Statistical methods applied to volcanology and volcanic hazard assessment

Rosa Sobradelo

Department of Statistics and Operations Research

Universidad Politécnica de Cataluña (UPC)

A thesis submitted for the degree of

Doctor of Philosophy

July 2011

To my family, my partner Addy and our beautiful daughter
Sabela Rose

Acknowledgements

First and foremost I offer my sincerest gratitude to my supervisor, Dr. Joan Martí, who has guided and supported me throughout my thesis with his patience and knowledge whilst allowing me the room to work in my own way. I want to also thank Dr. Martí for providing me with the financial support for this project while being a member of the Group of Volcanology of Barcelona, and allowing me to be part of this great Institution. I attribute the level of my PhD degree to his encouragement and effort and without him this thesis would not have been completed or written. One simply could not wish for a better supervisor. Deepest gratitude is also due to my tutor and co-director, Dr. Guadalupe Gómez, who has offered invaluable assistance, support and guidance. Special thanks go to Dr. Gordon Woo, without whose knowledge and assistance this study would not have been successful.

I would also like to thank Dr. Adelina Geyer, for her support and guidance with her technical knowledge and scientific expertise. Many thanks also to Dr. Ramón Ortiz, Dr. Gerardo Aguirre and once again Dr. Joan Martí for teaching me volcanology during our volcanic field trips.

I want to thank Dr. Christopher Kilburn for my Honorary Research Assistantship at the Aon Benfield UCL Hazard Centre and fellow students Lucy, Carina, Bob, Wendy, Melanie, Clare and Catherine. Also thanks to the Grup de Recerca en Análisi Estadística de la Supervivencia - GRASS from Barcelona, for their support. Thank you to Dr. Warner Marzocchi and Dr. Roberto Carniel for their thorough review of this thesis and their valuable comments, which helped improve the quality of this work.

Thank you to my friend Ricardo Martínez for queuing at the government offices in downtown Manhattan, sorting out all the paperwork required to transfer my masters credits from Courant Institute of Mathematical Sciences (NYU) to my PhD program at the UPC, while Paty, his wife, was days away from giving birth to Diego and Andrea. Also thanks to my dearest friend Monica, for her constant encouragement and support, and to all my friends and family, in particular my mother and brother. Last, but not least, special thanks go to my wonderful boyfriend Addy, to whom this thesis is dedicated, for his constant support throughout this project, and for giving me a most precious and beautiful daughter, Sabela Rose.

Contents

Chapter 1

Introduction

Disasters are large intractable problems that test the ability of communities and nations to effectively protect their populations and infrastructure, to reduce both human and property loss, and to rapidly recover. The randomness of impacts and problems, and uniqueness of incidents demand dynamic, real-time, effective and cost efficient solutions. For this reason, we need quantitative risk-based methods for decision-making under uncertainty to be developed and applied to volcanology.

Volcanic activity is a natural phenomenon that can turn into a disaster under certain conditions. It is a natural process that cannot be controlled, but its potentially disastrous effects can be mitigated. Volcanoes have implicit a natural hazard which can threaten human lives and properties of those communities living near by. The eruptions of volcanoes considered "dormant" or "inactive" have been liable for major disasters in the past. The volcanic hazard from volcanoes with a long term recurrence tends to be ignored, especially when little or no historical data exists. This is the case of Teide - Pico Viejo stratovolcanoes in the island of Tenerife.

Due to the limited scientific observability of the interior of a volcano, there is a lot of uncertainty in forecasting volcanic eruptions. During a volcanic crisis decision-makers need to take important life and death decisions under strict time and uncertainty constrains. They are afraid of getting a decision wrong, causing unnecessary economic disruption and public anxiety and distress.

There is an increasing recognition of the need of combining mathematical models, together with statistical and operations research methods to address

disaster management. The interdisciplinary science of mathematics applied to the study of volcanology and volcanic hazard is an important approach, which will help understand volcanic processes by integrating keen volcanological insights with sound statistical modeling and artful application of computational power.

De la Cruz-Reyna [2000] have defined the different phases of a volcanic crisis management. The first is the pre-event or pre-critical phase, and includes the sub-phases of risk assessment, hazard and risk mapping, and postulation of expected scenarios; volcano monitoring; and emergency planning. The second phase is the critical or decision-making phase, which includes alert, communication, and information procedures; response and implementation of emergency measures; and defining the end of the critical phase.

Volcanic hazard is defined as the probability of any particular area being affected by a destructive volcanic event within a given period of time. (Fournier d'Albe 1979). The aim in a volcanic crisis management is to save lives and minimize economic loss in the event of a volcanic eruption. To do this we need to quantify the volcanic hazard in order to build an evacuation model to use before the onset of the eruption. It is very important to note here that the hazard term was used until recently for the scientific community of volcanologists to refer to the different destructive volcanic events derived from a restless volcano. It is very recent that volcanologists started associating the term hazard with the probability of occurrence of a destructive volcanic event. It will take time before scientists become familiar with the probability concept associated to the volcanic hazard. In this sense, this thesis is also a contribution to make the scientific community of volcanologists more familiar with the probabilistic concept associated to hazard. When we talk about volcanic hazard, we now refer to the probability of a destructive volcanic event happening in a given period of time.

The decision to evacuate is easy for remote volcanoes, where the potential economic loss is small and there are no major consequences of a false alarm. The hardest decisions are for volcanoes near to densely populated regions of industrialized nations. In these cases, a prolonged mass evacuation costs millions, and even billions of dollars, and the consequences of an unnecessary evacuation are very costly and castigated.

Additionally, volcano hazard assessment is directly related with the estimation

of the evacuation time. There is a large degree of uncertainty in the evacuation process which needs to be taken into account as well as the volcanological uncertainties. There has been major progress in the study and understanding of the behavior of nonlinear complex systems, including the stochastic dynamics of mass population evacuation.

Traditionally, the scientific community of volcanologists elaborate volcano hazard maps and reports and communicate their scientific views on volcano hazard to the government and civil protection officials, leaving to them the problem of decision making. There is a need to support civil protection officials and government further by establishing probabilistic criteria for evacuation decision-making.

The stochastic nature of many volcanological data has rarely been exploited and the models do not generally produce the type of probabilistic outputs needed for forecasting. In the last decade or so scientists have begun to exploit a wide range of analytical and statistical methods for analyzing stochastic datasets. The primary aim is to develop rigorous methods for quantifying the likelihood of outcomes given the set of current and past observations.

The aim of this thesis is to work with volcanologists to try and address, with the appropriate statistical methods, those questions they raise, and have volcanologists collaborate with statisticians to learn about the advantages in the application of statistical techniques to the interpretation of volcanic data. Here, we propose and analyze different statistical methodologies to interpret volcanic data and assess volcanic hazard. The statistical technique will depend on the nature of the data and the type of problem we want to address. The models will be used to analyze and interpret the historical and geological volcanic data for Teide - Pico Viejo stratovolcanoes (TPV) and the Canary Islands archipelago.

1.1 Some useful definitions

Bayesian inference

Uses Bayes' formula to compute the conditional posterior probability that the hypothesis is true, given what was observed. The calculation involves the unconditional prior probability that the hypothesis is true, as well as the conditional probabilities of getting what was observed given the hypothesis and given the

alternative. The logic of Bayesian inference is sound, but there are practical difficulties in its application. In particular, the prior probability that the hypothesis is true may have to be determined subjectively, the alternative hypothesis must be completely specified.

This represents the subjectivist school. According to Bayesian practice, changes in belief upon receipt of new information are reflected in the transition from a prior to a posterior probability. Bayes' theorem is used as part of the process of learning from new data. Thus if A_i is a particular event, and X is some data, Bayes' theorem can be used to update the prior probability $P(A_i)$ to obtain the posterior probability $P(A_i|X)$ [Aspinall, 2006; Woo, 1999].

Frequentist inference

Is usually based on a p-value, which is the probability of getting what was observed or a result more extreme than that, given that the hypothesis is true. The weakness in the logic is that we are computing probabilities of events that didn't happen, assuming that the hypothesis is true, which it may not be. The alternative hypothesis is used only to determine what events are more extreme than the observed event and need not be completely specified.

In contrast with the subjectivist, the frequentist assigns a measure of uncertainty to an individual event by considering it to belong to a class of similar events having similar randomness properties, and associates probability with some notion of limiting frequency [Woo, 1999].

Event tree

A graphical, tree-like representation of events in which branches are logical steps from a general prior event through increasingly specific subsequent events (intermediate outcomes) to final outcomes. Event trees are used to show possible outcomes of volcanic unrest at progressively higher levels of detail. Probabilities are estimated for each event through the tree. The multiplicative product of probabilities along any one path will yield the probability of the terminal (rightmost) event. For graphical and conceptual simplicity, events at any given level of the tree need not be mutually exclusive or exhaustive [Newhall and Hoblitt, 2002].

Probability tree

A graphical, tree-like representation of the probabilities of comprehensive (ex-

haustive), mutually exclusive events. As above, events are progressively more specific as one moves outward along branches. As above, the multiplicative product of probabilities along any one path will be the probability of the most specific event. However, the requirement that events at any given level of specificity be comprehensive and mutually exclusive means that probabilities of events at that level will sum to 1. This sum of 1 is required if one wishes to know, for example, the total probability of an outcome (e.g., death) that might be reached along different possible paths [Newhall and Hoblitt, 2002].

Volcano

A volcano is a vent in the earth's crust through which molten rock (magma) and rock fragments and gases are ejected from the Earth's interior. A volcano is created when magma erupts onto the surface of the Earth. Volcanoes take many forms according to the chemical composition of their magma and the conditions in which the magma is erupted. When magma is erupted freely in a continuous way it is referred to as lava. Some lava known as 'basalt' is hot and fluid. Opposite of basalts are 'rhyolites', which are characterized by their inability to flow freely, erupt explosively or form steep domes. Midway in between are 'andesites' which are thick, flow slowly, and are mildly explosive [Blong, 2000].

Active Volcano

There are approximately 1,500 active volcanoes worldwide, of which only about 550 are monitored because they are normally next to populated areas and so considered a threat. Although there are no precise or generally accepted definitions for the terms active volcano, dormant volcano and extinct volcano, all volcanoes that have erupted in the last 10,000 years are commonly regarded as active, and to volcanologists this means that the volcano have the possibility of erupting again. Over 75% of the world's active volcanoes are located in the Pacific Ring of Fire, an arc of intense seismic and volcanic activity stretching form New Zealand, along the eastern edge of Asia, north across the Aleutian Islands of Alaska, and south along the coast of North and South America [Blong, 2000].

Extinct Volcano

Extinct volcanoes are those that scientists consider unlikely to erupt again, because the volcano no longer has a lava supply. Examples of extinct volcanoes are many volcanoes on the Hawaiian Emperor seamount chain in the Pacific

Ocean, Hohentwiel, Shiprock and the Zuidwal volcano in the Netherlands. Edinburgh Castle in Scotland is famously located atop an extinct volcano. Otherwise, whether a volcano is truly extinct is often difficult to determine. Since "supervolcano" calderas can have eruptive lifespans sometimes measured in millions of years, a caldera that has not produced an eruption in tens of thousands of years is likely to be considered dormant instead of extinct.

Dormant Volcano

It is difficult to distinguish an extinct volcano from a dormant one. Volcanoes are often considered to be extinct if there are no written records of its activity. Nevertheless, volcanoes may remain dormant for a long period of time. For example, Yellowstone has a repose/recharge period of around 700 ka, and Toba of around 380 ka. Vesuvius was described by Roman writers as having been covered with gardens and vineyards before its famous eruption of AD 79, which destroyed the towns of Herculaneum and Pompeii. Before its catastrophic eruption of 1991, Pinatubo was an inconspicuous volcano, unknown to most people in the surrounding areas. Two other examples are the long-dormant Soufriére Hills volcano on the island of Montserrat, thought to be extinct before activity resumed in 1995 and Fourpeaked Mountain in Alaska, which, before its September 2006 eruption, had not erupted since before 8000 BC and had long been thought to be extinct.

Volcanic Hazard

Volcanic hazard is defined as the probability of any particular area being affected by a destructive volcanic event within a given period of time. [Blong, 2000].

In mathematical notation. Let $N(t)$ = Number of destructive volcanic events arriving to a region $\mu(S)$ up to time t, where $t \in [0, \infty)$, $\mu(S)$ is a finite measure from the spatial region $S \subset V$ and V is some vector space (e.g. $\Re^2$ or $\Re^3$).

Then, $P_{\mu(S)}[N(t + \Delta t) - N(t) \geq 1]$ = Probability of at least one destructive volcanic event arriving to region $\mu(S)$ during the time interval Δt.

Under certain conditions we can say that the collection of random variables $\{N(t) : t \geq 0\}$ is a spatial Poisson Process where the intensity rate $\lambda(\vec{x}, t)$ $(\vec{x} \in V)$ depends on time and space, and is defined as: $\lambda_S(t) = \int_S \lambda(\vec{x}, t) d\mu(\vec{x})$.

It is important to clarify that the concept of hazard here is not equivalent to the hazard function known in Statistical Science and more specifically in Survival

Analysis. Further research is needed to establish the relationship between the hazard function and the concept of volcanic hazard.

Hazard Function

Let T be a random variable which denotes the elapsed time to an event E, the event could be a restless volcano, a pyroclastic flow, or simply the evacuation alarm. When T is a continuous random variable, the hazard function is defined to be:

$$\lambda(t) = \lim_{\delta t \to 0} \frac{1}{\delta t} Pr[t \leq T < t + \delta t | T \geq t]$$

The hazard function describes a particular aspect of a probability distribution function. Intuitively, $\lambda(t)\Delta t$ is the probability that the event E happens within the next Δt units of time, given that it has not happened by time t.

Volcanic Risk

Volcanic risk is the adverse effect of volcanic hazards, and can be defined as the product of (volcanic hazard) $*$ (vulnerability to those hazards) $*$ (value of what is at risk). In probabilistic assessments, "risk" means the probability or likely magnitude of loss, calculated by the same formula [Blong, 2000].

Types of volcanic hazards

According to the classification done by Martí and Ernst [2005], physical volcanic hazards as destructive volcanic events can be:

Direct volcano hazards: fall processes (tephra falls and ballistic projectiles), flowage processes (pyroclastic flows, pyroclastic surges and laterally directed blasts, debris avalanches, primary debris flows (lahars or mudflows) and floods), lava flows.

Indirect volcano hazards:volcanogenic tsunamis, secondary debris flows (lahars), post-eruption famine and disease, aircraft encounters with volcanic ash.

Long-term versus short-term volcanic hazard

For the purpose of this thesis long-term volcanic hazard is used when we assume the volcanic system shows no signs of unrest. Similarly, short term volcanic hazard is used when the volcanic system is clearly in unrest mode and continuous and frequent monitoring is needed to update the status of the volcanic hazard [Blong, 2000].

Classical Model for elicitation of expert judgment

The so-called Classical Model has been developed in Delft [Bedford and Cooke, 2001] for deriving uncertainty distributions over model parameters from expert judgments. The name "Classical Model" derives from an analogy between calibration measurement and classical statistical hypothesis testing, and the approach provides a basis for performance-based linear pooling or weighted averaging of subjective opinions from a group of experts. The weights are derived from the experts calibration and information performances, as measured against so-called "seed" variables.

As mentioned above, the performance-based expert weight uses two quantitative measures of performance: "calibration" and "informativeness". The overall individual expert weights W_j in the Classical Model are taken to be proportional to the product of the individuals calibration score C_j (i.e. statistical likelihood) and his informativeness score $I_j(S_j, P)$ (estimated from all variables jointly, that is, both seeds and target variables):

$$W_j = C_j \times I_j(S_j, P)$$

where

$$C_j = 1 - \chi^2_R(2 \times N \times I(s_j, p) \times Power)$$

and

$$I(s_j, p) = \frac{1}{n} \sum_{i=1}^{n} s_i ln \left(\frac{s_j}{p_i} \right)$$

where s_i is the distribution obtained from the expert on each of the seed variables, p_i is the background reference density function for each seed, scaled appropriately for the item in question. j denotes the expert, R is the number of quantiles, N is the number of seed variables used in calibration. Here C_j corresponds to the asymptotic probability, under the hypothesis, of seeing a discrepancy between s and p at least as great as $I(s_j, p)$ and, for N large, C_j is taken to be approximately χ^2 distributed. The number of degrees of freedom is related to the number of quantiles used to define the experts distribution.

Aleatory uncertainties

The aleatory (stochastic) uncertainty is a consequence of the intrinsic complexity of a system, hence our limitation in predicting the evolution of the system in a deterministic way. The aleatory uncertainty introduces a component of randomness in the outcomes, regardless of our physical knowledge of the system Woo [1999].

Epistemic uncertainties

The epistemic uncertainty is directly related to our knowledge of the system and the quality and quantity of data we have about the system. The more data we have, the better we know the system and the lower the epistemic uncertainty Woo [1999].

EXPLORIS

The EXPLORIS project (Explosive Eruption Risk and Decision Support for EU Populations Threatened by Volcanoes) was an initiative funded by the European Union's research programme: Energy Environment and Sustainable Development whose main aim consisted in the introduction of a new methodology and new tools in the quantitative assessment of volcanic risk in densely populated regions. This goal was fully achieved through the development of research facilities completely new for the volcanological community. The project also tried to stimulate new mitigation policies and decision support to the authorities responsible for the most dangerous European volcanoes.

1.2 State of the art

Assessing eruption risk scenarios in probabilistic ways has become a main challenge in modern volcanology [Aspinall, 2006; Marzocchi et al., 2004, 2006, 2008; Neri et al., 2008; Newhall and Hoblitt, 2002]. A logic-tree of volcanic events and impacts tends to be constructed on the basis of the volcanological scenarios that can be defined using the existing geological and historical volcanological records [Marzocchi et al., 2004, 2006; Newhall and Hoblitt, 2002].

Despite the limitations in the construction of an event tree usually imposed by the lack of knowledge on the past and present behavior of active volcanoes, it is clear from the works previously cited and experiences on volcanic crises [Aspinall

and Cook, 1998] that the construction of an event tree is a major step in the hazard assessment. Most of the research done so far is based on a deterministic approach for short-term forecasting (e.g., Hill et al. [2001]; Kilburn [2003]). The alternative approach is probabilistic (e.g., Aspinall and Woo [1994]; Marzocchi et al. [2004, 2006, 2008]; Newhall and Hoblitt [2002]). Newhall and Hoblitt [2002] proposed a general event tree scheme to estimate the probability of all the relevant possible scenarios of a volcanic crisis and, in general, to quantify the volcanic hazard and risk. Later, Marzocchi et al. [2008] developed a probabilistic tool for long- and short-term eruption forecasting based on Bayesian methodology and fuzzy logic using event trees.

Event trees developed using Bayesian methodology assume that unrest is caused by internal (magmatic) triggers only. However, there are volcanic systems where unrest episodes and, occasionally, eruptions may also be caused by external triggers (geothermal, seismic) [Carniel et al., 2008; Gottsmann et al., 2007; Tárraga et al., 2006]. In computing the long-term probability of an eruption if we only consider magmatic triggers as the source of the unrest we would be underestimating the total probability, since we need to account for the long-term probability that the eruption is originated by a geothermal unrest (when a hydrothermal system exists) or by a seismic unrest.

One chapter of this thesis presents a model to assess the volcanic hazard for Teide - Pico Viejo, following the 2004-2005 seismic volcanic crisis on Tenerife [Martí and Geyer, 2009], where an event tree was developed that accounts for different triggers of volcanic unrest, and uses elicitation of expert judgment to assign a probability of occurrence to each possible eruptive scenario. However, the nature of the methodology applied requires the event tree to be as simple as possible, grouping events which may require to be analyzed individually (e.g. origin of the unrest), and leaving out relevant nodes (e.g., type of composition of the magma); does not account for the epistemic and aleatory uncertainties and requires the elicitation team to meet in order to update the probabilities each time new data arrives. Also, despite the corrections applied according to the relative relevance (weight) of each expert, the method has still a strong human decision component which adds an additional source of bias to the final results. For this reason, we have included in another chapter of the thesis a model which uses

Bayesian Inference to assign the probabilities of occurrence to each volcanic scenario and accounts for all of the above mentioned disadvantages of the elicitation method. The method was applied considering several triggers for the volcanic unrest and a more detailed event tree.

Studies of volcanic time series have been done using stochastic principles to study eruption patterns on specific volcanoes or volcanic groups [De la Cruz-Reyna, 1991, 1993; Klein, 1982; Reyment, 1969; Wickman, 1976]. Bebbington and Lai [1996a] applied a Weibull renewal model to describe the patterns of New Zealand volcanoes. Other studies used transition probabilities of Markov chains [Aspinall, 2006; Bebbington, 2007; Carta et al., 1981], change-point detection techniques [Burt et al., 1994; Mulargia et al., 1987], Rank-order statistics [Pyle, 1998], Bayesian analysis of volcanic activity [Ho, 1990; Ho et al., 2006; Marzocchi et al., 2008; Newhall and Hoblitt, 2002; Solow, 2001], non-homogeneous models [Bebbington and Lai, 1996b; Ho, 1991a], a mixture of Weibull distributions [Turner et al., 2007], geostatistical hazard-estimation methods [Jaquet and Carniel, 2006; Jaquet et al., 2000], and a mixture of exponential distributions [Dzierma and Wehrmann, 2010a,b; Mendoza-Rosas and De la Cruz-Reyna, 2009, 2010]. Extreme-value methods have been applied to geological and historical eruption time series combined [Mendoza-Rosas and De la Cruz-Reyna, 2010, 2008] and historical series of large volcanic magnitudes [Coles and Sparks, 2006].

1.3 Aims of dissertation

The aim of this thesis is to work with volcanologists to try and address, with the appropriate statistical methods, those questions they raise, and have volcanologists collaborate with statisticians to learn about the advantages in the application of statistical techniques to the interpretation of volcanic data. Here, we propose and analyze different statistical methodologies to interpret volcanic data and assess volcanic hazard. The statistical technique will depend on the nature of the data and the type of problem we want to address. The models will be used to analyze and interpret the historical and geological volcanic data for Teide - Pico Viejo stratovolcanoes (TPV) and the Canary Islands archipelago.

The first statistical method is an Elicitation of Expert Judgment using the

Classical Model to assign probabilities of occurrence to each possible eruptive scenario that can be outlined from the eruption history of the volcano, and our knowledge of other analogous volcanoes. The aim was to assess the long-term volcanic hazard of TPV, following an unrest episode in 2004 which created discrepancies among scientists regarding the nature of the unrest and the level of hazard. The method helps formalize the way volcanic experts present scientific advice by using a performance-based procedure for eliciting opinions that relies on proper scoring rules. It uses concepts and principles of eliciting expert opinion and structured elicitation within a mathematical framework. My contribution came once the elicitation of expert opinion was done, to interpret and explain the probability results, and be responsible for the statistical part of the manuscript.

The second statistical method is a Bayesian Inference approach to compute the long-term probability for each volcanic scenario. The idea to use this method came after seeing the limitations on the Classical Model, this is, the human bias, the restrictions on the complexity of the tree, and the high dependence on the elicitation team to update the probabilities when new data arrives. Also, there were limitations in the application to TPV of other Bayesian event trees built for analog volcanoes (eg. Vesuvius), since they did not contemplate all the options in each node (eg. geothermal unrest), hence, underestimating the total probability of an eruption at TPV. To accomplish this, we need a more complete event tree with all the eruptive scenarios for TPV, then, to overcome the restrictions in the data we apply Bayesian Inference to estimate the long-term probability of occurrence for each scenario. Also, we build a computer program in R statistical language to easily compute and update the probabilities. My contribution is to construct the new probability tree, design the Bayesian model and write the computer program.

The third method is a Non-parametric one-way unbalanced ANOVA using the Kruskal-Wallis test. This method is used to study a unique dependent variable against one classification variable which has two or more categories, where each classification group has unequal number of observations. Additionally, when the observations in the response variable are assumed to be independent from each other, but we do not have enough evidence to assume a particular distributional form, such as the normal (due to insufficient data), we need to use non-parametric

procedures to perform the ANOVA analysis, in our case, the Kruskal-Wallis test. This method is used to study collapse calderas, another type of volcanic structures important for their hazard implications and their high energy potential and association with mineral deposits of high economic interest. This study was suggested following the publication for the first time of the World Collapse Caldera Database (WCCD) by the Group of Volcanology of Barcelona. The aim is to study how collapse calderas are formed by analyzing their size and identifying key variables which may help understand the geodynamic environment were volcanic calderas are generated. Given the nature of the data we use the non-parametric technique described above to address the problem.

The fourth statistical methodology NHGPPP (Non-homogeneous generalized Pareto-Poisson process) uses extreme value theory to study eruptive time series combining geological and historical records. This method accounts for the time-dependence of the series and includes rare or extreme events, in the form of few data of large eruptions, since these data require special methods of analysis. The method is not restricted by the small number of observations in the time series and takes into account the limitations inherent to the available data, including its short sample time, probable absence of larger events, and incomplete reporting of small and intermediate magnitudes, as well as uncertainties in the age and magnitude of the eruptions. Hence, taking into account the probable non-homogeneity, incompleteness and missing data in the eruptive series, allowing for a better estimation of the volcanic hazard. This methodology is applied to the Canary Islands eruptive time series to study volcanic recurrence.

1.4 Thesis structure

This thesis is divided into 6 chapters, the first chapter is the Introduction and the other five are as follow:

Chapter 2 proposes and describes a long-term volcanic hazard event tree for Teide - Pico Viejo stratovolcanoes using the Elicitation of Expert Judgment Classical Model. First, we do an analysis of the potential location of future vents. The second stage in the construction of the Teide - Pico Viejo event tree involved the identification and characterization of all the effusive and explosive eruption types

that have been associated with the Teide - Pico Viejo central complex during the last 35 Ka. The third stage included the generation of the event tree (nodes and branches) using the information obtained in the two previous stages, to later assign a probability of occurrence to each branch using the elicitation of expert judgment procedure. This methodology has been published in the Journal of Volcanology and Geothermal Research and constitutes the body of chapter two.

Chapter 3 presents an alternative methodology to assess volcanic hazard applied to Teide - Pico Viejo volcanic complex which uses Bayesian Inference to assign the probabilities of occurrence to each eruptive scenario. This method allows to define a more complex and complete event tree. We first define a probability event tree, then apply the Bayesian model to compute the long-term probability of each and all of the mutually exclusive and exhaustive events, and write a program in R which automatically updates and computes these probabilities. This methodology has been published in the Journal of Geophysical Research and constitutes the body of chapter three and the appendix.

Chapter 4 performs a statistical analysis on the worldwide Collapse Caldera Database (CCDB), currently formed by 473 calderas and 28 variables. First we do a descriptive analysis of each variable in the database to identify the response and the explanatory variables to enter the study. Then we use a non-parametric one-way unbalanced ANOVA with the Kruskal-Wallis test to identify groups of volcanic calderas statistically significantly different, according to area, which may belong to a particular geodynamic environment. This methodology has been published in the Journal of Volcanology and Geothermal Research and constitutes the body of chapter four.

Chapter 5 presents and describes a statistical methodology based on Extreme Value Theory and the non-homogeneous Poisson point process to estimate the volcanic hazard. First we explore the historical time series to assess independence and stationarity of the process. Then we use a Weibull distribution to analyze and adjust the waiting times between eruptions and last we apply the non-homogeneous Poisson process with a Generalized Pareto distribution as intensity function. We apply this methodology to the Canary Islands historical eruptive time series. The contents of this chapter are written in the form of a paper which has been submitted to Natural Hazards and Earth System Sciences

in February 2010, and constitutes the body of chapter five.

There is a manuscript submited for publication at the Bulleting of Volcanology under the title "Hazard assessment of phonolitic volcanism at Teide - Pico Viejo volcanic complex (Tenerife, Canary Islands)". The project was not included in this thesis because it uses the same NHGPPP statistical method presented in chapter five to analyze the historical data for the Canary islands, hence, not contributing with new statistical methodology. The results cannot be compared with those from the Canary Islands because TPV data is phonolitic and Canary Islands covers only effusive eruptions.

Chapter 6 presents the conclusions regarding the statistical methodologies presented in this thesis and talks about future research.

Chapter 2

A long-term volcanic hazard event tree for Teide - Pico Viejo stratovolcanoes (Tenerife, Canary Islands)

2.1 Introduction

Assessing eruption risk scenarios in a probabilistic way has recently become one of the main challenges of modern volcanology. The main reason for that is the need to look for 1) a straight forward way to assess the relative likelihoods of different ways in which a volcanic system may evolve in the future or more urgently when a new eruptive process starts; and 2) a simple way to transfer this information to the corresponding decision makers, without loosing essential information. In most cases, a logic-tree of volcanic events and impacts tends to be constructed on the basis of the volcanological scenarios that can be defined using the existing geological and historical volcanological records [Marzocchi et al., 2004, 2006; Newhall and Hoblitt, 2002]. Probability weights for the various logic-tree branches are assigned through statistical analysis of data or formal elicitation of expert volcanological judgment [Aspinall, 2006; Aspinall and Woo, 1994]. An elicited risk assessment undertaken during a live volcanic crisis was conducted

for Montserrat [Aspinall and Cook, 1998]. Although previous experience shows that probabilities are not always well understood by decision makers or even by scientists, this is a necessary discipline to forecast and predict the complex and random behavior of volcanic systems and quantify and explain the underlying uncertainty. Additionally, in the case of an actual volcanic crisis the statistical methodologies serve as a tool in the elaboration of a cost/ benefit analysis, in relation to the decisions to be made by the authorities (i.e. emergency plans, evacuation). This should help them understand the complexities of the problem and envisage the potential consequences of making poorly informed decisions.

The construction of a probability event tree to estimate the volcanic hazard is based on the existence of a good volcanological record, allowing a precise reconstruction of the past history of the volcano. This allows us to determine eruption scenarios that can quantitatively define the future eruptive behavior and potential impact of the volcano. However, problems arise when the knowledge of the past volcanological history is poor, the geochronological data are scarce and historical activity has not occurred or has not been recorded in the existing chronicles. Some recent eruptions such as those from Montserrat or Pinatubo have encountered this problem (Aspinall and Cook [1998], and references therein; Newhall and Punungbayan [1996]). In these cases, the lack of knowledge of previous unrest and, more crucial, of the precursors of previous eruptive events, precludes using repetitive patterns of precursors to anticipate new eruptions (see Sandri et al. [2004]).

This is the case of Teide - Pico Viejo twin stratovolcanoes, which form one of the largest active volcanic complexes in Europe. Since the current information on their past activity is scarce and they have not shown clear signs of activity in historical times, they could be classified as dormant volcanoes [Connor et al., 2006; Szakács, 1994]. However, they have produced several central and flank vents, effusive and explosive eruptions during the last 5000 years, the last one having occurred about 1000 years ago [Carracedo et al., 2003, 2007]. This, together with the presence of permanent fumarolic activity at the summit of Teide and the occurrence of a recent unrest episode [García et al., 2006; Gottsmann et al., 2006; Martí et al., 2009], reminds us that Teide - Pico Viejo are potentially active volcanoes that could erupt again in the near future. Despite the potential

risk that these volcanoes represent for the island of Tenerife, extensively populated and one of the main tourist destinations in Europe, much remains to be learned about Teide - Pico Viejo's eruptive history. The fact that many eruption mechanisms, magma compositions, and vent sites can be distinguished from their products [Ablay and Martí, 2000; Carracedo et al., 2007; Martí et al., 2008b], complicates the definition of eruption scenarios and the establishment of eruption patterns that could be used as a guide to predict the future behavior of these twin stratovolcanoes.

This paper presents a first attempt to construct a probability, longterm event tree for Teide - Pico Viejo stratovolcanoes. First, we study the possible location of future vents based on the available geological and geophysical data. Second, we analyze the different eruption types that have characterized the volcanic activity from Teide - Pico Viejo during the last 35 ka. And third, we create the event tree structure using the information obtained in the two previous steps and we define the formalized statistical procedure for the elicitation of expert judgment that should be used to assign a probability of occurrence to each branch of the event tree.

2.2 Background geology and past volcanic activity

Teide - Pico Viejo stratovolcanoes started to grow up about 180 190 ka ago at the interior of the Las Cañadas caldera (Fig. 2.1). This volcanic depression originated by several vertical collapses of the former Tenerife central volcanic edifice (Las Cañadas edifice) following explosive emptying of high-level magma chamber. Occasional lateral collapses of the volcano flanks also occurred and modified the resulting caldera depression [Martí and Gudmundsson, 2000; Martí et al., 1994b, 1997]. The construction of the present central volcanic complex on Tenerife encompasses the formation of these twin stratovolcanoes, which derive from the interaction of two different shallow magma systems that evolved simultaneously, giving rise to a complete series from basalt to phonolite [Ablay et al., 1998; Martí et al., 2008b].

The structure and volcanic stratigraphy of the Teide - Pico Viejo stratovolcanoes were characterized by Ablay and Martí [2000], based on a detailed field and petrological study. More recently, Carracedo et al. [2003, 2007] have provided the first group of isotopic ages from Teide - Pico Viejo products, and Martí et al. [2008b] analyses their explosive activity. The reader will find in these works a more complete description of the stratigraphic and volcanological evolution of Teide- Pico Viejo.

Teide - Pico Viejo stratovolcanoes mostly consist of mafic to intermediate products, being felsic materials volumetrically subordinate overall (see Martí et al. [2008b], Fig. 1). Felsic products, however, predominate in the recent output of the Teide - Pico Viejo system. Eruptions at Teide and Pico Viejo stratovolcanoes have occurred from their central vents but also from a multitude of vents distributed on their flanks (Fig. 2.1). Mafic and phonolitic magmas have been erupted from these vents. The Santiago del Teide and Dorsal rift axes, the two main tectonic lineations currently active on Tenerife, probably join beneath Teide - Pico Viejo complex [Ablay and Martí, 2000; Carracedo, 1994]. Some flank vents at the western side of Pico Viejo are located on eruption fissures that are sub-parallel to fissures further down the Santiago del Teide rift, and define the main rift axis. On the eastern side of Teide some flank vents define eruption fissures orientated parallel to the upper Dorsal rift.

The eruptive history of the Teide - Pico Viejo (see Fig. 7 in Martí et al. [2008b]) comprises a main stage of eruption of mafic to intermediate lavas that form the core of the volcanoes and also infill most of the Las Cañadas depression and the adjacent La Orotava and Icod valleys. About 35 ka ago the first phonolites appeared, and, since then, they have become the predominant composition in the Teide - Pico Viejo eruptions. Basaltic eruptions have also continued mostly associated with the two main rift zones. The available petrological data suggest that the interaction of a deep basaltic and a shallow phonolitic magmatic systems beneath central Tenerife controls their eruption dynamics (see Martí et al. [2008b]). Most of the phonolitic eruptions from Teide - Pico Viejo show signs of magma mixing, suggesting that eruptions were triggered by intrusion of deep basaltic magmas into shallow phonolitic reservoirs.

Phonolitic activity from Teide - Pico Viejo shows a recurrence of around 250

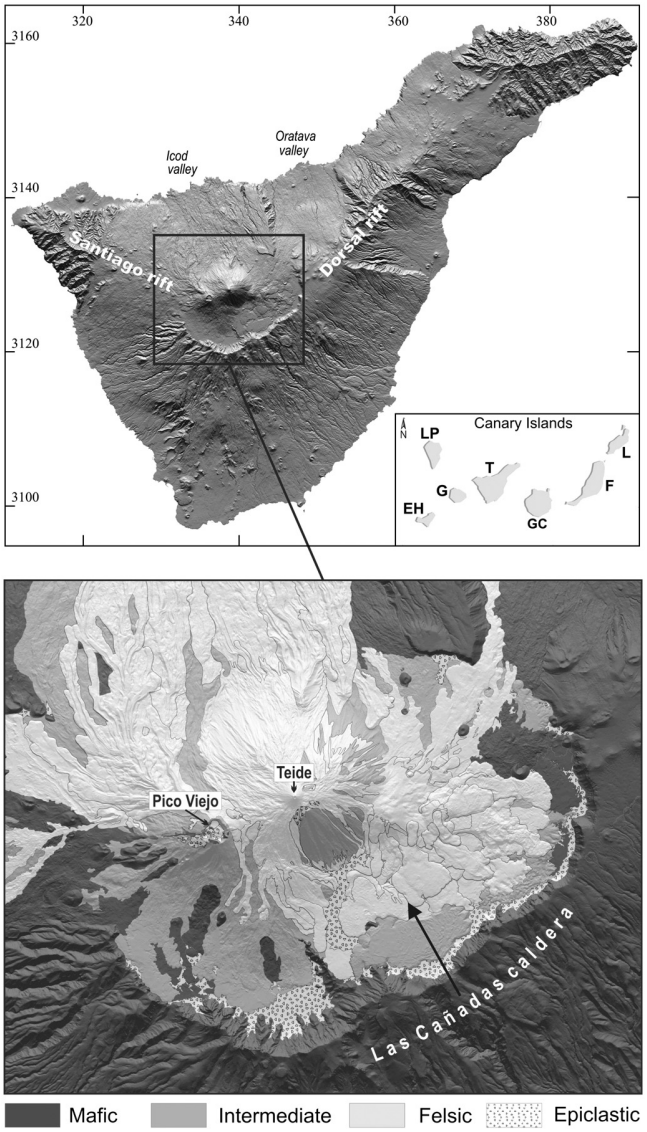

Figure 2.1: Simplified geological map of Teide, based on Ablay and Martí [2000]. Eruptive products from the Teide - Pico Viejo stratovolcanoes are identified according to their composition (mafic, intermediate and felsic). Pre - Teide - Pico Viejo rocks are undifferentiated (darkest grey). LP: La Palma; EH: El Hierro; G: Gomera; GC: Gran Canaria; T: Tenerife; L: Lanzarote;F: Fuerteventura.

