Statistical methods for assembling and incorporating volcanic records in hazard estimation

A thesis presented in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

in

Statistics

at Massey University, Palmerston North, New Zealand.

Rebecca Green

2015

Abstract

The estimation of hazard arising from volcanic eruptions is a research topic of great interest to New Zealand, given the number and location of active and dormant volcanoes. Probabilistic temporal models are required to handle the stochastic nature of observed records. Such models are usually assembled using point process techniques or renewal theory and most are purely temporal in the sense that they only consider the distribution of event or inter-event times as predictors of further volcanic activity. I demonstrate using a high-resolution eruption record from Mt Taranaki (New Zealand) how geochemical data can be incorporated, using a proportional hazards type approach, to improve the performance of current renewal-type models.

Probabilistic forecasting relies on the accuracy and completeness of historical eruption records. This poses the question of how to establish a detailed record of past volcanic events. Multiple sites are needed to build a composite tephra record, but correctly merging them by recognizing events in common and site-specific gaps remains complex. I present an automated procedure for matching tephra sequences, using stochastic local optimization techniques. Implausible matches are eliminated through careful reasoning, while heuristically searching over the remaining alternatives. Possible matches are verified using known tephra compositions and stratigraphic constraints. The method is applied to match tephra records from five long sediment cores in Auckland, New Zealand. The correlated record compiled is statistically more likely than previously published arrangements from this area.

In addition to the matching of tephras found in the Auckland region, the algorithm is applied to stratigraphic records obtained from Mt Taranaki. With more detailed geochemical information available, matches are constrained further by considering principle component analysis of titanomagnetite compositional data.

Finally, after combining the amalgamated record of Mt Taranaki events with point thickness measurements, the eruptive volume of Mt Taranaki events is estimated. Utilizing isopach maps and individual point observations a model is formulated, in a Bayesian framework, for the thicknesses of tephra deposits as a function of the distance and angular direction of each location. The model estimates, in addition to eruptive volume, the wind and site-specific effects on the thickness deposits. The findings lead on to methods of incorporating eruptive volumes in hazard estimation.

Mt Taranaki. View from the North Egmont Visitor Center, April 2012.

Acknowledgements

First and foremost, I would like to give my sincere thanks to my primary supervisor Professor Mark Bebbington for his invaluable guidance and support during this entire research process. Without his time and care this PhD would not have been as successful. I started this PhD journey with limited knowledge of volcanology and Mark has been exceedingly patient as I explore this new field. I have enjoyed working under his supervision. His confidence in my ability and encouragement throughout the whole PhD process has made this experience very rewarding.

I am also very grateful to my co-supervisors Professor Shane Cronin and Associate Professor Geoff Jones. Shane has helped with interpreting many exciting results and has shown interest in my work even during the more 'statistical' supervision meetings. Shane has greatly assisted in developing my understanding of volcanology, for which I am very thankful. I also wish to express my thanks to Geoff, whose input has strengthened many aspects of this research. His attention to detail and his critical eye has enhanced the finer details of this PhD while his statistical expertise has resulted in many improvements to the more theoretical aspects of this research. As Einstein once said "it is the supreme art of the teacher to awaken joy in creative expression and knowledge." Mark, Shane and Geoff have succeeded in creating an intellectually stimulating environment, and it has been a very enriching experience working with such a spectacular supervisory team.

I want to acknowledge all those who have provided insightful comments on my research or otherwise helped me throughout my PhD. I recognize that this research would not have been possible without the financial assistance of a Massey University Doctoral Scholarship. I also give thanks to the postgraduate statistics students who have accompanied me throughout this journey; their friendships have been invaluable. My special thanks goes to Emily Kawabata. We have had many fruitful discussions which I have benefitted from. It has been fun discovering statistical volcanology with Emily, and I have enjoyed being able to exchange knowledge while curiously awaiting results together. I would like to give special mention to Dr Jonathan Godfrey whose support with many aspects of university has been a great help. I have gained a lot from his enthusiasm and willingness to teach. His approachability and encouragement has seen him act as another informal supervisor and mentor.

I would also like to give a very special thanks to my partner Leigh whose hope, optimism and humor have been a virtue throughout this roller coaster of a PhD. Lastly, I give the greatest

thanks to my family whose love and support has known no bounds. To my sister Kim whose encouragement and company has been an immense help; and to my parents whose lifelong support and advice has helped me to achieve so many of my goals. xx

Publications arising from this thesis

- **Green, R.**, Bebbington, M., Cronin, S., and Jones, G. (2013). Geochemical precursors for eruption repose length. *Geophysical Journal International*, 193:855–873.
- **Green, R.**, Bebbington, M., Cronin, S., and Jones, G. (2014). Automated statistical matching of multiple tephra records exemplified using five long maar sequences younger than 75 ka, Auckland, New Zealand. *Quaternary Research*, 82(2):405–419.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This chapter aims to describe the role statistics plays in volcanology, as well as establishing the motivation for the research which follows. An overview of the objectives and thesis structure is also provided.

1.1 Motivation

The Smithsonian Institution's Global Volcanism Program (Siebert and Simkin, 2002-) maintains a database of current and past activity for ∼1,500 volcanoes that have been active during the last 10,000 years. An estimated 12% of the worlds population live within a 100 km radius of these Holocene active volcanoes (Small and Naumann, 2001), and that number is expected to increase with sustained population growth and rapid urbanization (Sumner, 2014). The impact from a volcanic eruption can have a catastrophic effect on society. The 1902 eruption of Mt Pelée, one of the most lethal in human history, resulted in the death of over 28,600 people (Tanguy et al., 1998) which wiped out the entire population of St Pierre. Closer to home, the 1953 Mt Ruapehu Tangiwai disaster saw 151 people lose their lives, as the result of volcanic hazard in the form of a lahar (Johnston et al., 2000). In the age of international air travel even those not living within the vicinity of the volcano are at risk. One only needs to recall the April 2010 eruption of the Icelandic volcano, Eyjafjallajökull, to understand the economic impact volcanic eruptions have on civil aviation. More than 8.5 million passengers were stranded and economic damage was estimated to be somewhere between 1.7 and 3.3 billion euros (Alexander, 2013).

Many volcanoes experience long periods of quiescence which can last for centuries. For that reason, they may not be viewed as an immediate threat by communities. However, the possibility of the volcano re-awakening still poses a serious risk. The 1980 Mt St Helens eruption and the 1991 Mt Pinatubo (Philippines) eruption are just two examples of catastrophic events that occurred, despite the volcanoes not having erupted for more than 120 and 500 years respectively (Siebert and Simkin, 2002-). Mt Taranaki (New Zealand), the volcano around which the majority of this thesis is based, is another typical example of a possible re-awakening volcano. Although it has not erupted since the mid-1800s, geological records suggest it has been frequently active in the past.

With more people settling around volcanoes, it is becoming increasingly necessary to develop mitigation strategies and procedures to reduce some of the effects of volcanic eruptions. This is where statistics plays its part in volcanology. Volcanoes are complex systems, therefore large uncertainties are associated with volcanic hazards. Probabilistic approaches are required to answer questions such as; *when*, *where*, and *how big* the next eruption will be. In order to establish long-term forecasts, data on past volcanic events is required and statistical models must be developed.

The advent in the 1950s of radiometric dating techniques, to estimate the ages of volcanic deposits, has meant that volcanologists are now able to delve further into past eruption histories than was previously possible with historical (observed) records. While there are catalogs of events for a large number of volcanoes (see, Siebert and Simkin, 2002-), it is common for large eruptions to be over-represented. While these events likely caused the most damage, an understanding of the smaller eruptions is just as important, especially in determining the true eruption frequency of past events and possible temporal patterns in the data.

There are different methods available to reconstruct the timing of events, depending on the type of eruption (i.e effusive or explosive), each of which involves a different set of challenges. Effusive eruptions are characterized by lava flows, whereas explosive eruptions involve the dispersal of ash and pyroclastic (volcanic) material high into the atmosphere. Lava flows are much easier to find, and correlate, than tephra (volcanic ash deposits). However, the problem with lava is that new flows obscure the evidence of older events. Challenges involve working out what lies beneath recent lava flows, and establishing the ages of events. Even the record for Mt Etna (Sicily, Italy), considered to be one of the more detailed volcanological records, contains large age uncertainties. Tanguy et al. (2009) found that most of the historically

dated Mt Etna lavas were several centuries older than believed, with discrepancies sometimes exceeding a millennium.

In contrast to lava flow events, this thesis focuses around explosive Mt Taranaki eruptions. In order to reconstruct the timing of small (and large) explosive events, long sediment cores that preserve volcanic deposits have been extracted from lakes or swamps near the volcano. Multiple sites are needed to construct accurate records and this involves many complicated factors. To give an example; if we have multiple cores, or combined data with 'near-source' data, these can contain events in common, or events not represented in the other cores. Furthermore, these events may or may not be from a particular volcano. There is also the question of events missing from the record altogether. The latter may be due to geography or the small size of the event. Uncertainty in age estimates, and the nonlinear conversion from radiocarbon years to calendar years, further clouds the water.

From a statistical perspective, the problem presents itself as a combinatorial optimization problem. Given the age and error associated with each volcanic deposit, and any number of cores extracted from different locations, there could be a substantial number of possible matches. An exhaustive search for the most likely representation of events across the various locations would be infeasible. Instead, stochastic local search algorithms could be employed to find the most likely representation of events observed across the different locations. The volcano's past behavior could be reconstructed, and records suitable for supporting hazard models could be compiled.

Quantitative approaches to forecasting the likelihood of future events may involve the timing, size/volume, location or style of eruptions. It is the first feature (timing), that the first part of this thesis is focused around. As volcanological data is inherently stochastic, a probabilistic temporal model is required to handle the imprecisions and inaccuracies inherent in age data. Such models are usually assembled using techniques from renewal theory, (marked) point processes, or hidden (semi-) Markov models (see Section 2.3). However, most models are purely temporal in that they only use information from the eruption onset times. In the past, there has been little data available beyond the catalog of eruption times, locations, and durations. However, recent improvements in technology and procedures have resulted in the acquisition of greater quantities of high quality ancillary data. This data relating to eruptions, could be incorporated into hazard estimation. For example, it has been hypothesized that there are cycles in magma geochemistry that could be linked to eruption rate (Turner et al., 2008b).

This type of geochemical data has never been used in probabilistic volcanic hazard forecasting. From a statistical point of view, a proportional hazards type modeling approach is very appealing. This would involve incorporating ancillary geochemical data into the hazard by attaching a vector of covariates to an underlying temporal baseline hazard function. A suitably well-fitting model, describing how the hazard varies in response to explanatory geochemical information, can provide a probabilistic forecast of future eruptions. Model selection can also reveal much about the geochemical dynamics of a volcano. It is hoped that the ability to understand the evolution of an eruption may lead to the provision of more accurate and detailed advice to emergency managers and, particularly, land use planners. Volcanic eruptions present significant risks not only to people, but surrounding infrastructure. Better insight into volcanic processes and their potential consequences will enable more effective structural and land use initiatives to be developed to alleviate the effects of forthcoming eruptions.

As well as estimating the timing of future events volcanologists strive to understand the likely size of impending eruptions. There is some indication that the bigger an eruption is the longer you will wait for the next one (see Section 2.3.6). Or conversely, that longer repose intervals tend to be followed by larger eruptions (the latter relation being much less observed). Testing these hypotheses requires some knowledge of the size/eruptive volume of previous events. Although there are some well documented eruptions and certain well-studied volcanoes (e.g. Etna, Kilauea, Mauna Loa) that have estimated volumes, for the majority of events (particularly the small ones) volume estimates are few and far between. Current estimation methods rely on geological interpretations of thickness deposits and dispersal areas (and shapes) of previous eruptions. However, most volume estimation approaches are empirical and do not consider model or data uncertainty. This is where Bayesian statistics can lend a hand. Given the disparate nature of volcanological datasets, existing empirical (parametric) models for dispersal, uncertainties inherent in field observations, and the general scarcity of data, Bayesian statistics provides a suitable framework for estimating eruptive volumes. Tephra dispersal depends on (but is not limited to) the height of the eruption plume and the wind speed, strength, and direction during an eruption. Tephra deposited at different locations can be preserved under different environmental conditions, thus site-specific effects for over- or under-thickening are also present. A high-dimensional complex model is required. Model development in a Bayesian framework, enables knowledge of these phenomena governing tephra attenuation to be incorporated in a probabilistic way, through prior distributions. Well constrained larger events can be leveraged into site-specific effects and volumes, while properly considering sampling error.

1.2 Objectives and thesis structure

The intention of this study was to develop advanced statistical models and tools for use in probabilistic volcanic hazard forecasting. As a result, I hoped to provide new insights into the eruptive process which may lead to the provision of more accurate and detailed predictions. This study predominately focuses on New Zealand volcanoes, due to the availability of data and significant interest. However, the methods developed can be generalized to volcanic processes worldwide.

The three core objectives of this thesis were to:

- Formulate a probabilistic temporal hazard model for Mt Taranaki, incorporating ancillary geochemical data, with the hope of improving forecasts of future Mt Taranaki eruptions.
- Create an automated statistical procedure, through application to a series of long sediment core records, to match tephra sequences observed at multiple deposition sites. To stimulate a wider interest in the methods of the thesis, this was developed on higher profile data from the Auckland Volcanic Field. However, the primary application is to sediment cores surrounding Mt Taranaki.
- Develop a statistical model, informed by the physical processes that govern tephra dispersal, to estimate the eruptive volume of Mt Taranaki events.

This thesis has a further five chapters. Chapter 2 provides a review of some of the current statistical approaches to modeling eruption occurrence. It starts with an overview of the complex nature and limitations of volcanological data. Temporal hazard models and existing methods of estimating eruptive volumes are discussed. The terminology used throughout the thesis is introduced. In addition, I highlight shortcomings of current modeling approaches that are relevant to motivating and understanding the subsequent thesis.

In Chapter 3, using a high-resolution volcanic event record from Mt Taranaki, I demonstrate how geochemical data can be used to modulate a renewal model for the estimated probability of a future eruption. Mt Taranaki has an activity record punctuated by long periods of quiescence and subsequent re-awakening. Therefore the distribution of inter-onset times is bimodal, with the possibility of anomalously long reposes. It is shown that a bimodal renewal distribution is outperformed for eruption forecasting by a proportional hazards type model.

The latter model uses magma geochemistry as a proxy for the state of the magmatic system. The most notable improvement is found in the forecasting of long reposes. In a bimodal renewal model, the estimated hazard is only updated to reflect the likelihood of a long repose after the present repose-length exceeds the shorter mode. In the geochemistry modulated model proposed, this predictive information is available at the beginning of the repose to be forecast.

Chapter 4 presents a sophisticated automated procedure created, for correlating tephra (volcanic ash deposit) records from multiple locations. The matching algorithm is developed using stochastic local search techniques and is illustrated through application to five long sediment cores from the Auckland volcanic region (New Zealand). A statistical objective function is defined and then used to search for the most statistically likely arrangement of tephra across the different locations. The age and associated age error of each event forms the basis of the procedure. Matches are further constrained by considering the geochemical signature of each deposit, in addition to the order in which deposits were observed in the sediment cores. When compared with a previously published arrangement, the compiled record is statistically far more likely. This has implications for hazard models which will be discussed. A statistical model is also formulated to quantify the variability in observance probabilities of tephras at given maar craters (closed depressions with a low catchment area).

