of cascading landslide dams," *International Journal of Sediment Research*, 30, 223 – 234.

Zhou, H.-J., Wang, X., and Yuan, Y. (2015b), "Risk assessment of disaster chain: experience from Wenchuan earthquake-induced landslides in China," *Journal of Mountain Science*, 12(5), 1169–1180.

Zhuang, J. (2011), "Next-day earthquake forecasts for the Japan region generated by the ETAS model," *Earth, Planets and Space*, 63(3), 207–216.

Zhuang, J., Harte, D., Werner, M., Hainzl, S., and Zhou, S. (2012), "Basic models of seismicity: temporal models," `http://www.corssa.org/export/sites/corssa/.galleries/articles-pdf/Zhuang-et-al-2012-CORSSA-Temporal-models.pdf`.

Zhuang, J., Ogata, Y., and Vere-Jones, D. (2004), "Analyzing earthquake clustering features by using stochastic reconstruction," *Journal of Geophysical Research: Solid Earth*, 109(5), B05301 1–17.

Zimmermann, M., Bichsel, M., and Kienholz, H. (1986), "Mountain hazards mapping in the Khumbu Himal, Nepal," *Mountain Research and Development*, 6(1), 29–40.

Zimmermann, M., Mani, P., and Romang, H. (1997), "Magnitude-frequency aspects of alpine debris flows," *Eclogae Geologicae Helvetiae*, 90(3), 415–420.

Zonno, G. and Montaldo Falero, V. (2009), "Analysis of strong ground motions to evaluate regional attenuation relationships," *Annals of Geophysics*, 45, 439–454.

Zuccaro, G., Cacace, F., Spence, R., and Baxter, P. (2008), "Impact of explosive eruption scenarios at Vesuvius," *Journal of Volcanology and Geothermal Research*, 178 (3), 416–453.