- 1000 years, according to the isotopic ages published by Carracedo et al. [2003, 2007]. Phonolitic eruptions from Teide and Pico Viejo range in volume from 0.01 to 1 km^3 and have mostly generated thick lava flows and domes, some of them associated with minor explosive phases, and some subplinian eruptions, such as the Montaña Blanca at the eastern flank of Teide, 2000 years ago.

Some significant basaltic eruptions have also occurred from the flanks or the central vents of the Teide - Pico Viejo stratovolcanoes. All basaltic eruptions have developed explosive strombolian to violent strombolian phases leading to the construction of cinder and scoria cones and occasionally producing intense lava fountaining and violent explosions with the formation of ash-rich eruption columns. Violent basaltic phreatomagmatic eruptions have also occurred from the central craters of the Teide - Pico Viejo stratovolcanoes, generating high-energy, pyroclastic density currents.

According to Martí et al. [2008b], the total volume of magma erupted in the last 35 ka is of the order of 1.5-3 km^3, 83% corresponding to phonolitic magmas, while the rest includes basaltic and intermediate magmas. Therefore, phonolitic eruptions have been less frequent but much more voluminous than basaltic eruptions. All phonolitic activity has been concentrated at the central vents and flanks of the Teide - Pico Viejo stratovolcanoes. Basaltic eruptions during this period have also occurred through the rift zones.

2.3 The Teide - Pico Viejo event tree

The event tree for the Teide - Pico Viejo stratovolcano was established using a methodology similar to that proposed by Newhall and Hoblitt [2002] and Marzocchi et al. [2004, 2006], and following a similar systematic and structure than in Neri et al. [2008]. However, a significant difference with regard to these previous models is due to the fact that in previous cases only the possibility of central vent eruptions is considered, i.e. there is only one possible vent, while in Teide - Pico Viejo many vent sites are possible, including central and flank vents, having significantly different hazard implications. In addition, the stratigraphic record from Teide and Pico Viejo shows that both volcanoes may either behave independently or erupt simultaneously [Ablay and Martí, 2000]. Therefore, we had

to include in our event tree, at initial stages in the estimation process, the possible vent locations, in addition to all possible outcomes of the volcanic unrest. We include all possible options for the evolution of volcanic unrest, even those that have a very low probability of occurrence but which have been recognized in the geological record of Teide - Pico Viejo. Compared to the previous models [Marzocchi et al., 2004; Neri et al., 2008; Newhall and Hoblitt, 2002], we built our probability event tree to include the first phases of the long-term volcanic hazard estimation, since we have only geological data and no relevant historical or monitoring data is available. Hence, the subsequent risk branches, such as sectors affected, distance of runouts, exposure or vulnerability cannot be properly estimated in this paper.

2.3.1 Location of future vents

The first stage involves an analysis of the potential location of future vents. Teide - Pico Viejo have undergone several flank and central vent eruptions and these have been of basaltic and phonolitic composition, without any apparent structural or petrological pattern that could explain such random eruption behavior [Martí et al., 2008b]. The lack of a good surveillance network makes the identification of future vents more challenging than in other similar volcanoes. It has been generally assumed that future eruptive activity on Tenerife, if any, should be of basaltic nature, far from the central Teide system, and generate short lava flows and small cinder cones [Carracedo et al., 2003, 2007]. However, there are no scientific reasons to rule out the possibility of an eruption, basaltic or phonolitic, from Teide - Pico Viejo. On the contrary, all the petrological, geochronological and volcanological data available, suggest that Teide - Pico Viejo cannot be considered as extinct volcanoes at all [Ablay and Martí, 2000; Martí et al., 2008b] and that they could erupt again in the near future, even in a violent way.

In addition to field studies [Ablay and Martí, 2000], numerical experiments have been used to investigate the possible factors that determine the occurrence of either a central or a flank eruption at Teide - Pico Viejo. We have considered a wide range of situations in which the main physical conditions of the volcanic system have been changed (topography, size, depth and shape of the magma

chambers, presence of deviatoric stresses, internal structure of the volcano) [Martí et al., 2006]. The results obtained show that the main control on the pathway that phonolitic magma will follow to leave the shallow chamber and reach the surface is exerted by the stress field distribution around and above the chamber, being this a function of the shape and depth of the magma chamber. An alternative explanation is that the magma follows paths around any massive blockage such as recent intrusions. In the case of basaltic eruptions, which are fed by much deeper magmas, we think that a structural control resulting from the interaction between the rift systems and the central complex, rather than the geometry and location of the pressure source at depth, determines the exact location of the vent(s) in each eruption. Comparison of these results with the available geological [Ablay and Martí, 2000] and geochronological [Carracedo et al., 2003, 2007] information suggests that the number of flank eruptions occurred on Teide - Pico Viejo during the time period considered is slightly higher than that of the central vent eruptions.

2.3.2 Eruption types and eruption scenarios

The second stage in the construction of the Teide - Pico Viejo event tree involved the identification and characterization of all the effusive and explosive eruption types that have been associated with the Teide - Pico Viejo central complex during the last 35 ka. This is the maximum period that we can investigate from surface geology and also represents an upper time limit for the appearance of the first phonolites on that volcano. Despite this could be regarded as a long time interval because the record of older events is incomplete, it gives us a minimum estimate. Teide - Pico Viejo's eruptive activity has been associated with both, mafic (basalts, tephri-phonolites) and felsic (phono-tephrites and phonolites) magmas, and has produced a large variety of eruption types: mostly effusive (lavas and/or domes), strombolian, violent strombolian and sub-plinian magmatic eruptions, as well as phreatomagmatic eruptions of mafic magmas and phreatic explosions. There is no any apparent control on the style of the eruption by the vent location, so that we assumed a priori that similar eruptions can occur from both flank and central vents on Teide - Pico Viejo. The set of eruptive types we have included

Table 2.1: *Description of the main characteristics of the eruption types (largest or most explosive event) that have occurred from the Teide - Pico Viejo complex during the last 35 ka (see Martí et al. [2008b] for more information)*

Eruptive type	VEI	DRE km^3	Composition	Vent location	Recurrence period	Eruption phases and products
Strombolian	1-2	0.001-0.03	Mafic/felsic	Any	80-200 years, $>$ 25 events	Fine ash/lapilli dispersion through short eruption columns (b 1 km), proximal coarse lapilli, scoria and spatter from fire fountains (up to 300 m), some ballistics, frequent lava flows
Violent strombolian	2-3	0.01-0.2	Mafic/felsic	Any	200-500 years ? , $>$ 5 event	Fine ash/lapilli dispersion through eruption columns up to 2 km high, proximal coarse lapilli, scoria and spatter from fire fountains (up to 600 m), abundant ballistics (up to 1 m across), frequent lava flows
Sub-plinian	3-4	0.01-0.5	Felsic	Any	400-2000 years ? , $\geq$ 3 events	Eruption column up to 10 km high, pumice fall deposit, some ballistics, emplacement of lava flows and domes at the beginning and end of the eruption
Phreatic	?	?	?	Pico Viejo Caldera	? , 1 event	Large ballistics (up to 1 m across) at distances of several km from the vent
Phreatomagmatic	2-3	0.001-0.01	Mafic	T-PV old and present central vents	? , $\geq$ 4 events	Radial/directed distribution of ballistic breccias and pyroclastic surges up to 10 km from the vent, lahars
Mainly effusive (lava flows, clastogenic lavas and domes)	0-1	0.01-1	Felsic	any	300-1000 years ? , $>$ 14 events	Thick (up to 40 m) massive lava flows (up to 15 km in length), clastogenic lavas and domes. Occasional occurrence of associated bock and ash deposits (and related debris flows and lahars) caused by gravitational collapse of lavas flows and/or domes.

in the event tree represents the largest or the most explosive event that occurs in the course of an eruption that might include many events and eruption styles. Table 2.1 summarizes the characteristics of the main Teide - Pico Viejo eruption types. A more detailed description of the eruptive activity of Teide - Pico Viejo is given in Martí et al. [2008b].

2.3.3 Event tree structure

The third stage included the generation of the event tree (nodes and branches) using the information obtained in the two previous stages, to later assign a probability of occurrence to each branch using the elicitation of expert judgment procedure presented in the next section. As for other volcanoes [Marzocchi et al., 2004, 2006; Neri et al., 2008; Newhall and Hoblitt, 2002], the estimation was carried out for all the branches of each node for the Teide - Pico Viejo event tree, progressing from general to specific events (Fig. 2.2). From the study of the different eruption types identified on Teide - Pico Viejo, we can deduce that all of them, including the phreatic episode from Pico Viejo [Ablay and Martí, 2000] require the presence of fresh magma, either mafic or felsic, at shallow depths in the volcanoes. However, we did not discard the possibility of starting an eruption process from an unrest directly associated with the hydrothermal system or event due to external triggers, such as regional tectonics, if eruptible magma is present in the system. Therefore, we assumed a precursory step as an unrest episode, regardless of it is magmatic, hydrothermal or tectonic, characterized by an anomalous increase of seismic activity, ground deformation, gravity changes, gas emissions, and so on.We started by considering whether the unrest episode could lead to a sector collapse that then could trigger an eruption or not. If an eruption is expected regardless of the existence of sector collapse, this could be from the central vent (s) (either Teide or Pico Viejo) or from any of the volcano's flanks. In any of these cases, there are several possibilities, mainly controlled by the composition (mafic or felsic) of the erupting magma, according to what has been observed in the period considered in our study. The construction of our event tree continued with the inclusion in the following sequence of nodes and branches of the different possible eruptions and related hazards that may occur

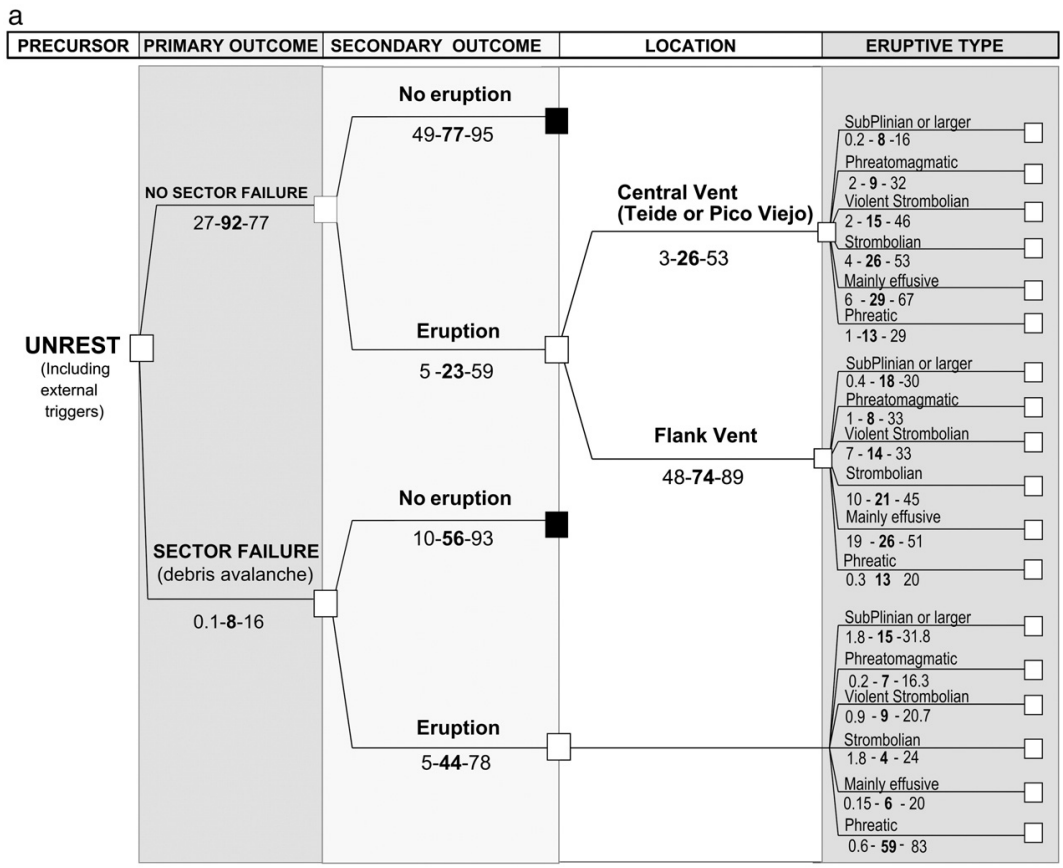

Figure 2.2: The volcanic event tree for Teide - Pico Viejo. Six stages of eruptive progression from general to more specific events (from left to right) are presented. Nodes from which subsequent multiple different outcomes (branches) of the eruptive process are possible at each stage are indicated by white squares. Nodes that represent the termination of a particular process are represented by black squares. a) shows an overview of the primary branches of Teide - Pico Viejo event tree as far as eruptive style. The set of eruptive types we have included in the event tree represents the largest or the most explosive event that occurs in the course of an eruption that might include many events and eruption styles. Expanded tree details including hazards are shown for: b) the scenario of a central vent eruption; c) the flank vents scenario; and, d) the case of a magmatic eruption triggered by sector failure. In addition to the authors of the paper, the following scientists have also participated in the experts elicitation judgement: J. Andújar, A. Bertagnini, R. Chioni, O, Cornellá, A. Folch, A. M. García, MJ. Jiménez, Neri, T. Ongaro, G. Queiroz, M. Rosi, L. Sandri, C. Soriano, R. Spence, F. Teixidé, M. Todesco, G. Toyos, G. Zuccaro.

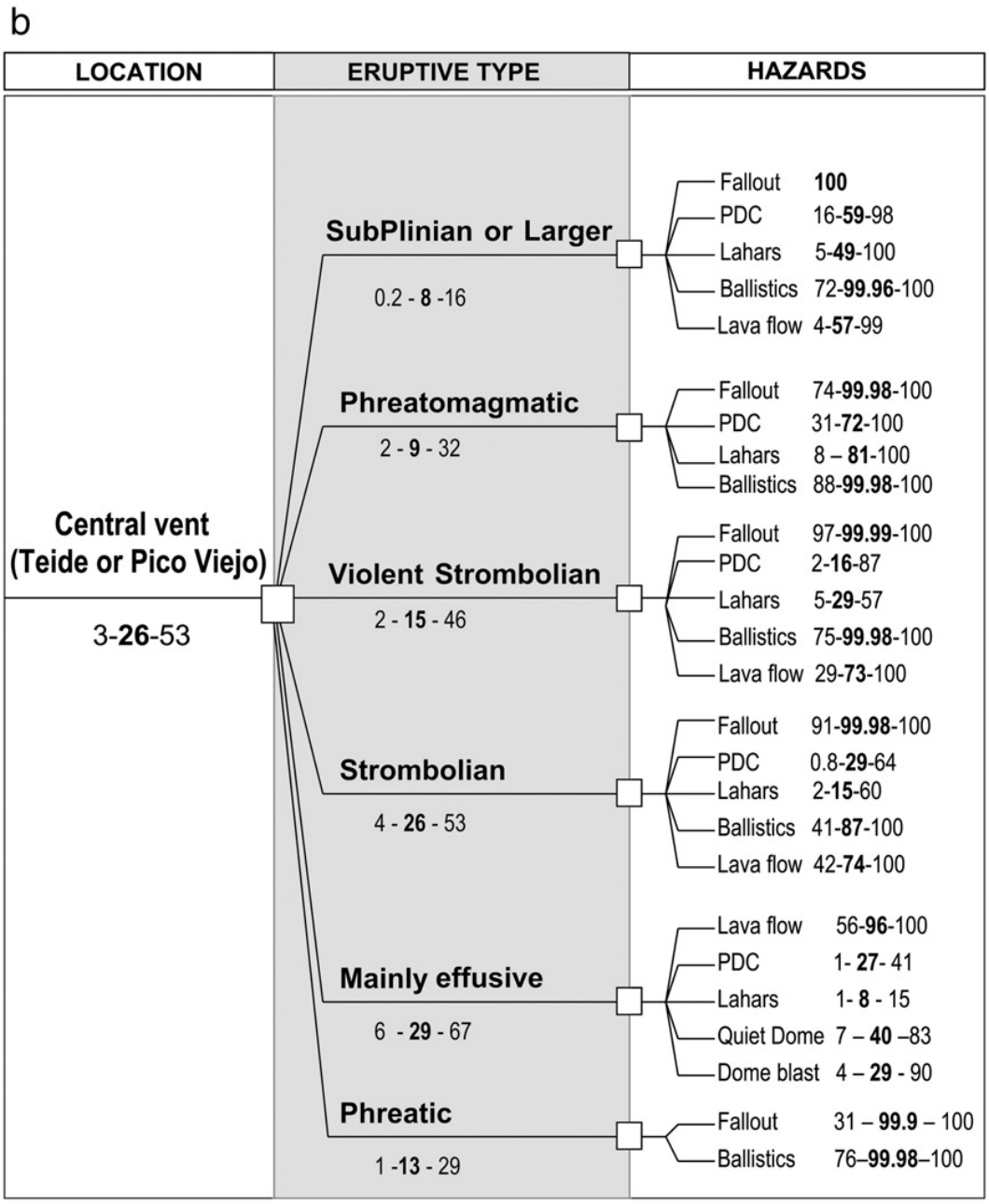

Fig. 2.2 (continued).

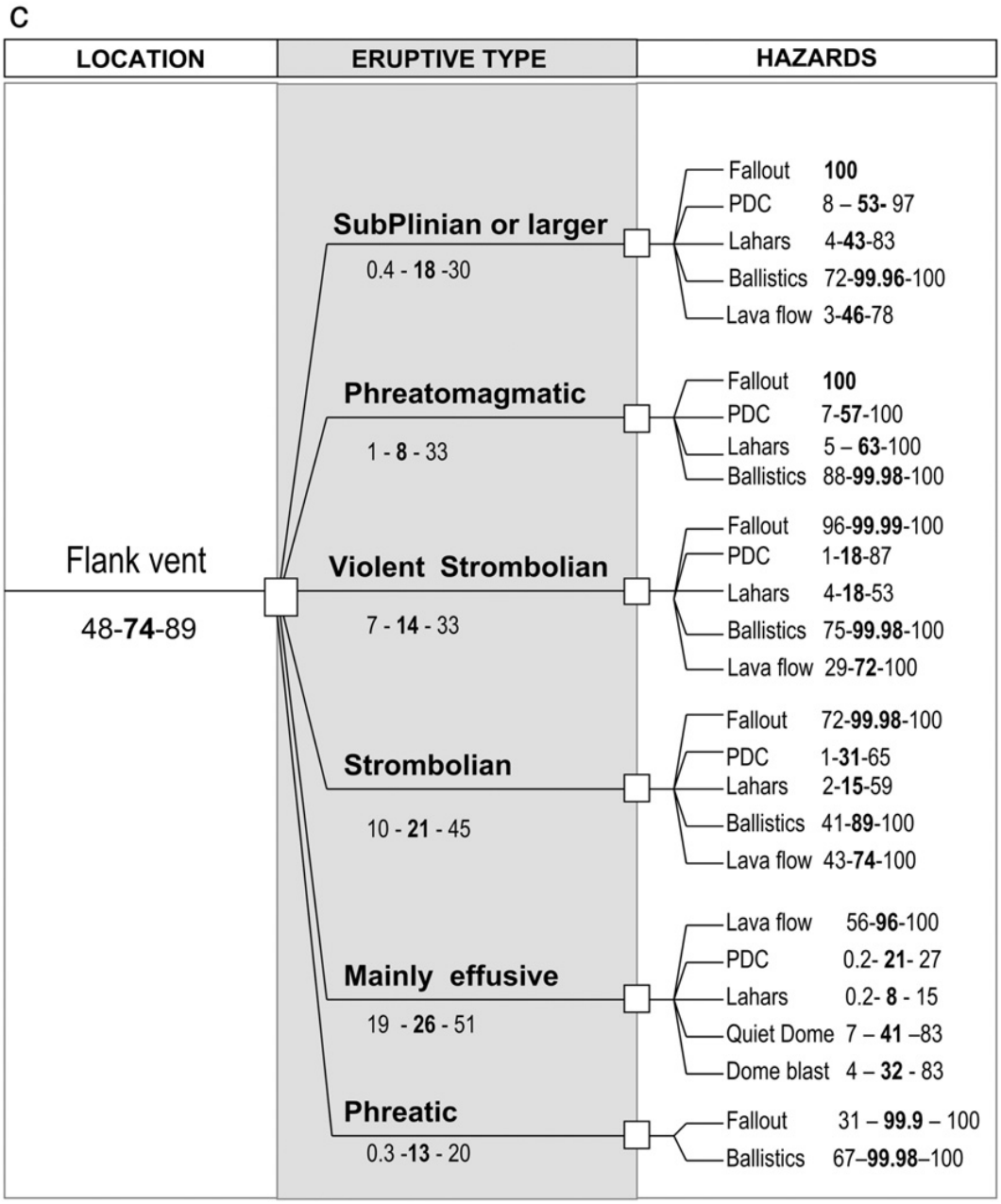

Fig. 2.2 (continued).

d

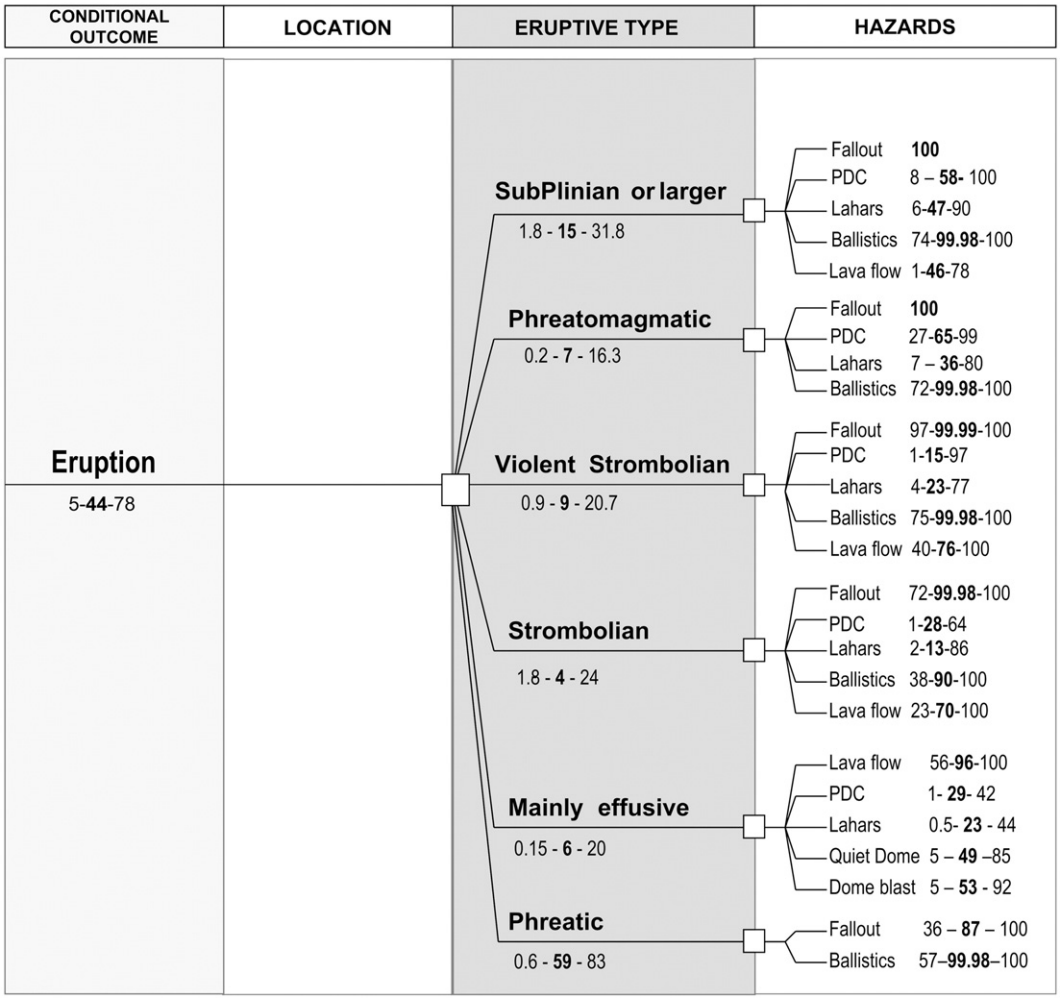

Fig. 2.2 (continued).

from Teide - Pico Viejo (see Table 2.1 and Fig. 2.2a-d).

As we have indicated before, it would be premature at this stage to continue the construction of the Teide - Pico Viejo event tree adding the corresponding steps concerning the potential impacts of each hazard, as there is still some basic information missing. However, the structure proposed will allow the event tree to be extended at any moment and to adapt it for a short-term hazard assessment when information from volcano monitoring from Teide - Pico Viejo will be available.

2.3.4 Statistical methodology: The Teide - Pico Viejo expert elicitation procedure

The ultimate aim of a volcanic hazard event tree is to assign probabilities to the different eruption possibilities or scenarios that we can envisage from the eruption history of the volcano and our knowledge of other analogous volcanoes. An event tree aims to quantitatively estimate both long- and short-term volcanic hazard. In order to achieve this objective, we use a methodology based on Elicitation of Expert Judgment. In particular, the so-called Classical Model developed in Delft [Bedford and Cooke, 2001]. Other statistical procedures as well as other methods to elicit expert judgments are available. The Classical Model is a performance-based formalized procedure for the elicitation of expert judgments which allows to derive uncertainty distributions over model parameters using expert judgment. This approach provides a basis for weighted averaging of subjective opinions. The weights are derived from the experts' calibration and information performances, measured by the so-called seed variables [Aspinall, 2006].

Within the EXPLORIS Project, the approach adopted for parameterizing event probability nodes on event trees was to elicit values from colleagues (see Fig. 2.2a-d) in a structured manner, and to use Cooke's (1991) classical model mentioned above implemented in the computer program EXCALIBR [Cooke and Solomatine, 1990] for the quantification of the corresponding collective scientific uncertainty (Note that the software has recently changed its name to EXCALIBUR). The classical model is unique in that it embodies a performance-based expert scoring scheme, by which weights are ascribed to individual experts on

the basis of empirically determined calibration and informativeness scores. The expert's assessments are treated as statistical hypotheses and the probability at which these hypotheses would be rejected is used to provide a score for calibration (under the assumption that the calibration variables are independent realizations of the experts' distributions). A second factor, the informativeness of the expert, is defined as the relative information of his or her distributions with respect to some specified background measure. The theory of strictly proper scoring rules is used to combine calibration and informativeness scores expressed as a product, and from these results the so-called performance- based decision-maker is formed from a weighted combination of the judgments of the group of experts involved.

In EXPLORIS, the processing of group elicitations utilized a variant of the EXCALIBR application in which the power of the statistical hypothesis test and the significance level for rejection were adjusted so that all participants obtained some positive weight for their opinions; this approach is termed constrained optimization weighting [Aspinall, 2006]. The purpose of doing this was to limit the chances that any individual expert might be unduly penalized in their measured performance just as a result of any particular property or bias in the set of seed questions that had been selected for calibration. Thus, the outcome of an EXCALIBR calibration exercise is a set of numerical scores for the panel of experts involved, each individual's score representing his or her empirical success in making uncertainty judgments about known parameters or values. Within all the EXPLORIS volcano groups, there was a tendency for the majority of participants to be over-confident in their judgments, and to receive reduced relative weighting accordingly this is a widespread characteristic of almost all groups of experts, in any scientific or engineering discipline. However, the EXCALIBR procedure accounts for such traits in an objective and traceable way and provides a rational and neutral pooling of diverse opinions.

In the case of Teide - Pico Viejo, more than 30 experts (Fig. 2.2) from the EXPLORIS project participated in the elicitation process. For every node on the Teide - Pico Viejo event tree, the experts provided their individual opinions as the relative likelihoods of occurrence of the alternative pathways for the way the course of the eruption could progress, and these opinions are pooled using the weights obtained from the EXCALIBR calibration procedure. The outcomes of

this process are recorded as numerical probability values on the event tree shown on Fig. 2.2a-d. On each branch, the results are given as three numbers: the median probability (i.e. 50 percentile value for the distribution of opinions provided by the group), together with corresponding 90% credible interval bounds (i.e. approximate 5 percentile and 95 percentile distributional values). This way of representing the collective scientific uncertainty associated with forecasting volcanic hazards is uniquely different to that of other approaches, and gives formal, quantitative expression to all the uncertainties involved, essential for any comprehensive probabilistic risk assessment.

In the particular case of Teide - Pico Viejo, due to the lack of information on their past geology and recent volcanic activity, two questions where posed to the referees regarding the nodes with two branches of the first steps of eruptive progression (outcome and location) (Fig. 2.2). This makes a significant difference compared to other event trees, such as the one for Vesuvius [Neri et al., 2008], from which there is a good set of data and most of the experts know the corresponding volcanoes very well. In such cases, only one question (three percentiles) is asked for one branch since the three percentiles of the other branch are simply the complement. In our case, the reason to elicit both probabilities and their complement in separate questions for binary branches is due to the general lack of knowledge on Teide - Pico Viejo and, consequently, as away to test the expert understanding of the method and the volcano, just looking at how stable the paired answers were. Consequently, in our example the credible intervals do not need to be complementary. Although, the resulting numbers from the Teide - Pico Viejo elicitation were not always very consistent, we decided to show them as a first indication of the great uncertainty we have to deal with when working with a poorly known volcano.

The following are examples of some of the group's judgements for scenarios relating to the next eruption of Teide - Pico Viejo. According to Fig. 2.2, if unrest occurs, it is considered 11.5 more likely that no sector failure will ensue (conditional probability given unrest=92%) than a sector failure (conditional probability=8%) will happen. Note that the probabilities we discuss refer to the 50 percentile confidence level. On this basis, it is concluded there is a 1-in-12.5 chance of a sector failure, given unrest. Of course, these probabilities would be

updated in the light of data, observations or evidence that is detected in relation to the unrest episode, when it develops.

Moving across the tree on Fig. 2.2, if an eruption should occur, the collective view is that a flank eruption (conditional probability=74%) is 2.8 more likely than a central vent eruption (conditional probability=26%). Again, these values are based solely on the available geological data and interpretations of that incomplete data, and could easily be modified by indicators from emerging unrest (e.g. seismological data suggesting dike propagation at a flank location).

If the next eruption of Teide - Pico Viejo is from a central vent (Fig. 2.2), the expert elicitation provides the following guidance about relative likelihoods of the eruptive style that may ensue, assuming that the possible eruption scenarios are mutually exclusive: the most likely scenario is either a mainly effusive eruption (conditional probability =29%) or a Strombolian intensity explosive eruption (conditional probability=26%). Less likely next are either a Violent Strombolian scenario (conditional probability=15%) or an eruption of predominantly phreatic activity (conditional probability=13%). The other possible scenarios are a Phreatomagmatic eruption (conditional probability=9%) or a sub-Plinian or larger eruption (conditional probability =8%). Although these last two are the least likely possibilities, neither is judged to be highly unlikely - either has a chance of about 1-in-10 of being experienced in a future central vent eruption.

On Fig. 2.2b-d the last columns of the Teide - Pico Viejo event tree in its present form, show the hazards that could accompany each of the different eruptive style scenarios. Here, the relevant individual conditional probabilities are not exclusive, as more than one hazard may be present in any one eruption episode. The overall likelihood of any particular hazard being manifest in the next eruption of Teide - Pico Viejo, whatever its eruptive style or vent location, can be calculated by summing up the compounded conditional probabilities across all those branches which lead to a termination node with that hazard.

The next stage in the enhancement of the Teide - Pico Viejo event tree should be to expand it to add further branches that represent factors important for risk assessment, such as flow directivity and run out distances, for example. As with the initial eruptive scenario stages of the Teide - Pico Viejo event tree, described here, these further elements can also be derived from expert judgment, but in this

case extensively informed by physical modeling results, as well as the geological record. This said, the rational quantification of scientific uncertainty remains the key reason for adopting a formalized expert elicitation approach.

With the data and knowledge currently available for Teide - Pico Viejo, the median probabilities and their uncertainty spreads, shown on Fig. 2.2a-d, are expected to be stable against most plausible alternative interpretations if no new information emerges, in which case they can be modified significantly. For instance, if unrest develops, many of the conditional probabilities on the Teide - Pico Viejo event tree could change in response to the occurrence of precursors that are robust indicators of certain aspects of forthcoming activity. In this sense, any volcanic event tree should be regarded as an evolving, organic object, which requires attention (and pruning) as new information emerges.

2.4 Discussion and conclusions

We propose a volcanic hazard event tree for Teide - Pico Viejo that shows all possible outcomes of volcanic unrest at progressively higher degrees of detail. The construction of the current version of the Teide - Pico Viejo event tree is based on the past eruptive history of the volcano and corresponds to its long-term volcanic hazard assessment. Moreover, the Teide - Pico Viejo event tree can be easily extended to account for short-term hazard assessment (in case of volcanic crisis) when precise monitoring data from the volcano will be available.

The Teide - Pico Viejo event tree has been constructed following previous models [Marzocchi et al., 2004, 2006; Neri et al., 2008; Newhall and Hoblitt, 2002] and it is based on similar concepts. It uses a statistical approach based on Expert Judgment Elicitation to ascribe the appropriate probabilities and to determine the corresponding uncertainty for the different possible events that could occur, based on available past geological data.We have conducted this process among the participants of the EXPLORIS project using the performance-based Classical Model for expert judgment elicitation. However, there is a significant difference between the present Teide - Pico Viejo event tree and the reference models we have used in its construction. This difference relates to the need to allow for several possible vent sites, instead of a single one as is the case in the previous

models. The actual location of the vent will condition the resulting hazards and, in particular, their potential impacts, as the vent position will determine the area affected by each volcanic process. This is clear in the case of Teide - Pico Viejo from which eruptions have occurred not only from the central vent area but also from any of its flanks. Therefore,we have included in our event tree a new step for the progression of the eruption that considers the location of the eruptive vent. However, it is important to remark that this is correct for the present situation but if we get to the future step of hazard to specific location, the central vs flank distinction will not be specific enough.

The Teide - Pico Viejo event tree considers all possible volcanic processes that could occur according to the available past information, even those with a low probability of occurrence. In this sense,we have to mention that the structure of the event tree has to be simple but as informative as possible. We have to keep in mind that a volcanic hazard event tree has to be prepared for showing to decision makers, who will not necessarily be familiar with volcanic or probabilistic terminology. The role of scientists should be just to advise decision makers, so that if a possible outcome is not shown in the event tree and it finally occurs, scientists could be regarded as at fault.

Volcanoes are complex, non-linear natural systems that rarely follow a constant pattern of behavior. Although we can establish some general eruptive patterns for certain group or types of volcanoes, each volcano will at the end behave in its own particular way, different from the others. In the case of Teide - Pico Viejo, as in many other volcanoes around the World, the information we have on its past eruptive history and on its present state of activity is incomplete, and requires much more effort before being more confident that we can precisely forecast its future behavior. This said, the event tree we have presented in this paper is a contribution to advancing and rationalizing our current knowledge on Teide - Pico Viejo, and to providing a useful tool to help the society increase its confidence that any future threat of the volcano is properly assessed, according to the main objectives of the EXPLORIS project.

Chapter 3

Bayesian event tree for long-term volcanic hazard assessment: Application to Teide - Pico Viejo stratovolcanoes, Tenerife, Canary Islands

3.1 Introduction

Assessing eruption risk scenarios in probabilistic ways has become a main challenge in modern volcanology (Newhall and Hoblitt [2002]; Marzocchi et al. [2004, 2006, 2007]; Aspinall [2006]; Neri et al. [2008]; Martí et al. [2008a]). Volcanic risk is usually defined as the product of volcanic hazard, value and vulnerability, where volcanic hazard is the probability of any particular area being affected by a destructive volcanic event within a given period of time; the value is the number of human lives at stake, or the capital value (land, buildings, etc.), or the productive capacity (factories, power plants, highways, etc.) exposed to the destructive events; the vulnerability is a measure of the proportion of the value likely to be lost as a result of a given event (Blong [2000]).

Short and long term eruption forecasting are defined based on the expected

characteristic time in which the process shows significant variations. During a quiet phase of the volcano the time variations occur in time intervals significantly longer than during unrest. For the purpose of this paper, long term forecasting refers to the time window before the volcanic system goes into unrest, and short term forecasting refers to the unrest phase. Consequently, long term forecasting is based on historical and geological data, and theoretical models, while short term forecasting is complemented with continuous monitoring data.