Chapter 5 is devoted to methods of estimating eruptive volumes. I begin by compiling an updated record of Mt Taranaki events using the statistical matching algorithm developed in Chapter 4. The matching algorithm is applied to a series of sediment cores extracted from distal and near-source unexposed locations. In merging the records, an additional matching constraint is introduced based on principal component analysis of magma compositional data. Utilizing isopach maps and individual point observations a model is formulated, in a Bayesian framework, for deposit thicknesses as a function of the distance and angular direction of each location. The model estimates, in addition to eruptive volumes, the wind and site-specific effects on deposit thicknesses. A series of sensitivity analyses are performed regarding the chosen model and the prior distributions for the model parameters. This is followed by a discussion of the implications for hazard forecasts and an initial investigation into predicting the size of Mt Taranaki eruptions.

Finally, Chapter 6 presents conclusions and suggestions for future research. This is followed by the bibliography, and then a series of appendices containing data files and Matlab code to accompany the analyses in each of the working chapters.

Chapter 2

Statistical volcanology background

The statistical modeling of eruption occurrence is a fundamental tool for developing a better understanding of the physics of eruptive processes, and to be able to make reliable forecasts (Marzocchi and Zaccarelli, 2006). Before making headway it is necessary to become familiar with the comprehensive range of literature available surrounding this topic.

2.1 Introduction

Quantitative forecasting of volcanic hazard is a main goal in volcanology. It is a subject of particular interest for communities, businesses, and governments that are potentially exposed. The occurrence of an eruption can have a catastrophic effect on society, posing a serious risk to the lives, property, environment, and economy of surrounding populations. Increasing urbanization results in higher levels of vulnerability and exposure to volcanic risk. As a result, there is a greater demand for disaster mitigation procedures (Marzocchi and Bebbington, 2012). Volcanic eruptions are inevitable but the ability to accurately forecast their occurrence can lead to considerable reductions in community risk.

Eruption forecasts are commonly divided into separate time scales: short-term forecasts, for time scales of hours to months; and long-term forecasts, for time scales of years to decades or longer. Short-term forecasts are prepared in response to real-time monitoring signals and related observations obtained during phases of volcanic unrest (see, e.g., Newhall and Hoblitt, 2002; Marzocchi et al., 2008). When a volcano becomes restless the past frequency of eruptions become less important than the warnings observed from the monitoring of physical

quantities (Marzocchi and Zaccarelli, 2006). Short-term forecasts aim to provide time for evacuating people and implementing related civil-defence health and safety measures. On the other hand, long-term forecasts are mostly established when the volcano in question is in a stationary state of quiet, where seismic activity and other monitored signals are at background levels. Long-term forecasts are established on the basis of past eruptive activity (Decker, 1986), with the objective of obtaining a better understanding of volcanic processes and their potential consequences. It is hoped that the ability to understand the evolution of a volcano may aid in the provision of more accurate and detailed forecasts that can be used for land-use planning and emergency management scenarios. As a result, many lives could be saved and countless property losses averted. The focus of this thesis is on *long-term* forecasts so any discussion of 'volcanic hazard' and 'eruption forecasts' will, hereafter, refer to this case.

Ideal models for forecasting eruption occurrence should consider the timing, location, size, and style (e.g. effusive or explosive) of potential eruptions. This study is restricted to polygenetic volcanoes, which are defined by having repeated eruptions from the same vent(s), as opposed to monogenetic volcanic fields which contain many volcanoes, each of which erupts only once. In contrast to polygenetic volcanoes, monogenetic volcanic fields require both spatial and temporal elements to forecast the location and timing of future events. Although some polygenetic volcanoes exhibit spatial structure via multiple vents or fissures, such as Mt Etna, attention is generally limited to the presence of flank eruptions (cf. Salvi et al., 2006) and lava flows. The applications in this thesis do not feature spatial aspects, the focus is solely on the timing of future events and estimates of eruptive volumes.

Volcanologists strive to quantify the likelihood and potential impact of future eruptions. Volcanic processes are inherently stochastic. The extreme complexity and limited knowledge of volcanic systems make deterministic approaches to forecasting practically impossible (Passarelli et al., 2010a). The indeterminacy in volcanic processes, and the large uncertainties associated with observed records, means probabilistic approaches are required. The randomness in the timings of eruptions can be quantified by modeling occurrences as a stochastic process.

This chapter aims to create the setting for the work which follows. Terminology used throughout the thesis is introduced here. I start with an overview of the complex nature and limitations of volcanological data, providing a brief explanation how eruptive records are compiled.

Some of the challenges faced when correlating records obtained from multiple sites are highlighted. A review of some of the statistical approaches for modeling eruption onsets and repose periods is given, which in turn, leads on to an introduction to methods used to estimate eruptive volumes.

2.2 Data

2.2.1 Defining an eruption

In order to avoid ambiguity, a clear understanding of what constitutes an eruption is required. Eruptions are not isolated point observations in time, each eruption episode can be a mix of both activity and quiescence. Eruptive periods can last from minutes, through to many months. Although the onset date of eruptions is generally well reported, eruption durations are much less frequently documented. Prehistoric eruption durations cannot be determined from geological records without very high precision dating methods (Connor et al., 2015). Difficulties in recognizing the gradual end of an eruption means that even historically observed eruptions often have end dates (and thus durations) unrecorded.

While there are a number of volcanoes that have duration data available, such as Etna and Kilauea, it is more commonly used in short-term forecasting (Bebbington, 2012). Since the focus of this thesis is on long-term forecasts of eruption *occurrence*, the usual convention of modeling eruptions as a point process of onsets is adopted (see, for example, Wickman, 1966; Klein, 1982; De la Cruz-Reyna, 1991; Bebbington and Lai, 1996a; Marzocchi and Zaccarelli, 2006). For a model that incorporates eruption durations, see Bebbington (2007).

Treating eruptions as point observations in time requires an arbitrary, but consistent, time limit of quiescence to be defined for separating successive eruptions. The Smithsonian Institution's Global Volcanism Program defines eruptive phases separated by at least three months of surface quiet, to be distinct events (Siebert and Simkin, 2002-).

2.2.2 Historical vs geological records

Most of the literature for hazard forecasting is based around the idea that a volcano's future activity is best predicted by its prior history. With this method, records of eruptions are studied and then projected to provide estimates of future activity (Decker, 1986). Statistical

models are required to handle the complex nature of volcanological data and to provide quantitative assessments of volcanic activity.

Eruption catalogs are compiled from historical observations and geological reconstructions of the eruptive history. Historical (or observed) records are usually short, spanning only a few centuries at most, with completeness dependent on the presence of observers and/or monitoring equipment. In contrast, geological records of pre-historic eruptions can span millennia. Chronological records of pre-historic records are constructed by identifying and dating volcanic deposits found on and around a volcano.

Both historical and geological records suffer from incompleteness and uncertainty. Some of the complexities of eruption records will be discussed shortly, but first a brief outline is given of some of the terminology required to understand how geological records are compiled.

2.2.3 Nomenclature

Tephra (derived from the Greek word $\tau\varepsilon\phi\rho\alpha$) is a collective term used to describe the fragmental material produced by volcanic eruptions. It was introduced by Sigurdur Thorarinsson in his doctoral thesis published in 1944 (Thorarinsson, 1944, 1981) as a general term for airborne volcanic fragments, such as 'ash', 'lapilli' or volcanic 'bombs/blocks', ejected during an eruption. Larger volcanic bombs/blocks fall to the ground quickest so are found closest to the vent, whereas smaller ash particles can travel hundreds to thousands of kilometers depending on eruption size (i.e. the eruptive column height), and the strength and direction of prevailing winds. Although 'tephra' is a collective noun which may be singular or plural, an 's' is commonly appended (forming 'tephras') for pluralization to avoid ambiguity (Froggatt and Lowe, 1990). Tephra dispersal from an eruption is relatively instantaneous with respect to geological time-scales. As a result, tephra layers (unless reworked, see below for explanation) will have the same age wherever they are deposited.

Each eruption deposits a layer of tephra to the ground surface. After a sufficient period of quiescence, new soils form. Subsequent eruptions conceal older tephras. Geological records can be established by extracting and analyzing sediment columns from unexposed locations, such as lakes or swamps, or by sampling from exposed locations, such as road cuttings and deep drains. Tephra layers appear sandwiched between organic material where older tephra are buried beneath younger tephra.

The relative position of each tephra layer within the sediment column is referred to as the *stratigraphic* position. Tephra layers can however, be *reworked* into different stratigraphic positions due to influences from surface processes, weather conditions, vegetational changes, and human or livestock interference. Postdepositional reworking can have adverse affects on age estimates if the geological record is misinterpreted.

Tephrochronology is the technique of identifying and correlating volcanic eruptions by analyzing discrete layers of tephra, to establish a chronological framework of geological and archaeological occurrences (Thorarinsson, 1981). The order, or *stratigraphy*, of tephra layers means the chronology can be transferred among deposition sites in order to correlate records. Undertaking tephrochronology requires *tephrostratigraphy*, which involves the characterization or 'fingerprinting' of tephras using physical, mineralogical, or geochemical properties (Lowe, 2011). Tephra layers found at multiple deposition sites can be correlated by considering the stratigraphic position, physical and chemical composition, and estimated ages of each tephra layer. A discussion of methods used for correlating tephras is given in Section 2.2.6.

Tephra studies play a pivotal role in reconstructing the history of Quaternary events throughout many parts of the world (Lowe, 2011). *Quaternary* events refer to those which occurred in the Quaternary Period; the geologic time period encompassing the most recent 2.6 million years. The Quaternary Period can be separated into two eras, the *Pleistocene* (2.6 million years ago to 11,700 years ago) and the *Holocene* (the latest 11,700 years). Tephra studies are being applied across a widening range of scientific disciplines and environments. As well as providing insight into volcanic processes, tephrochronology has been used to understand soil and environmental processes, paleoclimiatic changes (Sandiford et al., 2003), and human evolution (Tyron et al., 2009; Brown and McDougall, 2011). Projects such as INTIMATE (INTegration of Ice-core, MArine and TErrestrial records) have also emphasized the importance of developing a comprehensive history of tephras (Lowe et al., 2008).

Individual locations have incomplete records, primarily due to wind conditions at the time of eruption. A major focus of this thesis has been the development of a statistical procedure to compile a record of events by correlating a series of widely distributed sediment cores. In achieving this, many challenges were faced as a result of the complex nature of geological records. Uncertainties arise due to the variations in age estimates, incompleteness of records, imprecision in measurements, diversity of tephra preserving environments, and the general scarcity of data.

2.2.4 Age estimation

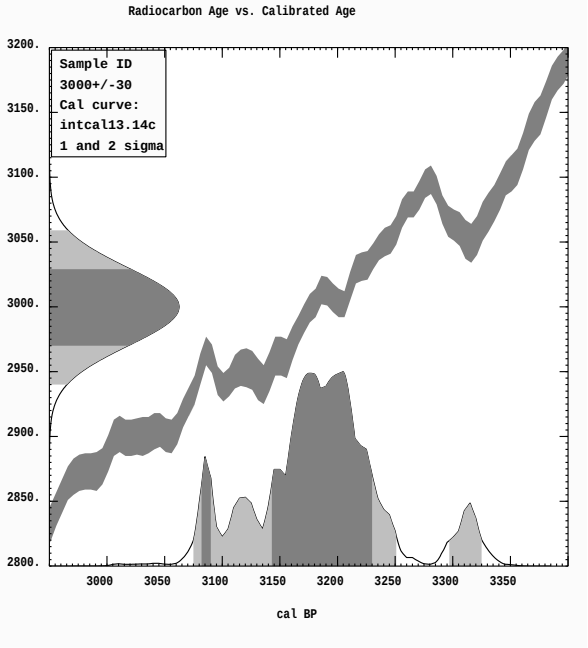
Estimated ages of tephra layers can be obtained through numerous dating methods. Where a vertical sequence of tephra deposits has been collected, tephra ages can be estimated by radiocarbon (C^{14}) dating of intervening organic (once-living) material. Radiocarbon (C^{14}) dating is feasible for organic material within the last 50,000 years. Beyond 50,000 years there are alternative dating techniques such as potassium-argon (K-Ar) and argon-argon (Ar-Ar). However, these are used to directly date lava and rocks, as opposed to intervening sediments. The precision of age estimates decreases further into the past.

Most of the prehistoric eruptions in this thesis were dated using the radiocarbon C^{14} dating method. The variable rate of radiocarbon concentrations over time requires C^{14} age estimates to be converted into calendar years. Radiocarbon C^{14} ages are reported in years 'BP', where 'BP' stands for 'before present' relative to the reference date of AD 1950. Calibrated/-calendar ages are reported as 'cal BP'. The conversion from radiocarbon C^{14} age estimates into calendar ages is complicated. C^{14} age estimates have associated normally distributed errors, but calibration into calendar years results in complex error distributions. There have been numerous calibration curves published over the last few decades, each with the aim of improving the accuracy and precision of previous calibration curves. The latest calibration curves are IntCal13 for the northern hemisphere (Reimer et al., 2013) and SHCal13 for the southern hemisphere (Hogg et al., 2013). Figure 2.1 illustrates two examples of the complex distributions obtained through the calibration from C^{14} to calendar ages.

Difficulties in finding suitable material for dating, in addition to the expense involved, means it is common to date only a portion of tephra samples (Bebbington, 2009). Remaining ages are subsequently obtained on the basis of their relative stratigraphic positions. Turner et al. (2008a) considered a spline fitting method, where dates are considered as a function of the depth at which the tephra layers are found in the core. Estimated mean ages and associated age errors, for each tephra layer, can be obtained through Monte-Carlo simulation of the normally distributed C^{14} ages ensuring the stratigraphy or order of events is maintained. Figure 2.2 shows the spline-fitting method of Turner et al. (2008a) for the age-depth relationship of a sediment core extracted from Lake Umutekai (New Zealand). Careful consideration of the errors associated with age estimates must be taken when modeling and forecasting eruptive activity.

Figure 2.1: Example calibrations of two C^{14} age estimates. Plots obtained using CALIB 7.0 (Stuiver et al., 2013) and the INTCAL13 calibration curve (Reimer et al., 2013). Vertical axes show the normally distributed C^{14} age estimates and horizontal axes show the calibrated/calendar age distributions.