The complexity of any volcanic system and its associated eruptive processes, together with the lack of data that characterize many active volcanoes, particularly those with long recurrences, make volcanic hazard quantification very challenging, as there is often not enough observational data to build a robust statistical model.

Despite the limitations in the construction of an event tree usually imposed by the lack of knowledge on the past and present behavior of active volcanoes, it is clear from the works previously cited and experiences on volcanic crises (Aspinall et al. [1998]) that the construction of an event tree is a major step in the hazard assessment. Most of the research done so far is based on a deterministic approach for short-term forecasting (e.g., Kilburn [2003]; Hill et al. [2001]). The alternative approach is probabilistic (e.g., Newhall and Hoblitt [2002]; Aspinall and Woo [1994]; Marzocchi et al. [2004, 2006, 2007]). Newhall and Hoblitt [2002] proposed a general event tree scheme to estimate the probability of all the relevant possible scenarios of a volcanic crisis and, in general, to quantify the volcanic hazard and risk. Later, Marzocchi et al. [2007] developed a probabilistic tool for long- and short- term eruption forecasting based on Bayesian methodology and fuzzy logic using event trees.

Event trees developed using Bayesian methodology assume that unrest is caused by internal (magmatic) triggers only. However, there are volcanic systems where unrest episodes and, occasionally, eruptions may also be caused by external triggers (geothermal, seismic) (Tárraga et al. [2006], Gottsmann et al. [2007], Carniel et al. [2008]). In computing the long term probability of an eruption if we only consider magmatic triggers as the source of the unrest we would be underestimating the total probability, since we need to account for the long term probability that the eruption is originated by a geothermal unrest (when a

hydrothermal system exists) or by a seismic unrest. On the other hand, event trees developed using Elicitation of Expert Judgment have a human decision component which adds an additional source of bias to the model, require the event tree structure to be as simple as possible, do not account for the epistemic and aleatory uncertainties and require the elicitation team to meet in order to update the probabilities each time new data arrives.

In order to show the limitations of the previous attempts and the need for a more extensive structure we use the example of Teide - Pico Viejo stratovolcanoes, as they present alternative scenarios (i.e., nodes and branches) than those included in the previous event tree structures and may experience unrest triggered by external causes such as regional seismicity or oscillations in the hydrothermal system. Teide - Pico Viejo stratovolcanoes form one of the largest volcanic complexes in Europe, situated on the island of Tenerife, extensively populated and one of the main tourist destinations in Europe. There is scarce information on its past activity and the volcanoes have not shown clear signs of activity in historical times. However, it has produced several central and flank vent, effusive and explosive eruptions during the last 5000 years, the last one about 1000 years ago (Carracedo et al. [2007]). It has permanent fumarolic activity at the summit of Teide volcano and the occurrence of a recent unrest episode (Martí et al. [2009]), reminds us that these volcanoes are presently quiescent, but potentially active and could erupt again in the near future. For this reason, research is needed to assess the volcanic hazard and forecast the range of potential volcanic eruptions. Since we rely on geological and geophysical data, aleatory (stochastic) and epistemic (data or knowledge limited) uncertainties are significant, and we need to find a way to minimize them.

The aleatory (stochastic) uncertainty is a consequence of the intrinsic complexity of a system, hence our limitation in predicting the evolution of the system in a deterministic way. The aleatory uncertainty introduces a component of randomness in the outcomes, regardless of our physical knowledge of the system. The epistemic uncertainty is directly related to our knowledge of the system and the quality and quantity of data we have about the system. The more data we have, the better we know the system and the lower the epistemic uncertainty (Woo [1999]).

A first attempt to asses the volcanic hazard for Teide - Pico Viejo has been made by Martí et al. [2008a], following the 2004-2005 seismic volcanic crisis on Tenerife (Martí et al. [2009]), who have proposed an event tree using Elicitation of Expert Judgment to assign a probability of occurrence to each possible eruptive scenario. However, the nature of the methodology applied required the event tree to be as simple as possible, grouping events which may require to be analyzed individually (eg. Origin of the unrest), and leaving out relevant nodes (eg. Type of composition of the magma). Also, despite the corrections applied according to the relative relevance (weight) of each expert, the method has still a strong human decision component which adds an additional source of bias to the final results.

In this paper, we present an event tree structure which accounts for external triggers (geothermal, seismic) as additional sources of volcanic unrest and looks at the hazard from different types of magma composition and different vent locations. We then take the available geological data for Teide - Pico Viejo from the last 8 ka and run it through a Bayesian model built following this new event tree structure. The result is an estimation of the long term probability for each possible scenario. We compare both Elicitation and Bayesian methods applied to Teide - Pico Viejo. Also, we compare the results from the Bayesian event tree developed here allowing for external triggers with previous Bayesian event tree structures. This is a new step in the development of useful tools for volcanic hazard assessment, but additional studies will be needed to define the precursors and monitoring parameters for each eruptive scenario in order to estimate the short term probabilities.

3.2 Background geology and past volcanic activity in the Teide - Pico Viejo volcanic complex

Teide - Pico Viejo stratovolcanoes started to grow up about 180-190 ka in the interior of the Las Cañadas caldera (Fig. 3.1). This volcanic depression originated by several vertical collapses of the former Tenerife central volcanic edifice

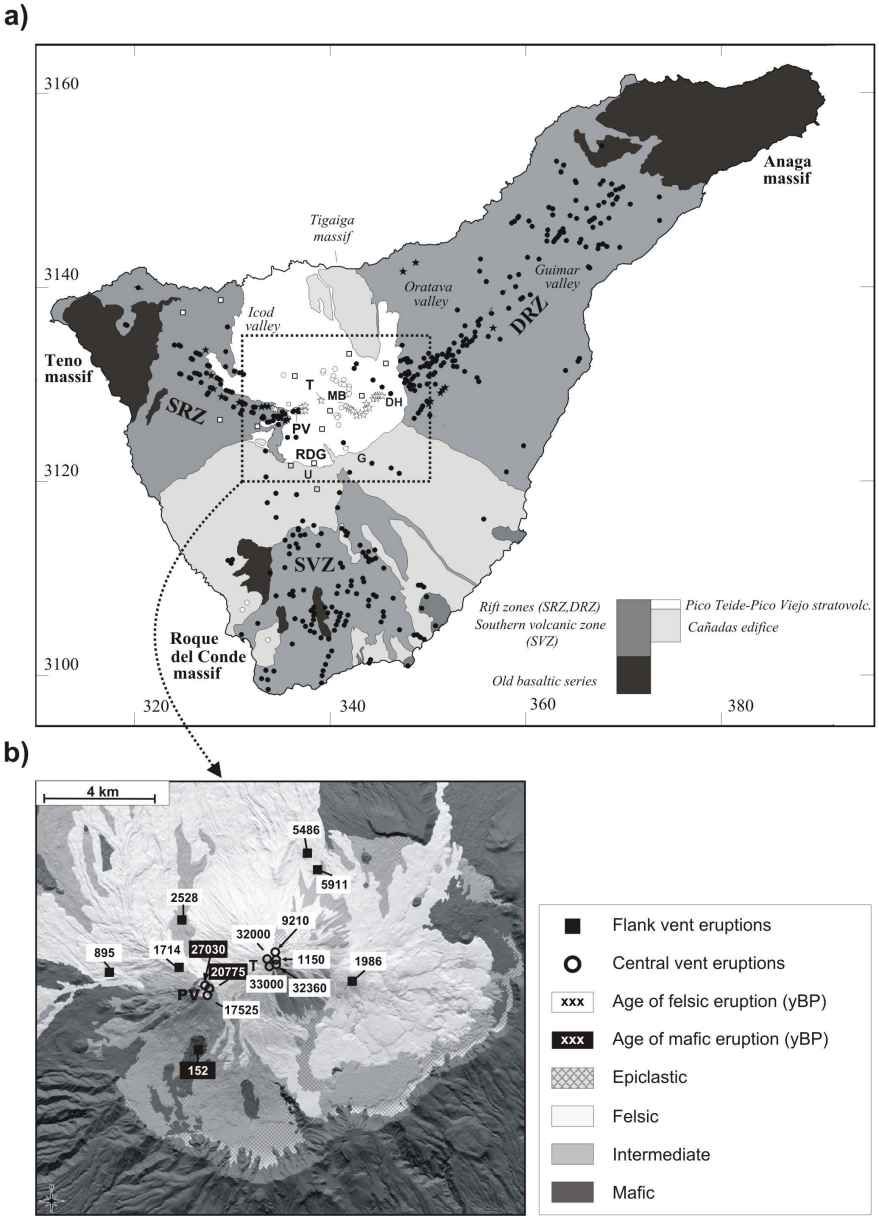

Figure 3.1: *(a) Simplified geological and topographic map of Tenerife illustrating the general distribution of visible vents. RDG: Roques de García; G: Guajara;. T: Teide volcano; PV: Pico Viejo volcano; MB: Montaña Blanca. SRZ: Santiago rift zone; DRZ: Dorsal rift zone; SVZ: Southern volcanic zone. Black symbols: mafic and intermediate vents; White symbols: felsic vents; Stars: historic and sub-historic vents; Circles: other vents. Names and locations of landslide valleys are also shown. Coordinates refer to 20 km squares of the Spanish national grid (UTM). (b) Simplified geological map of the central part of Tenerife Island. Black squares and rings indicate flank and central vent eruptions, respectively. White boxes include the age in yBP of phonolitic events and black boxes those of mafic events. The figure does not include all the eruptions from Teide - Pico Viejo stratovolcanoes, only those dated by Carracedo et al. [2007]).*

(Las Cañadas edifice) following explosive emptying of a high-level magma chamber. Occasional lateral collapses of the volcano flanks also occurred and modified the resulting caldera depressions (Martí et al. [1994a]). The construction of the present central volcanic complex on Tenerife encompasses the formation of these twin stratovolcanoes, which derive from the interaction of two different shallow magma systems that evolved simultaneously, giving rise to a complete series from basalt to phonolite (Martí et al. [2008b]).

Eruptions at Teide and Pico Viejo stratovolcanoes have occurred from their central vents but also from a multitude of vents distributed on their flanks (Fig. 1) (Martí and Geyer [2009]). Mafic and phonolitic magmas have been erupted from these central and flank vents. The Santiago del Teide and Dorsal rift axes, the two main tectonic lineations currently active on Tenerife, probably join beneath the Teide - Pico Viejo complex (Ablay and Martí [2000]). Some flank vents at the western side of Pico Viejo are located on eruption fissures that are sub-parallel to fissures further down the Santiago rift, and define the main rift axis. On the eastern side of Teide some flank vents define eruption fissures orientated parallel to the upper Dorsal rift.

The eruptive history of the Teide - Pico Viejo comprises a main stage of eruption of mafic to intermediate lavas that form the core of the volcanoes and also infill most of the Las Cañadas depression and the adjacent La Orotava and Icod valleys. About 35 ka the first phonolites appeared, and, since then, they have become the predominant composition in the Teide - Pico Viejo eruptions. Basaltic eruptions have also continued mostly associated with the two main rift zones, but also through some flank vents. The available petrological data suggest that the interaction of a deep basaltic and a shallow phonolitic magmatic systems beneath central Tenerife controls their eruption dynamics (Martí et al. [2008b]). Most of the phonolitic eruptions from Teide - Pico Viejo show signs of magma mixing, suggesting that eruptions were triggered by intrusion of deep basaltic magmas into shallow phonolitic reservoirs.

Phonolitic activity from Teide -Pico Viejo shows a recurrence of around 250-1000 years, according to the isotopic ages published by Carracedo et al. [2003, 2007]). Phonolitic eruptions from Teide and Pico Viejo range in volume from 0.01 to 1 km^3 and have mostly generated thick lava flows and domes, some of them

associated with minor explosive phases, and some with subplinian eruptions, such as the Montaña Blanca at the eastern flank of Teide, 2000 years ago (Martí et al. [2008b]).

Some significant basaltic eruptions have also occurred from the flanks or the central vents of the Teide - Pico Viejo stratovolcanoes, with a recurrence of around 80-150 years according to the historic record. All basaltic eruptions have developed explosive strombolian to violent strombolian phases leading to the construction of cinder and scoria cones and occasionally producing intense lava fountaining and violent explosions with the formation of ash-rich eruption columns. Violent basaltic phreatomagmatic eruptions have also occurred from the central craters of the Teide - Pico Viejo stratovolcanoes, generating high-energy, pyroclastic density currents.

According to Martí et al. [2008b], the total volume of magma erupted in the last 35 ka is of the order of 1.8-3 km^3, 83% corresponding to phonolitic magmas, while the rest includes basaltic and intermediate magmas. Therefore, phonolitic eruptions have been less frequent but much more voluminous than basaltic eruptions in the recent history of Teide - Pico Viejo.

In summary, several possible eruptive scenarios can be envisaged for the Teide - Pico Viejo stratovolcanoes according to their most recent volcanological history (Fig. 3.2). These include central and flank vent magmatic and phreatomagmatic eruptions of phonolitic and basaltic magmas, phreatic explosions, and sector collapses. All these scenarios may be preceded by unrest episodes of different origins and can generate a significant number of products (hazards). Each potential scenario is assigned a probability of occurrence.

3.3 Teide - Pico Viejo Bayesian Event Tree

An event tree is a tree graph representation of events in the form of nodes and branches. Each node represents a step and contains a set of possible branches (outcomes for that particular category). The nodes are alternative steps from a general prior event, state, or condition through increasingly specific subsequent events to final outcomes. The event tree includes all relevant possible outcomes of volcanic unrest at progressively higher degrees of detail. In the Bayesian model

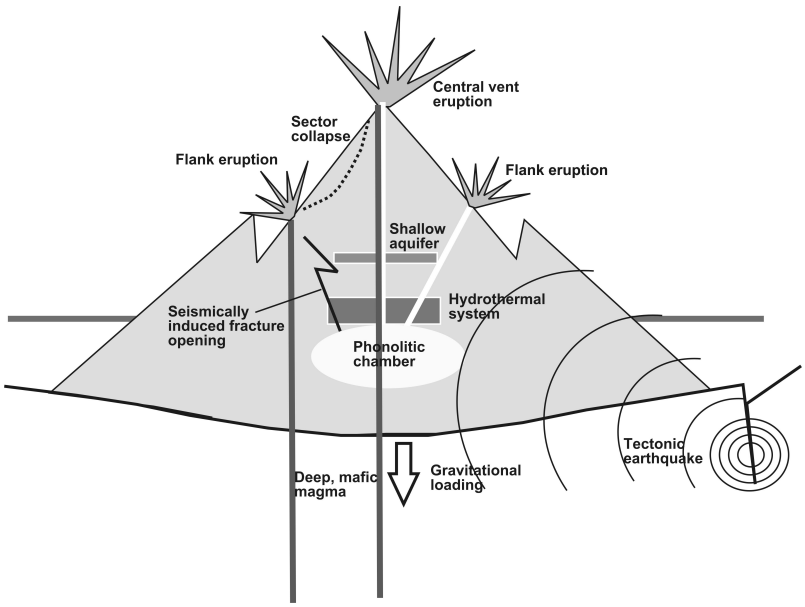

Figure 3.2: *Schematic representation of the most relevant potential unrest and eruptive scenarios for Teide - Pico Viejo stratovolcanoes (see text for more explanation).*

developed here, one condition is that the branches in each node are mutually exclusive and exhaustive.

To account for the possibility of flank vent eruptions, as opposed to only central eruptions, geothermal or seismic unrest, as opposed to only magmatic, phonolitic or basaltic composition, as opposed to no composition, and other relevant volcanic hazard possibilities for Teide - Pico Viejo, we have developed a new event tree structure which expands and complements that one previously proposed by Newhall and Hoblitt [2002] and Marzocchi et al. [2004, 2006, 2007], where Bayesian methodology was applied, and by Martí et al. [2008a], where Eliciting Expert Judgment was used.

Figure 3.3 shows the Bayesian event tree developed for Teide - Pico Viejo stratovolcanoes based on the geological information gathered from Figure 3.2. All events in each node are assumed mutually exclusive and exhaustive, this is, they do not happen simultaneously and the sum of probabilities of occurrence for different events in one node sums up to one.

3.3.1 Node 1: Unrest

Given that we have the capacity to differentiate the origin of the precursory signals, we define unrest as any modification of the background activity of the volcano recorded by the monitoring network and which may or may not be followed by an eruption of any kind.

The previous event tree (Martí et al. [2008a]) started with the volcano at unrest (including external triggers), to make the event tree as simple as possible. That leaves out the computation of the probability of unrest and the origin of this unrest. The Bayesian methodology used here allows to further expand this node into two new branches, "unrest" and "no unrest" in a given time window τ, and add a new node which accounts for the origin of this unrest. When there is no unrest we will compute the long term probabilities based on past data, expert judgment, scientific beliefs, etc., using the Bayesian methodology as explained in the next section. This will allow us to estimate the absolute probability of a specific event given past information, for example, the long term probability of a VEI 3 or less, basaltic, central vent, magmatic eruption with magmatic unrest in a given time window τ given information derived from past eruptions.

Short term probabilities could be computed in the event of some degree of unrest, provided that monitoring data was available. However, for the scope of this paper, we will only discuss long term probabilities.

3.3.2 Node 2: Origin of the Unrest

We define four types of unrest likely to happen in Teide - Pico Viejo stratovolcanoes: Magmatic, Geothermal, Seismic, and Other. Assuming in the future, in an optimal situation we can define the precursors which identify the source of the unrest, it is crucial in a complex system like Teide - Pico Viejo to differentiate between unrest caused by internal triggers or caused by external triggers, which ultimately may condition the outcome and further development of the system. The previous Teide event tree grouped all types of unrest in a unique branch called "unrest", and previous event trees defined for other volcanoes consider so far only one type of unrest of magmatic origin. From the study of the different eruption types identified on Teide - Pico Viejo, we can deduce that all of

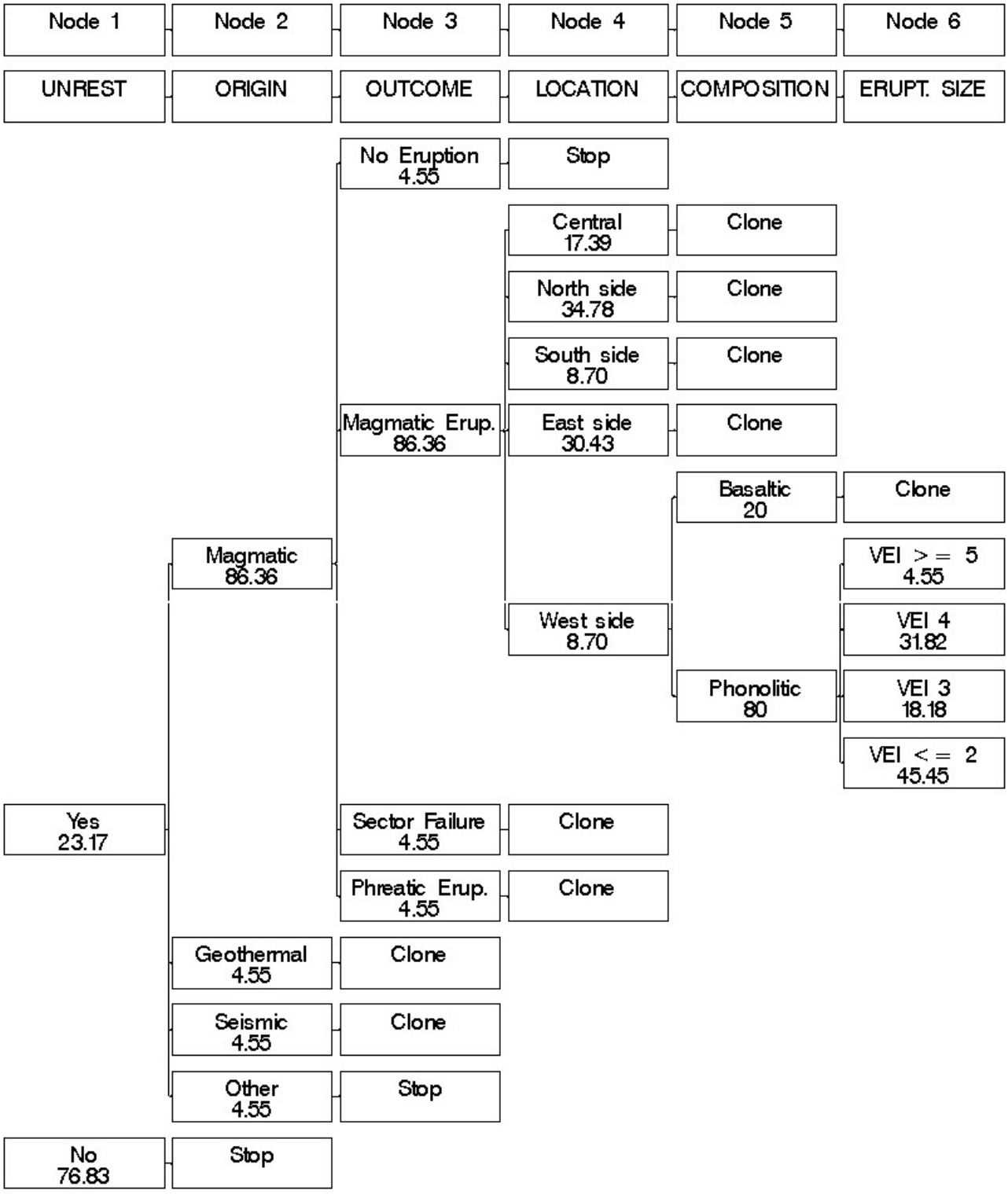

Figure 3.3: *Bayesian Event Tree for Teide - Pico Viejo. The six steps of estimation progress from general to more specific events (left to right). Any branch ending with "Clone" is identical to the detailed branch. Each branch contains the % long term absolute (posterior) probability of occurrence.*

them, including the phreatic episode (Ablay and Martí [2000]), require the presence of fresh magma, either mafic or felsic, at shallow depths in the volcanoes. However, we do not discard the possibility of starting an eruption process from an unrest directly associated with the hydrothermal system or even due to external triggers, such as regional tectonics, if eruptible magma is present in the system. In fact, the existence of an active hydrothermal system below Teide - Pico Viejo is evidenced by the presence of fumaroles and indirectly by geophysical data (Pérez et al. [1996], Coppo et al. [2008]). Volcanic unrest related to hydrothermal rather than to magmatic activity has been documented in similar volcanic systems (Gottsmann et al. [2003, 2007]). It is also important to mention that the interior of Tenerife is currently reacting to changes in the regional stress field or regional tectonics (Carniel et al. [2008], Tárraga et al. [2006], Martí et al. [2009]), so a seismic trigger for unrest cannot be ruled out.

Volcanic unrest and subsequent eruption represent an increase of the internal pressure of the volcanic system. This pressure increase may be caused directly by intrusion of new magma and/or pressurization of the associated hydrothermal system, or indirectly by reducing the external loading by a sector collapse or opening of a fracture during a tectonic episode. Because we define unrest based on geophysical and geochemical signals recorded by the monitoring network, we accept that the unrest (i.e the variation in this recorded signals) may derive from changes in the magma chamber due to intrusion of new (fresh) magma (magmatic unrest), changes in the hydrothermal system, having a magmatic origin or not (geothermal unrest) or changes in the host rock caused by regional seismicity (seismic unrest). However, there is also a possibility for a false unrest when non volcanic signals are recorded together with the volcanic ones. This is for example the case of variations in the recharge and extraction of meteoric water in/from the shallow aquifer inside the Las Cañadas caldera, which may cause changes in the gravity field, ground deformation, and even seismicity not related to any volcanic activity.

3.3.3 Node 3: Outcome of the Unrest

We consider here the outcome of the unrest being of four different types:

1 - Magmatic eruption: Triggered directly by a magmatic unrest, which may or may not be preceded by a sector failure, or triggered indirectly by a geothermal or seismic unrest, in which case, external decompression of the shallow volcanic system would be required. This could be achieved by sector failure or tectonic fracture opening.

When the unrest is geothermal or seismic, for a magmatic eruption to occur we would need a sector collapse first or fracture opening to decompress the whole system, so when we talk about a magmatic eruption originated by a geothermal or seismic unrest, we assume that a sector failure or a tectonically induced fracture opening have previously occurred.

2 - Sector failure: Triggered by a magmatic, geothermal or a seismic unrest. In this branch the outcome is the sector collapse itself, not being followed by an eruption. A sector failure followed by a magmatic eruption is considered in the previous branch (magmatic eruption), caused indirectly by a magmatic unrest which triggered a sector collapse.

3 - Phreatic eruption: Triggered by unrest of any type, where no magma is involved in the eruption.

4 - No eruption: There is unrest but no further outcome develops.

3.3.4 Node 4: Location

Teide - Pico Viejo has undergone several flank and central vent eruptions without any apparent structural or petrological pattern that could explain such random eruption behaviour (Martí et al. [2008b], Martí and Geyer [2009]). The lack of a good surveillance network and detailed knowledge on past activity makes the identification of future vents more challenging than in other volcanoes better known and monitored (eg. Vesuvius, Popocatépetl, etc).

Martí and Geyer [2009] show that the main control on the pathway of phonolitic magma between the shallow magma chamber and the surface is exerted by the stress field distribution around and above the chamber, this being a function of the shape and depth of the magma chamber. Comparison of these results with the available geological and geochronological information suggests that the number of flank eruptions that occurred on Teide - Pico Viejo during the time

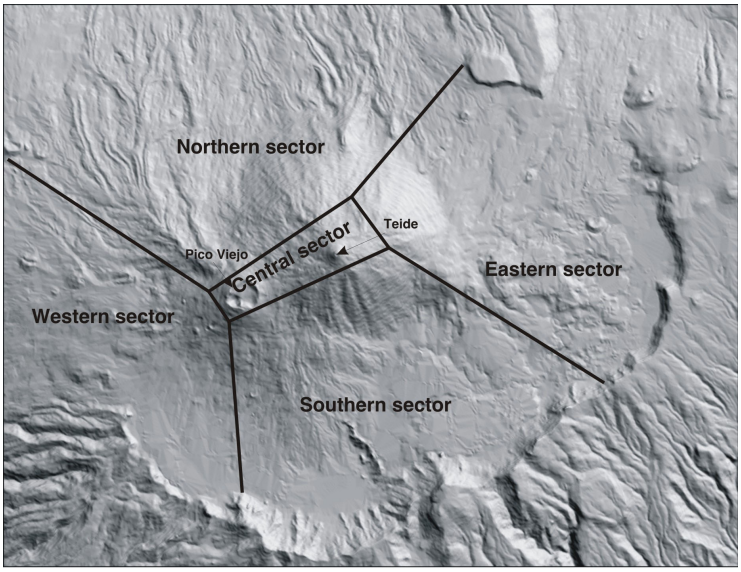

Figure 3.4: *DEM (Digital Elevation Model) of the central part of Tenerife Island showing the sectors distribution used to define the different source areas of the hazardous events (eruptions, sector collapses, lahars, etc) considered in the event tree.*

period considered (last 8000 years) is slightly higher than that of the central vent eruptions.

With respect to previous work (Martí et al. [2008a]), this node is an expansion of the node "Location" by segmenting the "Flank Vent" branch. Aside from the central vent location, we segmented the flank vent location into "North", "South", "East" and "West". Further segmentation is possible. The reason for this is that the impact of the different hazards that have occurred from each eruption may differ significantly depending on the exact location of the vent. The abrupt topography of the Teide - Pico Viejo and their surroundings, together with the presence of important topographic barriers such as the Cañadas caldera wall, impose a different level of hazard and risk depending on what side of the volcano the eruption occurs. The north side of the volcano is the area that poses the greater risk, due to the densely populated area and lack of any topographical protection from gravity driven flows.

Hence, node 4 "Location" segments the area around Teide - Pico Viejo volcano

complex into five sectors (Figure 3.4): Central, North, South, East, and West.

3.3.5 Node 5: Composition

Unlike previous event trees, the composition node is a new contribution. Teide - Pico Viejo's eruptive activity has been associated with both mafic (basaltic, tephri-phonolites) and felsic (phono-tephrites and phonolites) magmas. Hence, the magma composition will determine two main types of eruptions, basaltic or phonolitic, and they will have different hazard implications, as phonolitic magmas are associated with more violent eruptions than basaltic magmas.

There are two feeding systems in Teide - Pico Viejo that should be associated with different precursors. These will allow us to determine if the eruption is going to be associated with the phonolitic or the basaltic system. This conditions where the eruption is going to take place and the composition of the erupting magma, which will influence the level of hazard and the magnitude of the eruption. Phonolitic eruptions are more common in the central vents of Teide - Pico Viejo and their flanks, while basaltic eruptions are randomly distributed on the stratovolcanoes but also inside the caldera. It is obvious that outside and inside the caldera basaltic volcanism is tectonically controlled by the rift systems, which also affects the plumbing systems that allow deep magma to rise into the central complex (Ablay and Martí [2000], Martí et al. [2008b]). However, the structural constraints of the rift system that allow basaltic magma to reach the surface in Teide - Pico Viejo are still not clear. Therefore in terms of forecasting future eruptions from Teide - Pico Viejo it will be crucial to clearly distinguish between precursory activity of basaltic and phonolitic magmas. Thus, the composition in Node 5 to be of two types: basaltic or phonolitic. The importance in distinguishing these two outcomes for node 5 is the different level of hazard that is associated with each one (Martí et al. [2008b]).

3.3.6 Node 6: Size

Teide - Pico Viejo's eruptive activity has produced a large variety of eruption types and magnitudes. This node represents the magnitude of the eruption in

terms of the volcanic explosive index (VEI), categorized here in four possible outcomes: VEI ≥ 5, VEI $= 4$, VEI $= 3$, VEI ≤ 2.

Basaltic eruptions recorded in the geological record of Teide - Pico Viejo mostly correspond to strombolian to violent strombolian eruptions (VEI 2, VEI 3), while phonolitic eruptions may generate eruptions of VEI 5 or higher. The hazards that this wide range of eruptions have generated in the past and may produce in the future is rather variable and include lava flows and lahars of different volumes and run out distances, ash fallout, pumice-lapilli fallout, ballistic bombs and pyroclastic density currents.

3.4 Bayesian Model for Teide - Pico Viejo Event Tree

In Bayesian statistics, probability has the subjective interpretation. Bayesians use probability to make statements about the partial knowledge available concerning some underlying process or "state of nature" (un-observable or as yet unobserved) in a systematic way. The fundamental principle of Bayesian statistics is that what is known about anything that is incompletely or imperfectly known is described by a probability or probability distribution.

Bayesians regard both the observed data y and the unknown parameters θ as random variables. Posterior inference about θ is then conditional on the particular realization of y actually observed.

This is in contrast to classical inference, where only the data are regarded as random, while parameters θ are treated as fixed but unknown. Classical inference is not just conditional on the observed data, but on what might have been observed under repeated sampling.

Suppose we have data y and unknowns θ. We posit a model which specifies the likelihood $p(y|\theta)$.

From a Bayesian point of view, θ should have a probability distribution reflecting our uncertainty about it, and as y is known, should be conditional on θ.

Therefore our knowledge about θ is expressed through its posterior distribu-

tion $p(\theta|y)$:

$$p(\theta|y) \propto p(\theta) \times p(y|\theta) \tag{3.1}$$

$$\text{posterior} \propto \text{prior} \times \text{likelihood} \tag{3.2}$$

3.4.1 The prior distribution

The prior distribution, $p(\theta)$, expresses our uncertainty about θ before seeing the data. The posterior distribution, $p(\theta|y)$, expresses our uncertainty about θ after seeing the data.

The only requirement for the prior distribution is that it should represent the knowledge about θ before observing the current data. The prior can be specified entirely subjectively, depend on past data or be weak or non-informative.

We model the prior distribution for the j^{th} event at the k^{th} node with a Dirichlet distribution, which is the generalization of the Beta distribution, (Marzocchi et al. [2007]):

$$\theta_{prior_k} \approx Di_{J_k}(\alpha_{k1}, \alpha_{k2}, ..., \alpha_{kJ_k}) \tag{3.3}$$

Where J_k is the number of possible mutually exclusive and exhaustive events at the k^{th} node, and $\alpha_{k1}, \alpha_{k2}, ..., \alpha_{kJ_k}$ are the parameters of the distribution in that particular node. The choice of the Dirichlet (Beta) distribution is itself rather subjective. In general, theoretical models, a priori beliefs, and/or expert elicitation give estimation of the expected average of the prior distribution that represents the "best guess". Further details on this choice can be found in Marzocchi et al. [2004].

The Expected value (mean) E and variance V of the priori random variable (Eq. 3.3) from the k^{th} node and the n^{th} event, which follows a Dirichlet distribution, are:

$$E[\theta_{kn}] = \frac{\alpha_{kn}}{\left(\sum_{i=1}^{J_k} \alpha_{ki}\right)} \tag{3.4}$$

$$V[\theta_{kn}] = \frac{\alpha_{kn}\left(\sum_{i=1}^{J_k}\alpha_{ki} - \alpha_{kn}\right)}{\left(\sum_{i=1}^{J_k}\alpha_{ki}\right)^2\left(\sum_{i=1}^{J_k}\alpha_{ki} + 1\right)} \tag{3.5}$$

The expected value of those distributions represents an estimation of the aleatory uncertainty, i.e., the intrinsic (and unavoidable) random variability due to the complexity of the process. The dispersion around the average (i.e. the variance) represents an estimation of the epistemic uncertainty, due to our limited knowledge of the process. The estimation of the epistemic uncertainty is very important for correct comparison between the probabilities of different hazards, and the confidence limits that are ascribed to them (Woo [1999]). The variance can be seen as a sort of "confidence degree" of our a priori information, i.e., an evaluation of the epistemic uncertainties. The confidence degree is set up by writing the variance in terms of "equivalent number of data" (λ_k) (Marzocchi et al. [2007])

$$\lambda_k = \sum_{i=1}^{J_k}\alpha_{ki} - J_k + 1 \tag{3.6}$$

And then,

$$V[\theta_{kn}] = \frac{E[\theta_{kn}](1 - E[\theta_{kn}])}{\lambda_k + J_k} \tag{3.7}$$

The higher λ_k the larger our confidence on the reliability of the model, hence we need more past data to modify significantly the prior, but if we believe that the prior is poorly informative, λ_k is small, and so even a small number of past data can drastically modify the prior. In our case, we will use the minimum value for λ_k which is 1, this is the maximum possible epistemic uncertainty, since some of the past geological records we have are not accurate.

3.4.2 The likelihood function

The likelihood function allows us to use the past data (y_k) at node k to modify the a priori beliefs or priori distributions. In our model the data for each event in each node is a random variable that follows a multinomial distribution, which

is the generalization of the binomial distribution. In our case, we have:

$$[y_k|\theta_k] \approx Mu_{J_k}\left(y_{k1}, y_{k2}, ..., y_{kJ_k}; \theta_k\right) \tag{3.8}$$

Where J_k is the number of possible mutually exclusive and exhaustive events at the k^{th} node. Note this distribution assumes the data of the set y_k are iid, this is, independent and identically distributed.

3.4.3 The posterior Distribution

Since the Dirichlet and Multinomial are conjugate distributions, the posterior distribution for θ_k is still a Dirichlet:

$$\theta_{posterior_k} \approx Di_{J_k}\left(\alpha_{ki} + y_{ki}, \forall i = 1...J_k\right) \tag{3.9}$$

Where the parameter α_{ki} will be determined by:

$$\alpha_{ki} = E[\theta_{ki}]\left(\lambda_{ki} + J_k - 1\right) \tag{3.10}$$

As discussed before, $E[\theta_{kn}]$ is the central value inferred by a priori models and/or of the theoretical beliefs, and will account for the aleatory uncertainty, while λ_{ki} controls the confidence at which $E[\theta_{kn}]$ is considered a reliable estimate and will account for the epistemic uncertainty. Both these parameters will be inputs to the model.

3.4.4 The total probability

Once we have all the probability density functions for each branch in each node and the conditional probability assessment calculated, we combine all these probabilities to estimate the total long-term probability of a particular event. For example, the long term probability of having a magmatic unrest with central vent basaltic eruption in the time interval $(t_0, t_0 + \tau)$ is $P\left(U \cap Mo \cap Me \cap C \cap B\right) =$

$$\theta_{post_U}.\theta_{post_{Mo}}.\theta_{post_{Me}}.\theta_{post_C}.\theta_{post_B} \tag{3.11}$$

Where U is unrest, Mo is magmatic origin, Me is magmatic eruption, C is central and B is basaltic.

3.5 Long term hazard at the nodes for the Teide - Pico Viejo event tree based on existing data: Computation and Results

Suppose we have 80 time windows investigated ($n_1 = 80$) out of which 18 have had unrest ($y_{11} = 18$, $y_{12} = 62$), and we want to compute the long term probability of unrest in the next time window τ (each time window 100 years long) using Bayesian Inference. According to equation 3.9 the posterior probability follows a Dirichlet distribution of parameters (α_{11}+18) and (α_{12}+62), where $\alpha_{11} = \alpha_{12} = 1$ (Eq. 3.10), $E[\theta_{11}] = E[\theta_{12}] = 0.5$ (prior weight) and $\lambda_{11} = \lambda_{12} = 1$ (data weight). Hence, the long term probability of having an unrest in the next time window is the expected value of a random variable that follows a Dirichlet distribution with parameters (1 + 18) and (1 + 62), which by definition (Eq. 3.4) is:

$$\frac{\alpha_{11} + y_{11}}{(\alpha_{11} + y_{11}) + (\alpha_{12} + y_{12})} = \frac{1 + 18}{1 + 1 + 80} = 0.2317$$

Using R, a free language and environment for statistical computing and graphics (available at http://www.r-project.org/), a code has been developed to apply the above Bayesian model to the Teide - Pico Viejo data (code available upon request). Table 3.1 shows eighteen eruptions recorded geologically for the last eight thousand years, using a time window of one hundred years, we have eighteen events in the last eighty time windows. Table 3.2 shows the input data for the model.

There are no historical records to know if there where any false alarms. We assume that all the eruptions geologically documented here where preceded by volcanic unrest. Hence, during the 80 time windows of 100 years each, there were 18 episodes of unrest versus 62 of no unrest.

The probability function named $ltvh$, programmed in R, computes the long-term probability vector (last column of Table 3.2) for each node, and is defined

Table 3.1: *Geological records of eruptions at Teide - Pico Viejo volcanic complex for the last eight thousand years (data from Ablay and Martí [2000], Carracedo et al. [2007]; unpublished data). Since there are no historical records, we assumed that an episode of unrest correspond to each eruption geologically documented. (Mag.=Magmatic; E=Eruption; W=West side; N=North side; C=Central; S=South side; E=East side. Comp.=Composition; B=Basaltic; P=Phonolitic)*

		Node 1	**Node 2**	**Node 3**	**Node 4**	**Node 5**	**Node 6**
Eruption name	Year	Unrest	Origin	Outcome	Location	Comp.	VEI
Chahorra	1798	Yes	Mag.	Mag. E	W	B	≤ 2
Mta Reventada	895 bp	Yes	Mag.	Mag. E	N	B	≤ 2
Lavas Negras	1150 bp	Yes	Mag.	Mag. E	C	P	3
Roques Blancos	1714 bp	Yes	Mag.	Mag. E	N	P	4
Mta Blanca	2000 bp	Yes	Mag.	Mag. E	E	P	3
PV surges	(2528-2000) bp	Yes	Mag.	Mag. E	C	B	≤ 2
Hoya del Cedro	(2528-2000)bp	Yes	Mag.	Mag. E	N	P	4
Mta Majua	(2528-2000) bp	Yes	Mag.	Mag. E	S	P	≤ 2
Mta de la Cruz	(2528-2000) bp	Yes	Mag.	Mag. E	E	P	≤ 2
Arenas Blancas	(2528-2000) bp	Yes	Mag.	Mag. E	E	P	≤ 2
Mta Los Conejos	(2528-2000) bp	Yes	Mag.	Mag. E	E	P	≤ 2
Bocas de Maria	(2528-2000) bp	Yes	Mag.	Mag. E	E	P	≤ 2
Mta Las Lajas	(2528-2000) bp	Yes	Mag.	Mag. E	E	P	≤ 2
El Boqueron	2528 bp	Yes	Mag.	Mag. E	N	P	4
Cañada Blanca	(5911-2528) bp	Yes	Mag.	Mag. E	C	P	3
Abejera Baja	5911 bp	Yes	Mag.	Mag. E	N	P	4
Abejera Alta	5486 bp	Yes	Mag.	Mag. E	N	P	4
Pico Cabras	(7900-5486) bp	Yes	Mag.	Mag. E	N	P	4

Table 3.2: *Long term probability estimation for each branch of the event tree, using Bayesian Inference and geological records. Past events are based on the geological records for the last 80 time windows (each time window is 100 years) shown in Table 3.1. Weights for the prioris, aleatory and epistemic uncertainties are equally distributed across branchs of the same node based on the assumption of non-informative priors. (Prob.=Probability)*

Node #	**Node**	**Branch**	**Past events**	**Prior weight**	**Data weight**	**Prob. (%)**
1	unrest	yes	18	0.5	1	0.2317
1	unrest	no	62	0.5	1	0.7683
2	origin	magmatic	18	0.25	1	0.8636
2	origin	geothermal	0	0.25	1	0.0455
2	origin	seismic	0	0.25	1	0.0455
2	origin	other	0	0.25	1	0.0455
3	outcome	Magmatic Eruption	18	0.25	1	0.8636
3	outcome	Sector Failure	0	0.25	1	0.0455
3	outcome	Phreatic Eruption	0	0.25	1	0.0455
3	outcome	No Eruption	0	0.25	1	0.0455
4	location	Central	3	0.2	1	0.1739
4	location	North	7	0.2	1	0.3478
4	location	South	1	0.2	1	0.087
4	location	East	6	0.2	1	0.3043
4	location	West	1	0.2	1	0.087
5	composition	Basaltic	3	0.5	1	0.2
5	composition	Phonolitic	15	0.5	1	0.8
6	size	VEI ≥ 5	0	0.25	1	0.0455
6	size	VEI4	6	0.25	1	0.3182
6	size	VEI3	3	0.25	1	0.1818
6	size	VEI ≤ 2	9	0.25	1	0.4545

as:

$$f(pb1, pb2, pb3, pb4, pb5, pb6, y1, y2, y3, y4, y5, y6) \tag{3.12}$$

where:

$pb1$, ..., $pb6$ are the J_k*2 matrices containing the $E[\theta_{ki}]$, λ_{ki} parameter information for each node ($k = 1, ..., 6$) and each event within each node ($i = 1, ..., J_k$) (Table 3.2, *Prior* and *Data weight* columns). Weights are equally distributed based on the assumption of non-informative priors.

$y1, ..., y6$ are the vectors with the geological and geophysical data for each node, each vector is of dimension $1 * J_k$ (Table 3.2, *Past events* column).

Due to the fact that each volcano is unique, we cannot use validation data from analogs. To validate the model we used the previous model BET_EF already published (Marzocchi et al. [2007]) and verify that the new extended model yields the same results as the former one when the same parameters are given. After the model has been validated for accuracy of the results, below we present some examples and interpretations.

The event tree on Figure 3.3 shows the long term probabilities computed with the model presented above, this is, the posterior probabilities computed as shown in Section 4.3. This probabilities are also shown in the last column of Table 3.2 and are then used to compute the total long term probability for each eruptive scenario (Eq. 3.11) as explained in Section 4.4. Some examples are shown below.

According to our model, the long term probability of a basaltic central vent magmatic eruption with magmatic unrest in the next time window is: 0.2317 * 0.8636 * 0.8636 * 0.1739 * 0.20 = 0.6%. This is, the probability of an unrest (0.2317), times the probability that this unrest is magmatic (0.8636), times the probability that this magmatic unrest derives into a magmatic eruption (0.8636), times the probability that this magmatic eruption is central (0.1739), times the probability that the composition is basaltic (0.20), is 0.6%. Similarly, the long term probability of a phonolitic central vent magmatic eruption with magmatic unrest in the next time window is: 0.2317 * 0.8636 * 0.8636 * 0.1739 * 0.80 = 2.4%, four times the probability of a central magmatic eruption of basaltic composition.

The branch called "Yes" in the tree node "Unrest" refers to episodes of un-

rest that could either evolve into false alarm or into eruption. Even though we assume the 18 geologically documented eruptions followed unrest, the model also computes a probability of false alarm, this is, having an unrest which does not evolve into an eruption: 0.2317 * 0.0455 * (0.8636 + 0.0455 + 0.0455) = 1.01%.

The probability of "No Eruption" would be the sum of the probability of a false alarm, this is, 1.01%, plus the probability of a false unrest (0.2317 * 0.0455 = 1.05%), plus the probability of no unrest of the volcanic system (76.83%). Thus, the long term probability of not having an eruption in the next time window is 78.89%.

To compute the long term probability of having a magmatic eruption of any origin in the next time window, we have to consider all the sources that can trigger a magmatic eruption (magmatic, geothermal or seismic unrest). This is, the long term probability of having a magmatic eruption caused by (1) a magmatic unrest (0.2317 * 0.8636 * 0.8636 = 17.28%), (2) a geothermal unrest (0.2317 * 0.0455 * 0.8636 = 0.91%), (3) a seismic unrest (0.2317 * 0.0455 * 0.8636 = 0.91%). Thus, the long term probability of having a magmatic eruption in the next time window is 19.10%.

3.6 Discussion and conclusions

We have developed a new event tree model for long term volcanic hazard assessment based on Bayesian methodology and that represents a step forward with respect to previous attempts based on the same methodology or the elicitation of expert judgment. We have applied this new model to the particular case of Teide - Pico Viejo stratovolcanoes, but it may be applied to other active volcanoes.

In comparison with previous event trees based on Bayesian methodology (Newhall and Hoblitt [2002], Marzocchi et al. [2007]), the model presented here accounts for the possibility of the unrest being caused by external triggers (geothermal, seismic), and adds new nodes with two additional sources of volcanic hazard based on the composition of the magma and different vent locations. With respect to event trees based on Elicitation of Expert Judgment (Neri et al. [2008], Martí et al. [2008a]) the new model does not have the additional source of bias that the human decision component adds to the final results of the elicitation

method, controls for the epistemic and aleatory uncertainties, and allows the level of segmentation and complexity of the event tree structure to be as complete and extensive as needed, with the only requirements of mutually exclusive and exhaustive events in each node. It also permits to automatically update the probabilities when new data arrives or the system becomes active and monitoring data on precursors exists, as opposed to the eliciting method which requires the group of experts to meet each time new data arrives to update the probabilities. However, during a volcanic crisis, Elicitation and Bayesian models are needed and the elicitation team should provide input and interpretation to the probabilities from the updated Bayesian model.

In order to understand the differences between this new model of event tree and the previous ones, let us to consider the example presented in this paper. With respect to previous hazard assessment, we have expanded the existing event tree for Teide Pico Viejo (Martí et al. [2008a]) and estimated the long term probability of an eruption when this can be caused by internal as well as external triggers, consider magma composition as an additional source of hazard and segmented the flank vent location into four lateral locations. To do this we have used Bayesian methodology applied to a more detailed and complete event tree scheme. By doing this, we have been able to account for additional eruptive scenarios that were not contemplated before and more accurately estimate the long term probabilities of an eruption within a given time window.

Therefore, in a volcanic system like the one formed by the Teide - Pico Viejo stratovolcanoes we may consider different possibilities of unrest, each one with its associated probability. However, if we do not contemplate the possibility of an unrest caused by external triggers, parts (2) and (3) in the example presented in the previous section to compute the probability of having a magmatic eruption would be set to zero, and the total probability of having a magmatic eruption would be computed only for the case of the unrest being magmatic, in the example in previous section, this is (17.28%). In this case, the result should be the same as the one computed with the BET_EF (Marzocchi et al. [2007]), where the weight assigned to the branches "geothermal" and "seismic" in node 2 "origin" would be zero and all the weight would go for the remaining two branches "magmatic" and "other". However, in the future some volcanic systems may want to considered

the possibility of an eruption caused by external triggers, as we have done here for Teide - Pico Viejo, in which case this new event tree could be applied to that volcano. On the other hand, if we consider only the origin of the unrest being magmatic on a volcanic system with an hydrothermal system underneath or high seismic activity, then we are underestimating the volcanic hazard of an eruption, as the probability of an eruption originated by a geothermal or seismic unrest would be embedded in the residual probability of "other" origin and not taken into account. This would happen if we use the abovementioned BET_EF model to compute the probability of a magmatic eruption in Teide - Pico Viejo, which yields a 17.28% versus the 19.10% computed with our model, since the former only considers internal (magmatic) triggers.

Assuming the same probability of unrest than in our Bayesian event tree, and using the Eliciting event tree made for Teide - Pico Viejo to compute the probability of an eruption (Martí et al. [2008a]), we get: ((unrest) 0.2317 * (no sector failure) 0.92 * (eruption) 0.23) + ((unrest) 0.2317 * (sector failure) 0.08 * (eruption) 0.44) = 5.72% , while the probability of an eruption computed with the Bayesian model, as explained in the previous section, is 19.10%, more than three times higher.

Despite these differences between the different methods, we believe that Expert Elicitation and Bayesian Inference complement each other and must be used simultaneously during volcanic crises, where Bayesian approach provides a way to quick and automatically update the final probabilities, but the lack of information on precursors and triggers for each branch makes it impossible to automatically compute the short term probabilities. Further research is needed to define the precursors and automate the Bayesian event tree to be used in the short term.

Our event tree does not include a node with all the different geological hazards from the eruption (lahars, pyroclastic flow, ash fall, etc.) because these events can happen simultaneously, hence are not mutually exclusive, which is one of the conditions we made on the Bayesian model. Again, further work is needed to address this issue. However, the method accounts for the aleatory (intrinsic) and epistemic (due to scarce knowledge) uncertainties, allowing us to merge theoretical models, geological data and expert elicitation exercises to assign a long term probability to each eruptive scenario in a more realistic manner.

The new method allows us to estimate the long-term probability during a quiet period of the volcano, being useful for land use policy, and will be of use for estimating and automatically updating the short term probabilities when monitoring data are obtained during unrest.

Although this method is specifically applied to the Teide - Pico Viejo stratovolcanoes in Tenerife, it can be used with other similar volcanoes as it offers a wider structure in comparison with previous event trees that have a more restricted structure and do not include some relevant eruptive scenarios which are likely in the Teide - Pico Viejo but also in many other composite volcanoes.

Chapter 4

Statistical data analysis of the CCDB (Collapse Caldera Database): Insights on the formation of caldera systems

4.1 Introduction

Collapse calderas are defined as volcanic depressions that result from the disruption of the magma chamber roof due to down faulting during the course of an eruption [Branney, 1995; Druitt and Sparks, 1984; Gudmundsson, 1988, 1998; Lipman, 1984; Martí et al., 1994a; Smith, 1979; Smith and Bailey, 1968; Walker, 1984; Williams, 1941; Williams and McBirney, 1979]. The diameter of these volcanic depressions, usually more or less circular or elliptical in form, is many times greater than the diameter of the associated eruptive vents [Lipman, 1997, 2000]. Despite their low frequency of occurrence, large pyroclastic eruptions and associated collapse calderas represent one of the most catastrophic geologic events that have occurred on the Earth's surface during Phanerozoic times, causing considerable impacts on the environment (e.g. climate) and on human society (e.g. Tambora, 1815 [Newhall and Dzurisin, 1988], Krakatau, 1883 [Newhall and Dzurisin, 1988; Self and Rampino, 1981; Simkin, 1983] and Pinatubo, 1991 [Dartevelle

et al., 2002; Hattori, 1993; Lipman, 2000]). Additionally, collapse calderas have received considerable attention due to their link to ore deposits and geothermal energy resources [Lipman, 2000]. Calderas have been analyzed through field studies, analogue models and numerical simulations. Several attempts to determine the relationships between caldera size and various parameters such as magma chambers size have been made in previous studies [Hughes and Mahood, 2008; Smith, 1979; Spera and Crisp, 1981]. Also some reviews on collapse calderas have been published [Cole et al., 2004; Lipman, 1997; Martí et al., 2008c]. However, some important aspects on caldera dynamics and structure still remain uncertain and controversial.

Traditionally, field studies have constituted the most important way to investigate and understand volcanic processes. Since dealing with such a large amount of information from original sources is usually unfeasible, easier-to-use databases compiling the existing information on field studies of collapse calderas are required. Recently, Geyer and Martí [2008] have presented the "Collapse Caldera DataBase (CCDB)" (http://www.GVB-csic.es/CCDB.htm), a comprehensive catalogue including most of the known or identified collapse calderas produced by either explosive eruptions or effusive basaltic activity. The final aim of the CCDB is to update the current field based knowledge on calderas, by merging together the above mentioned databases and complementing them with the existing peer-reviewed articles on calderas. The current version of the CCDB contains more than 473 calderas worldwide and the included information comes from around 400 peer-reviewed papers. In addition to its intrinsic value as database, this extensive data compilation should become an accessible and useful tool for caldera studies, by the application of accurate and comprehensive numerical analyses.

The area of the caldera is a variable that results from the caldera process and depends on other variables that control it, in statistical terms, a function of some independent variables. These independent variables include the size and shape of the magma chamber, the depth of the magma chamber, strength of host rock, influence of regional and local tectonics, size of the caldera-forming eruption, etc. [Acocella et al., 2000, 2001, 2004; Folch and Martí, 2004; Geyer et al., 2006; Gudmundsson, 1998; Gudmundsson et al., 1997; Kennedy et al., 2004; Lavallée

et al., 2004; Martí et al., 1994a; Roche and Druitt, 2001; Roche et al., 2000]. Therefore, understanding how these variables determine the area of the caldera will broaden our understanding of the dynamics of the corresponding volcanic system.

The objective of this paper is to study the area of the caldera by means of a statistical analysis, on a homogeneous sample of the CCDB. We look for some pattern or statistical association between the size of the caldera and those key variables which define the geodynamic environment where caldera develops. Such a key variables are assumed to be directly related and have an impact on the size of a caldera structure, helping explain the dynamics of such volcanic systems. First, we extract a uniform sample from the CCDB database with the relevant variables for the study and then perform a statistical analysis using ANOVA (Analysis of Variance) on the selected variables.

Most of the findings of the statistical analysis performed here are already assumed from a geological point of view [Cole et al., 2004; Lipman, 1997; Martí et al., 2008c; Smith, 1979]. The aim of this paper is to study and test quantitatively some of those geological assumptions already shared by volcanologists, and to uncover other facts that would have gone unnoticed otherwise.

4.2 The Collapse Caldera DataBase (CCDB)

Information included in the CCDB was compiled from a wide range of primary and secondary sources of information, which are referenced for each database entry. These information sources include the IAVCEI and the Smithsonian Museum of Natural History databases (http://www.iavcei.org/ and http://www.volcano.si.edu/, respectively) and the works of Spera and Crisp [1981], Walker [1984] and Newhall and Dzurisin [1988]. Additionally, Geyer and Martí [2008] performed a comprehensive compilation of the most representative publications.

Originally the CCDB contained data on 361 calderas. Further updates have allowed to release a new version of the database that includes 473 calderas worldwide. The CCDB is available online at the website of the CSIC Group of Volcanology of Barcelona (http://www.GVB-csic.es/CCDB.htm). The CCDB architecture is based on the principle that all the information concerning the different

calderas included in the database has to be comparable and consistent enough for future comparisons and data analyses [Geyer and Martí, 2008]. The information included in the CCDB can be classified into the following classes: caldera depression (e.g. dimensions, morphology, age), caldera-forming deposits (e.g. volume and thickness of the deposits), associated magmatic system (e.g. magma composition), geodynamic setting where the caldera is located (e.g. crustal type, plate tectonic setting), type of pre-caldera volcanism, caldera-forming eruption sequence (deduced from the sequence of deposits) and post-caldera evolution (e.g. post-caldera volcanism, resurgence, caldera erosion). The reader is addressed to the original paper and the CCDB website (http://www.GVB-csic.es/CCDB.htm) to get more information on each class.

All the information included in the CCDB comes from published data and the references used in each case are adequately indicated. In some cases, due to the scarcity or ambiguity of the available information or the lack of consensus it is difficult or almost impossible to decide objectively on the specific information concerning the classes. This is, for example, the case of the calderas in Italy. The models that are generally invoked to explain the origin of volcanism in central Italy consider either subduction and back-arc mantle updoming [Di-Girolamo, 1978; Doglioni, 1991; Keller et al., 1994; Peccerillo, 1985; Thompson, 1977], or intracontinental rift environment unrelated to subduction [Cundari, 1979; Decandia et al., 1998; Stoppa and Lavecchia, 1992]. In these cases we have indicated the existing discrepancies by putting a "?" instead of the corresponding class. This implies that the information for some calderas is incomplete and we have not used these examples in our statistical study. The only exception is for those calderas associated with the Basin and Range tectonics. The last version of the CCDB includes the class "Basin and Range" to characterize the tectonic setting of many of the calderas located in North America (e.g. Platoro, Bachelor). There exist different hypotheses concerning the origin of the Basin and Range extension which include, beyond others, rifting processes related to the subducting slab (i.e. back-arc rifting) [Dickinson, 2002] or lithospheric extension without the influence of the preceding subduction process [Hawkesworth, 1995; Hooper et al., 1995]. In such case, we would assign to the tectonic setting a "?" instead of deciding between a plate tectonic class, however, this would leave more than 30 calderas

without information concerning the *Plate Tectonic Setting*. Thus, we have decided to include a special class "Basin and Range" to be able to include these calderas in the analysis. In the future, further information concerning the Basin and Range taphrogen will help to constrain better the tectonic setting of these collapse calderas.

4.3 The CCDB Sample Data for the study

A sample of the 473 calderas and four relevant variables have been selected for the analysis of the *Caldera Area*. The selection was done based on variable relevance and available data. The missing values are due to lack of information or consensus, so we look at variables with at least 50% of non missing data. The selected variables are: *Caldera Area*, *Rock Suite*, *Crustal Type*, *Plate Tectonic Setting* and *World Region*. Table 4.1 shows all variables in the CCDB and the percentage of missing information in each one (see Geyer and Martí [2008] for further detail on each variable). The variable *Caldera Area* has 402 calderas with the area informed. There is a considerable number of variables with high percentage of missing data, and therefore not suitable for the study. Of those variables with at least 50% of non missing data, variable *Age* is not relevant to explain the size, *Rock Suite* shows in better format the same information than *Magma Composition*. In the case of *Plate Tectonic Setting* and *Tectonic Faulting*, whereas the first variable describes the plate tectonic setting, the second gives information about the type of local and regional structures that may have influenced pre-caldera volcanism, caldera formation and/or the distribution of post-caldera volcanism. Since we are interested in defining the geodynamic environment in general terms we use the variable *Plate Tectonic Setting*, however further detailed analysis of specific caldera samples have to consider also the regional/local information contained in the variable *Tectonic faulting*. With the above, to analyze *Caldera Area*, the relevant variables available for the study, with sufficient information to carry out an analysis of the geodynamic environment where calderas form are: *Rock Suite*, *Crustal Type* and *Plate Tectonic Setting*. Last, for informative purposes, we will look at the distribution of calderas according to *World Region*. In the future, the analysis can be further expanded provided new information becomes available.

Table 4.1: *CCDB Variables with Caldera Area informed and their missing values (N = number of calderas). A detailed explanation of the different variables can be found in Geyer and Martí [2008] and the CCDB website (http://www.GVB-csic.es/CCDB.htm).*

Calderas with Area informed: 402	Missing N	Missing %	Include
Age	90	24	
Subsidence (km)	346	86	
Volume (km^3)	346	94	
Subsidence type/geometry	287	71	
Thickness of deposits (km)	348	87	
Volume of deposits (km^3)	239	80	
Volume magma DRE (km^3)	275	88	
Magma composition	77	19	
Rock suite	59	30	Yes
Chamber depth (km)	367	93	
Plate tectonic setting	75	28	Yes
Crustal type	55	22	Yes
Tectonic faulting	118	33	
Pre-caldera volcanism	220	54	
Timing caldera onset	311	77	
Post caldera volcanic activity	287	71	
Latitude	11	3	
Longitude	11	3	
World region	0	0	Yes
Subregion	0	0	
Max. diameter (km)	26	4	
Min. diameter (km)	27	4	
Regional pre-caldera doming	0	0	
Post-caldera resurgence	0	0	

Table 4.2 shows all five variables and their categories. The definition of the individual categories was established by Geyer and Martí [2008] and has been recently revised for the new version of the CCDB (http://www.GVB-csic.es/CCDB.htm). The different crustal types (Table 4.3) and plate tectonic settings have been defined on the basis of consulted references [Carey, 2005; Condie, 1993; Kearey and Vine, 1996; Turcotte and Schubert, 2002; Uyeda, 1982]. For the particular case of *Plate Tectonic Setting* we have included Figure 4.1 to illustrate each category. According to *Plate Tectonic Setting* 59% of the calderas are located in the Chilean-type subduction, Basin and Range and continental rift. Regarding the *Crustal Type* (excluding category "unknown") calderas are grouped in six different categories, with 75% of the calderas located in the continental silicic (thin, $\leq 30-35$ km, and thick, $> 30-35$ km) and transitional thick crust. Of the 402 calderas, 14% have missing *Crustal Type* information. Similarly, there are twelve different types of *Rock Suite* composition, apart from a 15% of calderas with unknown composition. Of these, 194 calderas (48%) are of calc-alkaline felsic composition.

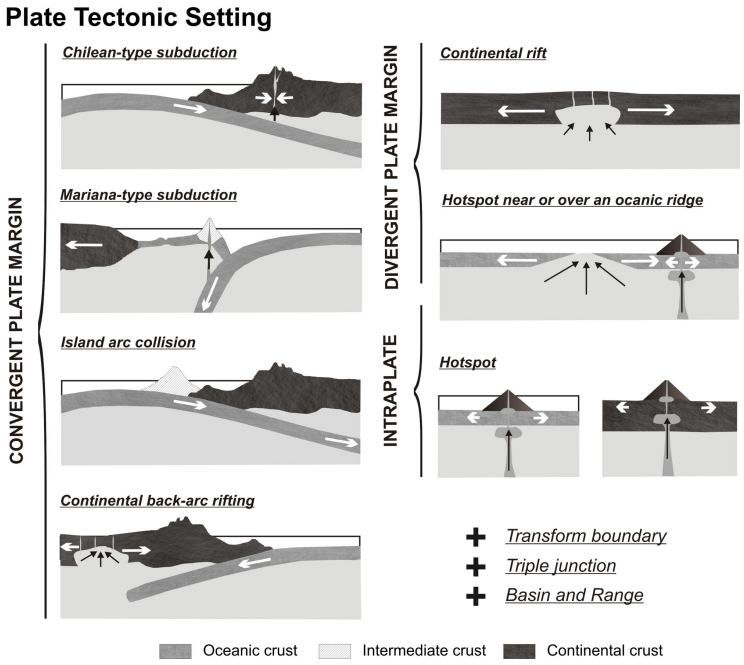

Figure 4.1: Sketch of the different plate tectonic settings.

Table 4.2: *Summary of variables from sample CCDB used for the analysis (Note: % are over the 402 calderas with the area informed. N = number of calderas)*

Caldera Area		
Calderas: 402		
Min: 0.03 km^2; Max: 4712 km^2		
Mean:206 km^2; Std dev:477 km^2		
Crustal Type	N	%
Unknown	55	14
Continental silicic thick crust	115	29
Continental silicic thin crust	105	26
Transitional thick crust	80	20
Oceanic basaltic thin crust	25	6
Oceanic basaltic thick crust	16	4
Transitional thin crust	6	1
Rock Suite	N	%
Unknown	59	15
Calc-alkaline felsic	194	48
Alkaline felsic	28	7
Peralkaline-alkaline	21	5
Calc-alkaline intermediate	17	4
Tholeiite	16	4
Calc-alkaline intermediate-felsic	15	4
Alkaline mafic	11	3
Calc-alkaline mafic	11	3
Calc-alkaline alkaline	10	2
Alkaline intermediate	5	1
Calc-alkaline mafic-felsic	5	1
Peralkaline-calc-alkaline	4	1
Alkaline mafic-felsic	3	1
Alkaline intermediate-felsic	2	0.5
Calc-alkaline mafic-intermediate	1	0.2

Plate Tectonic Setting	N	%
Unknown	75	19
Chilean-type subduction	159	40
Basin and Range	44	11
Continental rift	32	8
Hotspot	24	6
Hotspot near/over oceanic ridge	24	6
Island arc collision, transform boundary, triple plate junction	16	4
Mariana-type subduction	15	4
Continental back-arc rifting	13	3
World Region	N	%
North America	79	20
Japan and Mariana Islands	53	13
South America	36	9
Mediterranean	35	9
Africa	33	8
Central America	30	7
Kamtchatka and Mainland Asia	24	6
Indonesia	23	6
Alaska	17	4
Melanesia	14	3
New Zealand, Tonga and Kermadec Islands	10	2
Kurile Islands	10	2
Atlantic Islands	9	2
Iceland	8	2
Hawaii	7	2
Antartica	6	1
Philippines	4	1
Indian Ocean	2	0.5
Caribbean	2	0.5

Table 4.3: *Classes of the Crustal Type variable. The transitional crust has an intermediate composition between the continental and the oceanic crust. Although the main common crustal divisions are: oceanic, transitional and continental, it is typical to find zones of thinned or thickened crust for examples in areas of continental rifting or very thick transitional crust in highly evolved island arcs. Consequently, an accurate classification has to take into account both parameters: thickness and composition. The thickness thresholds from thick to (standard-) thin are established according to the definition of Condie [1993]. Important problems appear classifying island arcs. Their composition should be theoretically transitional (e.g. Izu-Osawa arc) [Condie, 1993]. However, these may range from almost oceanic (e.g. Western Aleutians) to practically continental (e.g. Japan) [Condie, 1993; Kearey and Vine, 1996].*

Crustal Type	Thickness (km)	Composition
Continental silicic thick (C)	$>$ 30-35	silicic (granitic)
Continental silicic (standard-) thin (Cd)	$\leq$ 30-35	silicic (granitic)
Oceanic basaltic thick (O)	$\geq$ 10-15	mafic (primarily basaltic)
Oceanic basaltic (standard-) thin (Od)	$<$ 10-15	mafic (primarily basaltic)
Transitional thick (T)	$\geq$ 20-25	intermediate composition
Transitional (standard-) thin (Td)	$<$ 20-25	intermediate composition

Table 4.4 shows the distributions of the various populations of *Caldera Area* to be compared. Calderas in continental silicic thick crust have the largest mean area (411 km^2), as well as those with peralkaline-calc-alkaline (357 km^2), calc-alkaline alkaline (333 km^2) and calc-alkaline felsic rock composition (322 km^2). Calderas associated with hotspot (863 km^2) and in the Basin and Range (368 km^2) plate tectonic setting are by far the largest on average.

4.4 The Statistical Methodology

Our hypothesis is that the area of the caldera depends on *Crustal Type*, and/or *Rock Suite*, and/or *Plate Tectonic Setting*, the main variables that define the geodynamic environment where the caldera develops, in addition to the size and depth of the magma chamber, two variables that are not included in the CCDB

Table 4.4: *Distribution of Caldera Area (km^2) according to Crustal Type, Rock Suite and Plate Tectonic Setting. (Note: values for the mean, standard deviation, minimum and maximum are in km^2. N = number of calderas)*

Crustal Type	N	Mean	Std Dev	Min.	Max.
Continental silicic thick crust	104	411	794	1	4712
Continental silicic thin crust	87	206	338	1	2000
Transitional thick crust	66	102	111	0	471
Oceanic basaltic thin crust	24	27	33	0	163
Oceanic basaltic thick crust	13	61	51	15	184
Transitional thin crust	3	44	30	24	79
Rock Suite	N	Mean	Std Dev	Min.	Max.
Calc-alkaline felsic	173	322	662	2	4712
Peralkaline-alkaline	21	49	51	3	170
Alkaline felsic	20	79	108	1	380
Tholeiite	15	52	48	10	184
Calc-alkaline intermediate	13	103	94	13	358
Alkaline mafic	11	36	48	0	163
Calc-alkaline intermediate-felsic	11	220	191	7	550
Calc-alkaline alkaline	10	333	266	20	628
Calc-alkaline mafic	10	46	49	0	151
Calc-alkaline mafic-felsic	4	55	45	31	123
Peralkaline-calc-alkaline	4	357	279	33	707
Alkaline mafic-felsic	3	68	115	1	201
Alkaline intermediate	1	20	.	20	20
Calc-alkaline mafic-intermediate	1	24	.	24	24
Plate Tectonic Setting	N	Mean	Std Dev	Min.	Max.
Chilean-type subduction	144	191	314	0	2160
Basin and Range	44	368	366	47	1500
Continental rift	31	101	111	1	434
Hotspot near/over oceanic ridge	24	38	43	0	184
Hotspot	21	863	1575	10	4712
Continental back-arc rifting	12	158	124	11	363
Island arc collision, transform boundary, triple plate junction	11	69	65	13	236
Mariana-type subduction	10	47	33	20	102

but which have been clearly identified as crucial in experimental and numerical models of collapse calderas (see Martí et al. [2008c] for a review). The question of interest is whether this dependence really exists. The statistical methodology used to analyze this hypothesis, given the number of calderas and the variables in the dataset, is a "Nonparametric one-way unbalanced ANOVA using the Kruskal-Wallis test" [Rice, 1995].

In analysis of variance (ANOVA), a continuous response variable, *Caldera Area*, known as a dependent variable, is measured under experimental conditions identified by classification variables (*Crustal Type*, *Rock Suite*, *Plate Tectonic Setting*), known as independent or explanatory variables. The variation in the response variable is assumed to be due to effects in the classification variables, with random error accounting for the remaining variation. In general, the purpose of analysis of variance (ANOVA) is to test for significant differences in the response variable among two or more classification groups, in our case by looking at the group mean *Caldera Area*. The name "Analysis of Variance" is derived from the fact that in order to test for statistical significance between means, we are actually comparing (i.e., analyzing) variances.

In our case we have the response variable, *Caldera Area*, and three classification variables, *Crustal Type*, *Plate Tectonic Setting* and *Rock Suite*. Each classification variable has two or more categories which make different groups of calderas. We want to look at *Caldera Area* in each of these groups and determine if the mean area is statistically different across different groups. The qualitative variables (*Crustal Type*, *Plate Tectonic Setting*, *Rock Suite*) serve to distinguish between (potentially) different populations (groups).

When we compare one unique dependent variable, *Caldera Area*, against one classification variable which has two or more categories we call the design one-way ANOVA. If each classification group has unequal number of calderas, we call the experiment unbalanced, as opposed to a balanced experiment where the number of calderas are equal for all groups. Additionally, when the observations (number of calderas) in the response variable (*Caldera Area*) are assumed to be independent from each other, but we do not have enough evidence to assume a particular distributional form, such as the normal (due to insufficient data), we then need to use nonparametric procedures to perform the ANOVA analysis, in our case,

the Kruskal-Wallis test [Rice, 1995]. Due to the small number of calderas in some of the groups we don't have enough evidence to assume a normal distribution of the data, that is why we have to use nonparametric procedures. With the Kruskal-Wallis test, the observations are pooled together and ranked. The data are replaced by their ranks. This replacement has the effect of moderating the influence of outliers.

The observations are pooled together and ranked. Let R_{ij} be the rank of Y_{ij} in the combined sample. Let

$$\bar{R}_{i.} = \frac{1}{J_i} \sum_{j=1}^{J_i} R_{ij} \tag{4.1}$$

be the average rank in the ith group. Let

$$\bar{R}_{..} = \frac{1}{N} \sum_{i=1}^{I} \sum_{j=1}^{J_i} R_{ij} = \frac{N+1}{2} \tag{4.2}$$

where N is the total number of calderas. Let

$$SS_B = \sum_{i=1}^{I} J_i (\bar{R}_{i.} - \bar{R}_{..})^2 \tag{4.3}$$

be a measure of the dispersion of the $\bar{R}_{i.}$. Under the null hypothesis that the probability distributions of the I groups are identical, the statistic

$$K = \frac{12}{N(N+1)} SS_B \tag{4.4}$$

is approximately distributed as a Chi-square random variable with $I-1$ degrees of freedom. This test statistic is then used to perform a hypothesis testing.

A statistical hypothesis testing is a method of making statistical decisions using experimental data. In frequency probability, these decisions are almost always made using null-hypothesis tests; that is, ones that answer the question "Assuming that the null hypothesis is true, what is the probability (p-value) of observing a value for the test statistic that is at least as extreme as the value that was actually observed?"

In practice, a statistical package is used to compute the K statistic and the p-value. Most statistical softwares can perform this type of analysis.

In statistical hypothesis testing, the p-value is the probability of obtaining a result at least as extreme as the one that was actually observed, assuming that the null hypothesis is true. It is computed using the K statistic defined above.

In statistics, a result is said to be "statistically significant" if is unlikely to have occurred by chance. Hence, after a result has been tested empirically using hypothesis testing, and proven to be statistically significant, we have statistical evidence to reject the null hypothesis that the differences observed are due to random variability alone in favor of the alternative that the differences are due to the specific characteristics of each group. The amount of evidence required to accept that an event is unlikely to have arisen by chance is known as the significance level or critical p-value. In statistics, popular levels of significance are 5% (0.05), 1% (0.01) and 0.1% (0.001), the lower the p-value falls below the significance level, the greater the statistical evidence. See Rice [1995] for further explanation on p-values. So at the 5% level of significance we can say that there is enough statistical evidence to reject the null hypothesis that the groups are identical, in favor of the alternative hypothesis that the groups are different.