(a) C^{14} age estimate 3000 ± 30 BP



(b) C^{14} age estimate 1000 ± 25 BP

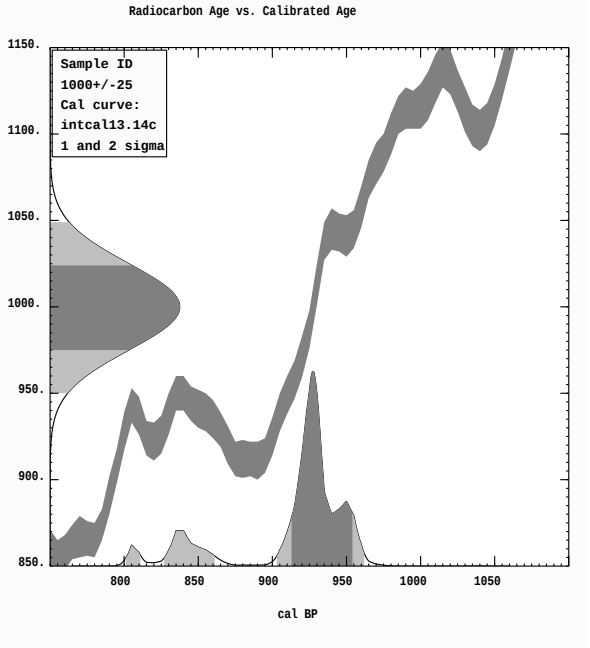
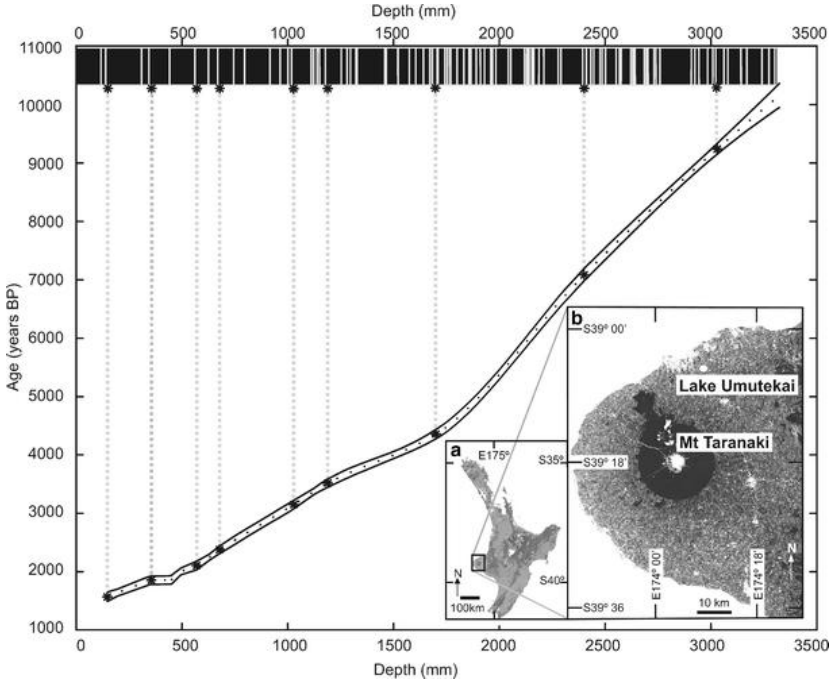


Figure 2.2: Plot of the radiocarbon age versus depth model (mm) for tephra layers in the sediment core from Lake Umutekai (inset location maps of: (a) North Island of New Zealand and (b) Taranaki). The core stratigraphy is displayed at the top of the plot with white-layers representing tephra deposits. The locations of radiocarbon dated layers are indicated by stars. Reprinted from Turner et al. (2008a), with kind permission from Springer Science and Business Media



2.2.5 Data complexities

Historical (observed) records of events are both short and incomplete as they depend on the presence of observers and the development of written records. Over the last few centuries there has been an increase in the frequency of observed eruptions. This is an apparent result of the global population increase and improvements in communication technologies. However, there are also noticeable drops in records of eruption onsets around the time of both world wars (Simkin, 1994), when documenting volcanic events was not of paramount importance.

Subsequent chapters in this thesis introduce a catalog of eruptive events from Mt Taranaki (New Zealand). Only one historically observed eruption is included in the Mt Taranaki catalog, an account from the diary entries of an early settler (Turner et al., 2008a). Some of the more detailed historical records (such as Etna, Kilauea/Mauna Loa, Vesuvius, and Colima) contain eruption durations and estimates of eruption size/volume. However, extracting detailed information of that kind from geological records is difficult. Prehistorical eruption durations cannot be determined from geological records. Although there are methods of estimating eruptive volumes, which are discussed in Section 2.4, they contain large uncertainties. Additionally, tephra from individual eruptions that occur in close succession will often be merged if there is no observable organic material to separate them.

Eruption records become more incomplete looking further back in time. This is not only a result of the decline in historical accounts of eruption activity, but is also a consequence of working with geological catalogs. Each eruption adds a new layer of tephra to the landscape. New soils form after a prolonged period of time. Therefore, series of eruptions are recognized in geological records as a sequence of layers interspersed with organic material. Older tephra layers become more deeply buried over time. Consequently, more is known about the younger eruptions as they correspond to the most accessible tephra layers. The oldest tephras are often unexposed or unrecognizable so the onset of the first eruption is rarely known.

Observations of eruptions also tend to be biased towards the largest events. In historical records, this tendency is likely because the larger events caused the most damage and were worth reporting. Whereas, in a geological setting, larger eruptions leave thicker deposits with greater chances of being preserved. In addition, the expense involved in establishing records of events means that often the longer and more detailed records belong to volcanoes from wealthier countries. The more money, resources, advanced technologies, and trained specialists, the more complete the record of events is likely to be.

Available records of eruptions often represent only partial listings of the actual eruptions that occurred (Wang and Bebbington, 2012). If catalog incompleteness is not properly detected, it may result in false patterns subsequently leading to wrong interpretations. Since incompleteness is primarily due to short historical records, it is a difficult function to quantify (Mulargia et al., 1987). Wang and Bebbington (2012) discussed the sensitivity of hazard estimates with respect to missing observations. They adopted a hidden Markov model (HMM) framework to reflect missing onsets, and to estimate record completeness. However, the HMM developed by Wang and Bebbington (2012) requires a steady state, such that the 'true' or 'underlying' rate of eruptions does not exhibit trends over time.

Eruption behavior patterns can have a tendency to change which causes past records to become misleading. Events may occur that are unusual for a specific volcano. Changes in the size, shape, and composition of a volcano over time, may change hazard estimates (Decker, 1986). Even volcanoes with well documented records of many historical eruptions can show a wide variation in repose times, and large variations in eruption characteristics (Decker, 1986). This problem was considered by Wickman (1966), who first illustrated that while some volcanoes show a random pattern in the timing of their historical eruptions, others show patterns of increasing or decreasing probability of eruption depending on the time window considered. Wadge (1982) suggested that while a volcano can experience periods of steady state behavior, where phases of quiescence and activity do not change for a few hundred years, over longer periods of time the assumption of steady state behavior can be unsustainable. Identifying and modeling departures from stationary behavior can further complicate methods of forecasting eruption occurrences.

Many eruptions do not leave a long-lasting deposit. Erosion can result in the complete disappearance of some tephras in different environments. *Unexposed* locations, such as lakes and swamps, provide ideal sites for observing very fine ash only millimeters thick. Sediment cores extracted from *unexposed* locations have a higher tephra preservation rate than *exposed* surfaces (such as cliff faces, outcrops, road cuttings and drains). However, the coring process for extracting sediment columns can disturb tephra layers and embedded deposits can be altered by twisting and compaction. Figure 2.3(a) shows an example of a small portion of a sediment core extracted from Lake Umutekai containing tephra sourced from Mt Taranaki. Figure 2.3(b) shows tephra beds exposed on the flanks of Mt Taranaki.

Exposed locations are vulnerable to rapid erosion and reforestation. Therefore, they only preserve thicker tephras from much larger eruptions. In addition, human impact on previously

Figure 2.3: Comparison between records from (a) unexposed and (b) exposed locations. Photos courtesy of Shane Cronin.

(a) Sample portion of a sediment core extracted from Lake Umutekai, Taranaki, New Zealand.

(b) Tephra section from the Kaupokonui Valley on the flanks of Mt Taranaki, at approximately 1400 m elevation.

unmodified landscapes can change the configuration of layers causing tephra to be reworked or displaced. Tephra thickness and grain-size are also affected by variations in topography. Different sized catchment areas in the case of unexposed sites, and diversity of surrounding landscapes, can result in over- or under-thickening of tephra layers.

Larger eruptions are expected to leave thicker deposits, thus tephra thicknesses are important for understanding the size of previous eruptions. Most models for estimating eruption size (or volume) assume thickness trends are dictated by eruptive and atmospheric conditions, rather than depositional processes (Engwell et al., 2013). Understanding and quantifying the uncertainty in thickness measurements is difficult. Little attention has been paid to variations in tephra preservation from site to site. Throughout this thesis particular attention is paid to these site-specific effects for over- or under-thickening of tephra.

Multiple sites are needed to build the most accurate composite tephra record. Not all tephra producing eruptions leave a deposit in the same place, so it is preferable to sample from

multiple locations in different dispersal directions, both proximal and distal to the vent. In practice, difficulties in finding suitable sites for tephra preservation and the associated costs, means establishing a comprehensive record of events is challenging. Merging records obtained from multiple sites is complex. Common events and site-specific gaps must be recognized to avoid over or under inflating the true frequency of events. Establishing age estimates for the tephras is useful. However, the error associated with age estimates allows possible overlap among multiple tephras deposited at different locations. Separating tephras on the basis of geochemistry can help, but it is not a fail safe solution. Some volcanoes exhibit cyclic variations in deposit chemistry, where tephra fingerprints are 'non-unique' (Turner et al., 2011a). Distal records may also contain tephra erupted from multiple volcanoes. This poses the additional task of identifying the source of each tephra.

It is important to recognize that available data is only a sample of the eruptions that have occurred. Many eruptions may have gone unnoticed or unrecorded. Even if numerous eruption records are available, merging them can cause eruption frequencies to be over-represented. Alternatively, if only one record is used the true frequency may be underestimated (Turner et al., 2009).

2.2.6 Statistical methods of correlating tephra records

Tephrochronology characterizes individual tephras using a number of techniques. Stratigraphic positions and physical characteristics can be used to identify widespread *marker* deposits. Tephras similar in age, observed in different locations, can be correlated by examining mineralogical and geochemical variations.

Numerical and statistical correlation techniques have been essentially limited to discriminant function analysis (DFA) and its applications. DFA was first used by Stokes and Lowe (1988) to classify the source of tephra deposits, using glass-shard major element chemistry determined by electron microprobe analysis. Highly discriminating variables, or oxides, are selected through a stepwise procedure. The source of tephra is subsequently distinguished using scatter plots of canonical variables. Since then DFA has been successfully applied in a range of studies with mixed eruption sources, to either separate tephra from different volcanoes (Stokes et al., 1992; Cronin et al., 1996, 1997), or in some cases individual tephras (Cronin et al., 1997). A comprehensive review of the chemical and analytical techniques for tephra 'fingerprinting' is given in Lowe (2011).

Turner et al. (2009) developed a merging procedure for correlating records from two lake cores containing Mt Taranaki sourced tephras, extracted from Lake Umutekai and Lake Rotokare. Radiocarbon ages and associated normally distributed age errors of tephras were used to identify candidate matches between tephras at each site. They proposed calculating the age distance ($d_{1,2}$) between two tephras in different sites with mean ages m_1 and m_2, and standard deviations s_1 and s_2

$$d_{1,2} = \frac{|m_1 - m_2|}{\sqrt{s_1^2/2 + s_2^2/2}}. \tag{2.1}$$

Pairs with an age distance less than three were classified as *candidate* matches, i.e. possibly from the same eruption. They identified mutually closest pairs of tephras, ensuring the stratigraphic ordering of tephras was not violated. Secondly, chemical compositions of *titanomagnetite* grains were used to check candidate matches.

Titanomagnetite is a mineral containing titanium and iron oxides. It is easy to identify and concentrate, and is especially useful for correlating fine grained distal tephras (Turner et al., 2011a). Turner et al. (2009) used stepwise discriminant analysis to establish the subset of oxide constituents of the titanomagnetite grains that gave the greatest separation between tephras (namely TiO_2, Al_2O_3 and MgO). They performed principal component analysis on the major oxide compositions, obtaining the principal components:

$$\text{PC1} = -0.195\ \text{TiO}_2 + 0.711\ \text{Al}_2\text{O}_3 + 0.675\ \text{MgO}, \tag{2.2}$$

$$\text{PC2} = -0.949\ \text{TiO}_2 + 0.036\ \text{Al}_2\text{O}_3 - 0.312\ \text{MgO}. \tag{2.3}$$

Geochemical data, in the form of titanomagnetite chemistry, is not available for all tephras in the two lake core records of Turner et al. (2009). Those that do have geochemical data available have multiple sample measurements. Turner et al. (2009) converted principal components for each available tephra into ellipses centered at the means, with axes length equal to four times the standard deviation.

Eruptions are an assemblage of multiple magmas, so each titanomagnetite sample reflects a different magma. Interest lies with the distribution of the titanomagnetite composition for a number of magmas, not that of the 'average magma'. For this reason, ellipses are constructed using standard deviations rather than standard errors (Turner et al., 2009).

Figure 2.4: Geochemistry of two Lake Rotokare tephras (R9 and R10) and two Lake Umutekai tephras (U7 and U8) from Turner et al. (2009).

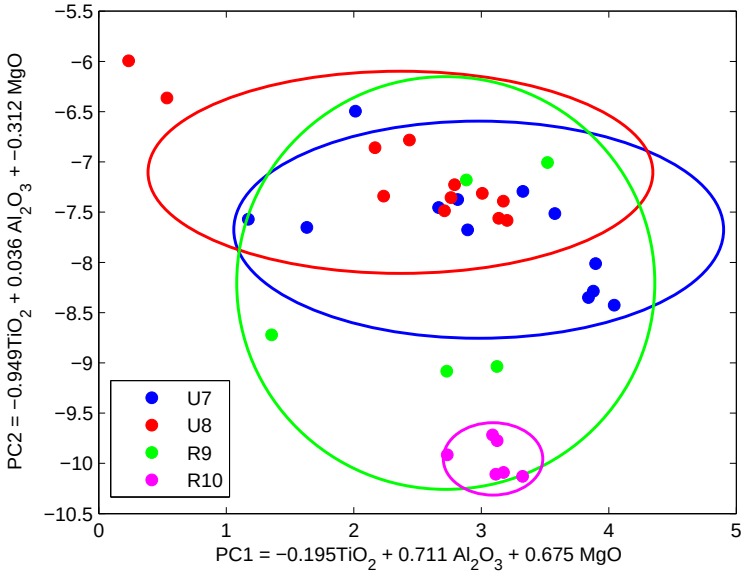


Figure 2.4 shows an example comparing the geochemistry of four tephras, two from Lake Umutekai (U7 and U8) and two from Lake Rotokare (R9 and R10). Candidate age matches can be confirmed by examining overlaps among the geochemical ellipses. R9 shows a possible correlation with both Umutekai tephras. R10, on the other hand, is compositionally distinctive showing no overlap with either Umutekai tephras. The ages and standard deviations of candidate matches that have the smallest age distances and compatible geochemistry can be pooled.

The method of Turner et al. (2009) is effective for merging tephra records from two sites. Extending the idea to the case of records from at least three sites, is more challenging. In Chapter 4 I propose a matching procedure for multiple cores using stochastic local search algorithms.