In our case the null hypothesis is that the caldera sizes are the same for each group, and the alternative hypothesis is the opposite, i.e., the caldera sizes are different for each group. The aim of the test is to find enough statistical evidence (using the Kruskal-Wallis test) to reject the null hypothesis in favor of the alternative. To do this, we compute and look at the p-value. If the p-value is less than or equal to 0.05 or 0.10 (the lower the better), we can then say that there is enough statistical evidence to reject the null hypothesis that the calderas have the same size for all groups, and accept the alternative hypothesis that the caldera sizes vary depending on what group they belong to (each group will be determined by the different categories in the classification variables explained earlier).

We use the statistical package SAS 9.1.3. Copyright (c) 2002-2003 by SAS Institute Inc., Cary, NC, USA. SAS (r) 9.1 (TS1M3) Licensed to DEPT. D ESTADISTICA I INVESTIGACIO OPERATIVA, Site 0081102002.

4.5 Statistical Analysis and Results

We start performing the analysis with one variable at the time, grouping the response variable *Caldera Area* according to the different categories of the explanatory variable. We can start with either *Crustal Type*, *Plate Tectonic Setting* or *Rock Suite* as the explanatory grouping variable. The final result is the same regardless of which grouping variable we start with. Of the 402 calderas with the area informed, a total of 297 calderas have data in variables *Crustal Type*, *Rock Suite* and *Plate Tectonic Setting* simultaneously. We use these 297 calderas for the study.

We are performing hypothesis testing on Caldera Area only and the explanatory variables are used independently to make the groups. They contribute with different and relevant information to explain the geodynamic environment where calderas form. Due to the nature of the methodology used, a correlation analysis does not apply as it would in regression analysis. The areas are assumed to be independent and hence, based on the area, the groups tested are assumed to be independent.

4.5.1 *Crustal Type*, *Rock Suite*, *Plate Tectonic Setting*

Table 4.5 shows an example of the Kruskal-Wallis test output explained in previous section, using SAS statistical software. Any statistical software can be used to easily compute the p-value, if not computed manually. We see that grouping the response variable according to the six different categories of *Crustal Type*, we obtain a p-value lower than 0.0001 (less than 5%), which means that there is enough statistical evidence at the 5% level of significance to say that the *Caldera Area* is not the same for all six different categories of *Crustal Type*, suggesting they have different caldera sizes. The mean score for oceanic basaltic thin and thick and transitional thin suggests these categories may have similar areas. When tested together, the p-value is greater than 5%, indicating the area is not significantly different across these three categories. We make a unique group of 40 calderas and call it oceanic basaltic (thin, thick), transitional thin (Table 4.6). We want to see which crustal type categories are significantly different from each other. For this, we need to do a pair-wise Kruskal-Wallis test on all different pairs. Table 4.6

Table 4.5: *Kruskal-Wallis test output using SAS Statistical software. See text for details.*

Crustal Type	Calderas	Sum of Scores	Expected Under H0	Std Dev Under H0	Mean Score
Continental silicic thick crust	104	19431	15496	706	187
Oceanic basaltic thin crust	87	13206	12963	674	152
Transitional thick crust	66	8621	9834	615	131
Continental silicic thin crust	24	1336	3576	403	56
Oceanic basaltic thick crust	13	1380	1937	303	106
Transitional thin crust	3	281	447	148	94

Kruskal-Wallis Test	
Chi-Square	56
DF	5
Pr > Chi-Square	<.0001

Table 4.6: *Pair-wise Kruskal-Wallis result at the 5% level of significance for Caldera Area classified according to Crustal Type (significant difference* < 0.05*, not significant* ≥ 0.05*)*

Crustal Type	C	Cd	T	Od, O, Td
Continental silicic thick (C)		0.0045	< .0001	< .0001
Continental silicic thin (Cd)	0.0045		0.1115	< .0001
Transitional thick (T)	< .0001	0.1115		< .0001
Oceanic (thin (Od), thick (Od)) Transitional thin (Td)	< .0001	< .0001	< .0001	

shows these combinations and the p-value of the test for each pair. The p-value is lower than 0.05 for most combinations. Calderas in continental silicic thick have an area significantly different than those in continental silicic thin, transitional thick and oceanic (thin, thick), transitional thin. Same applies to the other pairs, except when calderas in transitional thick and continental silicic thin are tested against each other, suggesting this two categories could be grouped. Using Kruskal Wallis test we have identified three groups of calderas with different area according to the crustal type environment where they develop (Table 4.7). There are 104 calderas in continental silicic thick crust with a mean area of 411 km^2, 153 calderas in continental silicic thin and transitional thick crust with a mean area of 161 km^2, and 40 calderas in oceanic basaltic (thin, thick), transitional thin crust with a mean area of 39 km^2.

Performing the same pair-wise analysis for *Plate Tectonic Setting* and *Rock Suite* we have identified three groups of plate tectonic setting and two groups of rock suite (Table 4.7). Calderas in Basin and Range, and hotspot form one group of 65 calderas with a mean area of 528 km^2. Calderas in continental rift, continental back-arc rifting, Chilean-type subduction, island arc collision, transform boundary and triple plate junction form a second group of 198 calderas with a mean area of 168 km^2. Calderas in Mariana-type subduction and hotspots near or over an ocean ridge form the third group of 34 calderas with a mean area of 41 km^2. According to *Rock Suite* there are two groups with clearly different caldera area, 211 calderas with calc-alkaline (felsic, intermediate, intermediate-felsic), calc-alkaline alkaline and peralkaline-calc-alkaline composition with a mean area of 304 km^2 (going forward we will refer to this group as *composition A*), and 86 calderas with other rock suite (alkaline, tholeiitic, other) which have a mean area of 55 km^2 (going forward we will refer to this group as *composition B*).

The Kruskal-Wallis test shows enough statistical evidence to say that calderas in different geodynamic environments yield groups with significantly different *Caldera Area* according to the three classification variables (*Crustal Type*, *Plate Tectonic Setting* and *Rock Suite*). This is an indication of some sort of pattern in the geodynamics of the volcanic setting.

To further explore this pattern, we now want to look at the interaction of these grouped variables all together and look for possible group combinations

Table 4.7: *Crustal Type, Plate Tectonic Setting and Rock Suite variables have been regrouped according to the size of the calderas in each category, based on Kruskal-Wallis Test (p-value* < 0.05*)*

	Calderas	Mean (km^2)
Crustal Type Groups		
1 - Continental silicic thick	104	411
2 - Continental silicic thin, Transitional thick	153	161
3 - Oceanic basaltic (thin, thick), Transitional thin	40	39
Plate Tectonic Setting Groups		
1 - Basin and Range, Hotspot	65	528
2 - Continental back-arc rifting, Continental rift, Island arc collision, triple plate junction, transform boundary, Chilean-type subduction	198	168
3 - Mariana-type subduction, Hotspot near/over oceanic ridge	34	41
Rock Suite Groups		
1 - Calc-alkaline (felsic, intermediate, intermediate-felsic), Calc-alkaline alkaline, Peralkaline-calc-alkaline	211	304
2 - Alkaline, Tholeiite, Calc-alkaline (mafic, mafic-felsic, mafic-intermediate), Peralkaline-alkaline	86	55

with significantly different areas. To do this, we cross the three explanatory variables and their new categories as shown in Table 4.7. We then outline all the different combinations as we have done earlier for each variable alone, and do a pair-wise Kruskal-Wallis test on them. Table 4.8 shows the result of this process, which suggests that we can classify calderas according to the geological conditions or geodynamic environment where they form. In this sense, we can distinguish three main groups or caldera settings (*GE1*, *GE2*, and *GE3*), with a relationship existing between these groups and the area of the associated calderas.

In particular, when *composition A* group is combined with continental silicic thick crust (whether in Basin and Range, hotspot or Chilean-type subduction) or combined with continental silicic thin crust (whether in Basin and Range, hotspot or continental rift), the result is a group of 110 calderas with significantly large area (460 km^2 on average). *Composition A* group in continental silicic thin

Table 4.8: *Geodynamic environments (GE) according to Rock Suite, Crustal Type and Plate Tectonic Setting. (Composition A: calc-alkaline (felsic, intermediate, intermediate-felsic), calc-alkaline alkaline, peralkaline-calc-alkaline. Composition B: alkaline, tholeiitic, calc-alkaline (mafic, mafic-felsic, mafic-intermediate), peralkaline-alkaline).*

GE	Rock Suite	Crustal Type	Plate Tectonic Setting	Calderas	Mean (km^2)
GE 1	Composition A	Continental silicic thick	Basin and Range, Hotspot, Chilean-type subduction	110	460
		Continental silicic thin	Basin and Range, Hotspot, Continental rift		
GE 2	Composition A	Continental silicic thin	Continental back-arc rifting, Chilean-type subduction	92	142
		Transitional thick	Island arc collision, Transform boundary, Triple plate junction, Chilean-type subduction		
GE 3	Composition A	Transitional (thin, thick) Oceanic basaltic (thin, thick),	Mariana-type subduction	95	54
	Composition B	any	any		

crust (either continental back-arc rifting or Chilean-type subduction), as well as transitional thick (whether in Chilean-type subduction, island arc collision, transform boundary or triple junction) create a second geodynamic environment of 92 medium size calderas (142 km^2 on average). There is a third geodynamic environment with relatively small calderas (95 with 54 km^2 on average) formed by those with *composition A* in either transitional (thin, thick) or oceanic basaltic (thin, thick) crust located in Mariana-type subduction, and those calderas of *composition B* type regardless of the crustal type and plate tectonic setting where they develop.

Hence using the Kruskal-Wallis analysis of variance we have identified three main groups of caldera settings or geodynamic environments with statistically different areas of the associated calderas. Table 4.9 shows some examples in each group and Figure 4.2 summarizes step-by-step the segmentation process used to identify the three groups.

To describe *GE1*, *GE2* and *GE3* in more detail we have included a boxplot of *Caldera Area* according to the geodynamic environment group (Figure 4.3). A boxplot is a graphical display that shows in a glimpse, several very important elements describing the data: median (50th percentile or second quartile) *Q2*), interquartile range *IQR*, lower quartile (25th percentile) *Q1*, higher quartile (75th percentile) *Q3*, and smallest and largest observation. In addition, a boxplot is very useful to identify abnormal data (outliers). Horizontal lines are drawn at the median and at the upper and lower quartiles and are joined by vertical lines to produce the box. Then a vertical line is drawn up from the upper quartile to the most extreme data point that is within a distance of 1.5 (IQR) of the upper quartile. A similarly defined vertical line is drawn down from the lower quartile. Short horizontal lines are added to mark the ends of these vertical lines. Each data point beyond the ends of the vertical lines is marked with a circle, and they are considered abnormal or unusual data (outliers) for this particular distribution.

Group *GE1* has 110 calderas with a mean of 461 km^2, a standard deviation of 784 km^2, median of 201 km^2, *Q1* 79 km^2 and *Q3* 510 km^2. According to the median, 50% of the calderas in *GE1* group have an area above 201 km^2. According to *Q3*, 75% have an area below 550 km^2. The fact that the mean

Table 4.9: *Some examples of calderas according to the GE group where they develop. CALC: Calc-alkaline; PERALK: Peralkaline; int: intermediate. (Not all calderas are included, only a few examples in each group)*

GE	**Caldera**	**Area**	**Rock Suite**	**Crustal Type**	**Plate Tectonic Setting**
1	Blacktail	4712	CALC felsic	C	Hotspot
1	Kilgore	3770	CALC felsic	C	Hotspot
1	Yellowstone II	2435	CALC felsic	C	Hotspot
1	Cerro Galan	550	CALC felsic	C	Chilean-type subduction
1	Platoro	358	CALC int	C	Basin and Range
1	Vilama II	550	CALC int-felsic	C	Chilean-type subduction
1	Mc Dermitt	707	PERALK-calc-alkaline	Cd	Hotspot
1	Santana	628	CALC alkaline	C	Basin and Range
2	Ranau	251	CALC felsic	Cd	Chilean-type subduction
2	Kutcharo	430	CALC int-felsic	C	Chilean-type subduction
2	Rotorua	363	CALC felsic	Cd	Continental back-arc rifting
2	Bulusan	95	CALC felsic	T	Island arc collision
3	Las Cañadas	22	Alkaline felsic	Od	Hotspot
3	Sierra Quemada	28	Alkaline felsic	C	Chilean-type subduction
3	Nabro	50	Alkaline felsic	Cd	Continental rift
3	Tofua	20	CALC felsic	O	Mariana-type subduction
3	Witori	32	CALC felsic	T	Mariana-type subduction
3	Longonot	3	PERALK-alkaline	Cd	Continental rift
3	Tousside	28	PERALK-alkaline	C	Hotspot
3	Cinco Picos	38	PERALK-alkaline	Cd	Hotspot near/over oceanic ridge
3	Cinque Denti	28	PERALK-alkaline	Cd	Continental rift
3	Cerro Azul	10	Tholeiite	Od	Hotspot near/over oceanic ridge

is significantly higher than the median indicates the existance of calderas with unusually large size shifting the distribution towards higher values. As we see in Figure 4.3, there is a group of 8 calderas with area ranging from 1402 to 4712 km^2. These *Caldera Area* values are significantly larger than the values of the other 102 calderas in this group. The inclusion of these 8 calderas in group *GE1* explains the large mean with respect to *Q2*. Also, the disperssion of the values around the mean, this is, the standard deviation, is high, 784 km^2, suggesting the observation values are not concentrated around the mean but rather disperse towards large values. This measures of location (*Q1*, *Q2*, *Q3*, mean) and dispersion (standard deviation) suggest the existance of outliers which deserve further attention. These outliers could be due to measurement errors or to observations with exceptional geological characteristics apart from the type of rock composition, the crustal type or the plate tectonic settings defined in these groups that make them larger than usual. The boxplot technique in Figure 4.3 has identified 8 calderas as outliers in group *GE1*, 5 calderas in group *GE2* and 9 calderas in group *GE3*. Similarly for group *GE2*, a mean of 142 km^2 versus a median of 79 km^2 and a standard deviation of 228 km^2 is indicating the distribution of the *Caldera Area* is slightly skewed towards high values, suggesting the existence of some calderas with unusually large values for this group. Similar effect is observed in group *GE3*, with a mean of 54 km^2 versus a median of 28 km^2 and a standard deviation of 66 km^2, suggesting once more the existance of large caldera sizes for this group.

Clearly, calderas in *GE1* are the largest in size, group *GE2* hosts medium size calderas and *GE3* calderas of small size with respect to the other groups. Also, we can see from the shape of the boxplots that the central part of the distribution is somewhat skewed towards high values. Further attention should be given to the outliers, especially those with an unusually large size which should respond to particular structural constraints other than the geodynamic environment where they form.

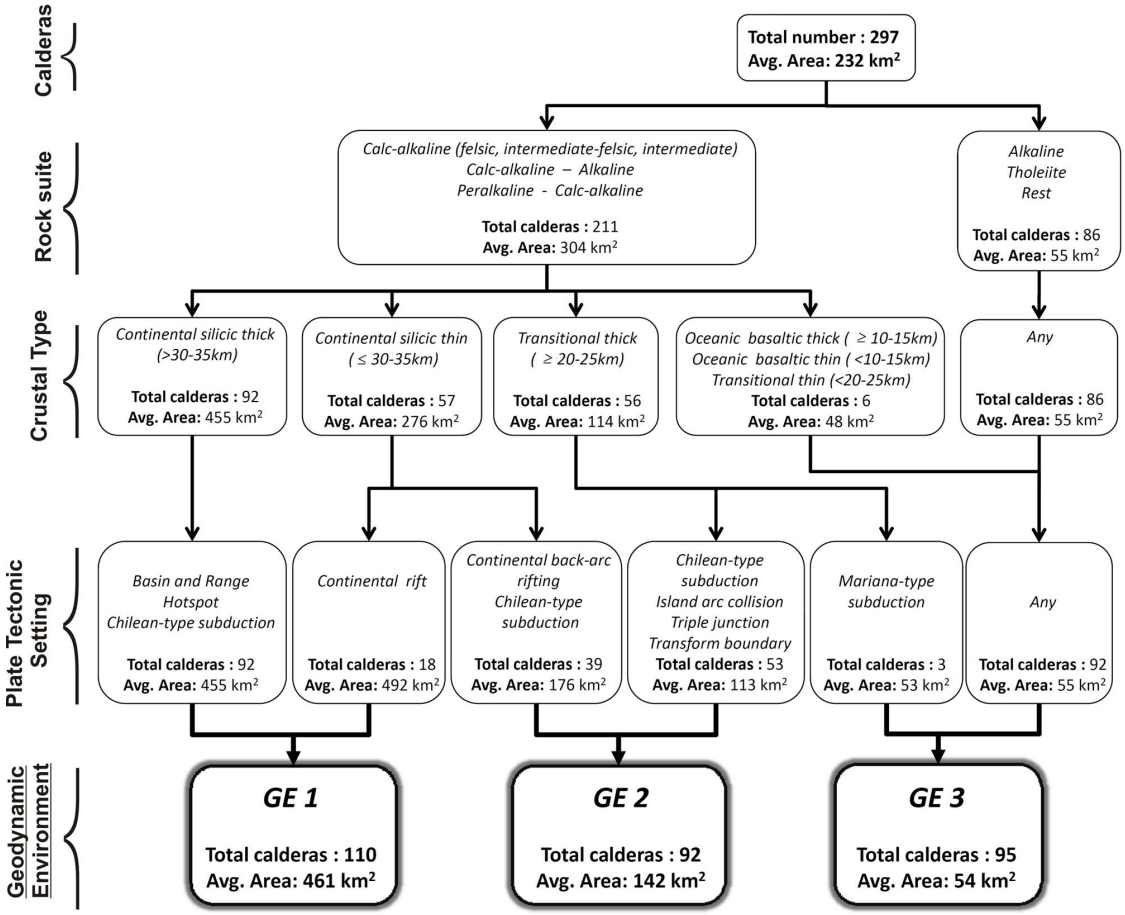

Figure 4.2: Stepwise segmentation of variables by *Caldera Area* using Kruskal-Wallis test. Three groups of calderas with average areas significantly different can be distinguished according to the characteristics of the geodynamic environment where they form. We distinguish three main geodynamic environments, named *GE1*, *GE2*, and *GE3*, which preferentially host large, medium, and small calderas, respectively.

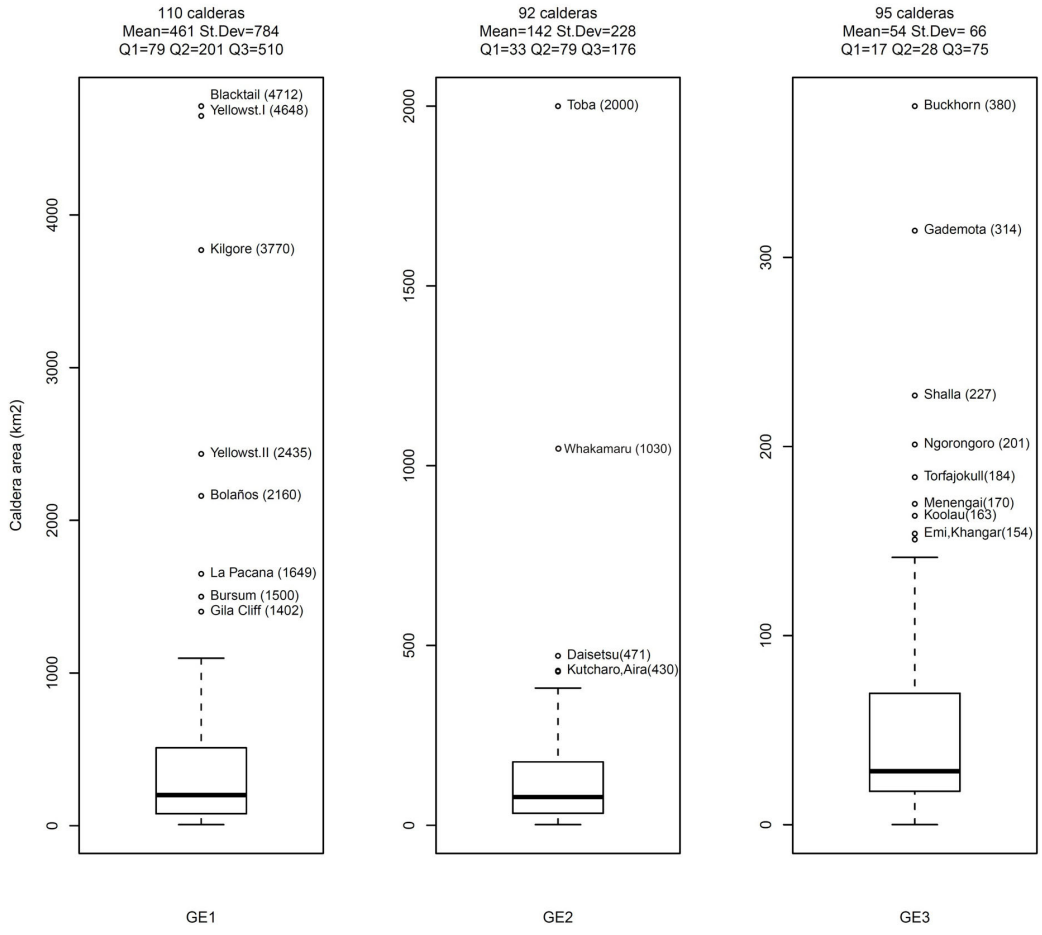

Figure 4.3: Boxplots of geodynamic environment groups for *Caldera Area* (Note the different scales in the y axes. See text for explanation.)

4.5.2 Distribution of the Caldera Groups according to World Region

For information purposes, lets look at the location of the calderas in each geodynamic environment. The map on Figure 4.4 shows the world distribution of calderas in each group, and Table 4.10 gives more detail of the regions in each group. In particular, 65 out of 110 calderas from group *GE1* are located in North America, and 30 out of 92 calderas in group *GE2* are located in the Japan and Mariana Islands.

Table 4.10: *World Region distribution of calderas according to their geodynamic environment*

GE 1		GE 2		GE 3	
South America	25	Japan and Mariana		Africa	29
North America	65	Islands	30	Atlantic Islands	9
Mediterranean	2	Alaska	12	Melanesia	8
Kamtchatka, Asia	6	Indonesia	10	Iceland	8
Central America	12	Central America	10	South America	7
		Kurile Islands	8	North America	6
		Kamtchatka, Asia	8	Hawaii	6
		New Zealand, Tonga		Alaska	4
		and Kermadec Islands	5	Central America	4
		Mediterranean	2	Mediterranean	3
		Melanesia	2	Japan and Mariana	
		Philippines	2	Islands	3
		North America	2	Kamtchatka, Asia	3
		Antartica	1	Indean Ocean	2
				New Zealand, Tonga	
				and Kermadec Islands	2
				Indonesia	1
Total	110		92		95

As mentioned earlier, this is only for information purposes as the geographical location is not by itself a variable that may be combined with the others we have considered in our statistical analysis. However it is interesting to see where the different types of calderas are preferentially located, as this, in addition to its possible relation with a specific plate tectonic setting, could also be related to

the different degree of knowledge of the geology of each region, which is something totally unrelated to the geodynamic environment where calderas form. As an example, we can see that the number of calderas in North America is considerably higher than in the other studied regions (see Fig. 4.4). It is reasonable to think that this high number of calderas is due to the fact that this area has been more intensively studied and thus more calderas have been detected and/or more information is available. This unusual high number of calderas could also be due to better preservation conditions and observation possibilities of the area or because this area is especially prone to the occurrence of caldera-forming eruptions. Hence, it is worth analyzing the tectonic and magmatic evolution of North America in this context during future work.

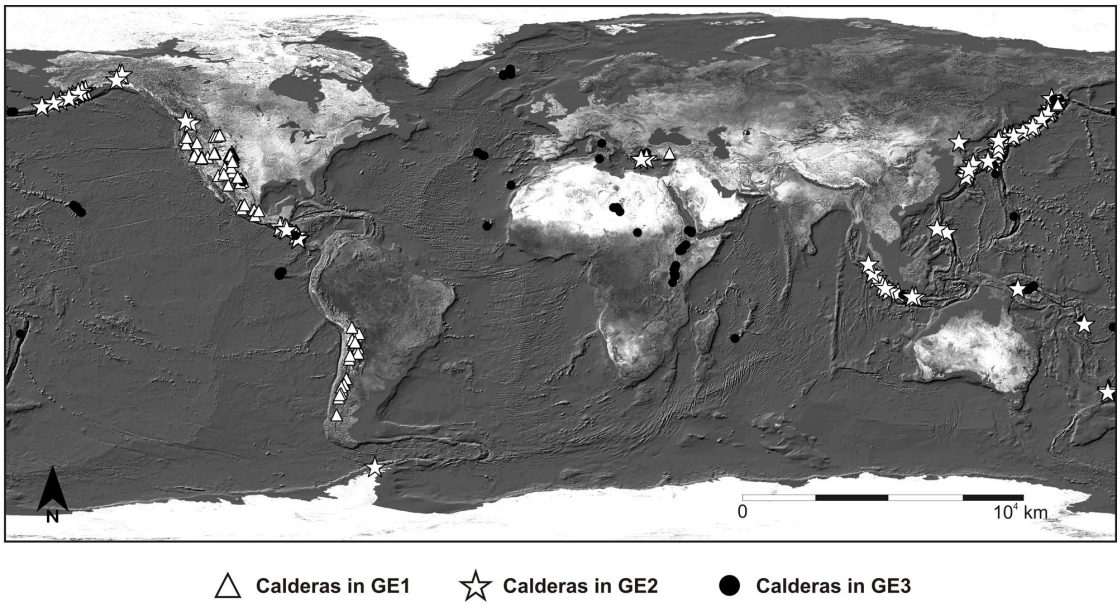

Figure 4.4: Geographical location of each group.

4.6 Discussion and conclusions

One of the main restrictions of the statistical analysis presented in this paper is the CCDB itself. Despite being a comprehensive catalog of most known calderas, it is still incomplete. Some of the examples included are poorly constrained, in-

formation on some classes (i.e.: tectonic setting, rock suite, type of crust, etc.) is lacking, and probably a large number of representative examples are still not included. The construction of the CCDB is an ongoing process that requires permanent updates with the inclusion of the new relevant information published in the scientific literature. For this reason, our statistical analysis has only considered a sample of 297 calderas (a 63% of the total), for which the information required to undertake this study was available. Therefore, it would be logical to realize that completion of the missing data and/or inclusion of new examples in future updates could imply changes in the results of the present study. However, the statistical methodology used here is appropriate for the data considered and provides reliable and accurate results about the different caldera types and related geodynamic environments. Therefore, future modifications of the CCDB should reinforce the results obtained here and allow to explore further extension and segmentation of the current groups.

Field studies and analogue and numerical modeling suggest that the original area of a collapse caldera is directly related to the size and depth of the associated magma chamber (see Martí et al. [2008c] and references herein), although further increases of the caldera size, not directly associated with the collapse process itself, may occur due to erosion, landsliding or later tectonics [Acocella and Rossetti, 2002; Acocella et al., 2004; Geshi, 2009; Holohan et al., 2005; Lipman, 1997; Spinks et al., 2005]. The results obtained from our statistical analysis show that caldera size is strongly dependent on the geodynamic environment where caldera develops, i.e. type and nature of crust, magma composition and plate tectonic setting. This dependency of caldera area on the characteristics of the associated magma chamber and on the geodynamic environment where it develops supports the idea of Jellinek and DePaolo [2003] that the caldera size is related to magma supply. This is governed by the mantle thermal regime, the crust composition and thickness, and the tectonic regime (extensional or compressional), three of the main aspects that define a geodynamic environment.

From the data analysed, we are able to group collapse calderas in three different geodynamic environments (Table 4.8): *GE1*, *GE2* and *GE3*. Despite the fact that collapse calderas may develop in any geodynamic environment and show a wide range of sizes, our statistical analysis reveals that calderas formed in *GE1*

have on average an area of 461 km^2, in *GE2*, around 142 km^2 and in *GE3* 54 km^2 (Table 4.8). If we arbitrarily classify the studied calderas according to their size in *large* ($\geq$ 300 km^2), *medium* (300 - 100 km^2) and *small* (<100 km^2), we can see that *GE1* calderas are characteristically the largest ones. *GE2* preferentially hosts *medium* to *small* calderas, whereas in *GE3* we mostly find *small* calderas (Fig. 4.5). It is evident from Figure 4.5 that the size of calderas decreases from *GE1* to *GE3*. However, as seen in the boxplot analysis (Figure 4.3), there are several outliers in each group. The reasons for such remarkable exceptions are not studied in this work but deserve further investigation as they should respond to particular structural and petrogenetical constraints in addition to the geodynamic environment where they develop.

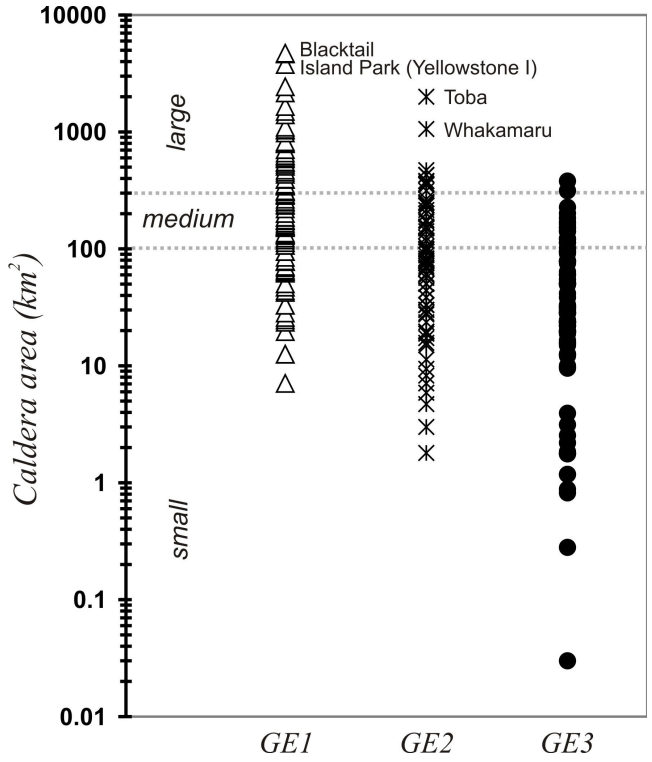

Figure 4.5: Area of the calderas for each defined geodynamic environments: *GE1*, *GE2* and *GE3*. The additional classification of *large* ($\geq$ 300 km^2), *medium* (300 - 100 km^2) and *small* ($\leq$ 100 km^2) calderas is also indicated. Data from the CCDB [Geyer and Martí, 2008]. (Note: Refer to Figure 4.3 and corresponding text for bandwidth and outliers explanation in each group.)

The tendency of calderas of a particular size to form depending on a preferential geodynamic environment, therefore suggests that there is a link between the physicochemical conditions and time scales to generate magma and accumulate it at shallow depths. In other words, the amount of magma able to be erupted during a caldera-forming eruption depends on the geodynamic environment. This statement is relevant when discussing the formation of large calderas. It is generally accepted that there is a positive linear relationship between the area of the caldera and the volume of material extruded during the eruption [Smith, 1979]. So, large calderas, which are always associated with calc-alkaline silicic magmas (CCBD in Geyer and Martí [2008]), are related to the eruption of large volumes of magma. Thus, they require large magma chambers, i.e. the geodynamic environments where they occur must allow the production and accumulation of great amounts of silicic magmas. The ultimate source of magma is the upper mantle. Whether the silicic magma comes from crystal fractionation of mantle derived magmas, or directly from melting of continental crust induced by thermal inputs of basaltic magma [Huppert and Sparks, 1988], or from partial melting of metasediment-MORB melanges in subduction zones [Castro et al., 2010], accumulation and storage of large amounts of silicic magma in the crust will require a long-term and a relative high mantle magma supply [Jellinek and DePaolo, 2003]. Therefore, *GE1* must preferentially enhance these conditions in comparison with *GE2* and *GE3*. Although the purpose of this paper is not to take further this discussion on the link between caldera area and the key geodynamics it is important to remark that the relationship between the driving force in the caldera system (i.e. magma) and its response (i.e. caldera area) is ultimately depending on the magma supply. Magma supply, accumulation and storage in each of the three geodynamic environments we have identified to group collapse calderas deserve a detailed investigation in order to get a more comprehensive picture on how calderas form.

Numerical and analogue models have demonstrated that the formation of large calderas is restricted to volcanic systems with a magmatic reservoir located at shallow depths [Folch and Martí, 2004; Geyer et al., 2006; Gudmundsson, 1987, 2007; Martí et al., 1994a; Roche and Druitt, 2001]. Thus, in addition to the previous considerations on the relationship between geodynamic environment and

caldera size, another aspect we should take into account to understand the formation of large collapse calderas is to figure out the approximate size of the magma chambers required to generate them. For this, we may assume that the volume of magma extruded during the eruption corresponds to all the eruptible magma inside the reservoir [Jellinek and DePaolo, 2003; Martí et al., 2000]. However, this volume does not necessarily correspond to the total volume of the magma chamber, i.e. it may be smaller. In fact, after the caldera-forming eruption there may be non-eruptable magma (i.e. non-compressible) remaining in the magma chamber. Non-eruptable magma may include mafic magma but also crystal rich near-to-solidus mush. Figure 4.6(a) shows the volume of extruded magma for all examples of calderas, which have this information in the CCDB [Geyer and Martí, 2008], in each geodynamic environment we have considered. We can see here that the most voluminous magma chambers (volumes $>$ 100 km^3) preferentially occur in *GE1*. We can also see how the volumes of extruded magma and hence of the total volumes of the corresponding magma chambers differ between a few orders of magnitude when passing from *GE1* to *GE3*. In the first case, the magma chamber volumes are of the order of hundreds to thousands of cubic kilometers. It is evident that to get such amount of magma it is necessary a very productive magma source and very efficient accumulation and storage mechanisms, probably acting over a long time period (e.g.: 10^5 to 10^6 years, for 1000 km^3 of silicic magma, Jellinek and DePaolo [2003]). In addition to extensive fractional crystallization of basaltic magma generation of large batholitic bodies has to be associated with melting and differentiation of continental crust [Annen and Sparks, 2002; Huppert and Sparks, 1988; Johnson et al., 1989; Lipman, 1984, 1988; Seager and McCurry, 1988], or generation of primary silicic magmas by sub-lithospheric melting of tectonic mélanges composed of a mixture of subducted oceanic crust and sediment [Castro et al., 2010], if we want to satisfactorily explain the volumes of silicic magmas involved in large caldera eruptions. This implies that the generation of batholitic-size magma chambers that should be related to large calderas requires a relatively thick continental crust in order to achieve the physical conditions required to accumulate such a large amount of silicic magmas. Additionally, the emplacement of silicic magma within the shallow continental crust marks the final stage in magma chamber

formation. However, the way in which magma accumulates in the shallow crust has been a point of discussion during the last decades and is still a matter of considerable debate [Burov et al., 2003; Castro et al., 2010; Cruden, 1998; Petford and Clemens, 2000; Petford et al., 1993, 2000; Petraske et al., 1978; Watanabe et al., 1999; Wickham, 1987].

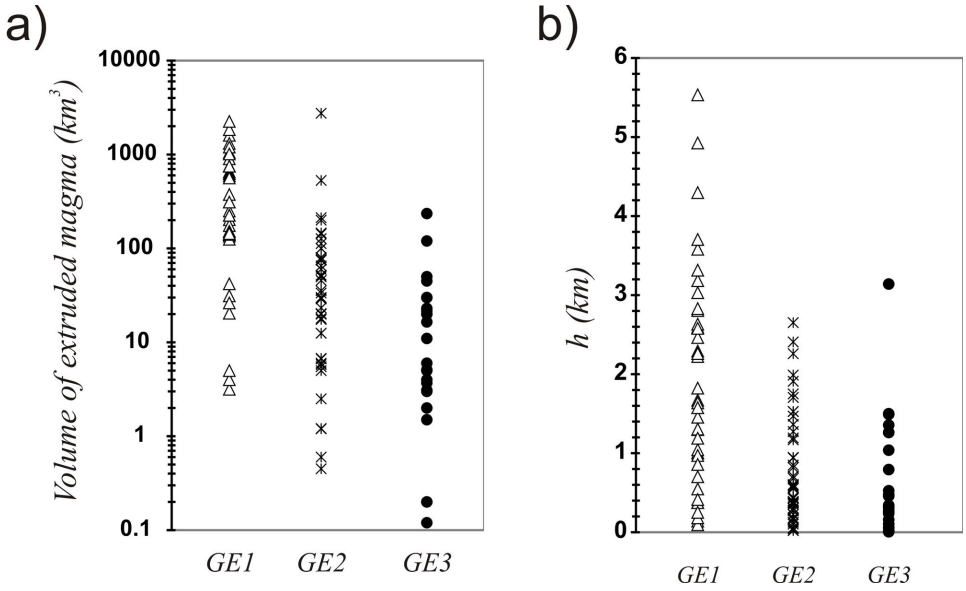

Figure 4.6: (a) Volume of extruded magma for calderas located in the different geodynamic environments: *GE1*, *GE2* and *GE3*. Data from the CCDB. (b) Values of magma chamber thickness (h) for *small*, *medium* and *large* calderas. Assuming that the magma chamber shape is proxy to a lentil, its minimum thickness h (vertical extension) can be calculated considering that $V_e = 4/3\pi(a/2)(b/2)(h/2)$, where a and b are the axes of the caldera (i.e. the horizontal axes of the magma chamber according to our approximation), and V_e the volume of extruded deposits. (Note: Refer to Figure 4.3 and corresponding text for bandwidth and outliers explanation in each group.)

Despite of the fact that the emplacement mechanisms of silicic magmas into the shallow crust is beyond the main purpose of this paper, we consider important to "visualize" at least the dimensions of the magma chambers we are dealing with. For this, we may calculate the thickness (vertical extension) of the magma chambers related to the caldera-forming eruptions analysed in the paper. Assuming that the area of the caldera is approximately equivalent to the projection at

surface of the magma chamber [Folch and Martí, 2004; Martí et al., 1994b] and that the magma chamber shape is proxy to a lentil, and knowing the extruded volume of magma, we can calculate a minimum value of h. Results are illustrated in Figure 4.6(b) for all calderas of all three geodynamic environments. We can observe how the thickness of the magma chamber decreases from *GE1* to *GE3*. The values illustrated in Figure 4.6(b) suggest a direct relationship between the thickness of the crusts and the maximum thickness of the magma chambers. In other words, *GE1* is more suitable to accommodate magma in thicker chambers, not necessarily larger in volume, than the other geodynamic environments.

In summary, the results obtained in this study confirm quantitatively the generally accepted idea that the characteristics of collapse calderas are influenced by the characteristics of the geodynamic environment where they develop, in particular the mantle thermal regime, the crust composition and thickness, and the tectonic regime. The conditions to form large calderas are closely related to those controlling the emplacement and evolution of silicic plutons in the shallow crust, so both processes should be studied together as part of the same geodynamic process. However, the environmental conditions that allow a volcanic systems to host a collapse caldera do not influence the dynamics of caldera subsidence, as it is indicated by the fact that similar caldera collapse mechanisms are recognized in all geodynamic environments.

Chapter 5

Volcanic hazard assessment for the Canary Islands (Spain) using Extreme Value Theory

5.1 Introduction

The Canary Islands are one of the major oceanic island groups of the world and have a long magmatic history, which began at the bottom of the ocean more than 40 million years ago [Araña and Ortiz, 1991]. The Canary Islands are an active volcanic region where all islands except for La Gomera show Holocene volcanic activity. Historical volcanism (last 600 years) has been reported on the islands of La Palma (1430, 1585, 1646, 1677, 1712, 1949, 1971), Tenerife (1492, 1704, 1706, 1798, 1909) and Lanzarote (1730-1736, 1824). In all cases, they have been eruptions of basaltic magmas characterized by emission of lava flows and construction of scoria cones.

The Canary Islands are a populated ultra-peripheral Spanish region and one of the most popular touristic destinations in Europe. The presence of recurrent historical volcanism in this region is a good reason to undertake volcanic hazard assessment in order to guarantee the safety of its inhabitants and of its numerous visitors. Volcanic hazard is the probability of any particular area being affected by a destructive volcanic event within a given period of time [UNESCO, 1972]. As

Table 5.1: *Original table of volcanic eruptions historically documented in the Canary Islands [Romero, 1991]*

Año*	Denominación	Lugar	Fecha (Inicio/ Final)	Duración (días)	Daños	Fuentes	Intervalos entre erupciones (años)	Observaciones
1430/1440	Tacande o Miña Quemada	Cumbre Vieja (La Palma)	?	?	Sin datos	Recogida en las obras de los primeros historiadores Canarios, como Abreu Galindo		Referencias orales recogidas por los primeros historiadores canarios
							52	
1492	Erupción de Colón	Sin localización documental precisa	?	?	Sin datos	Cristobal Colón		En actividad el 24 o 25 Agosto de 1492
							93	
1585	Tehuya	Cumbre Vieja (La Palma)	19 Mayo/10 Agosto	84	Destrucción de campos cultivados Incendio en el pinar	Fr. Alonso de Espinosa L. Torriani Nuñez de La Peña Ms Biblioteca Nacional		Extrusión forzada de varios monolitos fonolíticos
							61	
1646	Volcán Martín o de Tigalate	Cumbre Vieja (La Palma)	2 Oct/ 21 Dic	82	Destrucción de viviendas y estanques de agua Pérdida de ganado Daños en colmenas Daños en lugares de pasto Incendios en el pinar de Fuencaliente	Nuñez de La Peña Ms Biblioteca Nacional		Formación de varios conjuntos eruptivos tanto en cumbre, como en las cercanías de la costa Las cenizas de este volcán se recogieron también en Tenerife
							31	
1667/1678	Volcán de San Antonio	Cumbre Vieja (La Palma)	17 Nov/21 Ene	66	4 víctimas, al menos 1 de ellas por inhalación de gases tóxicos Destrucción de la Fuente Santa Destrucción de viviendas familiares Destrucción de tanques de agua Destrucción de pasto y pérdida de ganado caprino Pérdida de campos de cultivos	D. Juan Pinto de Guisla Archivo Parroquial de La Concepción La Laguna Nuñez de La Peña		Fases de contacto agua/magma
							26	
1704/1705	V. de Siete fuentes	Dorso Cañadas (Tenerife)	31 Dic/4 ó 5 Ene	5	16 víctimas por los terremotos asociados a la actividad volcánica Desplome de al menos unas 70 casas y daños graves en estructuras religiosas en Güimar y La Orotava Sin daños, por desarrollarse en el sector de cumbres	Ms Biblioteca Central Cataluña Nuñez de La Peña Prensa Local Canaria		Erupción triple Importante actividad sísmica asociada
	V. de Fasnia	Ladera Sur de Izaña (Tenerife)	5 Ene/ 16 Ene	12	Sin daños, por encauzamiento de coladas en un barranco			Columna eruptiva visible desde La Orotava
	V. de Arafo	Dorsal de Pedro Gil	2 Feb/27 Marz	54	Daños en tierras de cultivo			Desplomes y desprendimientos Apertura de grietas y fisuras Cambios en el caudal y curso de aguas corrientes Evacuación de la población de Arafo
							1	
1706	E. de Garachico V. de Arenas Negras	Dorsal de Abeque (Tenerife)	5 May/13 Jun	40	Incendio en el pinar Destrucción de viviendas familiares civiles y eclesiasticas Destrucción de campos de cultivo de subsistencia y de exportación Destrucción del Puerto de Garachico	Descripciones de Fray Juan García Fray Domingo José Cassares		Es la erupción que provocó mayores daños y pérdidas económicas en Tenerife. Garachico constituía en ese momento el puerto comercial más importante de Canarias, esencial en la ruta entre Europa y América Evacuación de Garachico
							6	
1712	E. del Charco	Cumbre Vieja (La Palma)	9 Oct/ 3 Dic	56	Destrucción de viviendas Destrucción de campos cultivados	Descripción de D. Juan Agustín de Sotomayor y Massieu		Es la erupción que cuenta con menor documentación
							18	
1730/1736	E. de Timanfaya	Sector centrooccidental (Lanzarote)	1 Set 1730 16 Abr 1736	2055	Destrucción de caseríos Destrucción de sistemas de captación de aguas Destrucción de campos de cultivos Destrucción de sectores de pasto Pérdida de ganado por inhalación de gases	Archivo de Simancas Lepoldo Von Buch (descripción del Cura de Yaiza) Actas Cabildo eclesiástico		Actividad volcánica discontinua temporalmente Apertura de unas 180 bocas eruptivas Construcción de unos 30 edificios Emigración de la población Es la primera erupción que cuenta con el nombramiento de una comisión para el seguimiento de una crisis volcánica en Canarias. Primer mapa de riesgo volcánico de Canarias elaborado durante la crisis Es la erupción de mayor envergadura en tiempos históricos en Canarias
							62	
1798	E. Narices del Teide V. de Chahorra	Laderas Pico Viejo (Teide, Tenerife)	9 Jun/14-15 Sep	99	Sin daños, por desarrollarse circunscrita en el ambito del Teide	Nicolás Segundo de Franqui (Anales de Historia Natural) Bernardo Cologán Fallow (Archivo Zárate Cólogan)		Fases de contacto agua/magma
							26	
1824	V. de Tao o del Clérigo Duarte V. Nuevo del Fuego o del Chinero V. Nuevo o de Tinguatón	Sector centrooccidental (Lanzarote)	31 Jul/31 Julio 29 Sep/5 Oct 10 Oct/24 Oct	86	Destrucción de tierras y de aljibes de agua Destrucción de tierras de "islotes" Sin daños, por ocupar terrenos afectados durante la erupción de 1730/36	Correspondencia personal de Pedro Barceló José Cabrera Carreño (Museo Canario) Diario de D. Ginés de Castro Alvarez, (Archivo Histórico Nacional y Prensa		Emisión tranquila de agua Emisión de agua en forma de geiser
							85	
1909	V. del Chinyero	Dorsal de Abeque (Tenerife)	19 Nov/27 Nov	10		Lucas Fernández Navarro Antonio de Ponte y Cólogan Prensa local, nacional e internacional		Primera erupción con documentación fotográfica Primera erupción con estudios de carácter científico
							40	
1949	Erupción de San Juan V. de Nambroque, Duraznero y Llano del Banco	Cumbre Vieja (La Palma)	24 Jun/30 Jul	47	Destrucción de viviendas por sismos Destrucción de tierras de labor	Prensa local, nacional e internacional Martel San Gil Benitez Padilla Bonelli Rubio		Fases de contacto agua/magma Formación de flujos piroclásticos Formación de pequeños lahares al finalizar la erupción Importante actividad sísmica
							22	
1971	V. del Teneguía	Cumbre Vieja (La Palma)	26 Oct/18 Nov	24	1 víctima por inhalación de gases	Prensa local, nacional e internacional Volumen especial de Estudios Geológicos		Formación de flujos a partir del desplome de coladas

* Solo se consideran las erupciones con referencias documentales constatadas y donde se describen procesos inequivocamente eruptivos. Existen datos documentales acerca de otras posibles erupciones, sin fechas concretas o sin localización espacial precisa que no han sido incluidas en esta tabla hasta no se confirmen con otras fuentes documentales

Tabla elaborada para el IGN por la Dra. Carmen Romero, Departamento de Geografía, Universidad de La Laguna. 2007

for any active volcanic region, volcanic hazard assessment at the Canary Islands requires knowing how volcanism has behaved in the past and determining the recurrence of volcanic eruptions. The first can be approached by detailed physical volcanology studies of past eruptions, in particular of those for which there exist historical chronicles (Table 5.1). The recurrence or eruption frequency needs to be based on historical records and precise dating of older events. Unfortunately, this is not an easy task as the reconstruction of the geological record of this volcanic region is far from accurate, lacking systematic dating of recent eruptions, and the historical records are imprecise and lack detail in some cases.

Despite these limitations we still can do an analysis of the volcanic hazard using the available historical data, covering the last 600 years since the Spaniards occupied the archipelago. A set of fifteen relatively well documented eruptions form the historical volcanism of the Canary Islands (Table 5.2). A few other eruptions have also been reported in historical times but their age and location is imprecise and do not constitute reliable information. Most of the historical eruptions in the Canary Islands have been short lived (from few weeks to few months) basaltic, strombolian to violent strombolian eruptions, which have generated scoria cones of different sizes and lava flows of different extent [Romero, 1991]. All the eruptions occurred in historical time, which goes from 1402 to present, have typically been separated a few tens of years but in some cases some have occurred in a very narrow period of time (e.g. Arafo, Fasnia, Siete Fuentes in Tenerife), or have lasted for some years (Timanfaya eruption in Lanzarote, 1730-1736).

Studies of volcanic time series have been done using stochastic principles to study eruption patterns on specific volcanoes or volcanic groups [De la Cruz-Reyna, 1991, 1993; Klein, 1982; Reyment, 1969; Wickman, 1976]. Bebbington and Lai [1996a] applied a Weibull renewal model to describe the patterns of New Zealand volcanoes. Other studies used transition probabilities of Markov chains [Aspinall et al., 2006; Bebbington, 2007; Carta et al., 1981], change-point detection techniques [Burt et al., 1994; Mulargia et al., 1987], Rank-order statistics [Pyle, 1998], Bayesian analysis of volcanic activity [Ho, 1990; Ho et al., 2006; Marzocchi et al., 2008; Newhall and Hoblitt, 2002; Sobradelo and Martí, 2010; Solow, 2001], non-homogeneous models [Bebbington and Lai, 1996b; Ho, 1991a], a mixture of Weibull distributions [Turner et al., 2007], geostatistical hazard-

estimation methods [Jaquet and Carniel, 2006; Jaquet et al., 2000], and a mixture of exponential distributions [Dzierma and Wehrmann, 2010a,b; Mendoza-Rosas and De la Cruz-Reyna, 2009, 2010]. Extreme-value methods have been applied to geological and historical eruption time series combined [Mendoza-Rosas and De la Cruz-Reyna, 2010, 2008] and historical series of large volcanic magnitudes [Coles and Sparks, 2006].

In this paper we use the historical volcanism to perform hazard assessment at the Canary Islands. Due to the limitations inherent to the available data, including its short sample time, and incomplete reporting of small and intermediate magnitudes, as well as uncertainties in the age, intensity and magnitude of the eruptions, we will use a method for the best estimate of the volcanic hazard based on a Non-Homogeneous Poisson process with a Generalized Pareto Distribution (GPD) as intensity function (NHGPPP) [Coles, 2001; Mendoza-Rosas and De la Cruz-Reyna, 2010, 2008]. This method has already been applied to other volcanoes for which little or incomplete data exists, like the Citlaltepetl volcano database with only six eruptions, or El Chichón volcano with 12 eruptions [Mendoza-Rosas and De la Cruz-Reyna, 2010, 2008]. This is the case with our time series of volcanic records for the Canary Islands. The methodology do not require stationarity or completeness for the full eruptive series, since it depends on the number of excesses of eruptions large enough to represent the behavior of the studied volcanoes.

First, we analyze the historical eruptive series to assess independence and homogeneity of the process. Second, we perform a Weibull analysis of the distribution of repose time between successive eruptions. Third, we analyze the non-homogeneous Poisson process with a generalized Pareto distribution as intensity function.

5.2 Geological setting

The Canary Islands are a roughly linear 500 km long chain grown on the passive margin of the African Plate within the eastern Central Atlantic Ocean (Fig. 5.1). The Canarian archipelago is the result of a long volcanic and tectonic activity that started at around 60 Ma ago [Le Bas et al., 1986; Marinoni and Pasquaré,

1994; Robertson and Stillman, 1979].

Several contrasting models have been proposed to explain the origin of the Canary Islands. These include a hotspot origin [Carracedo et al., 1998; Hoernle and Schmincke, 1993; Schmincke, 1982], a propagating fracture from the Atlas [Anguita and Hernán, 1975; Le Pichon and Fox, 1971], and mantle decompression melting associated with uplift of tectonic blocks [Araña and Ortiz, 1991]. However, each and every one of the latter hypotheses presents some inconsistencies with the local and regional geology. A unifying model has been proposed by Anguita and Hernán [2000] who considers the existence of a residual of a fossil plume under North Africa, the Canary Islands, and western and central Europe defined through seismic tomography [Hoernle et al., 1995]. Thus, volcanism is assumed to occur there where an efficient fracture system allows the magma to ascent [Anguita and Hernán, 2000], i.e. the central European rift system, the volcanic provinces of the westernmost Mediterranean (Balearic and Alborán basins), Iberia, the Canary Islands and Cape Verdes [Hoernle et al., 1995].

Although all islands except for La Gomera show Holocene volcanic activity, historical volcanism has been restricted to the La Palma, Lanzarote and Tenerife islands (see Fig. 5.1). In all cases historical eruptive activity has been related to basic magmas ranging in intensity from strombolian to violent strombolian, and has originated scoria cones and lavas. In most cases the historical eruptions have occurred on the active rift zones along eruptive fissures occasionally generating alignments of cones. The duration of the eruptions ranges from a few weeks to a few months, except in the case of the Timanfaya eruption in 1730 that lasted for six years. The total volumes of extruded magma range from 0.01 to >1.5 km3 (DRE), in the case of Timanfaya. The eruption sequences that may be deduced from the successions of deposits differ from one eruption to another and reveal that eruptions did not follow a common pattern. In all cases the resulting volcanic cones were constructed during single eruptive episodes (i.e.: they must be referred to as monogenetic) commonly including several distinctive phases that do not show significant temporal separations between them.

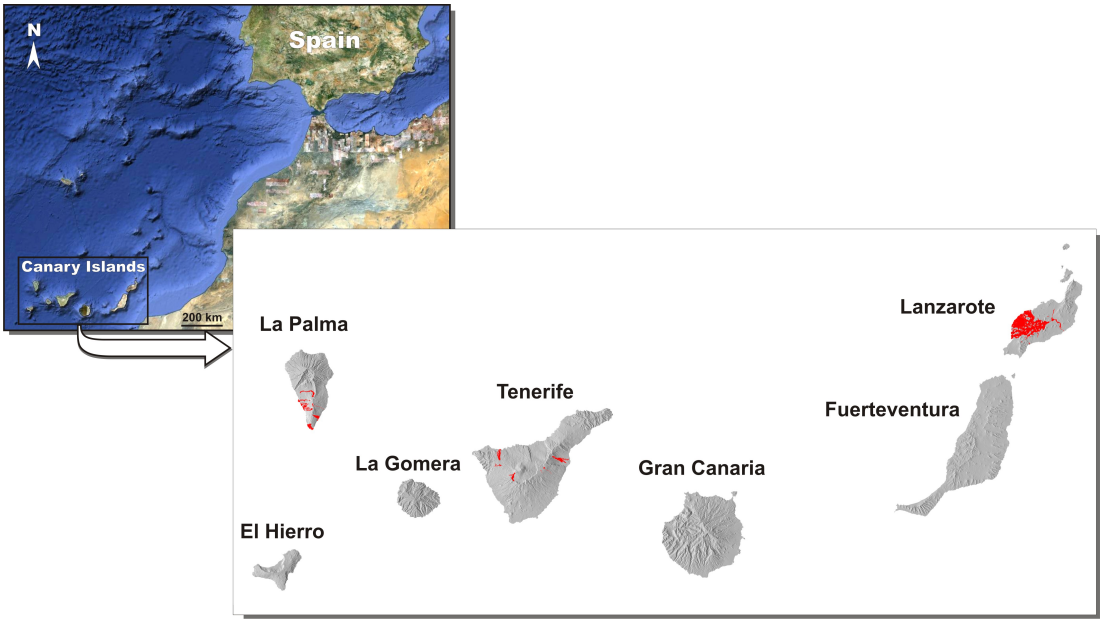

Figure 5.1: Distribution of historical volcanism in the Canary Islands.

5.3 Historical records of volcanic eruptions in the Canary Islands

Table 5.2 shows the data used for this study. It includes fifteen clearly different volcanic eruptions historically documented between 1430 and 1971 for which the eruption magnitude has been computed using existing information on lava and pyroclast volume. These data have been compiled from the original information on Table 5.1 and complemented with data on surface extent and volume of erupted products calculated from the geological maps at 1:25000 of IGME (Spanish Geological Survey, www.igme.es) and from a field revision of the historical eruptions that we have undertaken in this study.

In compiling the historical dataset of volcanic eruptions for the Canary Islands only those eruptions with well documented references and clearly described eruptive processes have been considered. There are documents that make reference to other possible eruptions, but the date and/or location is not clear. These eruptions have not been included until further documentation sources are

confirmed.

Table 5.2: *Volcanic Eruptions dataset used in the study for the volcanic hazard assessment of the Canary Islands (M=Magnitude)*

Event ID	Location	Volcano	Date (start/end)	Lava volume (Km^3)	Pyroclast volume (Km^3)	M
1	La Palma	Tacante	1430/1440 ?	0.038	0.026	4
2	Tenerife	Colon eruption	1492 ?	0.051	0.006	4
3	La Palma	Tehuya	1585 May19/Aug10	0.026	0.004	4
4	La Palma	Tigalate	1646 Oct2/Dec21	0.029	0.000	4
5	La Palma	Sant Antonio	1678 Nov17/Jan21 1677	0.037	0.006	4
6	Tenerife	Sietefuentes /Fasnia	1704 Dec31/Jan16 1705	0.006	0.002	3
7	Tenerife	Arafo	1704 Feb2/Mar27 1705	0.070	0.008	4
8	Tenerife	Arenas Negras	1706 May5/Jun13	0.045	0.014	4
9	La Palma	Charco	1712 Oct9/Dec3	0.034	0.021	4
10	Lanzarote	Timanfaya	1730 Sep1/Apr16 1736	3.804	0.000	6
11	Tenerife	Chahorra	1798 Jun9/Sep15	0.028	0.009	4
12	Lanzarote	Tao/ Nuevo del Fuego/ Tinguatón	1824 Jul31/Oct24	0.001	0.000	2
13	Tenerife	Chinyero	1909 Nov19/Nov27	0.010	0.005	4
14	La Palma	Nambroque, Duraznero and Llano del Banco	1949 Jun24/Jul30	0.029	0.050	4
15	La Palma	Teneguía	1971 Oct26/Nov18	0.018	0.005	4

The original dataset includes fifteen volcanic eruptions historically documented between 1430 and 1971 in three different islands (Lanzarote, Tenerife, La Palma). We have also considered the eruption of Montaña Cangrejo in Tenerife (*Erupción de Colón* in Table 5.1), that is supposed to have been observed by Columbus in his way to America and that has now been confirmed by Carracedo et al. [2007, 2010]. We have considered as one unique event the eruptions of Siete Fuentes (from 31/12/1704 to 05/01/1705) and Fasnia (from 05/01/1705 to 16/01/1705) in Tenerife. The Arafo eruption (from 02/02/1705 to 27/03/1705) happens later in time and the materials have a slightly different composition than Fasnia and Siete Fuentes, suggesting that this could be a different eruption. The eruptions of Tao, Nuevo del Fuego and Tiguantón in Lanzarote are considered as one unique event. They have been listed in Romero [1991] as different episodes but they are clearly related in terms of timing, petrology and location of vents on the same

eruptive fissure.

In order to classify the eruptions and apply the NHGPPP, we have calculated for each case the total volume of extruded magma (DRE) based on the volume of exposed materials (lavas and pyroclastic deposits), so our total volumes are minimum estimates (Table 5.2). Although rapid erosion of tephra and uncertain lava flow thicknesses may cause problems in making accurate calculations, order of magnitude determinations still provide a useful comparison between eruptions. We have made the volume estimates calculating separately the volumes of tephra and lavas for each eruption. Tephra and lava volumes have been calculated with the help of a DEM at a resolution of 5 m, the digital geological maps of IGME at 1:25000, and checking extension and thicknesses variations of the deposits and lavas in the field.

Volcanic eruptions are natural phenomena where the frequency of the events decreases as their size or magnitude increases. The fact that the magnitude distribution is irregular is not necessarily an indication of incompleteness in the catalog. When we are dealing with a historical catalog it is very difficult that a high magnitude eruption goes unnoticed. Compared with other volcanoes, a catalog of 15 eruptions in 600 years seems consistent. There are no records of any more high magnitude eruptions in historical times, other than the high magnitude eruption of Timanfaya. For this reason, we assume that the catalog for high magnitude eruptions in historical times is complete. On the other hand, the historical records for the oldest low magnitude eruptions are not as clear and accurate as for the most recent low magnitude eruptions.

To deal with the difficulties derived from the possible lack of catalog completeness for the Canary Islands, the hazard estimates for the full series are computed using the magnitude of the volcanic eruption. This is based on the logarithmic scale for magnitude [Pyle, 2000]:

$$M = \log_{10}[\text{mass(tephra + lava)erupted(kg)}] - 7$$

We have calculated the total magnitude of each eruption (in Kg) assuming a density for the basaltic magma of 2850 kg/m3 (Table 5.2).

5.4 The method: Extreme value theory (EVT)

Volcanic eruption datasets are usually small and the eruptive recurrence is usually very long, and as it happens with other natural phenomena like earthquakes, tsunamis, etc. the higher the magnitude the longer the time interval in between events. To face this problem of working with small datasets, and to be able to obtain a mathematical quantification of the volcanic hazard as accurate as possible, we look for methods that allow us to work with databases which are small and sometimes incomplete [Beguería, 2005; Coles, 2001; Davison and Smith, 1990]. These methods are part of a branch of statistics called Extreme value theory, where as the name implies, extreme values are atypical and rare events located at the tail of the distribution.

Just as the normal distribution proves to be the important limiting distribution for sample sums or averages, as is made explicit in the central limit theorem, another family of distributions proves important in the study of the limiting behaviour of sample extrema. This is the family of extreme value distributions. Extreme value theory and the central limit theory are derived in a similar manner. Both consider the limiting distributions of independent identically distributed (iid) random variables under an affine transformation. In the absence of empirical or physical evidence for assigning an extreme level to a process, an asymptotic argument is used to generate extreme value models. But extreme values are scarce, making it necessary to estimate levels that are much higher than those that have already been observed. In fact, the goal of an extreme value analysis is to quantify the statistical behavior of processes at unusually high levels. In particular, extreme value analysis requires an estimation of the probability of events that are more extreme than any that have ever been observed. This implies an extrapolation from observed levels to unobserved levels. Extreme value theory provides a family of models to make such extrapolation. In fact there are no more serious competitor models than those provided by extreme value theory (Coles 2001).

There are different Extreme value theory methods [Coles, 2001]. Depending on how we define our extreme values we select the method. In our case is more convenient to define our values as peaks or exceedances over a threshold, and so

we use the Exceedances over a threshold (EOT) method to sample the original data, i.e. $X_i > u$ for some value of i. This method is based on a limiting function called GPD, as opposed to the Annual Maximum method with is based on the Gumbel distribution as the limiting function.

The family of GPDs describes the behavior of individual extreme events. It considers observations from a collection of iid random variables where we keep those that exceed a fixed threshold u. As we increase the threshold, the two-parameter GPD family represents the limiting behavior of this new collection of random variables. This makes the family of GPDs a suitable choice for modeling extreme events.

The EOT method includes all the values of the variable that exceed an a-priori established threshold, u, fixed according to the model needs, providing a physically based definition of what must be considered an extreme event. The choice of the threshold value has a strong subjective component. This random variable is defined by the transformed random variable

$$Y = X - u, \text{ for all } X > u$$

where Y is the excess over the threshold u.

The parameter that will be used as random variable to estimate the probability of occurrence of a future eruption, and thus the volcanic hazard, will be the time interval T between eruptions , also called repose period, together with the magnitude M.

The generalized Pareto distribution can be fitted to data on excesses of high thresholds by a variety of methods including the maximum likelihood method (ML) and the method of probability weighted moments (PWM). We use the Davison and Smith [1990] graphical method. The NHGPPP is appropriate for our time series because it is less sensitive to the possible time dependence of the large-magnitude eruption sequence, since it only considers the number of exceedances over a threshold of a series that may be stationary or not. In this respect, it takes into account the limitations in the data mentioned earlier and provides an accurate and reliable estimate. Additionally, this methodology is appropriate for linking historical and geological records should they become available in the

future. So this method sets the base for future analyses and updates should geological records were found.

As a first step before model fitting is undertaken, a number of exploratory graphical methods provide useful preliminary information about the data and in particular their tail. We explain these methods in the next section. To apply the NHGPPP for volcanic hazard assessment we first need to examine the data to assess independence between successive events and homogeneity of the process. To test for independence we will use a serial correlation scatterplot and to test for homogeneity we first assess the stationarity by using the autocorrelation function (ACF) and the Dickey-Fuller unit root test. These tests should be done on a portion of the time series in which no significant eruption data are missing, which in most cases is the historical eruption dataset of intermediate-to-high eruption magnitudes.

After independence and homogeneity have been assessed, we do a Weibull analysis of the repose periods between eruptions to quantitatively describe the stationarity of the series through the distribution shape parameter. The further from 1 the shape parameter is, the more evidence that the process is not stationary.

After the data has been analyzed we apply the NHGPPP to estimate the volcanic hazard. The method is applied to an independent, non-overlapping series of events occurring in a space B with an intensity density $\lambda(x_i)$, where x_i are the B-domain variables where the process develops. In our case x_i are the coordinates T (time) and eruption magnitude M of a two-dimensional space.

5.5 Statistical analysis of the Canary Islands historic volcanic data using EVT

Assuming that past history of a volcano should reflect at least some relevant features of its expected future behavior, we look at the time series of historical volcanic eruptions in the Canary Islands. The time series dataset has fifteen volcanic events historically documented since 1402 [Romero, 1991].

5.5.1 Exploratory analysis of the Canary Islands volcanic data

5.5.1.1 Independence and stationarity assessment

The time independence of the repose periods is a necessary condition for the methods applied here to be appropriate, otherwise they would not satisfy the renewal process definition. A time series where the inter-event times are independent and identically distributed is considered a renewal process [Cox and Lewis, 1966]. To test the independence of successive repose intervals we use a serial correlation scatterplot, where each repose interval is plotted against the previous. The diagram in Fig. 5.2 shows a large dispersion of points and the correlation coefficient between consecutive repose times is 0.3062, indicating a very low serial correlation.

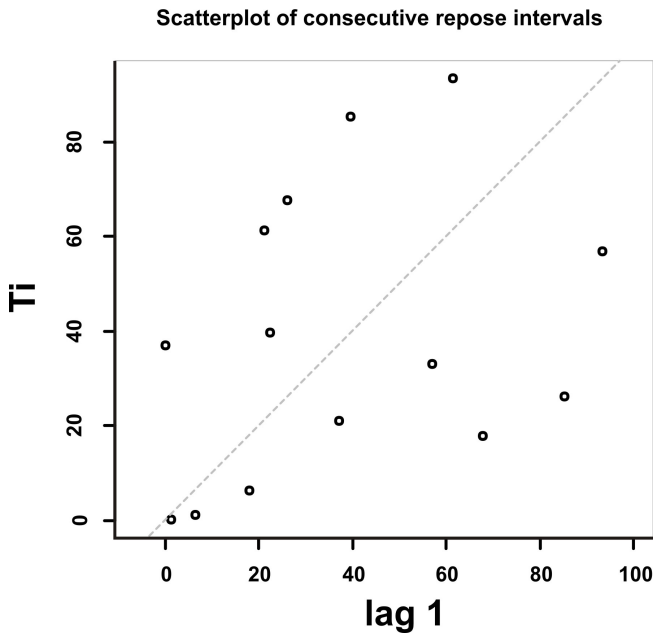

Figure 5.2: Scatterplot of consecutive repose intervals for the Canary Islands time series.

We do not have enough evidence to say that consecutive repose intervals are time-dependent. If new geological data arrives in the future we do not rule out

the possibility of a new outcome for the time-dependence analysis, but for the time being, based on the available data, we assume independence of repose times based on the above mentioned tests.

Next, we look at the stationarity of the process. A time series is stationary if its underlying statistical structure does not evolve with time. The correlogram is a simple diagram which can help diagnose non-stationarity. If a series is non-stationary then the theoretical autocorrelations will be nearly 1 for all lags k. Thus, if the estimated correlogram fails to die down (or dies down very slowly), the series is non-stationary. The theoretical correlogram is a plot of the theoretical autocorrelations between consecutive repose periods of lag k, $corr(x_t, x_{t-k})$, against k. Fig. 5.3 shows the autocorrelation function (ACF) of the Canary Islands time series.

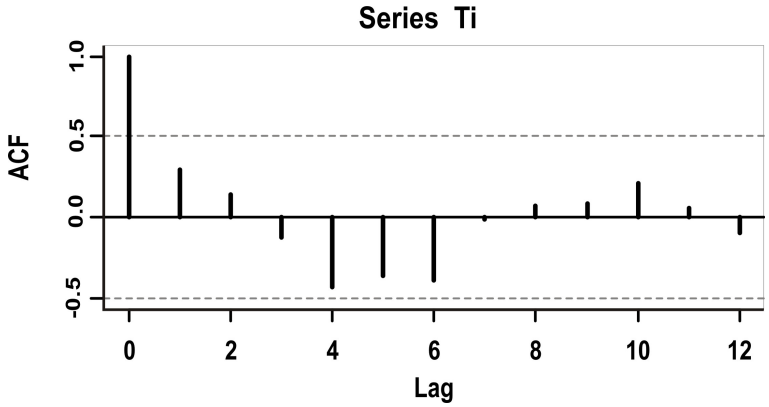

Figure 5.3: Autocorrelation function for the repose periods of the Canary Islands time series.

The argumentation of the non-stationarity based in the shape of the ACF is arguable since the ACF is sensible to seasonal variations which at the same time could correspond to a stationary process. For this reason, to assess stationarity we complement the visual ACF analysis in Fig.5.3 with the Dickey-Fuller unit root test.

The Dickey-Fuller unit root test was proposed by Dickey and Fuller [1979]. In its most basic form, the test compares the null hypothesis $H_0 : x_t = x_{t-1} + \epsilon_t$, i.e., that the series is a random walk without a drift, against the alternative

hypothesis $H_1 : x_t = c + \alpha x_{t-1} + \epsilon_t$ where c and α are constants with $\mid \alpha \mid < 1$. According to H_1, the process is a stationary $AR(1)$ with mean $\mu = c/(1-\alpha)$. To see this, note that, under H_1 we can write $x_t = \mu(1-\alpha) + \alpha x_{t-1} + \epsilon_t$, so that $x_t - \mu = \alpha(x_{t-1} - \mu) + \epsilon_t$. Table 5.3 shows the SAS output for the Dickey-Fuller test. The test statistic has a value of -2.5224, and is associated with a p-value of 0.3734, indicating that there is not enough statistical evidence to reject the null hypothesis that the series is not stationary. This result is consistent with the visual analysis of the ACF, where the series fails to die down.

Table 5.3: *Dickey-Fuller Unit Root test*

Test setup:
H_0: Series is not stationary
H_1: Series is stationary
reject H_0 if p-value < 0.05
Results:
Dickey-Fuller = -2.5224, p-value = 0.3734
p-value $> 5\%$, not enough statistical evidence to reject
the hypothesis that the series is not stationary

In this preliminary analysis of the time series no significant correlation was found, thus we can assume independence of consecutive repose periods. Additionally, we found no evidence to assume that the series is stationary, so based on the ACF and the Dickey-Fuller test for stationarity, we can say that the Canary Islands volcanic eruptions time series is not stationary.

5.5.1.2 Distribution of the repose periods: Weibull versus Exponential

We look at the Weibull distribution to analyze the characteristics of consecutive repose periods and quantitatively describe the stationarity characteristic of the time series through the distribution shape parameter.

The Weibull distribution has been widely applied in statistical quality control, earthquake hazard assessment, and many other applications. It has also been used to model volcanic eruption sequences [Bebbington and Lai, 1996a; Ho, 1991b, 1995]. The 2-parameter cumulative Weibull distribution and survival functions

are given by

$$F(t) = 1 - exp^{-(t/\alpha)^k}$$

and

$$S(t) = 1 - F(t)$$

respectively, where α is a scale parameter, and k is a shape parameter.

The shape parameter reflects the stationary or non-stationary character of the time series [Yang and Xie, 2003]. In the present paper, we obtain the distribution parameters using a fairly simple graphical method [Bebbington and Lai, 1996b]. The probability of having a repose period of duration greater than t has been obtained from the survival function $1 - F(t)$.

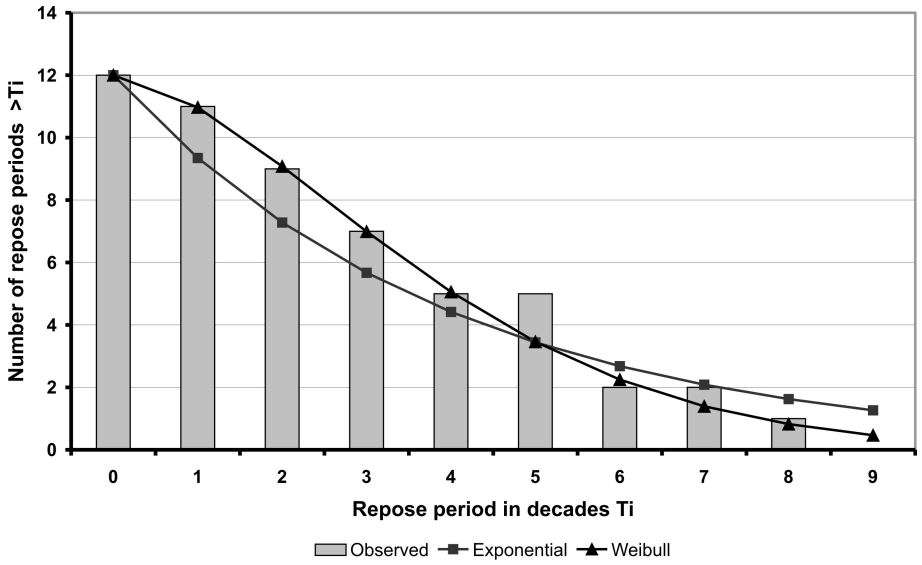

Figure 5.4: Distribution of observed repose intervals with duration greater than T decades (bars) for the Canary Islands since 1402. The survival Weibull distribution shows a much better fitting than the exponential distribution.