2.3 Modeling eruption occurrence

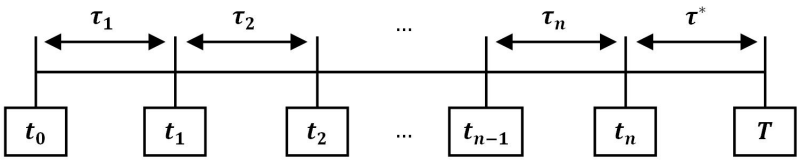
2.3.1 Notation

The beginning of an eruption is the *onset* commonly referred to as the *event*. The *onset time* of the ith eruption believed to occur in the observation period $[S, T]$ is denoted by t_i for $i = 0, \ldots, n$, where T usually represents the present. Most of the time the start of the

observation period is unknown in which case we let $S = t_0$, the oldest observed onset time. Models are constructed from the n onsets $t_1, \ldots, t_n$.

Time between events is referred to as *inter-event* times or *repose* periods. Strictly speaking repose periods refer to non-eruptive periods, denoted by the time between the end of one eruption and the start of another. However, the general lack of duration data available means it is common to model the time between successive onsets, $\tau_i = t_i - t_{i-1}$, as repose periods (cf. Klein, 1982), where t_i denotes the onset time of the ith eruption. A right-censored time is observed for the present unfinished interval, $\tau^* = T - t_n$, where t_n is the time of the last known eruption and T denotes the present time (see Figure 2.5).

Figure 2.5: Diagrammatic representation of the sequence of events.



For the analysis of repose data we are often interested in the hazard function or intensity, $\lambda(t)$, which specifies the instantaneous probability of event occurrence. Let $N(s,t)$ denote the number of events of events in a time interval (s,t), where the shorthand $N(t)$ is used if $s = 0$. The (point process) intensity of a process is

$$\lambda(t \mid H_t) = \lim_{\Delta t \downarrow 0} \frac{\Pr(N(t, t+\Delta t) = 1 \mid H_t)}{\Delta t}, \tag{2.4}$$

such that for very small Δt, the probability of eruption occurrence in the time interval $(t, t+\Delta t)$ is approximately $\lambda(t)\Delta t$. The intensity is dependent on the volcano's event history H_t, prior to time t. This can be characterized in terms of onset times or repose periods, and perhaps other variables such as eruption size or geochemistry. The intensity $\lambda(t)$ can be specified by

$$\lambda(t) = \frac{f(t)}{S(t)} \tag{2.5}$$

for a distribution of onset times with density function $f(t)$ and survival function

$$S(t) = 1 - F(t) = \int_t^\infty f(x)\,\mathrm{d}x \tag{2.6}$$

which denotes the probability of having a repose period of duration greater than t. Different density functions $f(t)$ characterize the time varying hazard in different ways.

2.3.2 Poisson process

The simplest model, for eruption (onset) forecasting is the *homogeneous* Poisson point process. This assumes that no more than one event can occur at any instant in time, and that events occur at a constant average rate given by the intensity function $\lambda(t) = \lambda$. This process means that regardless of the time interval selected, the number of events or eruptions, $N(t)$, occurring in the time interval is Poisson distributed with mean $\mu = \lambda t$. The times between events, assumed to be independent, are exponentially distributed with mean $\mu = 1/\lambda$.

The homogeneous Poisson process is a stationary stochastic process in time. The distribution of the number of points in an interval does not change when the interval is shifted in time. The number of eruption occurrences in any time interval only depends on the length of the interval. The process is considered memoryless, since the time since the last eruption provides no information about the time of the next eruption. Supposing some time s has elapsed since the most recent onset, the probability of extending the current repose period for a further time t is:

$$\begin{aligned} \Pr(T > s + t \mid T > s) &= \frac{\exp(-\lambda(t+s))}{\exp(-\lambda s)} \\ &= \exp(-\lambda t) \\ &= \Pr(T > t). \end{aligned} \tag{2.7}$$

The assumption of a constant recurrence rate λ suggests that the eruptive activity is relatively uniform, and the events occur at the same average rate at all times. Although this may sound unlikely, there are many volcanoes with long inter-event times that do exhibit such behavior (Marzocchi and Zaccarelli, 2006). However, as a result of the constant recurrence rate assumption, the homogeneous Poisson process is not suitable to model volcanic trends that are increasing or decreasing. De la Cruz-Reyna (1991) notes that, for a considerable number of volcanoes the distribution of reposes does not follow a simple Poisson process. Often the distribution shows a complex pattern of behavior departing from simple models. For example, Bebbington and Lai (1996a) considered the fit to five volcanic datasets considered in Ho (1991) (Aso, Etna, Kilauea, St Helens, and Yake-Dake) only one of which (Mt Etna) is satisfactorily modeled by a homogenous Poisson process.

If the volcanic trend fluctuates the model should be generalized to allow the intensity to vary over time. A *non-homogeneous* Poisson process is one such process by which the number of events, $N(t)$, in a given interval $(0, t)$, is Poisson distributed with mean $\mu = \int_0^t \lambda(s)ds$. Rather than the intensity being constant in time, as is the case with the homogeneous Poisson process, the intensity is some function of event time t. Ho (1991) considered a non-homogeneous Poisson process, namely the Weibull process (Bain and Engelhardt, 1986), for modeling volcanic hazard

$$\lambda(t) = \alpha\beta(\beta t)^{\alpha-1}. \tag{2.8}$$

This has the special case that for $\alpha = 1$ it is equivalent to the homogeneous Poisson process and is, thus, stationary. However, for $\alpha \neq 1$ the process is non-stationary. The point process intensity is either monotonically increasing ($\alpha > 1$) or decreasing ($\alpha < 1$).

Bebbington and Lai (1996a) highlight some undesirable theoretical and practical features of Ho's model which make it unsatisfactory for volcanic hazard modeling. Values of $\alpha > 2$ allow the hazard rate to accelerate constantly, imposing an undesirable monotonic trend on the volcano. In addition to this, the parameter estimates are sensitive to the selected time origin $t = 0$. Often the intensity of a volcanic process exhibits both increasing and decreasing volcanic activity over time. The Weibull distribution does not allow both types of trends to be modeled. It forces a strictly monotonic increasing or decreasing trend on the intensity which can give rise to a model that exhibits behavior that is physically unrealistic (Bebbington and Lai, 1996a). As an alternative, Bebbington and Lai (1996a) proposed a Weibull *renewal* model for volcanic processes as an extension to the homogeneous Poisson point process.

2.3.3 Renewal process

For the homogeneous Poisson point process, the times between events (τ) are independent exponential variables. Generalizing this to a process in which the time intervals are independent with some known distribution function $f(\tau)$, that is not restricted to the exponential, gives rise to a renewal process (Cox and Isham, 1980). The term 'renewal' stems from the fact that the process starts afresh after each event (Lindsey, 2004).

The intensity or hazard function of a renewal process is given by

$$\lambda(t) = \lambda(\tau) = \frac{f(\tau)}{1 - F(\tau)}, \tag{2.9}$$

where $f(\tau) = F(\tau)'$ is a density function and $\tau = t - s$, assuming that the most recent eruption occurred at a time $s < t$. In contrast to the nonhomogeneous Poisson process, which has intensity function depending on the onset time t, the renewal process intensity depends only on the inter-event time τ. Therefore, the time to the next eruption is controlled by the time since the last eruption. Previous eruptions only influence the hazard through their contribution to the parameter estimates involved in $f(\tau)$.

With a renewal process we come across an interesting problem related to defining stationarity. A stochastic process is defined as strictly stationary if the joint probability distribution does not change when shifted in time. In this sense a renewal process is 'stationary' since, by definition, the random intervals between events are independently and identically distributed (Lindsey, 2004). However, the mean number of counts of events in a small time interval of fixed width will not be constant. In other words the renewal process is not 'stationary' in terms of counts of events, it is only stationary in terms of the interval sequence (Cox and Isham, 1980). This ruling is applicable, except for the special case in which the renewal function is linear in time, i.e. it a homogeneous Poisson process with no observed trend in the eruption sequence over time. Unlike the point process, the renewal process is not memoryless. The time to the next eruption is no longer independent of the time elapsed since the last eruption.

Bebbington and Lai (1996a) considered a Weibull renewal process for modeling repose periods which has the density function

$$f(\tau) = \alpha\beta(\beta\tau)^{\alpha-1}\exp[-(\beta\tau)^{\alpha}], \tag{2.10}$$

and intensity function

$$\begin{aligned} \lambda(\tau) &= \frac{f(\tau)}{S(\tau)} \\ &= \frac{\alpha\beta(\beta\tau)^{\alpha-1}\exp[-(\beta\tau)^{\alpha}]}{\exp[-(\beta\tau)^{\alpha}]} \\ &= \alpha\beta(\beta\tau)^{\alpha-1}, \end{aligned} \tag{2.11}$$

where α and β are the shape and scale parameters, respectively.

For the special case when $\alpha = 1$, the Weibull distribution reduces to the exponential distribution and the renewal process becomes a homogeneous Poisson process. It is evident that the intensity function for the Weibull renewal process and the nonhomogeneous Poisson process

are similar. The only difference is that the intensity of the renewal process is dependent only on the elapsed time since the last eruption, τ, and not on the absolute time of event, t.

The renewal process has also been adopted by De la Cruz-Reyna and Carrasco-Nunez (2002) to model hazard expectations for Citaltépetl Volcano. A gamma density

$$f(\tau) = \frac{\beta^\alpha}{\Gamma(\alpha)} \tau^{\alpha-1} \exp[-\beta\tau], \tag{2.12}$$

was suggested as a suitable alternative generalization of the Poisson point process. The gamma distribution provides similar flexibility in the shape of the intensity function as the Weibull, but the computational details of the likelihood function are far more complex. The advantage of both the Weibull and gamma renewal processes is that they have the capacity to model either under-dispersed (i.e. semi-periodic) or over-dispersed (i.e. clustered) inter-event times (Turner et al., 2009). However, for processes in which the likelihood of 'clustering' is small, the lognormal distribution is also a reasonable alternative to the Weibull distribution (Bebbington and Lai, 1996a; Turner et al., 2008a). The lognormal distribution has a density function

$$f(\tau) = \frac{1}{\tau\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln\tau - \mu)^2}{2\sigma^2}\right). \tag{2.13}$$

The lognormal distribution was also considered in the Generalized Time Predictable Model of Marzocchi and Zaccarelli (2006). Instead of a constant μ, they make μ the result of a linear regression incorporating the volume/size v of the previous eruption

$$f(\tau) = \frac{1}{\tau\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln\tau - (a + bv))^2}{2\sigma^2}\right). \tag{2.14}$$

Other distributions that have been considered in volcanology include: the Pareto (Medina Martinez, 1983) and the power law (Pyle, 1998) distributions, for modeling monotonically decreasing intensity; and the log-logistic (Connor et al., 2003) and inverse Gaussian (Garcia-Aristizabal et al., 2012) distributions, for modeling unimodal inter-event times.

Turner et al. (2008a) considered an additional model to represent eruptive sequences, for the special case in which the distribution of inter-event times is *bimodal*. In particular, where the distribution has a strong peak followed by a long tail with a, generally smaller, additional mode in the distribution. They fitted this phenomenon using a mixture of Weibull

distributions formulated as a renewal process:

$$\begin{aligned} f(\tau) &= pf_1(\tau) + (1-p)f_2(\tau) \\ &= p\alpha_1\beta_1^{\alpha_1}\tau^{\alpha_1-1}\exp(-(\beta_1\tau)^{\alpha_1}) + (1-p)\alpha_2\beta_2^{\alpha_2}\tau^{\alpha_2-1}\exp(-(\beta_2\tau)^{\alpha_2}) \end{aligned} \tag{2.15}$$

where p is the mixing proportion ($0 < p < 1$). They noted that this mixture distribution appeared to perform better than both the lognormal and the Weibull distributions alone.

Although Turner et al. (2008a) only considered a mixture of Weibull distributions, variations can be made to include alternative density functions. For example, Mendoza-Rosas and De la Cruz-Reyna (2009) proposed a mixture comprising of a sum of exponential distributions. Parameter estimates were obtained graphically, and components of the mixture were arbitrarily weighted. The mixture of exponential distributions constructs a longer tailed monotonically decreasing density function, similar to that which could be achieved with a Weibull with $\alpha < 1$ (Bebbington, 2012). There is also no restriction on the components $f_1(\tau)$ and $f_2(\tau)$ in (2.15) having the same distribution. Different shapes in the distribution of inter-event times may require alternative density functions.

Given any density function and the observed inter-event times τ, the parameters of the given distribution can be estimated by maximum likelihood. Parameter values involved in $f(\tau)$ are chosen either algebraically or numerically to maximize the likelihood

$$L(\tau) = [1 - F(\tau^*)] \prod_{i=1}^{n-1} f(\tau_i), \tag{2.16}$$

where τ^* denotes the interval between the last event and the present time. Given difficulties in determining the start of the observation period for volcanic eruptions, it is common to consider observation to have started at the first recorded onset.

Parameters could also be estimated using other methods, such as the method of moments or least squares. The method of moments involves constructing estimators of the parameters based on matching sample moments with theoretical moments. For a two-parameter case this could involve matching the mean and standard deviation. With mixture distributions estimating the parameter values, using the method of moments, would be more complex. Different distributions are better suited to different methods of parameter estimation. For example, the method of moments is particularly attractive for estimating the parameters of the gamma distribution in (2.12), due to the simplified form of the mean ($E(\tau) = \alpha/\beta$)

and variance ($V(\tau) = \alpha/\beta^2$). The Weibull distribution (2.10), on the other hand, is better suited to maximum likelihood estimation. The method of moments would involve numerical calculation of the gamma function Γ in the mean and variance

$$E(\tau) = \frac{1}{\beta}\Gamma\left(1 + \frac{1}{\alpha}\right) \tag{2.17}$$

$$V(\tau) = \frac{1}{\beta^2}\left[\Gamma\left(1 + \frac{2}{\alpha}\right) - \Gamma\left(1 + \frac{1}{\alpha}\right)^2\right]. \tag{2.18}$$

The other alternative, least squares estimation, involves approximating parameter values by minimizing the sum of squared differences between observed curves, such as the survival function, and those produced by the model. Although least squares estimation is typically done using linear regression techniques, parameter values can be also found using numerical optimization.

Different models of eruption occurrence involve a different number of parameters. The question is, which model best describes the data and offers an improvement in fit sufficient enough to justify the inclusion of additional parameter values. Models fitted using maximum likelihood techniques, can be compared using the Akaike Information Criterion (Akaike, 1977)

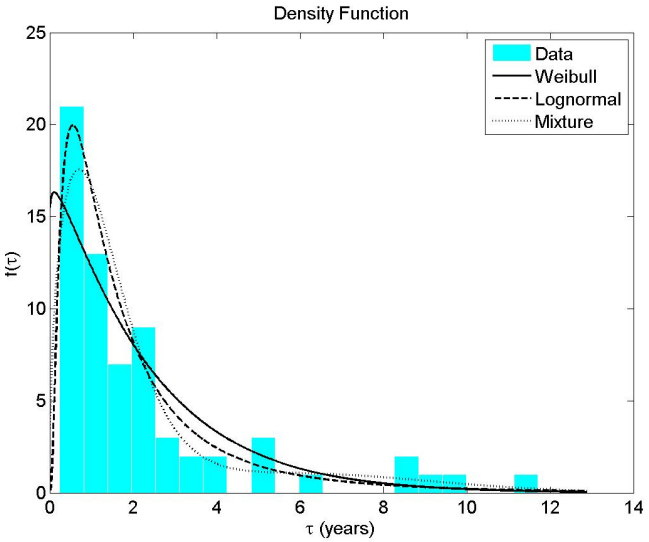
$$\mathrm{AIC} = -2\log L + 2k, \tag{2.19}$$

where k is the number of parameters in a model, and $\log L$ denotes the log-likelihood of the model. AIC compensates for the effect of additional parameters and thus avoids over-fitting. Given any two estimated models, the model with the lower AIC value is to be preferred. Note that we cannot in general use the corrected AIC (Hurvich and Tsai, 1989). Unlike for linear regression and autoregressive models, there is no proof of its validity for point process models (Claeskens and Hjort, 2008). Datasets are also not usually large enough to consider BIC.