The resulting Weibull distribution parameters are 1.63 for the shape parameter and 4.37 for the scale parameter. Fig. 5.4 shows the comparison between Exponential and Weibull distributions. We see that the Weibull survival function provides better fit to the repose periods than the exponential function, because the shape parameter accounts for the non-stationarity of the time series. Addi-

tionally, the shape parameter value being far from one confirms once more the non-stationarity of the process, as a shape of one would correspond to an exponential, which models very well stationary data, which is not the case here.

5.5.2 Volcanic hazard assessment for the Canary Islands

We then estimate the volcanic hazard for the Canary Islands using the NHGPPP. A Poisson process is a collection $\{N(t) : t \geq 0\}$ of random variables, where $N(t)$ is the number of events that have occurred up to time t (starting from time 0). These events occur continuously and independently of one another. The number of events between time a and time b, $N(b) - N(a)$, is said to have a Poisson distribution of intensity λ. When the rate parameter, or intensity, of the process is not constant, the Poisson process is said to be non-homogeneous, and the generalized rate function is given by $\lambda(t)$. As seen in a preliminary analysis of the data, the Canary Islands time series is non-stationary, and we will model the volcanic hazard with a non-homogeneous Poisson process (NHPP). Since we use the EOT method to sample the original data, we can use the GPD to model the intensity of the NHPP. Hence, we will be using a NHGPPP to estimate the volcanic hazard for the Canary Islands. (See Mendoza-Rosas and De la Cruz-Reyna [2008] for further details on this methodology).

To calculate the probabilities of occurrence of an eruption in the different magnitude classes we use the number of excesses inferred from the eruption occurrence rate of each class magnitude, this is, the events above a certain threshold u ($X_i - u$, for some i).

For the particular case of volcanic eruptions, the magnitude of the eruptions and the time of their occurrence are viewed as points in a two-dimensional space, which formally is the realization of a point process [Cox and Isham, 1980]. The intensity measure $\Lambda(B)$ of this two-dimensional Poisson process on the space $B = [t_1, t_2]$ with $[t_1, t_2] \subset [0, 1]$ is given by

$$\Lambda(B) = (t_2 - t_1)\left[1 - \frac{\beta(x-u)}{\theta}\right]^{1/\beta} \tag{5.1}$$

where β, and θ are the parameters of the *GPD*.

The *GPD* is described by a shape parameter β, a scale parameter θ, and a location parameter u (threshold), and has the following cumulative distribution function:

$$G_{\beta,\theta}(y) = 1 - (1 - \frac{\beta y}{\theta})^{1/\beta} \text{ for } \beta \neq 0$$

$$G_{\beta,\theta}(y) = 1 - e^{-y/\theta} \text{ for } \beta = 0$$

Another related property of the *GPD* refers to the mean excess: if $Y = X - u$ is a GP-distributed variable, then the mean excess over threshold u is:

$$E(X - u|X > u) = \frac{\theta - \beta u}{1 + \beta} \tag{5.2}$$

for $\beta > -1$, $u > 0$ and $\theta - u\beta > 0$

The sample mean excess plot is given by:

$$\bar{x}_u = \frac{\sum_{i:x_i>u}(x_i - u)}{N_u} \tag{5.3}$$

where N_u is the number of excess x_i over a threshold u

Davison and Smith [1990] introduced a diagnostic plot to decide how well the model fits the data. The mathematical basis for this method is Eq. (5.2), where the key feature is that if Y is GPD then the mean excess over a threshold u, for any $u > 0$, is a linear function of u [Beguería, 2005; Coles, 2001; Lin, 2003]. In Fig. 5.5 we plot the mean of the excesses, obtained with Eq. (5.3), vs their thresholds.

The x-axis is the threshold and the y-axis is the sample mean of all excesses over that threshold. As we can see, the mean excess follows a nearly straight line, with a R^2 of 0.8415, suggesting a good fit. A regression line of mean of exceedances over the threshold has been added to confirm the series follows the GPD. Since we are working with effusive eruptions only, we assumed an upper bound of six for the estimation of the magnitudes, and mapped the probabilities to the [1,6] magnitude interval. Hence, according to Davison and Smith [1990], the preceding results indicate that the NHGPPP is satisfactory and appropriate to model the Canary Islands eruptive time series.

The Pareto generalized parameters for the process, derived from the regres-

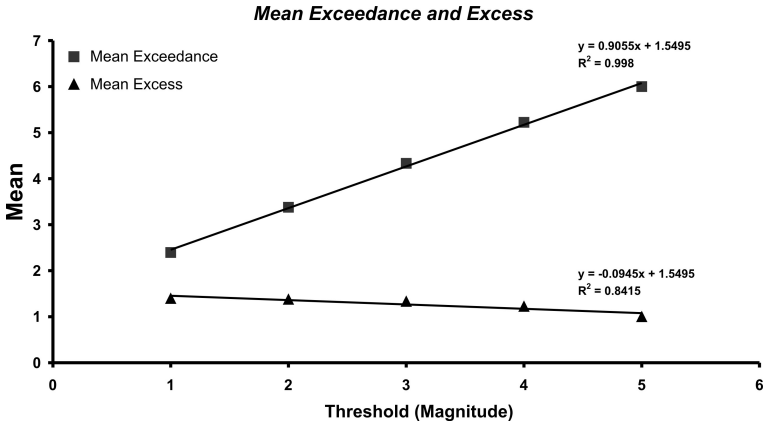

Figure 5.5: Plot of exceedance and excess mean vs. threshold for the Canary Islands.

sion parameters on Fig. 5.5 and Eq. (5.2) are 0.104 for shape and 1.711 for the scale. Using Eq. (5.1) we calculate the intensity of the NHGPPP and obtain the eruption occurrences. Table 5.4 and Fig. 5.6 show the probability of occurrence of at least one eruption exceeding a specific magnitude over different time periods. Table 5.5 displays the intensity of the two-dimensional Poisson process, computed using Eq.(5.1), which also serves as a measure of parameter uncertainty and the volatility of the probability results. This is the expected value of the distribution (μ_i), where $i = 1, 2, 3, 4$ are the number of eruptions, and represents an estimation of the intrinsic (and unavoidable) random variability due to the complexity of the process; and the standard deviation (σ_i, $i = 1, 2, 3, 4$), which represents an estimation of the epistemic uncertainty due to our limited knowledge of the process.

Based on the existing historical data, the probability of an event in the Canary Islands increases more rapidly in the first 20 years, with a 99.84% chance of an event of magnitude greater than one in the next 20 years and leveling out after that at 99.99 %. There is a probability of 27.58% of an event of magnitude between 1 and 6 in the next 12 months and 3.71% of an event of magnitude between 4 and 6 for the same period. There is ongoing work to assess more accurately the data in the volcanic eruptions catalog for the Canary Islands. In this respect, these probability results may vary should new geological records be

Table 5.4: *Probability of at least one event of Magnitude $> x$ in the next t years, based on the intensity rates from Table 5.5.*

Years	Magnitude> 1	Magnitude> 2	Magnitude> 3	Magnitude> 4
1	27.58%	15.80%	8.20%	3.71%
20	99.84%	96.79%	81.92%	53.06%
50	99.99%	99.98%	98.61%	84.90%
75	99.99%	99.99%	99.84%	94.13%
100	99.99%	99.99%	99.98%	97.72%

available.

Table 5.5: *Measures of parameter uncertainty: the intensity of the two-dimensional Poisson process, computed using Eq.(5.1), is the expected value of the distribution (μ_i), where $i = 1, 2, 3, 4$ are the number of eruptions, and represents an estimation of the intrinsic (and unavoidable) random variability due to the complexity of the process; and the standard deviation (σ_i, $i = 1, 2, 3, 4$), which represents an estimation of the epistemic uncertainty due to our limited knowledge of the process.*

Years	μ_0	σ_0	μ_1	σ_1	μ_2	σ_2	μ_3	σ_3	μ_4	σ_4
1	0.32	0.57	0.17	0.41	0.09	0.29	0.04	0.19	0.01	0.11
20	6.45	2.54	3.44	1.85	1.71	1.31	0.76	0.87	0.25	0.50
50	16.13	4.02	8.60	2.93	4.28	2.07	1.89	1.38	0.63	0.79
75	24.20	4.92	12.90	3.59	6.41	2.53	2.84	1.68	0.95	0.97
100	32.27	5.68	17.19	4.15	8.55	2.92	3.78	1.94	1.26	1.12

5.6 Discussion and conclusions

The Canary Islands are an active volcanic region densely populated and visited by several millions of tourists every year. Nearly twenty eruptions have been reported by written chronicles in the last 600 years. This gives an average of an eruption every 25-30 years, which suggests that the probability of having a new eruption in the near future is not so low. Under these circumstances and in order to reduce the potential volcanic risk of this region, it is highly recommendable to undertake hazard assessment, and determine the eruption recurrence for the area.

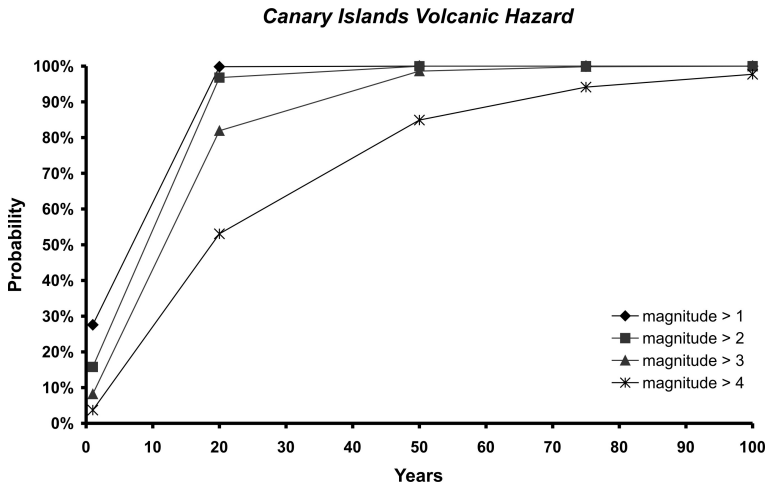

Figure 5.6: Probabilities calculated with NHGPPP of at least one eruption, with a Magnitude greater than one.

Recent volcanism in the Canary Islands is not well known and is poorly constrained in terms of age of the eruptions. For the island of Tenerife alone, Carracedo et al. [2007, 2010] have conducted a systematic geochronological study for Teide and the volcanism associated with the rifts zones, but this study is still far from being complete. Therefore, the data catalog to be used for statistical and probabilistic assessment of the Canary Islands to establish the eruption recurrence is formed uniquely by historical records. The model can be easily updated in the future should new volcanic records be dated.

As in any data analysis, we should be aware of various layers of uncertainty, perhaps magnified in an extreme value analysis. On one level, there is the parameter uncertainty, even if we had abundant, good quality data to work with and a good model, our parameter estimates are still subject to a standard error. Model uncertainty is also present - we may have good data but a poor model. Using extreme value methods we are at least working with a good class of models, but they are applicable over high thresholds and we must decide where to set the threshold. If we set the threshold too high we have few data and we introduce more parameter uncertainty. If we set the threshold too low we lose our theoretical justification for the model. But even more serious than parameter and model uncertainty is the data uncertainty. It is never possible to have enough

data in an extreme value analysis. Table 5.5 shows an estimation of the parameter uncertainty for our model, the expected value of the distribution (μ_i), where $i = 1, 2, 3, 4$ are the number of eruptions, which is an estimation of the aleatoric uncertainty due to the complexity of the process, and the standard deviation (σ_i, $i = 1, 2, 3, 4$) which represents an estimation of the epistemic uncertainty, due to our limited knowledge of the process.

Extreme value methods do not predict the future with certainty, but they do offer good models for explaining the extreme events we have seen in the past [McNeil, 1997]. Even with a good tail estimate we cannot be sure that the future does not hold some unexpected catastrophic volcanic eruption. The extreme value method used in this paper to assess the volcanic hazard for the Canary Islands do not predict the future with certainty, but it is a model based on rigorous mathematical theory concerning the behaviour of extrema. Based on past experience [Mendoza-Rosas and De la Cruz-Reyna, 2008], the GPD is a good approximation in the tail for our volcanic data, and the probability results yielded by the extreme value method used here to assess the volcanic hazard for the Canary Islands should not be neglected. It may well be that, by trial and error, some other distribution can be found which fits the data even better in the tail, but such a distribution would be an arbitrary choice, and we would have less confidence in extrapolating it beyond the data.

The probability results obtained are very high. This is partly due to the fact that the area of study is the quasi linear 500 Km long chain grown on the passive margin of the African Plate containing the actual Canarian archipelago, this is, several possible vent locations for an eruption. Also, we must consider the fact that we are measuring magnitudes (total erupted volumes) and not VEI (Volcanic Explosivity Index) Newhall and Self [1982]. The VEI is a combination of volume of products, eruption cloud height and qualitative observations. It is mainly applied to explosive eruptions and is not appropriate for eruptions which are mainly effusive. This is the reason why we have used magnitude instead of VEI and limited the study to level of magnitudes up to six. However, the eruption magnitude, measured as total erupted volume only takes into account one of the three measures of the VEI, hence, the probability estimates for volume alone are expected to be higher than the estimates for volume of products, eruption

cloud height and qualitative observations combined, since one is a subset of the other. With this in mind, given the current data, it is not surprising to observe a probability of 27.58% of having a volcanic event of magnitude greater than 1 in the next year in the Canary Islands, most likely in any of the three for which historical data exist (Lanzarote, Tenerife, La Palma) but without excluding the other four islands. Even if there are no historical records documented for all the islands, we cannot rule out the probability of an event forming there since they are part of the same archipelago and there are traces of previous volcanic eruptions. We do not have enough data to do an individual hazard analysis for each island alone.

It is important to highlight the fact that the Canary Islands have a probability greater than zero of undergoing a new volcanic event in the upcoming years. Hence, these results should be taken into account in the assessment of volcanic risk and in the design of prevention and response measures, particularly of major eruptions to which larger areas may be one hundred per cent vulnerable.

The results obtained only apply to the probabilities of having a basaltic eruption in the near future, as historical volcanism has been always represented by this kind of eruptions. However, the existence of several eruptions of phonolitic magmas from Teide in Holocene times on the island of Tenerife, the last one having occurred only 1000 year ago [Carracedo et al., 2007], reminds us that hazard assessment should also consider phonolitic eruptions. Despite being concentrated during Holocene exclusively on Tenerife, these eruptions may generate hazards that could have a much greater impact than basaltic eruptions, so their potential effects should not be neglected in a more complete volcanic hazard assessment for the Canary region.

Chapter 6

Conclusions

Volcanoes are complex, non-linear natural systems that rarely follow a constant pattern of behavior. Due to the limited scientific observability of the interior of a volcano, there is a lot of uncertainty in forecasting volcanic eruptions. Although we can establish some general eruptive patterns for certain group or types of volcanoes, each volcano will at the end behave in its own particular way, different from the others. In the case of Teide - Pico Viejo, as in many other volcanoes around the world, the information we have on its past eruptive history and present state of activity is incomplete, and requires much more effort before being more confident that we can forecast its future behavior precisely. For this reason, we need quantitative risk-based methods for decision-making under uncertainty to be developed and applied to volcanology.The interdisciplinary science of mathematics applied to the study of volcanology and volcanic hazard is an important approach, which helps understand volcanic processes by integrating keen volcanological insights with sound statistical modeling and artful application of computational power.

In this work we have presented different statistical methodologies and used different statistical software (R, SAS) to interpret volcanic data and assess volcanic hazard, using Teide - Pico Viejo stratovolcanoes (TPV) and the Canary Islands archipelago as the case study for the historical and geological volcanic data. The statistical methods, based on the nature of the data, aim to extract information about the future behavior of the volcano by looking at the geological and historical activity of the volcanic system, considering that the volcanic

database is incomplete and inaccurate.

The first methodology uses the so-called Classical model, a statistical approach based on expert judgment elicitation, to assign probabilities and determine the corresponding uncertainty for the different possible events that could occur, based on available past geological data. This method is very appropriate when none or little past data exists. It relies strongly on the judgment and knowledge of the scientific experts invited to the elicitation, and because of this there is a strong subjective component in the final probabilities. This method was used to successfully outline for the first time a probability event tree with all the possible eruptive scenarios of TPV, and assign a probability of occurrence to each scenario. This manuscript "A long-term volcanic hazard event tree for Teide - Pico Viejo stratovolcanoes (Tenerife, Canary Islands)" has been published on October 17th, 2008 in the Journal of Volcanology and Geothermal Research 178 (2008) 543-552. Rank 45 over 155 journals from multidisciplinary Geosciences, with impact factor 1.921 in the JCR. Authors: J. Martí, W.P. Aspinall, R. Sobradelo, A. Felpeto, A. Geyer, R. Ortiz, P. Baxter, P. Cole, J. Pacheco, M.J. Blanco, C. Lopez. Authors contribution: J. Martí is the leader volcanologist of the project and participated as member of the elicitation team. W.P. Aspinall coordinated the Elicitation of expert judgment sessions where the volcanic experts J. Martí, A. Felpeto, A. Geyer, R. Ortiz, P. Baxter, P. Cole, J. Pacheco, M.J. Blanco, and C. Lopez participated to design the event tree and collect the data. J. Martí developed the volcanology part of the manuscript, W.P. Aspinall and R. Sobradelo were responsible for the statistical part of the exercise.

The second methodology uses Bayesian Inference to develop a new event tree model for long term volcanic hazard assessment, that represents a step forward with respect to previous attempts based on the same methodology or the elicitation of expert judgment. In comparison with previous event trees based on Bayesian methodology, this model accounts for the possibility of the unrest being caused by external triggers (geothermal, seismic), and adds new nodes with two additional sources of volcanic hazard based on the composition of the magma and different vent locations. With respect to event trees based on Elicitation of expert judgment the new model does not have the additional source of bias that the human decision component adds to the final results of the elicitation

method, controls for the epistemic and aleatory uncertainties, and allows the level of segmentation and complexity of the event tree structure to be as complete and extensive as needed, with the only requirements of mutually exclusive and exhaustive events in each node. It also permits to automatically update the probabilities when new data arrives or the system becomes active and monitoring data on precursors exists, as opposed to the eliciting method which requires the group of experts to meet each time new data arrives to update the probabilities. However, during a volcanic crisis, Elicitation and Bayesian models are needed and the elicitation team should provide input and interpretation to the probabilities from the updated Bayesian model. The new method allows us to estimate the long-term probability during a quiet period of the volcano, being useful for land use policy, and will be of use for estimating and automatically updating the short term probabilities when monitoring data are obtained during unrest. Although this method is specifically applied to the Teide - Pico Viejo stratovolcanoes in Tenerife, it can be used with other similar volcanoes as it offers a wider structure in comparison with previous event trees that have a more restricted structure and do not include some relevant eruptive scenarios which are likely in the Teide - Pico Viejo but also in many other composite volcanoes. This manuscript "Bayesian event tree for long-term volcanic hazard assessment: Application to Teide - Pico Viejo stratovolcanoes, Tenerife, Canary Islands" has been published on May 21st, 2010 in the Journal of Geophysical Research vol 115, B05206 (2010), doi:10.1029/2009JB006566. Rank 18 over 155 journals from multidisciplinary Geosciences, with impact factor 3.082 in the JCR. Authors: R. Sobradelo and J. Martí. Authors contribution: R. Sobradelo developed, implemented and interpreted the Bayesian methodology. J. Martí contributed with the volcanological part as well as the design of the volcanic event tree.

The third method is a non-parametric one-way unbalanced ANOVA with the Kruskal-Wallis test to identify groups of volcanic calderas statistically and significantly different, according to area, which may belong to a particular geodynamic environment. This method is used to study a unique dependent variable against one classification variable which has two or more categories, where each classification group has unequal number of observations. Additionally, since the observations in the response variable are assumed to be independent from each other, but

we do not have enough evidence to assume a particular distributional form, such as the normal (due to insufficient data), we need to use non-parametric procedures to perform the ANOVA analysis, in our case, the Kruskal-Wallis test. The results obtained from our statistical analysis show that caldera size is strongly dependent on the geodynamic environment where caldera develops, i.e. type and nature of crust, magma composition and plate tectonic setting. From the data analyzed, we are able to group collapse calderas in three different geodynamic environments: GE1, GE2 and GE3. The results obtained confirm quantitatively the generally accepted idea that the characteristics of collapse calderas are influenced by the characteristics of the geodynamic environment where they develop, in particular the mantle thermal regime, the crust composition and thickness, and the tectonic regime. This manuscript "Statistical data analysis of the CCDB (Collapse Caldera Database): Insights on the formation of caldera systems" has been published on September 16th, 2010 in the Journal of Volcanology and Geothermal Research 198 (2010) 241-252. Rank 45 over 155 journals from multidisciplinary Geosciences, in the JCR with impact factor 1.921. Authors: R. Sobradelo, A. Geyer and J. Martí. Authors contribution: R. Sobradelo performed the statistical analysis. A. Geyer provided the volcanic caldera dataset as well as, together with J. Martí, the volcanological insights and interpretation.

The fourth statistical methodology NHGPPP (Non-homogeneous generalized Pareto-Poisson process) uses extreme value theory to study eruptive time series combining geological and historical records. This method has been chosen to analyze the Canary Islands volcanic time series given the limitations in the data, this is, short time interval of data, probable absence of large events, incomplete reporting of small and intermediate magnitudes, and uncertainties in the age and magnitude of the eruptions. The probability results obtained are very high. This is partly due to the fact that the area of study is the quasi linear 500 Km long chain grown on the passive margin of the African Plate containing the actual Canarian archipelago, this is, several possible vent locations for an eruption. Also, we must consider the fact that we are measuring magnitudes (total erupted volumes) and not VEI (Volcanic Explosivity Index). Even if there are no historical records documented for all the islands, we cannot rule out the probability of an event forming there since they are part of the same archipelago

and there are traces of previous volcanic eruptions. We do not have enough data to do an individual hazard analysis for each island alone. This manuscript "Volcanic hazard assessment for the Canary Islands (Spain) using Extreme Value Theory" has been submitted for publication to Natural Hazards Earth System Sciences on 15th February 2011. Rank 70 over 155 journals from multidisciplinary Geosciences, with Impact Factor 1.357 in the JCR. Authors: R. Sobradelo, J. Martí, A.T. Mendoza-Rosas and G. Gómez. Authors contribution: R. Sobradelo designed and implemented the statistical method with the help and supervision of A.T. Mendoza-Rosas and G. Gómez, whilst J. Martí provided the data for the study and the volcanological explanation and interpretation.

Future research work includes: completing the Bayesian event tree with a short term probability model to be used during a volcanic crisis; building an evacuation model to link the Bayesian probability event tree and the decision making process; update models when new volcanic calderas data and new geological records on volcanic eruptions for the Canary Islands become available; and making the computer program written to compute the Bayesian probabilities user friendly and available for the scientific community to use.

Appendix: R Code for the Bayesian event tree

We include here the computer program written in R statistical language, to automatically compute the long-term probability of an event within the next t years, using Bayesian inference. This code was developed for the probability tree model presented in chapter three.

```
#ltvh        function to compute the absolute probability for the
#            Long-Term Volcanic Hazard Assessment
#            for a generic explosive volcano GIVEN a time window t
#            allowing for several outcomes in each node.
#
#            This is, since the event has not happened yes, we compute
#            the absolute probability P(Y_1^N = 1,Y_2^N = 1, Y_3^N =
1|Y_1,Y_2,Y_3)
#            given the observed data Y_1, Y_2, Y_3
#            Uses the Dirichlet function (generalization of the Beta)
#            for the prior and the Multinomial for the likelihood.
#            The conjugate prior is a Dirichlet(alpha).
#
#            ASSUMPTIONS:
#            - multiple outcomes in each node.
#            - outcomes are mutually exclusive and exhaustive.
#            - uniform priors
#
#
# IMPUT
#  pb1, ..., pb6 = the J_k*2 are the matrices of past belief, containing the
E[theta_ki], lamda_ki
#  parameter information for each node (k = 1, ..., 6) and each event within
each node (i = 1, ..., Jk)
#  (Prior and Data weight). Weights are equally distributed based on the
assumption of non-informative priors.
#
#
#  y1, ..., y6 = the vectors with the geological and geophysical data for each
node, each
#  vector is of dimension 1*J_k (Past events).
#
# OUTPUT
#  probvector     vector of long term probabilities in each node
#


ltvh<- function(pb1,pb2,pb3,pb4,pb5,pb6,y1,y2,y3,y4,y5,y6)
{

  #Node1:Yes No
  #Node2:Magmatic Geothermal Seismic Other
  #Node3:Magmatic_Eruption Sector_Failure Phreatic_Eruption No_Eruption
  #Node4:Central North South East West
  #Node5:Basaltic Phonolitic
  #Node6:VEI 5+ VEI 4 VEI 3 VEI 2-

  #apbcv = a priori belief central value. Is the central value inferred by a
priori models
  #and/or of the theoretical beliefs

  #cre = confidence reliable estimate. Is the  confidence at which apbcv is
considered a
  #reliable estimate, this is an evaluation of the epistemic uncertainties. The
confidence
```

```
  #degree is set up by writing the variance in terms of "equivalent number of
data". This is
  #a more friendly measure that the variance of the confidence on the prior
distribution, or,
  #in other terms, of the epistemic uncertainty. In general, the higher the
value, the larger
  #our confidence on the reliability of the model, so that the number of past
data needed to
  #modify significantly the prior must be larger. On the contrary, if we believe
that the prior
  #is poorly informative, then the value must be small, so that even a small
number of past data
  #can drastically modify the prior. The minimum possible value is 1 and this
number represent
  #the maximum possible epistemic uncertinty. In the limit case of maximum
ignorance, the value is 1,
  #this is the reliability of our model is comparable to the information given
by on datum, and
  # a mean equal to .5 we obtain a prior uniform distribution.

  #NOTE: these two parameters (apbcv and cre) are part of the alphas value.
  #So, in node k, for each of the ith events:
  #alpha_ki = apbcv_ki (cre_ki + J_k - 1), where apbcv_ki and cre are imputs,
and
  #apbcv_ki = alpha_ki/sum(alpha_ki;i=1...J_k) and
  #cre=sum(alpha_ki;i=1...J_k) - J_k +1
  #so, past_belief, pb: pb_k = matrix of, pb[,1] = apbcv and pb[,2] = cre for
node K.


  assign("pb1",pb1 )
  assign("pb2",pb2 )
  assign("pb3",pb3 )
  assign("pb4",pb4 )
  assign("pb5",pb5 )
  assign("pb6",pb6 )
  assign("y1",y1 )
  assign("y2",y2 )
  assign("y3",y3 )
  assign("y4",y4 )
  assign("y5",y5 )
  assign("y6",y6 )


#node1: UNREST
  #random variable = number of unrest episodes in 1 time window
  #compute probability of unrest, given observed data

  outcome1<-c("Yes","No")
  alpha1 <- matrix(1,ncol=length(y1)) #alpha of the Dirichlet distribution
  a1  <- matrix(0,ncol=length(y1))
  eoy1<- matrix(0,ncol=length(y1))
  E1  <- matrix(0,ncol=length(y1))
  for (j in 1:length(y1))
  {
    alpha1[j]= pb1[j,1]*( pb1[j,2] + length(y1) - 1)
    a1[j]    =alpha1[j]+y1[j]
```

```
  }
  for (j in 1:length(y1))
  {
    eoy1[j]=a1[j]/(sum(alpha1)+ sum(y1))
    E1[j]  =eoy1[j]
  }
#node2: ORIGIN
  #random variable=number of event episodes in 1 time window
  #compute probability of the event, given observed data

  outcome2<-c("Magmatic","Geothermal","Seismic","Other")
  alpha2 <- matrix(1,ncol=length(y2))#alpha of the Dirichlet distribution
  a2  <- matrix(0,ncol=length(y2))
  eoy2<- matrix(0,ncol=length(y2))
  E2  <- matrix(0,ncol=length(y2))
  for (j in 1:length(y2))
  {
    alpha2[j]= pb2[j,1]*( pb2[j,2] + length(y2) - 1)
    a2[j]    =alpha2[j]+y2[j]
  }
  for (j in 1:length(y2))
  {
    eoy2[j]=a2[j]/(sum(alpha2)+ sum(y2))
    E2[j]  =eoy2[j]
  }
#node3: OUTCOME
  #random variable=number of events in 1 time window
  #compute probability of an event, given observed data

  outcome3<-c("Magmatic Eruption","Sector Failure","Phreatic Eruption","No
Eruption")
  alpha3 <- matrix(1,ncol=length(y3))#alpha of the Dirichlet distribution
  a3  <- matrix(0,ncol=length(y3))
  eoy3<- matrix(0,ncol=length(y3))
  E3  <- matrix(0,ncol=length(y3))
  for (j in 1:length(y3))
  {
    alpha3[j]= pb3[j,1]*( pb3[j,2] + length(y3) - 1)
    a3[j]    =alpha3[j]+y3[j]
  }
  for (j in 1:length(y3))
  {
    eoy3[j]=a3[j]/(sum(alpha3)+ sum(y3))
    E3[j]  =eoy3[j]
  }
#node4: LOCATION
  #random variable = number of events in this location in 1 time window
  #compute probability of an event in this location, given observed data

  outcome4<-c("Central","North","South","East","West")
  alpha4 <- matrix(1,ncol=length(y4))#alpha of the Dirichlet distribution
  a4  <- matrix(0,ncol=length(y4))
  eoy4<- matrix(0,ncol=length(y4))
  E4  <- matrix(0,ncol=length(y4))
  for (j in 1:length(y4))
  {
    alpha4[j]= pb4[j,1]*( pb4[j,2] + length(y4) - 1)
```

```
    a4[j]    =alpha4[j]+y4[j]
  }
  for (j in 1:length(y4))
  {
    eoy4[j]=a4[j]/(sum(alpha4)+ sum(y4))
    E4[j]  =eoy4[j]
  }
#node5: COMPOSITION
  #random variable = number of events in this location in 1 time window
  #compute probability of an event in this location, given observed data

  outcome5<-c("Basaltic","Phonolitic")
  alpha5 <- matrix(1,ncol=length(y5))#alpha of the Dirichlet distribution
  a5  <- matrix(0,ncol=length(y5))
  eoy5<- matrix(0,ncol=length(y5))
  E5  <- matrix(0,ncol=length(y5))
  for (j in 1:length(y5))
  {
    alpha5[j]= pb5[j,1]*( pb5[j,2] + length(y5) - 1)
    a5[j]    =alpha5[j]+y5[j]
  }
  for (j in 1:length(y5))
  {
    eoy5[j]=a5[j]/(sum(alpha5)+ sum(y5))
    E5[j]  =eoy5[j]
  }
#node6: ERUPTION SIZE
  #random variable=number of events in 1 time window
  #compute probability of an event, given observed data

  outcome6<-c("VEI 5+","VEI 4","VEI 3","VEI 2-")
  alpha6 <- matrix(1,ncol=length(y6))#alpha of the Dirichlet distribution
  a6  <- matrix(0,ncol=length(y6))
  eoy6<- matrix(0,ncol=length(y6))
  E6  <- matrix(0,ncol=length(y6))
  for (j in 1:length(y6))
  {
    alpha6[j]= pb6[j,1]*( pb6[j,2] + length(y6) - 1)
    a6[j]  =alpha6[j]+y6[j]
  }
  for (j in 1:length(y6))
  {
    eoy6[j]=a6[j]/(sum(alpha6)+ sum(y6))
    E6[j]  =eoy6[j]
  }

  column1=matrix(c(outcome1,outcome2,outcome3,outcome4,outcome5,outcome6),21,1)
  column2=format(matrix(c(E1,E2,E3,E4,E5,E6),21,1),digits=3)
  probvector=matrix(c(column1,column2),21,2)

  #OUTPUT
  return(probvector)
 }


  # EXAMPLE:
```

```
  # This example reads the past data from a csv file previously saved in C, and
runs the ltvh.r program on this data.
  # write the address on your C drive where the past_data.csv file is. Make sure
you keep the same template and structure,
  # otherwise it will affect the matrix order when the program reads in the
data:

  ltvh_data = read.csv("C:/event_tree/past_data.csv", header = TRUE, sep = ",")


  # DO NOT MODIFY CODE FROM HERE ON:

  y1=c(ltvh_data[1,4],ltvh_data[2,4])
  y2=c(ltvh_data[3,4],ltvh_data[4,4],ltvh_data[5,4],ltvh_data[6,4])
  y3=c(ltvh_data[7,4],ltvh_data[8,4],ltvh_data[9,4],ltvh_data[10,4])

y4=c(ltvh_data[11,4],ltvh_data[12,4],ltvh_data[13,4],ltvh_data[14,4],ltvh_data[1
5,4])
  y5=c(ltvh_data[16,4],ltvh_data[17,4])
  y6=c(ltvh_data[18,4],ltvh_data[19,4],ltvh_data[20,4],ltvh_data[21,4])

  z1=c(ltvh_data[1,5],ltvh_data[2,5])
  z2=c(ltvh_data[3,5],ltvh_data[4,5],ltvh_data[5,5],ltvh_data[6,5])
  z3=c(ltvh_data[7,5],ltvh_data[8,5],ltvh_data[9,5],ltvh_data[10,5])

z4=c(ltvh_data[11,5],ltvh_data[12,5],ltvh_data[13,5],ltvh_data[14,5],ltvh_data[1
5,5])
  z5=c(ltvh_data[16,5],ltvh_data[17,5])
  z6=c(ltvh_data[18,5],ltvh_data[19,5],ltvh_data[20,5],ltvh_data[21,5])

  w1=c(ltvh_data[1,6],ltvh_data[2,6])
  w2=c(ltvh_data[3,6],ltvh_data[4,6],ltvh_data[5,6],ltvh_data[6,6])
  w3=c(ltvh_data[7,6],ltvh_data[8,6],ltvh_data[9,6],ltvh_data[10,6])

w4=c(ltvh_data[11,6],ltvh_data[12,6],ltvh_data[13,6],ltvh_data[14,6],ltvh_data[1
5,6])
  w5=c(ltvh_data[16,6],ltvh_data[17,6])
  w6=c(ltvh_data[18,6],ltvh_data[19,6],ltvh_data[20,6],ltvh_data[21,6])

  pb1=cbind(z1,w1)
  pb2=cbind(z2,w2)
  pb3=cbind(z3,w3)
  pb4=cbind(z4,w4)
  pb5=cbind(z5,w5)
  pb6=cbind(z6,w6)


  # RUN THE LTVH.R PROGRAM ON THE DATA:


  ltvh(pb1,pb2,pb3,pb4,pb5,pb6,y1,y2,y3,y4,y5,y6)
```

References

G. J. Ablay and J. Martí. Stratigraphy, structure, and volcanic evolution of the Pico Teide-Pico Viejo formation, Tenerife, Canary Islands. *Journal of Volcanology and Geothermal Research*, 103, 2000. doi: 10.1016/S0377-0273(00)00224-9. 18, 19, 20, 21, 22, 23, 25, 41, 46, 49, 55

G.J. Ablay, M.R. Carroll, M.R. Palmer, J. Martí, and R.S.J. Sparks. Basanite-phonolite lineages of the Teide - Pico Viejo volcanic complex, Tenerife, Canary Islands. *Journal of Petrology*, 39:905–936, 1998. 18

V. Acocella and F. Rossetti. The role of extensional structures on pluton ascent and emplacement: the case of Southern Tuscany (Italy). *Tectonophysics*, 354: 71–83, 2002. 87

V. Acocella, F. Cifelli, and R. Funiciello. Analogue models of collapse calderas and resurgent domes. *Journal of Volcanology and Geothermal Research*, 104 (1-4):81 – 96, 2000. ISSN 0377-0273. doi: 10.1016/S0377-0273(00)00201-8. 63

V. Acocella, F. Cifelli, and R. Funiciello. Formation of non intersecting nested calderas: Insights from analogue models. *Terra Nova*, 13:58–63, 2001. doi: 10.1046/j.1365-3121.2001.00317.x. 63

V. Acocella, R. Funiciello, E. Marotta, G. Orsi, and S. de Vita. The role of extensional structures on experimental calderas and resurgence. *Journal of Volcanology and Geothermal Research*, 129:199–217, 2004. doi: 10.1016/S0377-0273(03)00240-3. 63, 87

F. Anguita and F. Hernán. A propagating fracture model versus a hot-spot origin for the Canary Islands. *Earth Planetary Science Letters*, 27:11–19, 1975. 97

F. Anguita and F. Hernán. The Canary Islands origin: a unifying model. *Journal of Volcanology and Geothermal Research*, 103:1–26, 2000. 97

C. Annen and R. S. J. Sparks. Effects of repetitive emplacement of basaltic intrusions on thermal evolution and melt generation in the crust. *Earth and Planetary Science Letters*, 203:937–955, 2002. 90

V. Araña and R. Ortiz. The Canary Islands: Tectonics, magmatism, and geodynamic framework. In A.B. Kampunzu and R.T. Lubala, editors, *Magmatism in Extensional Structural Settings and the Phanerozoic African Plate*, pages 209–249. Springer, New York, 1991. 93, 97

W. P. Aspinall. Structured elicitation of expert judgment for probabilistic hazard and risk assessment in volcanic eruptions. In H. M. Mader et al., editors, *Statistics in Volcanology*. Special Publication of IAVCEI, Geological Society of London, 2006. 4, 9, 11, 16, 30, 31, 36

W. P. Aspinall and G. Woo. An impartial decision-making procedure using expert judgment to assess volcanic hazards. Academia Nazionale del Lincei-British Council Symposium Large Explosive Eruptions, Rome, 24-25 May 1993, Atti dei Convengi Lincei 112: 211-220, 1994. 10, 16, 37

W.P. Aspinall and R.M. Cook. Expert judgement and the Montserrat Volcano eruption. In A. Mosleh et al., editors, *Proceedings of the 4th International Conference on Probabilistic Safety Assessment and Management PSAM4, 13-14 September 1998*, pages 2113–2118. Springer, New York, USA, 1998. 9, 17

W.P. Aspinall, L.L. Lynch, R.E.A. Robertson, K. Rowley, R.S.J. Sparks, B. Voight, and S.R. Young. The Sufriere Hills eruption, Montserrat, British West Indies: introduction to. Special Section, Part I. *Geophysical Research Letters*, 25:3387–3388, 1998. 37

W.P. Aspinall, R. Carniel, O. Jaquet, G. Woo, and T. Hincks. Using hidden multi-state markov models with multi-parameter volcanic data to provide empirical

evidence for alert level decision-support. *Journal of volcanology and geothermal research*, 153:112–124, 2006. 95

M.S. Bebbington. Identifying volcanic regimes using hidden markov models. *Geophysical Journal International*, 171:921–942, 2007. 11, 95

M.S. Bebbington and C.D. Lai. Statistical analysis of new zealand volcanic occurrence data. *Journal of volcanology and geothermal research*, 74:101–110, 1996a. 11, 95, 106

M.S. Bebbington and C.D. Lai. On nonhomogeneous models for volcanic eruptions. *Mathematical Geology*, 28:585–600, 1996b. 11, 95, 107

T.J. Bedford and R.M Cooke. *Probabilistic risk analysis, foundations and methods*. Cambridge University press, Cambridge, 2001. 8, 30

Beguería, S.: Uncertainties in partial duration series modeling of extremes related of the choice of the threshold value, Journal of Hydrology, 303, 215–230, 2005. 101, 109

R. Blong. Volcanic hazards and risk management. In H. Sigurdsson et al., editors, *Encyclopedia of Volcanoes*, pp 1417. Academic Press, San Diego, California, 2000. 5, 6, 7, 36

M.J. Branney. Downsag and extension at calderas: new perspectives on collapse geometries from ice-melt, mining, and volcanic subsidence. *Bull. Volcanol.*, 57: 303–318, 1995. 62

E. Burov, C. Jaupart, and L. Guillou-Frottier. Ascent and emplacement of bouyant magma bodies in brittle-ductile upper crust. *Journal of Geophysical Research*, 108(B4) ECV 1, 2003. 91

M.L. Burt, G. Wadge, and W.A. Scott. Simple stochastic modelling of the eruption history of a basaltic volcano: Nyamuragira, Zaire. *Bull. Volcanol.*, 56: 87–156, 1994. 11, 95

S.N. Carey. Understanding the physical behaviour of volcanoes. In J. Martí, , and G. Ernst, editors, *Volcanoes and the Environment*, pages 1–54. Cambridge University Press, Cambridge, 2005. 68

R. Carniel, M. Tárraga, O. Jaquet, R. Ortiz, and A. García. The seismic noise at Las Cañadas volcanic caldera, Tenerife, Spain: Persistence characterization, and possible relationship with regional tectonic events. *Journal of Volcanology and Geothermal Research*, 173(1-2):157–164, 2008. ISSN 0377-0273. doi: 10.1016/j.jvolgeores.2007.12.044. 10, 37, 46

J.C. Carracedo. The canary islands: an example of structural control on the growth of large oceanic-island volcanoes. *Journal of Volcanology and Geothermal Research*, 60:225–241, 1994. 19

J.C. Carracedo, S. Day, H. Guillou, E. Rodríguez, J.A. Canas, and F.J. Pérez. Hotspot volcanism close to a passive continental margin. *Geological Magazine*, 135:591–604, 1998. 97

J.C. Carracedo, M. Paterne, H. Guillou, F. J. Pérez-Torrado, R. Paris, E. Rodríguez-Badiola, and A. Hansen. Dataciones radiométricas (14c y k/ar) del Teide y el rift noroeste, Tenerife, Islas Canarias. *Estudios Geológicos*, 59: 15–29, 2003. 17, 19, 21, 22, 23, 41

J.C. Carracedo, E. Rodríguez-Badiola, H. Guillou, M. Paterne, S. Scaillet, F. J. Pérez-Torrado, R. Paris, U. Fra-Paleo, and A. Hansen. Eruptive and structural history of Teide volcano and rift zones of Tenerife, Canary Islands. *Geological Society of America*, 119:1027–1051, 2007. 17, 18, 19, 21, 22, 23, 38, 40, 41, 55, 99, 112, 114

J.C. Carracedo, H. Guillou, S. Nomade, E. Rodríguez-Badiola, F.J. Pérez-Torrado, A. Rodríguez-González, R. Paris, V.R. Troll, S. Wiesmaier, A. Delcamp, and J.L. Fernández-Turiel. Evolution of ocean-island rifts: The northeast rift zone of Tenerife, Canary Islands. *Geological Society of America Bulletin*, 2010. doi: 10.1130/B30119.1. 99, 112

S. Carta, R. Figari, G. Sartoris, E. Sassi, and R. Scandone. A statistical model for Vesuvius and its volcanological implications. *Journal of volcanology and geothermal research*, 44:129–151, 1981. 11, 95

A. Castro, T. Gerya, A. García-Casco, J. Fernandez, J. Diaz-Alvarado, I. Moreno-Ventas, and I. Low. Melting relations of MORB-sediment melanges in underplated mantle wedge plumes; implications for the origin of cordilleran-type batholiths. *Journal of Petrology*, 51:1267–1295, 2010. 89, 90, 91

J.W. Cole, D.M. Milner, and K.D. Spinks. Calderas and caldera structures: a review. *Earth-Science Reviews*, 69:1–26, 2004. 63, 64

S. Coles, editor. *An Introduction to Statistical Modeling of Extreme Values*. Springer-Verlang, London, 2001. 96, 101, 109

S.G. Coles and R.S.J. Sparks. Extreme value methods for modelling historical series of large volcanic magnitudes. In H.M. Mader, S.G. Coles, Connor C.B., and Connor L.J., editors, *Statistics in Volcanology*, pages 47–56. Geological Society of London for IAVCEI, London, 2006. 11, 96

K.C. Condie, editor. *Plate tectonics and crustal evolution*. Pergamon Press, New York, 1993. 68, 70

C.B. Connor, A.R. McBirney, and C. Burlan. What is the probability of explosive eruption at a long-dormant volcano? In H.M. Mader, S.G. Coles, C.B. Connor, and L.J. Connor, editors, *Statistics in Volcanology*. Geological Society of London. Special Publication of IAVCEI,vol. 1, pp. 39-46, 2006. 17

R. Cooke and D. Solomatine. EXCALIBR - software package for expert data evaluation and fusion and reliability assessment. Report to the Commission of the European Communities, Delft, 1990. 30

N. Coppo, P.-A. Schnegg, W. Heise, P. Falco, and R. Costa. Multiple caldera collapses inferred from the shallow electrical resistivity signature of the Las Cañadas caldera, Tenerife, Canary Islands. *Journal of Volcanology and Geothermal Research*, 170(3-4):153 – 166, 2008. ISSN 0377-0273. doi: 10.1016/j.jvolgeores.2007.09.013. 46

D.R. Cox and V. Isham, editors. *Point Process*. Chapman and Hall, London, 1980. 108

D.R. Cox and P. Lewis, eds.: The statistical analysis of series of events, Methuen and Co., London, 1966. 104

A. R. Cruden. On the emplacement of tabular granites. *Journal of the Geological Society of London*, 155:853–862, 1998. 91

A. Cundari. Petrogenesis of leucite bearing lavas in the Roman volcanic region, italy: the Sabatini lavas. *Contributions to Mineralogy and Petrology*, 70:9–21, 1979. 65

S. Dartevelle, G. Ernst, J. Stix, and A. Bernard. Origin of the Mount Pinatubo climactic eruption cloud: Implications for volcanic hazards and atmospheric impacts. *Geological Society of America*, 30:663–666, 2002. 62

A.C. Davison and R.L. Smith. Models for exceedances over high thresholds. *Journal of the Royal Statistical Society*, 52(B):393–442, 1990. 101, 102, 109

R. De la Cruz-Reyna. Volcanic crises management. In H. Sigurdsson et al., editors, *Encyclopedia of Volcanoes*, pages 1215–1227. Academic Press, San Diego, California, 2000. 2

S. De la Cruz-Reyna. Poisson-distributed patterns of explosive eruptive activity. *BULLETIN OF VOLCANOLOGY*, 54:57–67, 1991. 11, 95

S. De la Cruz-Reyna. Random patterns of occurrence of explosive eruptions at Colima volcano, México. *Journal of volcanology and geothermal research*, 55: 51–58, 1993. 11, 95

F.A. Decandia, A. Lazzarotto, and R. Nicolich. The CROP 03 traverse: insights on post-collisional evolution of northern Apennines. *Memorie Societá Geologica Italiana*, 52:275–294, 1998. 65

D. Dickey and W. Fuller: Distribution of the Estimators for Autoregressive Time Series With a Unit Root, Journal of the American Statistical Association, 74, 427–431, 1979. 105

P. Di-Girolamo. Geotectonic setting of Miocene-Quaternary volcanism in and around the eastern Tyrrhenian sea border (Italy) as deduced from major element geochemistry. *Bulletin Volcanologique*, 41:229–250, 1978. 65

W.R. Dickinson. The Basin and Range Province as a Composite Extensional Domain. *International Geological Review*, 44:1–38, 2002. 65

C. Doglioni. A proposal of kinematic modeling for W-dipping subductions - possible applications to the Tyrrhenian-Apennines system. *Terra Nova*, 3:423–434, 1991. 65

T. H. Druitt and R. S. J. Sparks. On the formation of calderas during ignimbrite eruptions. *Nature*, 310:679–681, 1984. 62

Y. Dzierma and H. Wehrmann. Eruption time series statistically examined: Probabilities of future eruptions at Villarrica and Llaima Volcanoes, Southern Volcanic Zone, Chile. *Journal of Volcanology and Geothermal Research*, 193:82–92, 2010a. doi: 10.1016/j.jvolgeores.2010.03.009. 11, 96

Y. Dzierma and H. Wehrmann. Statistical eruption forecast for the chilean southern volcanic zone: typical probabilities of volcanic eruptions as baseline for possibly enhanced activity following the large 2010 Concepción earthquake. *Natural Hazards and Earth System Sciences*, 10:2093–2108, 2010b. doi: 10.5194/nhess-10-2093-2010. 11, 96

A. Folch and J. Martí. Geometrical and mechanical constraints on the formation of ring-fault calderas. *Earth and Planetary Science Letters*, 221:215–225, 2004. 63, 89, 92

A. García, J. Vila, R. Ortiz, R. Maciá, R. Sleeman, J.M. Marrero, N. Sánchez, M. Tárraga, and A.M. Correig. Monitoring the reawakening of Canary Islands Teide volcano. *EOS Trans. AGU*, 87:61–65, 2006. 17

N. Geshi. Asymmetric growth of collapsed caldera by oblique subsidence during the 2000 eruption of Miyakejima, Japan. *Earth and Planetary Science Letters*, 280:149–158, 2009. doi: 10.1016/j.epsl.2009.01.027. 87

A. Geyer and J. Martí. The new worldwide collapse caldera database (CCDB): A tool for studying and understanding caldera processes. *Journal of Volcanology and Geothermal Research*, 175:334–354, 2008. 63, 64, 65, 66, 67, 68, 88, 89, 90

A. Geyer, A. Folch, and J. Martí. Relationship between caldera collapse and magma chamber withdrawal: An experimental approach. *Journal of Volcanology and Geothermal Research*, 157:375–386, 2006. 63, 89

J. Gottsmann, G. Berrino, H. Rymer, and G. Williams-Jones. Hazard assessment during caldera unrest at the Campi Flegrei, Italy: a contribution from gravity-height gradients. *Earth and Planetary Science Letters*, 211(3-4):295–309, 2003. ISSN 0012-821X. doi: 10.1016/S0012-821X(03)00225-5. 46

J. Gottsmann, L. Wooller, J. Martí, J. Fernandez, A. Camacho, H. Rymer, and A. García. Gravity changes mark reawakening of spanish volcano after 100 years of dormancy. *Geophysical Research Letters*, 33, 2006. doi: 10.1029/2006GL027523. 17

J. Gottsmann, R. Carniel, N. Coppo, L. Wooller, S. Hautmann, and H. Rymer. Oscillations in hydrothermal systems as a source of periodic unrest at caldera volcanoes: Multiparameter insights from Nisyros, Greece. *Geophys. Res. Lett.*, 34, 2007. doi: 10.1029/2007GL029594. 10, 37, 46

A. Gudmundsson. Formation and mechanics of magma reservoirs in Iceland. *Royal Astronomical Society Geophysical Journal*, 91:27–41, 1987. 89

A. Gudmundsson. Formation of collapse calderas. *Geology*, 16:808–810, 1988. 62

A. Gudmundsson. Formation and development of normal-fault calderas and the initiation of large explosive eruptions. *Bulletin of Volcanology*, 60:160–171, 1998. 62, 63

A. Gudmundsson. Conceptual and numerical models of ring-fault formation. *J. Volcanol. Geotherm. Res.*, 164:142–160, 2007. 89

A. Gudmundsson, J. Martí, and E. Turon. Stress fields generating ring faults in volcanoes. *Geophysical Research Letters*, 24:1559–1562, 1997. 63

K. Hattori. High-sulfur magma, a product of fluid discharge from underlying mafic magma: Evidence from Mount Pinatubo, Philippines. *Geology*, 21:1083–1086, 1993. 63

C. Hawkesworth. Calc-alkaline magmatism, lithospheric thinning and extension in the Basin and Range. *Journal of Geophysical Research*, 100(B6):10271–10286, 1995. 65

D. P. Hill, D. Dzurisin, W. L. Ellsworth, E. T. Endo, D. L. Galloway, T. M. Gerlach, M. S. J. Johnston, J. Langbein, K. A. McGee, C. D. Miller, D. Oppenheimer, and M. L. Sorey. Response plan for volcano hazards in the Long Valley Caldera and Mono craters region California. Bull. 2185, U.S. Geol. Surv., 2001. 10, 37

C.H. Ho. Bayesian analysis of volcanic eruptions. *Journal of volcanology and geothermal research*, 43:91–98, 1990. 11, 95

C.H. Ho. Time trend analysis of basaltic volcanism for the Yucca mountain site. *Journal of volcanology and geothermal research*, 46:61–72, 1991a. 11, 95

C.H. Ho. Nonhomogeneous poisson model for volcanic eruptions. *Math. Geol.*, 23:167–173, 1991b. 106

C.H. Ho. Sensitivity in volcanic hazard assessment for the Yucca mountain high-level nuclear waste repository site: the model and the data. *Math. Geol.*, 27: 239–258, 1995. 106

C.H. Ho, E.I. Smith, and D.L. Keenan. Hazard area and probability of volcanic disruption of the proposed high-level radioactive waste repository at Yucca mountain, Nevada,USA. *Bull. Volcanol.*, 69:117–123, 2006. 11, 95

K. Hoernle and H.U. Schmincke. The role of partial melting in the 15-ma geochemical evolution of Gran Canaria: a blob model for the Canary hotspot. *Journal of Petrology*, 34:599–626, 1993. 97

K. Hoernle, Y.S. Zhang, and D. Graham. Seismic and geochemical evidence for large-scale mantle upwelling beneath the eastern Atlantic and western and central Europe. *Nature*, 374:34–39, 1995. 97

E. P. Holohan, V. R. Troll, T. R. Walter, S. Munn, S. McDonnell, and Z. K. Shipton. Elliptical calderas in active tectonic settings: an experimental approach. *Journal of Volcanology and Geothermal Research*, 144:119–136, 2005. 87

P.R. Hooper, D.G. Bailey, and G.A.M. Holder. Tertiary calc-alkaline magmatism associated with lithospheric extension in the Pacific Northwest. *Journal of Geophysical Research,*, 100(B6):10303–10319, 1995. 65

G.R. Hughes and G.A. Mahood. Tectonic controls on the nature of large silicic calderas in volcanic arcs. *Geology*, 36:627– 630, 2008. doi: 10.1130/G24796A.1. 63

H. Huppert and R. S. J. Sparks. The generation of granitic magmas by intrusion of basalt to continental crust. *Journal of Petrology*, 29:599–624, 1988. 89, 90

O. Jaquet and R. Carniel. Estimation of volcanic hazard using geostatistical models. In H. M. Mader et al., editors, *Statistics in Volcanology*, pages 89–103. IAVCEI Publications, Geological Society of London, 2006. 11, 96

O. Jaquet, S. Low, B. Martinelli, V. Dietrich, and D. Gilby. Estimation of volcanic hazards based on cox stochastic processes. *Physics and Chemistry of the Earth*, 25:571–579, 2000. doi: 10.1016/S1464-1895(00)00087-9. 11, 96

M.A. Jellinek and D.J. DePaolo. A model for the origin of large silicic magma chambers: precursors of caldera-forming eruptions. *Bulletin of Volcanology*, 65: 363–381, 2003. 87, 89, 90

C. M. Johnson, J. R. Shannon, and C. J. Fridrich. Roots of ignimbrite calderas: batholithic plutonism, volcanism, and mineralization in the southern Rocky Mountains, Colorado and New Mexico. *New Mexico Bureau of Mines and Mineral Resources Memory*, 46:275–302, 1989. 90

P. Kearey and F.J. Vine, editors. *Global tectonics*. Blackwell Science, Oxford, 1996. 68, 70

J.V.A. Keller, G. Minelli, and G. Pialli. Anatomy of late orogenic extension: the northern Apennines case. *Tectonophysics,*, 238:275–294, 1994. 65

B. Kennedy, J. Stix, J.W. Vallance, Y. Lavallée, and M.-A. Longpre. Controls on caldera structure: results from analogue sandbox modeling. *GSA Bull.*, 116: 515–524, 2004. 63

C.R.J. Kilburn. Multiscale fracturing as a key to forecasting volcanic eruptions. *Journal of Volcanology and Geothermal Research*, 125(3-4):271–289, 2003. doi: 10.1016/S0377-0273(03)00117-3. 10, 37

F.W. Klein. Patterns of historical eruptions at hawaiian volcanoes. *Journal of volcanology and geothermal research*, 12:1–35, 1982. 11, 95

Y. Lavallée, J. Stix, B. Kennedy, M. Richer, and M.-A. Longpre. Caldera subsidence in areas of variable topographic relief: results from analogue modeling. *J. Volcanol. Geotherm. Res.*, 129:219–236, 2004. 63

M.J. Le Bas, D.C. Rex, and C.J. Stillman. The early magmatic chronology of Fuerteventura, Canary Islands. *Geological Magazine*, 123:287–298, 1986. 96

X. Le Pichon and P.J. Fox. Marginal offsets, fracture zones and the early opening of the North Atlantic. *Journal of Geophysical Research*, 76:6294–6308, 1971. 97

Lin, X.: Statistical modelling of severe wind gust, International Congress on Modelling and Simulation, Townsville 14-17 July, 2,, 620–625, 2003. 109

P. W. Lipman. The roots of ash flow calderas in western North America: Windows into the tops of granitic batholiths. *Journal of Geophysical Research*, 89(B10): 8801–8841, 1984. 62, 90

P. W. Lipman. Evolution of silicic magma in the upper crust. the mid-Tertiary Latir Volcanic field and its cogenetic granitic batholith, northern New Mexico, USA. *Transactions of the Royal Society of Edinburgh*, 79:265–288, 1988. 90

P. W. Lipman. Subsidence of ash-flow calderas: relation to caldera size and magma-chamber geometry. *Bulletin of Volcanology*, 59:198–218, 1997. 62, 63, 64, 87

P. W. Lipman. Calderas. In H. Sigurdsson et al., editors, *Encyclopedia of Volcanoes*, pages 643–662. Academic Press, San Diego, California, 2000. 62, 63

L.B. Marinoni and G. Pasquaré. Tectonic evolution of the emergent part of a volcanic ocean island: Lanzarote, Canary Islands. *Tectonophysics*, 239:111–135, 1994. 96

J. Martí and G. Ernst, editors. *Volcanoes and the Environment*. Cambridge University Press, Cambridge, 2005. 7

J. Martí and A. Geyer. Central vs flank eruptions at Teide-Pico Viejo twin strato-volcanoes (Tenerife, Canary Islands). *Journal of Volcanology and Geothermal Research*, 181(1-2):47–60, 2009. ISSN 0377-0273. doi: 10.1016/j.jvolgeores.2008.12.010. 10, 41, 47

J. Martí and A. Gudmundsson. The Las Cañadas caldera (Tenerife, Canary Islands): an overlapping collapse caldera generated by magma-chamber migration. *Journal of Volcanology and Geothermal Research*, 103(1-4):161–173, 2000. ISSN 0377-0273. doi: 10.1016/S0377-0273(00)00221-3. 18

J. Martí, G. J. Ablay, L. T. Redshaw, and R. S. J. Sparks. Experimental studies of collapse calderas. *Journal Geological Society London*, 151:919–929, 1994a. 41, 62, 64, 89

J. Martí, J. Mitjavila, and V. Ara na. Stratigraphy, structure and geochronology of the Las Cañadas caldera (Tenerife, Canary Islands). *Geol. Mag.*, 131:715–727, 1994b. 18, 92

J. Martí, M. Hurlimann, G.J. Ablay, and A. Gudmundsson. Vertical and lateral collapses on tenerife (canary islands) and other volcanic ocean islands. *Geology*, 25:879882, 1997. 18

J. Martí, A. Folch, G. Macedonio, and A. Neri. Pressure evolution during caldera forming eruptions. *Earth and Planetary Science Letters*, 175:275–287, 2000. 90

J. Martí, A. Geyer, and A. Felpeto. Susceptibility analysis of phonolitic eruptions at Teide volcano (Canary Islands). *Geophysical Research Abstracts*, 8:40–47, 2006. 23

J. Martí, W.P. Aspinall, R. Sobradelo, A. Felpeto, A. Geyer, R. Ortiz, P. Baxter, P. Cole, J. Pacheco, M.J. Blanco, and C. Lopez. A long-term volcanic hazard event tree for Teide-Pico Viejo stratovolcanoes (Tenerife, Canary Islands). *Journal of Volcanology and Geothermal Research*, 178(3):543 – 552, 2008a. ISSN 0377-0273. doi: 10.1016/j.jvolgeores.2008.09.023. 36, 39, 43, 44, 48, 58, 59, 60

J. Martí, A. Geyer, J. Andújar, F. Teixidó, and F. Costa. Assessing the potential for future explosive activity from Teide-Pico Viejo stratovolcanoes (Tenerife, Canary Islands). *Journal of Volcanology and Geothermal Research*, 178(3): 529–542, 2008b. ISSN 0377-0273. doi: 10.1016/j.jvolgeores.2008.07.011. Evaluating Explosive Eruption Risk at European Volcanoes - Contribution from the EXPLORIS Project. 18, 19, 21, 22, 24, 25, 41, 42, 47, 49

J. Martí, A. Geyer, A. Folch, and J. Gottsmann. A review on collapse caldera modelling. In J. Gottsmann and J. Martí, editors, *Caldera Volcanism: Analysis, Modelling and Response, Development in Volcanology*, pages 233–284. Elsevier, Amsterdam, 2008c. 63, 64, 72, 87

J. Martí, R. Ortiz, J. Gottsmann, A. García, and S. De-La-Cruz-Reyna. Characterising unrest during the reawakening of the central volcanic complex on Tenerife, Canary Islands, 2004-2005, and implications for assessing hazards and risk mitigation. *Journal of Volcanology and Geothermal Research*, 182(1-2):23 – 33, 2009. ISSN 0377-0273. doi: 10.1016/j.jvolgeores.2009.01.028. 17, 38, 39, 46

W. Marzocchi, L. Sandri, P. Gasparini, C. Newhall, and E. Boschi. Quantifying probabilities of volcanic events: The example of volcanic hazard at Mount Vesuvius. *J. Geophys. Res.*, 109, 2004. doi: 10.1029/2004JB003155. 9, 10, 16, 21, 22, 25, 34, 36, 37, 43, 51

W. Marzocchi, L. Sandri, and C. Furlan. A quantitative model for volcanic hazard assessment. In H. M. Mader et al., editors, *Statistics in Volcanology*. Special Publication of IAVCEI, Geological Society of London, 2006. 9, 10, 16, 21, 25, 34, 36, 37, 43

W. Marzocchi, L. Sandri, and J. Selva. BET_EF: a probabilistic tool for long- and short-term eruption forecasting. *Bulletin of Volcanology*, 70(5):623–632, 2007. doi: 10.1007/s00445-007-0157-y. 36, 37, 43, 51, 52, 57, 58, 59

W. Marzocchi, L. Sandri, and J. Selva. BET_EF: a probabilistic tool for long- and shortterm eruption forecasting. *Bulleting of Volcanology*, 70:623–632, 2008. 9, 10, 11, 95

McNeil, A.: Estimating the tails of loss severity distributions using extreme value theory., ASTIN Bulletin, 27, 117–137, 1997. 113

A. T. Mendoza-Rosas and S. De la Cruz-Reyna. A mixture of exponentials distribution for a simple and precise assessment of the volcanic hazard. *Natural Hazards and Earth System Sciences*, 9:425–431, 2009. 11, 96

A. T. Mendoza-Rosas and S. De la Cruz-Reyna. Hazard estimates for El Chichón volcano, Chiapas, México: a statistical approach for complex eruptive histories. *Natural Hazards and Earth System Sciences*, 10:1159–1170, 2010. 11, 96

A.T. Mendoza-Rosas and S. De la Cruz-Reyna. A statistical method linking geological and historical eruption time series for volcanic hazard estimations : Applications to active polygenetic volcanoes. *Journal of volcanology and geothermal research*, 176(2):277–290, 2008. ISSN 0377-0273. 11, 96, 108, 113

F. Mulargia, P. Gasperini, and S. Tinti. Identifying different regimes in eruptive activity: an application to Etna volcano. *Journal of volcanology and geothermal research*, 34:89–106, 1987. 11, 95

A. Neri, W. P. Aspinall, R. Cioni, A. Bertagnini, P. J. Baxter, G. Zuccaro, D. Andronico, S. Barsotti, P.D. Cole, T. Esposti-Ongaro, T. K. Hincks, G. Macedonio, P. Papale, M. Rosi, R. Santacroce, and G. Woo. Developing an event tree for probabilistic hazard and risk assessment at Vesuvius. *Journal of Volcanology and Geothermal Research*, 178(3):397–415, 2008. doi: 10.1016/j.jvolgeores.2008.05.014. Evaluating Explosive Eruption Risk at European Volcanoes - Contribution from the EXPLORIS Project. 9, 21, 22, 25, 32, 34, 36, 58

C. G. Newhall and D. Dzurisin. Historical unrest at large calderas of the world. *U.S. Geological Survey Bulletin*, 1855:1–1108, 1988. 62, 64

C. G. Newhall and R. P. Hoblitt. Constructing event trees for volcanic crisis. *Bull. Volcanol.*, 64:3–20, 2002. 4, 5, 9, 10, 11, 16, 21, 22, 25, 34, 36, 37, 43, 58, 95

C.G. Newhall and R.S. Punungbayan, editors. *Fire and mud: Eruptions and Lahars of Mount Pinatubo, Philippines*. PHIVOLCS. Quezon City and University of Washington Press, Seattle, 1996. 17

C.G. Newhall and S. Self. The volcanic explosivity index (VEI): an estimate of the explosive magnitude for historical eruptions. *Journal of Geophysical Research*, 87:1231–1238, 1982. 113

A. Peccerillo. Roman magmatic provice (central Italy): Evidence for subduction related to magma genesis. *Geology*, 13:103–106, 1985. 65

N. M. Pérez, N. C. Sturchio, G. Arehart, P. A. Hernández, J. M. Salazar, and H. Wakita. Short-term secular variations of carbon and radon isotopes of fumarolic discharges from Teide volcano, Tenerife, Canary Islands. *Bull. Earthquake Chem.*, 7:31–33, 1996. 46

N. Petford and J. D. Clemens. Granites are not diapiric! *Geology Today*, September-October:180–184, 2000. 91

N. Petford, R. C. Kerr, and J. R. Lister. Dike transport of granitoid magmas. *Geology*, 21(9):845–848, 1993. 91

N. Petford, A. R. Cruden, K. J. W. McCaffrey, and J. L. Vigneresse. Granite magma formation, transport and emplacement in the Earth's crust. *Nature*, 408:669–673, 2000. 91

A. K. Petraske, D. S. Hodges, and R. Shaw. Mechanics of emplacement of basic intrusions. *Tectonophysics*, 46(1-2):41–63, 1978. doi: 10.1016/0040-1951(78)90104-X. 91

D.M. Pyle. Forecasting sizes and repose times of future extreme volcanic events. *Geology*, 26:367–370, 1998. 11, 95

D.M. Pyle. Sizes of volcanic eruptions. In H. Sigurdsson, editor, *Encyclopedia of Volcanoes*, pages 263–269. Academic, San Diego, California, 2000. 100

R.A. Reyment. Statistical analysis of some volcanologic data. regarded as series of point events. *PAGEOPH*, 74(3):57–77, 1969. 11, 95

J. A. Rice, editor. *Mathematical Statistics and Data Analysis, 2nd Edition*. University of California, Berkeley, 1995. 72, 73, 74

A.H.F. Robertson and C.J. Stillman. Submarine volcanic and associated sedimentary rocks of the Fuerteventura Basal Complex, Canary Islands. *Geological Magazine*, 116:203–214, 1979. 97

O. Roche and T. H. Druitt. Onset of caldera collapse during ignimbrite eruptions. *Earth and Planetary Science Letters*, 191:191–202, 2001. 64, 89

O. Roche, T. H. Druitt, and O. Merle. Experimental study of caldera formation. *Journal of Geophysical Research*, 105:395–416, 2000. 64

C. Romero, editor. *Las manifestaciones volcánicas históricas del archipiélago canario*. Consejería de política territorial. Gobierno autónomo de Canarias. 2 tomos, Santa Cruz de Tenerife, 1991. 94, 95, 99, 103

L. Sandri, W. Marzocchi, and L. Zaccarelli. A new perspective in identifying the precursory patterns of eruptions. *Bulletin of Volcanology*, 66:263–275., 2004. 17

H.U. Schmincke. Volcanic and chemical evolution of the Canary Islands. In U. von Rad, K. Hinz, M. Sarnthein, and E. Seibold, editors, *Geology of the Northwest African Margin*, pages 273–306. Springer, Berlin, 1982. 97

W. R. Seager and M. McCurry. The cogenetic organ cauldron and batholith, South Central New Mexico: Evolution of a large volume ash flow cauldron and its source magma chamber. *Journal of Geophysical Research*, 93(B5):4421–4433, 1988. 90

S. Self and M. R. Rampino. The 1883 eruption of Krakatau. *Nature*, 294:699–704, 1981. 62

H. Sigurdsson and R. S. J. Sparks. Rifting episode in North Iceland in 1874-1875 and the eruptions of Askja and Seinagja. *Bulletin of Volcanology*, 41(3):1–18, 1978.

H. Sigurdsson and R. S. J. Sparks. Petrology of rhyolitic and mixed magma ejecta from the 1875 eruption of Askja, Iceland. *Journal of Petrology*, 22(1):41–84, 1980.

T. Simkin, editor. *Krakatau 1883: The Volcanic Eruption and its Effects*. Smithsonian Institute Press, Washington, DC, 1983. 62

R.L. Smith. Ash-flow magmatism. *Special Paper - Geological Society of America*, 180:5–27, 1979. 62, 63, 64, 89

R.L. Smith and R.A. Bailey. Resurgent cauldrons. *Geol. Soc. Am. Memories*, 116:613–622, 1968. 62

R. Sobradelo and J. Martí. Bayesian event tree for long-term volcanic hazard assessment: Application to Teide - Pico Viejo stratovolcanoes, Tenerife, Canary Islands. *Journal of Geophysical Research*, 115:B05206, 2010. doi: 10.1029/2009JB006566. 95

R. Sobradelo, A. Geyer, and J. Martí. Statistical data analysis of the CCDB (Collapse Caldera Database): Insights on the formation of caldera systems. *Journal of Volcanology and Geothermal Research*, 198:241–252, 2010.

R. Sobradelo, J. Martí, A.T. Mendoza-Rosas, and G. Gómez. Volcanic hazard assessment for the Canary Islands (Spain) using Extreme value theory. *Natural Hazards and Earth System Sciences*, submitted 2011.

A.R. Solow. An empirical bayes analysis of volcanic eruptions. *Math. Geol.*, 33 (1):95–102, 2001. 11, 95

F. Spera and J. A. Crisp. Eruption volume, periodicity, and caldera area: Relationships and inferences on development of compositional zonation in silicic

magma chambers. *Journal of Volcanology and Geothermal Research*, 11:169–187, 1981. 63, 64

D. K. Spinks, V. Acocella, J. W. Cole, and K. N. Bassett. Structural control of volcanism and caldera development in the transtensional Taupo Volcanic Zone, New Zealand. *Journal of Volcanology and Geothermal Research*, 144(1-4):7–22, 2005. ISSN 0377-0273. doi: 10.1016/j.jvolgeores.2004.11.014. 87

F. Stoppa and G. Lavecchia. Late-Pleistocene ultra-alkaline magmatic activity in the Umbria-Latium region: an overview. *Journal of Volcanology and Geothermal Research*, 52:277–293, 1992. 65

A. Szakács. Redefining active volcanoes: a discussion. *Bulletin of Volcanology*, 56:321–325., 1994. 17

M. Tárraga, R. Carniel, R. Ortiz, J. M. Marrero, and A. García. On the predictability of volcano-tectonic events by low frequency seismic noise analysis at Teide-Pico Viejo volcanic complex, Canary Islands. *Natural Hazards and Earth Systems Science*, 6:365–376, 2006. 10, 37, 46

R.N. Thompson. Primary basalts and magma genesis. III Alban Hills, Roman comagmatic province, central Italy. *Contributions to Mineralogy and Petrology*, 60:91–108, 1977. 65

D.L. Turcotte and G. Schubert, editors. *Geodynamics*. Cambridge University Press, Cambridge, 2002. 68

M.B. Turner, S.J. Cronin, M.S. Bebbington, and T. Platz. Developing probabilistic eruption forecasts for dormant volcanoes: a case study from Mt Taranaki, New Zealand. *Bulleting of Volcanology*, 70:507–515, 2007. doi: 10.1007/s00445-007-0151-4. 11, 95

UNESCO. Report of consultative meeting of experts on statistical study of natural hazards and their consequences. Technical Report SC/WS/500, New York, 1972. 93

S. Uyeda. Subduction zones: an introduction to comparative subductology. *Tectonophysics*, 81:133–159, 1982. 68

G. P. L. Walker. Downsag calderas, ring faults, caldera sizes, and incremental growth. *Journal of Geophysical Research*, 89:8407–8416, 1984. 62, 64

T. Watanabe, T. Koyaguchi, and T. Seno. Tectonic stress controls on ascent and emplacement of magmas. *Journal of Volcanology and Geothermal Research,*, 91:65–78, 1999. 91

S. M. Wickham. The segregation and emplacement of granitic magmas. *Journal of the Geological Society of London*, 144:281–297, 1987. 91

F. E. Wickman. Markov models of repose-period patterns of volcanoes. In DF Merriam, editor, *Random Processes in Geology*, pages 135–161. Springer-Verlag, Berlin, 1976. 11, 95

H. Williams. Calderas and their origin. *Bulletin of the Department of Geological Sciences*, 25(6):239–346, 1941. 62

H. Williams and A.R. McBirney, editors. *Volcanology*. Freeman, Cooper and Co., San Francisco, California,, 1979. 62

G. Woo, editor. *The mathematics of natural catastrophes*. Imperial College Press, London, 1999. 4, 9, 38, 52

Z.L. Yang and M. Xie. Efficient estimation of the Weibull shape parameter based on a modified profile likelihood. *Journal of Statistical Computation and Simulation*, 73:115–123, 2003. 107