The suitability of different distributions depends on the sequence of inter-event times at hand. Figure 2.6 illustrates the fit of Weibull, lognormal, and mixture of Weibull distributions to inter-event sequences from Ruapehu and Ngauruhoe (New Zealand). For the case of Mt Ruapehu, the second mode (at approximately 8 years) in the inter-event sequence is not prominent enough to require a mixture distribution. The lognormal distribution is preferred. The same can be said for the Ngauruhoe inter-event times. However, the results are more peculiar. The current repose period for Ngauruhoe is the longest in the catalog ($\sim$ 35 years)

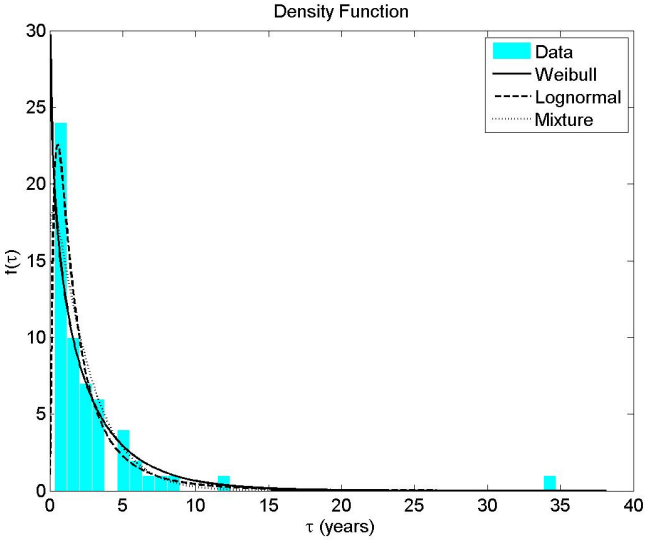
Figure 2.6: Density functions and estimated parameter values for a range of different distributions fitted to the inter-event times of (a) Mt Ruapehu and (b) Mt Ngauruhoe eruptions.

(a) Ruapehu inter-event times



Model	Weibull	Lognormal	Mixture of Weibulls
Parameters	$\alpha = 1.041$	$\mu = 0.338$	$p = 0.983$
	$\beta = 0.436$	$\sigma = 0.966$	$\alpha_1 = 1.484$
			$\beta_1 = 0.676$
			$\alpha_2 = 2.450$
			$\beta_2 = 0.137$
$\log L$	-119.5	-113.49	-114.3
AIC	243.01	**230.97**	238.61

(b) Ngauruhoe inter-event times



Model	Weibull	Lognormal	Mixture of Weibulls
Parameters	$\alpha = 0.869$	$\mu = 0.462$	$p = 0.983$
	$\beta = 0.369$	$\sigma = 1.042$	$\alpha_1 = 1.080$
			$\beta_1 = 0.413$
			$\alpha_2 = 0.559$
			$\beta_2 = 0.000$
$\log L$	-117.42	-107.84	-110.32
AIC	238.85	**219.68**	230.64

and is particularly unusual. All other inter-event times are less than 12 years. The mixture distribution forces the second mode out at this unusual value. However, $\beta_2 \approx 0$ suggests that the second mode is not required. In this instance, the mixture of Weibull distributions breaks down. The likelihood for the mixture distribution is maximized by assigning the last repose period to one component of the mixture by itself and as it is only a bound, the likelihood term is $\Pr(T > t)$, which is maximized by sending the mode to infinity.

AIC is a useful means of comparing between models but it does not explain whether a model is actually a reasonable representation of the data. That requires goodness-of-fit tests.

2.3.4 Goodness-of-fit tests

A stochastic process $\{X_t\}$ is said to be *strictly* stationary if, given any sequence of eruption onset times $t_1, \ldots, t_n$, and any time shift s, the joint probability distribution of both subsets $\{X_{t_1}, \ldots, X_{t_n}\}$ and $\{X_{t_1+s}, \ldots, X_{t_n+s}\}$ are identical. In other words, shifting a fixed-width time observation window along a strictly stationary series will always yield the same distribution (Lindsey, 2004). With the exception of the homogenous Poisson process, renewal processes generally do not have stationary increments; the time since the last event carries some information. Unlike the homogeneous Poisson process, where the intensity function is always constant, the intensity of a renewal process is dependent on the elapsed time since the last event τ. Therefore, renewal processes are stationary from event to event, but not necessarily from interval to interval. Renewal processes should however, be *weakly* stationary in the sense that there should be no trend in the occurrence rate with time. The process should exhibit constant variance. Henceforth, any discussion of stationarity will refer to this weaker case.

If a process is stationary onset times should be random, i.e. distributed uniformly within the observation period $[S, T]$ at the level at which a test can be done. To see if a sequence of events is a stationary process, the Kolmogorov-Smirnov test can be applied. The Kolmogorov-Smirnov test is easily implemented, and is well suited to volcanic datasets because of their characteristically small size. For a sequence of eruption onset times t_i $(i = 1, \ldots, n)$, occurring within the observation period $[S, T]$, the empirical distribution for the cumulative number of eruptions is

$$F_n(t) = \frac{\#(t_i \leq t)}{n}, \quad i = 1 \ldots n, \quad S < t < T. \tag{2.20}$$

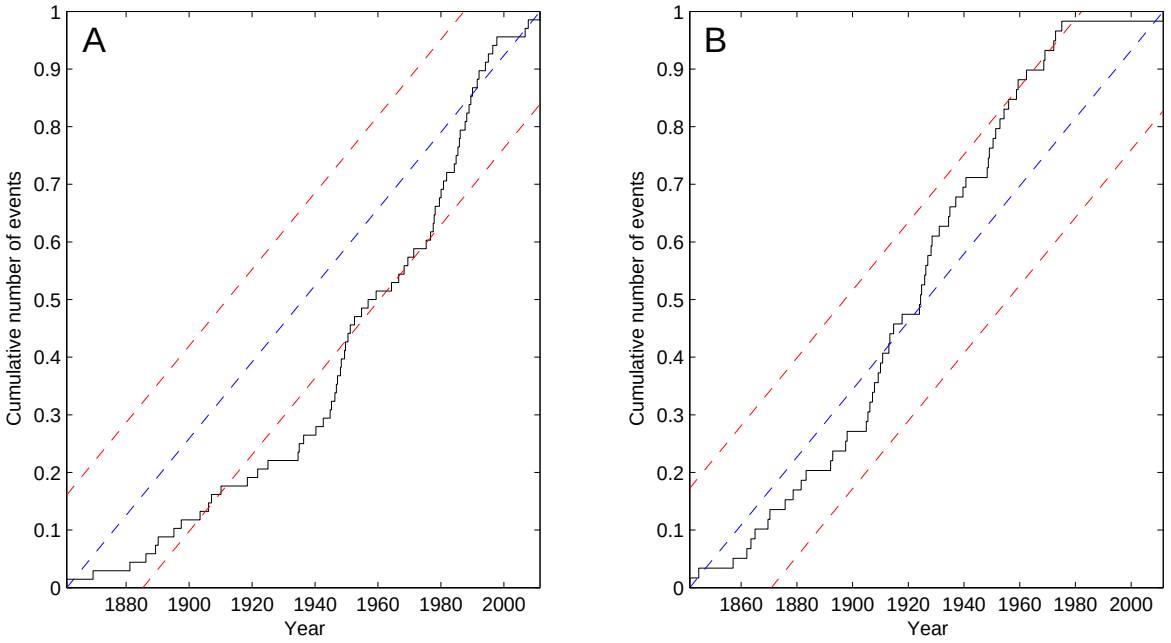
The Kolmogorov-Smirnov test statistic (D_n), based on the maximum distance between the theoretical (uniform) and empirical cumulative distribution functions, is then calculated as:

$$D_n = \max_{t \in [S,T]} \left[\left| F_n(t) - \frac{t - S}{T - S} \right| \right]. \tag{2.21}$$

Critical values for the Kolmogorov-Smirnov test statistic, in the uniform distribution case, have been tabularized by Miller (1955) for small sample sizes $n \leq 20$. For larger sample sizes the 5% critical value is $1.36/\sqrt{n}$.

The Kolmogorov-Smirnov test can be performed using the kstest function in Matlab. The results, when testing the Mt Ruapehu and Mt Ngauruhoe eruption records introduced in

Figure 2.7: Kolmogorov-Smirnov test for stationarity of the eruption records for (A) Mt Ruapehu and (B) Mt Ngauruhoe. Stationary eruption rate in blue with 95% confidence bands in red.



Section 2.3.3, are shown in Figure 2.7. The greatest distance between the observed and expected distributions gives the test statistic. For the Mt Ruapehu record $D_n = 0.257$ (P-value < 0.001), and for the Mt Ngauruhoe record $D_n = 0.197$ (P-value 0.016). Thus, in both situations the null hypothesis is rejected, suggesting that the frequency of eruptions has not been constant over time. For Mt Ngauruhoe, this result is due to the last repose period being uncharacteristically long. The eruption frequency prior to 1980 appears relatively constant. The frequency of Mt Ruapehu eruptions on the other hand appears to increase around 1940, consistent with observations of Bebbington and Lai (1996b).

Alternative tests for stationarity are discussed in Bebbington and Lai (1996a,b), such as the Pearson's χ^2 test for comparison of observed and expected eruptions in time intervals, or trend detection using a moving average of the inter-event times (Klein, 1982). Reyment (1969) also suggested a number of tests to check whether a series of eruptions is consistent with a renewal process, which are summarized in Bebbington (2009).

The renewal models fitted to the inter-event times of Mt Ruapehu and Mt Ngauruhoe eruptions in Figure 2.6, rely on the assumptions that; the eruption activity does not change over time (i.e. the process is stationary), inter-event times are independent, and the chosen renewal distribution is a reasonable fit to the inter-event times. Although the lognormal

renewal model provided a superior fit to the inter-event sequences of both Mt Ruapehu and Mt Ngauruhoe, the process is not stationary in time, emphasizing the importance of testing goodness-of-fit. Ruapehu actually appears to exhibit strong non-stationarity, whereas Ngauruhoe may be suitably modeled by a renewal process, with the exception of the last repose period. Suitable alternative models may also include a regime model (Bebbington, 2007) or a trend renewal process (Bebbington, 2010) (see Section 2.3.5).

Independence of inter-event times can be checked by calculating the Spearman rank correlation ρ, of successive repose periods.

$$\rho = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}}, \tag{2.22}$$

where x_i is the rank of the repose length τ_i $(i = 1, \ldots, n-2)$, and y_i is the rank of the subsequent repose length τ_{i+1} $(i = 1, \ldots, n-2)$.

Using the notation from Bebbington (2012), if R_i is the rank of the repose length τ_i from shortest $(R = 1)$ to longest $(R = n-1)$, then the Spearman rank correlation can be rewritten as

$$\rho = \frac{\sum_{i=1}^{n-2} \left(R_i - (n-2)^{-1} \sum_{j=1}^{n-2} R_j \right) \left(R_{i+1} - (n-2)^{-1} \sum_{j=2}^{n-1} R_j \right)}{\sqrt{\sum_{i=1}^{n-2} \left(R_i - (n-2)^{-1} \sum_{j=1}^{n-2} R_j \right)^2 \sum_{i=2}^{n-1} \left(R_i - (n-2)^{-1} \sum_{j=2}^{n-1} R_j \right)^2}}. \tag{2.23}$$

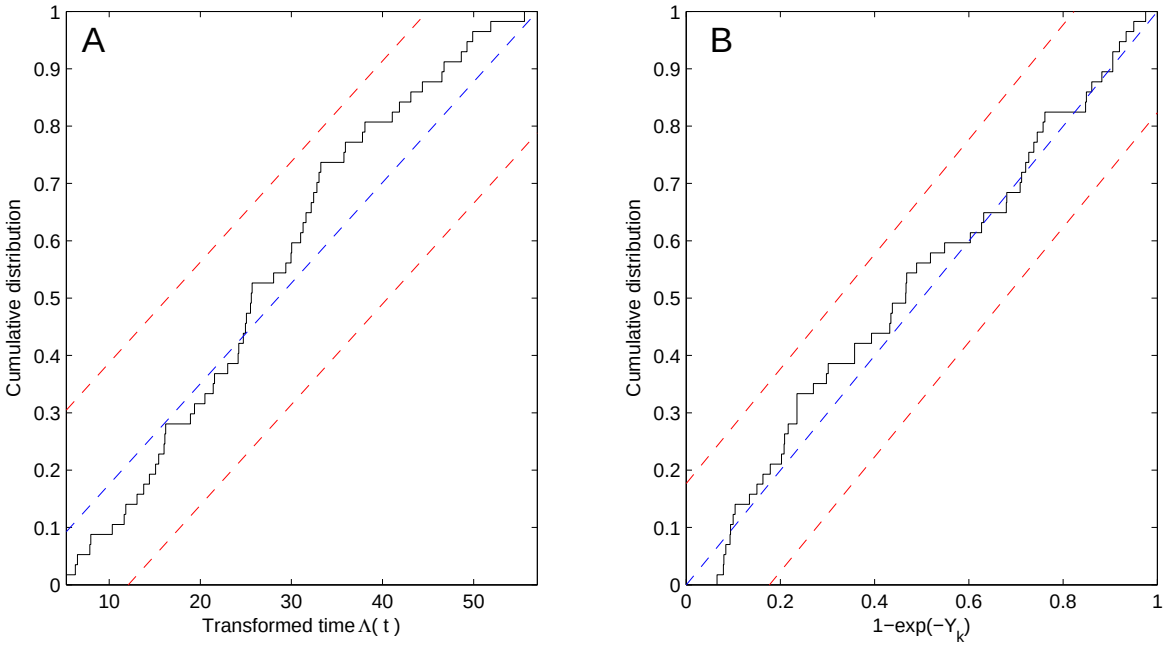
For the Mt Ruapehu and Ngauruhoe records $\rho = 0.450$ (P-value < 0.001) and $\rho = -0.092$ (P-value $= 0.502$), respectively. This suggests the inter-event times for Mt Ruapehu are not independent, which again gives concern with the analysis in Figure 2.6.

In order to assess if a particular distribution provides a suitable fit to the observed record, the point-process residual theory of Ogata (1988) can be utilized. This involves calculating the 'compensator'

$$\Lambda(t) = \int_0^t \lambda(t) dt. \tag{2.24}$$

If the model $\lambda(t)$ satisfactorily accounts for the structure of the data, the transformed event times $\{\Lambda(t_i)\}$ will be from a Poisson process of rate one. For a renewal process the inter-event times of the transformed process $Y_i = \Lambda(t_{i+1}) - \Lambda(t_i)$ should be independent, the cumulative counts $\{\Lambda(t)\}$ should follow a $U[0, N]$ distribution, and none of the $\{1 - \exp(-Y_i)\}$ should differ significantly from a $U[0, 1]$ distribution.

Figure 2.8: (A) Stationarity and (B) exponential inter-event diagnostic plots for the Mt Ngauruhoe record. Centerline in blue with 95% confidence bands in red.



For Mt Ngauruhoe, the lognormal distribution was proposed as a better fit to the inter-event times than both the Weibull and mixture of Weibull distributions (Figure 2.6(b)). Point-process residual diagnostics plots for the lognormal model are shown in Figure 2.8. We can assume that the transformed process is stationary (Figure 2.8A) and has exponential inter-event times (Figure 2.8B). None of the $\{Y_{i+1}\}$ and $\{Y_i\}$ have significant correlations, hence the transformed process is independent. The proposed lognormal distribution in Figure 2.6(b) is a good representation of the data. However, slight caution should be taken as the last repose period for Ngauruhoe suggests a change in volcanic activity.

While various alternative goodness-of-fit tests are available, as discussed in Bebbington and Lai (1996a,b) and Bebbington (2012), the point-process compensator of Ogata (1988) is particularly useful. It can assess a wide variety of distributions and is not restricted solely to the renewal processes discussed in Section 2.3.3. In Chapter 3 the point-process residual theory of Ogata (1988) is used to assess proportional hazard type models which incorporate geochemical covariate data into the intensity function. It can also be used to assess the fit of the nonhomogeneous Poisson processes (Marzocchi and Bebbington, 2012).

2.3.5 Time-varying behavior

The drawback of renewal models is that they assume events occur at the same average rate at all times, and therefore commonly fail to explain variations in eruption rates due to changes in activity level (Bebbington, 2010). With the exception of the nonhomogeneous Poisson process, all of the models discussed so far have been weakly stationary, where the underlying level of eruptive activity is assumed to be constant over time. Wickman (1966) however, suggested that the activity of a volcano could change over time.

The eruption records encountered throughout this thesis do not concern changes in activity level. However, a brief discussion is provided of some of the techniques available to identify and model such changes. For a more detailed review see Mulargia et al. (1987), Bebbington (2007, 2010), and references therein.

A process that is thought to change in activity level can be split into different regimes by separating the points in which the intensity of the process changes. These regimes may represent changes in the eruption mechanism which transports magma to the surface, or changes in eruptive style (Bebbington, 2009). The idea of different regimes of volcanic activity was discussed by Mulargia et al. (1987). They highlighted the importance of objectively identifying the different regimes. They noted that the number of regimes is unknown, different regimes may follow different distributions, and the sample sizes for volcanic activity are generally small.

The algorithm, used by Mulargia et al. (1987), for detecting the different regimes starts by making use of Kolmogorov-Smirnov two-sample statistics to determine the most significant change-point in the entire series. If a change point is identified the series is split into two. Each part is examined to see if any further change points are revealed. This procedure continues successively until no more significant change points are identified. The procedure followed by Mulargia et al. (1987) is a simple statistical technique used to identify changes in regime. It does not however, provide a stochastic model for forecasting future activity (Bebbington, 2007).

The suggestion that a volcano may switch between various discrete regimes or states (Wickman, 1966), prompted Bebbington (2007) to consider hidden Markov models for describing eruptive processes. Future eruption activity is a function of the current regime. Therefore, probabilistic models must encompass both future activity in the current regime, and possible changes in regime. Given an eruptive process which has a finite number of possible discrete

states, $S_1, S_2, \ldots, S_N$, the system may be described as being in any one of these N distinct states at a given time. The system experiences a change of state according to a set of probabilities associated with the state. A full probabilistic description of the system would, in general, require specification of not only the current state, but all previous states (Rabiner, 1989). If the regimes or states are not directly identifiable, the process can be automated using hidden Markov models, where the volcano is assumed to occupy an underlying, unobservable (hidden) sequence of states that follow a Markov chain. The process cycles between states or regimes in a stochastic fashion (see, Bebbington, 2007, for details).

Bebbington (2010) suggested a method by which varying levels of eruption activity could be introduced to renewal models, proposing the use of a *trend renewal process* to model inter-event times. The trend renewal process is capable of incorporating monotonic trends, waxing and waning, and cyclic patterns; and includes stationary renewal processes and the nonhomogeneous Poisson process, as special cases. Given an increasing function $\Psi(t)$, $t \geq 0$, and a renewal distribution F, the trend renewal process is defined by the values $\tau_i^* = \Psi(t_i) - \Psi(t_{i-1})$ being independent and identically distributed random variables according to the distribution function $F(\tau^*)$. If

$$\Psi(t) = \int_0^t \psi(s)\,\mathrm{d}s \tag{2.25}$$

for some $\psi(s) \geq 0$, where $\Psi(T) = T$, then for $\Psi(t) = t$ (i.e $\psi(t) = 1$) the trend renewal process is a stationary process. Values of $\psi(t) < 1$ model lower than average onset rates and values of $\psi(t) > 1$ model higher than average onset rates. If $F(\tau^*) = 1 - \exp(-\beta\tau)$, the exponential distribution, then the trend renewal process is equivalent to the nonhomogeneous Poisson process.

2.3.6 Time and size predictable models

Eruptions are usually modeled simply using functions of onset times. However, to incorporate the idea of eruption size into the models there has been a parallel line of development into models based on eruptive volume.

De la Cruz-Reyna (1991) proposed a general *load-and-discharge* model, where the energy level of a volcano is assumed to increase at a constant rate σ between eruptions until a critical threshold H_i is reached. The ith eruption occurs when the threshold H_i is exceeded, at which point the energy of the system drops to a threshold level L_i. The change in energy during an

eruption is denoted by $\xi_i = H_i - L_i$ and the time between events is given by

$$\tau_i = t_i - t_{i-1} = \frac{H_i - H_{i-1} + \xi_{i-1}}{\sigma}. \tag{2.26}$$

In this representation the inter-event times are not independent as they include a common term H_i/σ. There is a tendency for short repose periods to be followed by long ones and vice versa.

In general, the threshold energy levels can be random functions of time. However if a constant threshold H_i is assumed, then the model is *time-predictable*. The inter-event times $\tau_i = t_i - t_{i-1} = \xi_{i-1}/\sigma$ become independent, proportional only to the energy of the previous eruption. Whereas if a constant final energy threshold $L_i = L$ is assumed, then the relationship is *size-predictable*; $H_i = \xi_i + L$ and from (2.26) the inter-event times $\tau_i = t_i - t_{i-1} = (H_i - L)/\sigma$ are independent. The energy from the eruption is proportional solely to the previous inter-event time.

Time-predictable behavior is most simply explained by assuming a constant rate of magma input; where eruptions occur only when a threshold of the magma volume in the reservoir is reached. In contrast, size-predictable behavior assumes that the output of each eruption is determined only by the magma accumulated since the last event, and that the volcano's reservoir is drained to the same initial level during each eruption.

Martin and Rose (1981) noted that inter-event times in eruption sequences from Fuego Volcano (Guatemala) appear to be proportional to the volume of the preceding eruption. Evidence of time-predictable behavior has also been observed for intervals between eruptions in the Coso Range, California (Bacon, 1982), the Somma-Vesuvius volcanic complex, Italy (Santacroce, 1983), and Kilauea and Mauna Loa Hawaii (Klein, 1982), among others.

Burt et al. (1994) suggested a volcano could be tested for time-predictability by performing a regression analysis of the size or volume erupted during an eruption on the inter-event time following it. This relationship implies that $\tau_{i+1} \propto v_i$, where v_i is the volume erupted during the ith eruption occurring at time t_i, and τ_{i+1} is the inter-event time between the onset of the ith and $(i+1)$th eruption, $\tau_{i+1} = t_{i+1} - t_i$.

Sandri et al. (2005) extended the regression analysis for time-predictability to $\log(v_i)$ on $\log(\tau_{i+1})$

$$\log(v_i) = a + b\log(\tau_{i+1}), \tag{2.27}$$

which implies a power law relationship $v_i \propto \tau_{i+1}^b$. If $b = 1$ the relationship between size and time of an eruption will be linear. Hence the method of Sandri et al. (2005) is a generalization of the time-predictable model by Burt et al. (1994). Replicating the analysis done by Sandri et al. (2005), for the occurrence of flank and summit eruptions at Mt Etna, the logarithm of the erupted volume is obtained as a function of the logarithm of the inter-event time following it (see Figure 2.9A). The regression model rejects the null hypothesis $H_0 : b = 0$ (P-value 0.004), implying a significant linear relationship between the logarithm of eruptive volume and the following inter-event time. Note, however, that the analysis in Sandri et al. (2005) uses the inter-event time $\tau_{i+1} = t_{i+1} - t_i$ as the independent variable when predicting the eruptive volume v_i. The alternative representation proposed by Marzocchi and Zaccarelli (2006):

$$\log(\tau_{i+1}) = a + b \log(v_i), \tag{2.28}$$

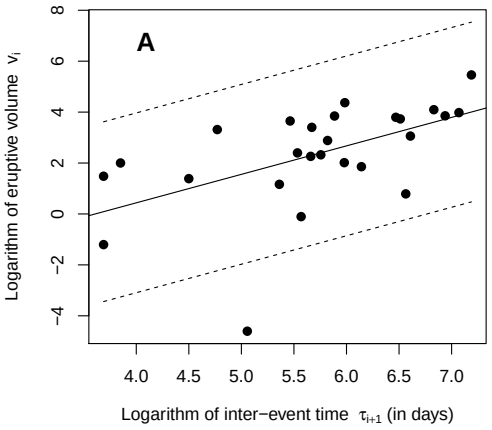
in which v_i is the independent variable, does not change the conclusions reached with respect to the significance of the slope (see Figure 2.9B).

We have seen an application of a time-predictable model in Section 2.3.3 where Marzocchi and Zaccarelli (2006) used the lognormal distribution from (2.14) as a renewal process for eruption occurrence, where a, b denote the intercept and slope regression coefficients from the regression of $\log v_i$ on $\log \tau_{i+1}$. The time-predictable model was also established in a Bayesian framework by Passarelli et al. (2010a) to allow additional covariates, such as eruption duration, to be incorporated while properly considering any uncertainties.

As well as the time-predictable model, there is the size-predictable model. The latter serves to incorporate eruption size into the model by assuming a relationship between the size/volume of an event and the previous repose time. Burt et al. (1994) proposed a 'water-butt' model for size-predictability entailing a regression analysis of τ_{i+1} on v_{i+1}. This was adapted by Marzocchi and Zaccarelli (2006) who formulated the size-predictable model through a regression analysis of $\log \tau_{i+1}$ on $\log v_{i+1}$, resulting in the relationship $v_{i+1} \propto \tau_{i+1}^b$. Since the distribution of inter-event sequences and volumes are typically highly skewed, logarithmic transformations are preferred to reduce the high leverage of some of the tail points.

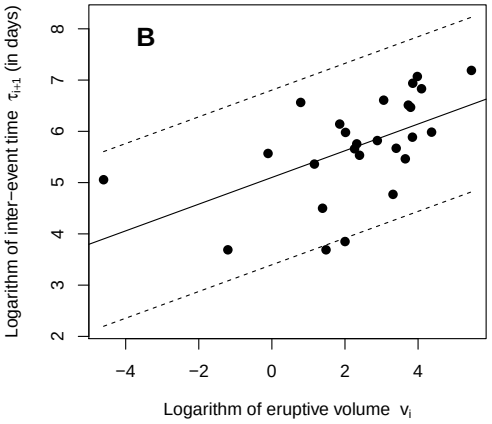
Applying the size-predictable model from Marzocchi and Zaccarelli (2006), to the Mt Etna records considered by Sandri et al. (2005), suggests that there is no significant linear relationship between the logarithm of erupted magma volume and the previous inter-event time

Figure 2.9: Linear regression analysis of the inter-event times versus the eruptive volumes of Mt Etna flank and summit eruptions (Sandri et al., 2005). Solid lines represent the fitted regression equations, dashed lines represent 95% confidence levels.



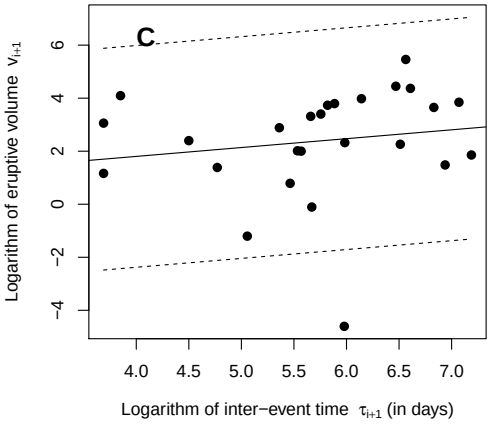
Regression analysis of $v_i \sim \tau_{i+1}$

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -4.0367     2.0637  -1.956  0.06219 .
Slope         1.1185     0.3561   3.141  0.00443 **
---
Residual standard error: 1.765 on 24 degrees of freedom
Multiple R-squared:  0.2913, Adjusted R-squared:  0.2618
F-statistic: 9.867 on 1 and 24 DF, p-value: 0.004427
```



Time-predictable regression analysis $\tau_{i+1} \sim v_i$

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   5.10061    0.25693  19.852  < 2e-16 ***
Slope         0.26049    0.08293   3.141  0.00443 **
---
Residual standard error: 0.8518 on 24 degrees of freedom
Multiple R-squared:  0.2913, Adjusted R-squared: 0.2618
F-statistic: 9.867 on 1 and 24 DF, p-value: 0.004427
```



Size-predictable regression analysis $v_{i+1} \sim \tau_{i+1}$

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.4668     2.4442   0.191    0.850
Slope         0.3343     0.4217   0.793    0.436
---
Residual standard error: 2.091 on 24 degrees of freedom
Multiple R-squared: 0.02551, Adjusted R-squared: -0.01509
F-statistic: 0.6283 on 1 and 24 DF, p-value: 0.4358
```

(see Figure 2.9C, P-value for the slope = 0.44 >> 0.05). Hence, eruptions at Mt Etna are consistent with time-predictability but do not follow a size-predictable pattern.

2.4 Methods of estimating eruptive volumes

While temporal forecasting of eruption episodes has been widely studied, forecasts of eruption size are less common. Data on eruption size/volume is scarce and size-predictable behavior appears to be much rarer than time-predictable behavior (Bebbington, 2014). Improving forecasts for eruption size is an important challenge facing volcanic hazard analysis (Marzocchi and Bebbington, 2012). Models that incorporate eruption size depend on the availability of reasonable estimates of eruptive volumes. However, there are very few volcanoes that have sufficient volume data available. Addressing this is one of the major foci of this thesis.

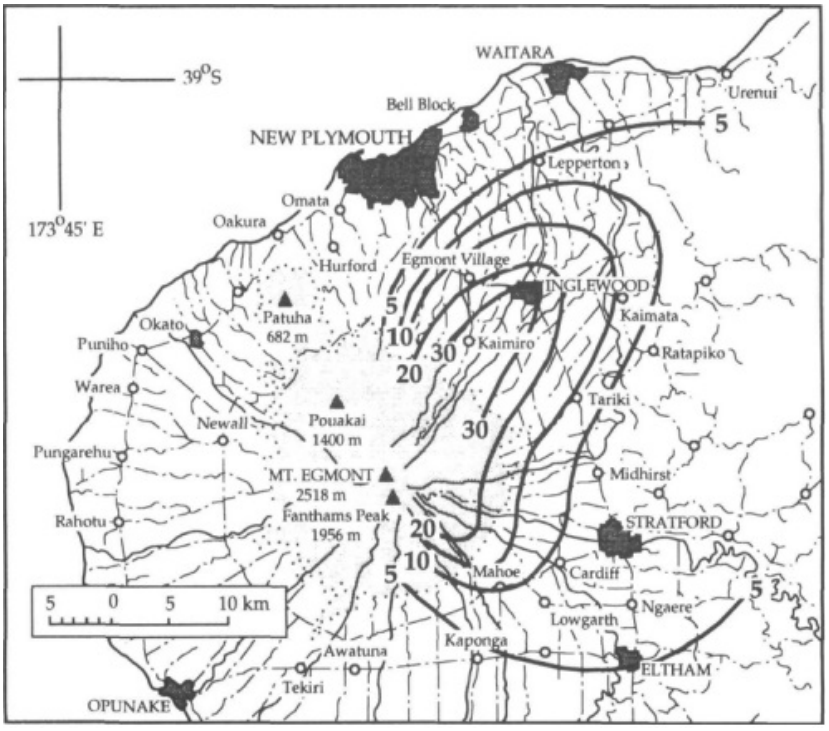
Estimates of tephra volumes are usually obtained by mapping the distribution of thickness deposits found after an eruption. Deposits are measured at selected field locations, on or around the vents. The resulting point thickness measurements are then converted into hand-contoured maps, termed *isopach* maps, according to geologist prior experience and field observations. Tephra dispersal is influenced by wind, thus isopach contours commonly show elliptical or distorted semi-elliptical shapes. Figure 2.10 shows an example of an isopach map, constructed by Alloway et al. (1995), for the Inglewood tephra erupted from Mt Taranaki/Egmont (New Zealand). Prominent south-west and north-westerly winds suggest two semi-elliptical lobes north-east and south-east of the vent.

The most widely used empirical methods for deriving volumes estimates, involve integration of various thinning relationships using the thickness and area of isopach maps. Difficulties in finding sites suitable for tephra preservation, limited data, and the subjectivity involved in the construction of isopach maps, means volume calculations are not straight forward.

2.4.1 Exponential thinning

Thorarinsson (1954) first proposed that the thickness of tephra deposits decays exponentially with distance from the vent. Pyle (1989) suggested that this exponential thinning could be utilized by studying isopach maps, and plotting the logarithm of thickness ($\log T$) against the square root of the area enclosed by an isopach contour ($\sqrt{A}$). He demonstrated that many

Figure 2.10: Isopach map for the Inglewood Tephra. Isopachs in centimeters. Reprinted from Alloway et al. (1995), with kind permission from Taylor and Francis.



deposits display straight line behavior on a plot of $\log T$ versus $\sqrt{A}$. The exponential thinning relationship between thickness and the square root of area can be described by the function

$$T = T_0 \exp(-k\sqrt{A}), \tag{2.29}$$

where T_0 represents the maximum thickness extrapolated at $A = 0$, and k is the slope of the line on a plot of $\log T$ versus $\sqrt{A}$. The maximum thickness T_0 need not be at the vent, it can be displaced downwind.

Pyle (1989) assumed that isopachs were elliptical in shape and calculated the volume by integrating (2.29) with respect to thickness, establishing the volume relationship

$$V = \frac{2\pi T_0 b_t^2}{(\log 2)^2 \alpha} = 13.08 \frac{T_0 b_t^2}{\alpha}, \tag{2.30}$$

where $b_t = k/\log 2$ is the thickness half-distance and α relates to the eccentricity of the elliptical isopachs. The derivation of (2.30) is given in Appendix 1 of Pyle (1989). The thickness-half distance b_t was defined by Pyle (1989) as the distance, measured along the minor axis of the isopach ellipse, at which the deposit thickness halves (analogous to half-life in radioactive decay). For the case of circular isopachs $\alpha = 1$ and (2.30) reduces to

$V = 13.08 T_0 b_t^2$.

The volume equation derived by Pyle (1989) assumes isopachs are elliptical. Fierstein and Nathenson (1992) relaxed that assumption by integrating (2.29) with respect to area rather than thickness. Therefore, volume is represented by the integral

$$V = \int_0^\infty T \, \mathrm{d}A = \int_0^\infty T_0 \exp(-k\sqrt{A}) \, \mathrm{d}A, \tag{2.31}$$

which can be evaluated by applying a change in variable $\omega = \sqrt{A}$ such that $\mathrm{d}A = 2\omega \, \mathrm{d}\omega$

$$\begin{aligned} V &= \int_0^\infty 2T_0 \omega \exp(-k\omega) \, \mathrm{d}\omega \\ &= \left[\frac{-2T_0(k\omega + 1)}{k^2} \exp(-k\omega) \right]_0^\infty \\ &= \frac{2T_0}{k^2}, \end{aligned} \tag{2.32}$$

where k is the slope of the straight line on a plot of $\log T$ versus $\sqrt{A}$. Equation (2.32) is valid for any shape of isopach. Therefore, it provides more flexibility than the representation given in (2.30).

Pyle (1989) extended his method presented above, by recognizing that some tephra deposits are better characterized by two straight line-segments on a plot of $\log T$ vs $\sqrt{A}$, where the slope of the proximal line-segment is steeper than that of the distal line-segment. This was explained theoretically by Bonadonna et al. (1998) as the result of differences in grain-size.

Multiple line-segments can be incorporated into the volume estimation by separating the integral in (2.32). This involves calculating the limits from 0 to Z and Z to infinity, where Z denotes the area at which the break-in-slope occurs. Break-in-slope points can be arbitrarily chosen by eyeballing the data. Alternatively an iterative approach, searching for the break-in-slope point which gives the lowest residual mean square error, could be used. Break-in-slope points can also be estimated using the segmented package in R (Muggeo, 2008; R Core Team, 2014). The importance of properly accounting for changes in slope was illustrated by Fierstein and Hildreth (1992). They noted that, for deposits that exhibit a break-in-slope, using a single line-segment can significantly underestimate the volume erupted.

Volume estimation is complicated by a number of things, mainly due to the general scarcity of data. Proximal deposits can be difficult to obtain due to burial from more recent eruptions

or vent collapses, whereas distal tephra deposits are very thin and can be rapidly eroded. Difficulties in accessing older tephra deposits, and poor preservation rates, makes estimating the eruptive volume of prehistoric events more problematic. Tephra deposits can be widespread and isopach maps rarely constrain the total area. Therefore extrapolation of field data is required to integrate the exponential thinning relation in (2.29) across infinite area (to zero thickness). Deciding how far from the vent isopach data must be accounted for is a crucial problem. Integrating beyond the preserved limits is risky as thinning rates may change (Fierstein and Nathenson, 1992). Significant underestimations in eruptive volume can occur when dealing with proximal and distal extrapolation. Limitations in isopach data means that methods of Pyle (1989) and Fierstein and Nathenson (1992) can generally only provide minimum volume estimates. The volume of tephra lost beyond areas constrained by isopach maps cannot be quantified. But some understanding of minimum volume may still be useful (Legros, 2000).

Legros (2000) proposed a method to calculate minimum eruptive volumes, for the extreme case when only one reliable isopach is available. The method assumes exponential thinning and can be applied to deposits which are characterized by one or more straight-line segments on a plot of $\log T$ versus $\sqrt{A}$, provided the proximal segments have steeper slopes. For an isopach of thickness T, there exists a slope ($k_{\min}$) corresponding to a minimum volume ($V_{\min}$). Rearranging (2.29) and substituting into (2.32) gives

$$V = \frac{2T\exp(k\sqrt{A})}{k^2}. \tag{2.33}$$

The slope corresponding to the minimum volume can be found by differentiating (2.33) with respect to k and equating to zero:

$$\frac{\mathrm{d}V}{\mathrm{d}k} = \frac{2T\sqrt{A}\exp(k\sqrt{A})}{k^2} - \frac{4T\exp(k\sqrt{A})}{k^3} = 0, \tag{2.34}$$

such that $k_{\min} = 2/\sqrt{A}$, and

$$V_{\min} = \frac{e^2}{2}TA = 3.69TA. \tag{2.35}$$

Although an estimate of the total eruptive volume would be preferred, this is never really known due to sparse data and difficulties with extrapolating thickness measurements from distal to proximate sites. The simple approximation in (2.35) provides a constraint on the minimum volume.

2.4.2 Power-law relationship

Bonadonna and Houghton (2005) illustrate that many well preserved deposits do not show a simple exponential decay, and would instead be better represented by two *or more* straight-line segments on a plot of $\log T$ versus $\sqrt{A}$. Although piecewise exponential functions may enhance the fit, extra uncertainty is involved in subjectively deciding the number of segments to fit and where the break-in-slope points should occur.

Bonadonna et al. (1998) showed that large changes in thinning rate occur with distance from the vent. They suggested that thinning of very small distal tephras is better described by a power law. Bonadonna and Houghton (2005) proposed the relationship

$$T = c(A^{1/2})^{-m} = cA^{-\frac{m}{2}}, \tag{2.36}$$

between the square root of the area ($A^{1/2}$) enclosed by an isopach line of thickness T, where c and m are free parameters. The volume integral for the power law relationship is given by:

$$\begin{aligned} V &= \int_0^\infty T \,\mathrm{d}A = \int_0^\infty cA^{-\frac{m}{2}} \,\mathrm{d}A \\ &= \left[\frac{2c}{2-m} A^{\frac{2-m}{2}}\right]_0^\infty \end{aligned} \tag{2.37}$$

Although the power-law function avoids the need to select the number of segments and break-in-slope points, it requires arbitrary integration limits to be defined for the calculation of volume estimates. Equation (2.36) is not capable of being integrated between zero and infinity, therefore alternative upper and lower integration limits for (2.37) must be specified for meaningful volume estimates to be obtained. For the lower limit, Bonadonna and Houghton (2005) suggested using the area at which the thickness in (2.36) reaches the maximum thickness T_0 derived under exponential thinning. The upper limit should be set large enough to cover the downwind extent of the whole deposit.

2.4.3 Weibull function

While the power-law method better describes the natural thinning of tephra deposits, in comparison to the exponential thinning method, it is sensitive to the choice of finite integration limits. These limits are largely specified according to the availability of proximal and distal data. To overcome the problems associated with the selection of arbitrary integration limits,

or the number and location of break-in-slope points, Bonadonna and Costa (2012) suggested a Weibull function to model the relationship between the thickness (T) and the square root of isopach area ($\sqrt{A}$). The thickness is calculated by

$$T = \theta \left(\sqrt{A}/\lambda\right)^{n-2} \exp\left(-\left(\sqrt{A}/\lambda\right)^{n}\right), \tag{2.38}$$

where θ, λ, and n are free parameters chosen by minimizing the weighted residual sum of squares

$$\sigma^2 = \sum_{i=1}^{N} w_i \left[T_i - \hat{T}_i\right]^2, \tag{2.39}$$

where w_i are weighting factors, N is the number of isopachs, and T_i and $\hat{T}_i$ denote the observed and calculated thicknesses, respectively. Bonadonna and Costa (2012) applied a proportional (quadratic) weighting scheme $w_i = 1/{T_i}^2$. Alternatives, such as no weighting ($w_i = 1$) and linear weighting ($w_i = 1/T_i$), were also been considered by Klawonn et al. (2014b). They found estimated volumes increased with the degree of weighting (from no weighting to quadratic weighting), but associated standard deviations decreased. Therefore, the choice of weighting scheme can introduce extra uncertainty into estimates.

The volume of the Weibull function is calculated by integrating (2.38) with respect to area:

$$\begin{aligned} V &= \int_0^\infty T \,\mathrm{d}A \\ &= \int_0^\infty \theta \left(\sqrt{A}/\lambda\right)^{n-2} \exp\left(-\left(\sqrt{A}/\lambda\right)^{n}\right) \mathrm{d}A \\ &= \left[-\frac{2\theta\lambda^2}{n} \exp\left(-\left(\sqrt{A}/\lambda\right)^{n}\right)\right]_0^\infty \\ &= \frac{2\theta\lambda^2}{n} \end{aligned} \tag{2.40}$$

The exponential thinning relationship proposed by Pyle (1989) and Fierstein and Nathenson (1992) is a special case of the Weibull function for $n = 1$.

Figure 2.11 illustrates the fit of exponential, Weibull, and power-law models, to the 18 May 1980 Mt St Helens eruption (Sarna-Wojcicki et al., 1981), and the 1996 Mt Ruapehu eruption (Bonadonna and Houghton, 2005). Arbitrary integration limits of 0.3 km and 1,000 km were chosen for the power law, for consistency with the analysis in Bonadonna and Houghton (2005). Model equations and estimated total volumes are given in Table 2.1. While the two-segment exponential and Weibull models appear to fit well to the Mt St Helens eruption, the deposits cannot be accurately described by a power law. This observation was also made

Figure 2.11: Semilog plots of thickness versus square root of isopach area for (A) Mount St Helens 1980 and (B) Ruapehu 1996. Comparisons between best fit exponential, Weibull, and power law models obtained using AshCalc (Daggitt et al., 2014).

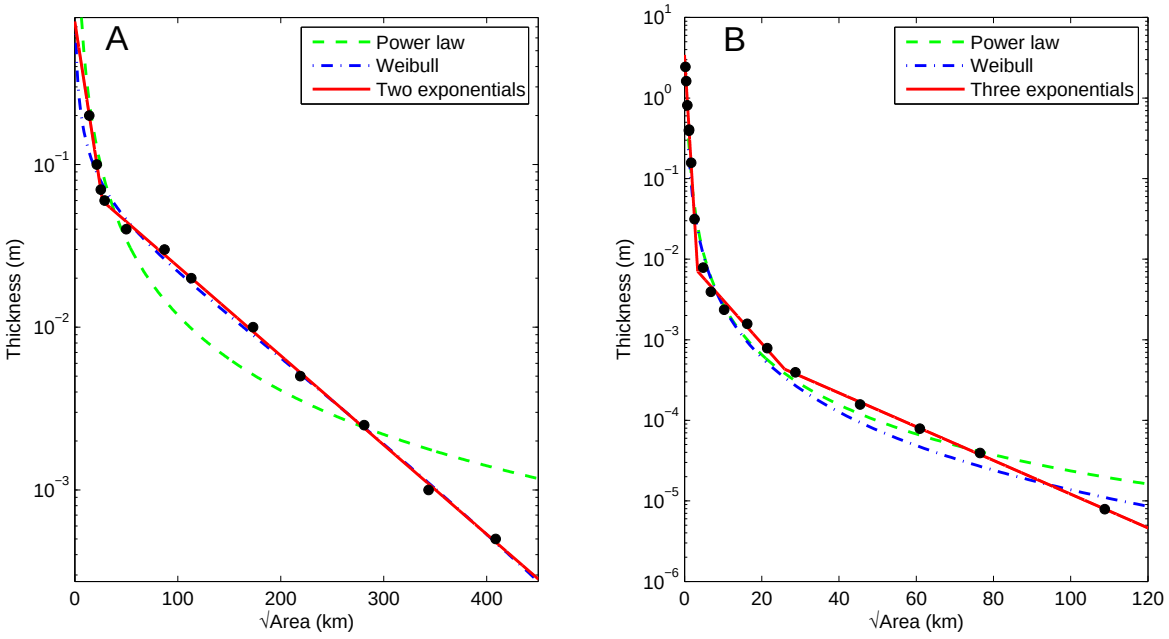


Table 2.1: Comparison between total volume estimates of the exponential, Weibull, and power law models fitted to the 18 May 1980 Mt St Helens eruption and the 1996 Ruapehu eruption using AshCalc (Daggitt et al., 2014). RMSE denotes the relative mean square error, the relative error associated with the empirical fit of observed thicknesses.

	Mount St Helens		**Volume (± RMSE)**
Exponential	$T = 0.757\exp(-0.094\sqrt{A})$,	$\sqrt{A} \leq 26.97$	
	$T = 0.084\exp(-0.013\sqrt{A})$,	$\sqrt{A} > 26.97$	1.124 km^3 (0.3%)
Power law	$T = 14.230\sqrt{A}^{-1.539}$		1.453 km^3 (40.1%)
Weibull	$T = 0.026((0.006\sqrt{A})^{-0.622})\exp(-(0.006\sqrt{A})^{1.379})$		1.058 km^3 (2.5%)
	Mount Ruapehu		**Volume (± RMSE)**
Exponential	$T = 3.435\exp(-1.818\sqrt{A})$,	$\sqrt{A} \leq 3.402$	
	$T = 0.011\exp(-0.124\sqrt{A})$,	$3.402 < \sqrt{A} \leq 25.95$	
	$T = 0.002\exp(-0.048\sqrt{A})$,	$\sqrt{A} < 25.95$	0.004 km^3 (1.4%)
Power law	$T = 0.315\sqrt{A}^{-2.063}$		0.006 km^3 (35.7%)
Weibull	$T = 0.023((0.178\sqrt{A})^{-1.663})\exp(-(0.178\sqrt{A})^{0.337})$		0.004 km^3 (15.9%)

by Bonadonna and Houghton (2005), where it was explained as a result of the particle-aggregation processes significantly affecting the thinning trend of fine ash particles. Models fitted to the Mount Ruapehu deposits appear to give similar volume estimates (Table 2.1). However, both the power-law and Weibull models appear to overestimate distal thickness deposits while underestimating deposits at medial distances ($\sqrt{A} \sim 20-80$ km)(Figure 2.11).

The suitability of the various empirical methods for estimating volumes depends on the data and deposit-thinning relationship at hand. The two examples illustrated in Figure 2.11 are for particularly well known eruptions. Greater uncertainty in model selection and volume estimates will arise for deposits that have less available data. All of the models so far make use of hand-drawn contours, which also introduces uncertainty into volume estimates. The location of sampling sites, the subjective process of constructing isopach maps, and variation in empirical fitting methods, can all lead to differences in volume estimates.

2.4.4 Use of isopach data

The construction of isopach maps involves subjective choices with respect to the number and spacing of contour lines. Individuals with varying degrees of experience may adopt different approaches in constructing isopach maps. Klawonn et al. (2014a) investigated the variability in volume estimates introduced by hand-drawn isopach maps. The effect of different empirical methods on volume estimates was later explored by Klawonn et al. (2014b). Identical maps of raw thickness measurements for the 1959 Kilauea Iki fall deposit were distributed to 101 volcanologists worldwide, who were subsequently instructed to create hand-drawn isopach maps.

The effect of sampling density was investigated by distributing four maps to each participant. Two were densely populated with 273 and 150 thickness measurements respectively. The other two were sparsely populated with only 75 thickness measurements each. One retained thickness measurements located along roads and trails, while the other had roughly evenly spaced thickness measurements. Klawonn et al. (2014a) found the volume estimates obtained using exponential smoothing were relatively consistent across all sampling densities, and surprisingly, that there was no significant relationship between the number or spacing of contours and the estimated volumes.

The later study by Klawonn et al. (2014b) calculated volume estimates using the exponential, power-law, and Weibull functions discussed above. They found that volume estimates from a

single map varied up to a factor of 4.9 with the use of different empirical estimation methods. They also found large uncertainties for all volume estimation methods for the proximal ($s = 62\%$) and distal ($s = 53\%$) deposits, whereas uncertainties for medial deposits was relatively small ($s = 8\%$). These uncertainties were largely introduced by participants' choices for the smallest and largest isopachs. The relative uncertainty impacts total volume estimates differently. Spread within the small distal deposits has a small impact on total volume, whereas large uncertainties in the thicker proximal deposits strongly affects volume estimates.

The findings of Klawonn et al. (2014a,b) emphasize the problems associated with using isopach maps to derive volume estimates. Even with a large number of data points, estimated volumes can differ greatly with the use of different empirical approaches. Chapter 5 focuses on estimating eruption volumes for a series of Mt Taranaki events. Data preserved in unexposed locations around Mt Taranaki has as few as one measurement for some eruptions. With so few data, the construction of isopach maps is not possible. Hence a new paradigm is required to deal with sparse data.

Despite the numerous uncertainties involved in estimating volumes, the empirical methods discussed so far remain widely used. The main issue with the empirical methods is the predictive capabilities in proximal and distal regions, where thickness measurements are sparse. The empirical approaches also avoid the question of sampling errors and variations in thicknesses. Measurements are taken as given, and the effect of post-depositional processes is largely ignored.

2.4.5 Fully statistical approaches

More recent model developments allow for uncertainty in field measurements by modeling raw thickness data without the construction of isopach maps. Rhoades et al. (2002) utilized a data set of eruptions from Taupo Volcano (New Zealand), to propose a linear relationship between the logarithm of tephra thickness (T in cm) and eruptive volume (V in km^3)

$$\log T = a + \frac{c+1}{3}\log V - c\log(r + dV^{1/3}) + f(\theta), \tag{2.41}$$

where r and θ denote the distance (in km) and angular direction (in radians) from the center of the deposit, respectively. The term $dV^{1/3}$ was included to ensure that thicknesses are finite at the vent (i.e. when $r = 0$).

The directional function $f(\theta)$ is a semi-elliptical term for wind

$$f(\theta) = \sum_{j=1}^{n} \alpha_j \sin(j\theta) + \beta_j \cos(j\theta). \tag{2.42}$$

The parameters a, c, d, and wind-based terms involved in (2.42), were estimated by Rhoades et al. (2002) along with their standard errors. Only those α_j and β_j involved in (2.42) that are significantly different to zero were included in the model fit.

The resulting parameter estimates have been used by Bebbington et al. (2008) to provide a probabilistic hazard assessment for Mt Taranaki (New Zealand). They translated observed tephra thicknesses into eruptive volumes by inverting the relationship proposed by Rhoades et al. (2002). Bebbington and Cronin (2011) also used the model of Rhoades et al. (2002) as a means of determining the likely source of tephra observed in the Auckland region (New Zealand), from among the many vents in the Auckland Volcanic Field (New Zealand). They isolated r to find the distance and volume expected to produce observed tephra thicknesses.

The relationship proposed by Rhoades et al. (2002) assumes the variability in thicknesses is lognormally distributed. Gonzalez-Mellado and De la Cruz-Reyna (2010) proposed an alternative relationship for tephra thickness

$$T(r, \theta) = \gamma r^{-\alpha} \exp\left[-\beta U r(1 - \cos\theta)\right], \tag{2.43}$$

which is dependent on the distance r and angle θ from the center of the deposit. U denotes the mean wind speed (km h^{-1}), assumed to be constant. If there is no significant wind, then (2.43) is equivalent to the power law relationship $T(r, \theta) = \gamma r^{-\alpha}$. The parameter α determines the rate at which the deposit thickness decays with distance. It is inversely related to column height. The parameter β is inversely related to the effective diffusion coefficient, and can be regarded as a proxy for the grain-size distribution (Kawabata et al., 2013).

Gonzalez-Mellado and De la Cruz-Reyna (2010) adjusted the model parameters recursively, to fit sets of thickness data obtained from reported isopachs, by minimizing the weighted residual sum of squared differences. Kawabata et al. (2013) developed a method to statistically fit the relationship from Gonzalez-Mellado and De la Cruz-Reyna (2010) to actual point thickness measurements. Kawabata et al. (2013) used standard maximum likelihood techniques, which avoided having to choose a weighting scheme in the least-squares minimization procedure.

The semi-empirical model proposed by Gonzalez-Mellado and De la Cruz-Reyna (2010), as adapted by Kawabata et al. (2013), has its advantages. It does not assume a specific wind profile, and uncertainties in thickness measurements are incorporated into the model fit. However, (2.43) does not include volume as a parameter. Therefore, complex numerical integration would be required to extract volume estimates suitable for supporting models for hazard forecasting. For this reason the model of Rhoades et al. (2002) (2.41) may be preferred, particularly for sedimentary data where over-thickening consistent with the lognormal distribution may be expected.

Burden et al. (2013) also approached the idea of estimating eruptive volumes without the construction of isopach maps. They proposed a linear regression model for the logarithm of thickness with distance from the vent. They used Bayesian methods to account for data and model uncertainty. However, a major drawback of their proposed linear regression model is that it ignores direction. Expected thickness, and its variance, is a function of distance *and* direction from the vent. Therefore the relationship cannot be explained by a single linear regression. Every thickness measurement lies on a different slope, which is dependent on the direction from the vent. The model of Burden et al. (2013) suggests *one* slope governs the entire distribution of thickness measurements. The 4600 BP deposit of Fogo A (Azores) as a case study. The thickness measurements of this eruption show little wind interaction. Therefore, the tephra dispersal is relatively symmetric around the vent. Due to the disregard for direction, the linear regression model would become highly biased with even slightly asymmetric tephra dispersal, or sparse data that is not evenly distributed around the vent. Moreover Burden et al. (2013) do not perform any residual analysis to validate their regression assumptions. Although they say that their estimates "show good correlation with previous estimates of volume" (p. 707), none of their confidence intervals constructed include the best estimates determined by other methods.

Establishing standard and reliable approaches for volume estimation, especially where data is sparse, remains a major scientific challenge facing volcanologists. Incorporating the uncertainty in data and parameter values into volume estimates is not straightforward. The estimation of tephra volumes without the need for constructing isopachs is a new development.

In Chapter 5, I investigate estimating eruptive volumes in a Bayesian framework following an adaptation of the model from Rhoades et al. (2002). Point thickness observations of Mt Taranaki sourced tephras are utilized, as opposed to making inferences from isopach maps.

Unlike the early empirical methods above, which focus on well documented deposits from one eruption, the analysis in Chapter 5 estimates the volume of *multiple* events from a long spanning (> 70,000 yrs) eruption record. Measurements from both exposed and unexposed locations are utilized. The major advance considered in Chapter 5 is the impact of site-specific effects on volume estimates, resulting from varying levels of suitability for tephra preservation.

2.5 Summary

This chapter has introduced some of the terminology and concepts that are referred to throughout this thesis. Some of the major complexities inherent in historical and geological records of events have been highlighted. Current approaches to compiling eruption records were reviewed. Existing models for forecasting the timing of eruptions, and developments in methods of estimating eruptive volumes, were discussed. Specific theories and models introduced in this chapter lend themselves to the work which follows. More detail will be provided in subsequent chapters where relevant.

Chapter 3

Incorporating ancillary geochemical data into probabilistic hazard estimation

Most probabilistic hazard models for eruption occurrence are purely temporal, in the sense that they only consider the distribution of event or inter-event times as predictors of further volcanic activity (see Section 2.3). In statistical seismology these are called 'catalog' models. It has been hypothesized, however, that there are underlying cycles in geochemistry data that could be linked to eruption rate (Turner et al., 2008b). This chapter demonstrates, using a high-resolution Holocene volcanic event record from Mt Taranaki, how geochemical data can be used to modulate a renewal model for the estimated probability of a future eruption. Work discussed in this chapter is published in Green et al. (2013).

3.1 Introduction

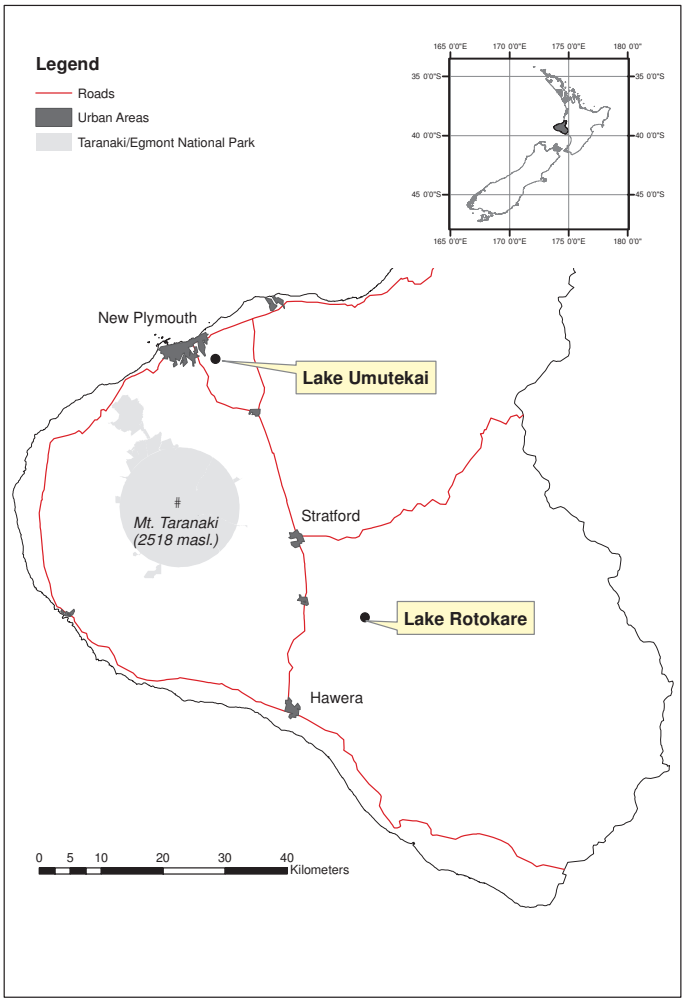
The advantages of forecasting future volcanic eruptions in a probabilistic framework have been outlined by Marzocchi and Bebbington (2012). Over a time scale of years or more, probabilistic forecasts are important for land-use planning and comprehensive risk management. In contrast to short-term eruption forecasting, which is typically based on geophysical monitoring information such as earthquakes, gas concentrations and ground deformation; long-term forecasts must be derived from the history of the volcano, especially for those volcanoes that exhibit long periods of repose (Decker, 1986). Reasonable models cannot predict long repose

periods unless they are present in the historical records. Therefore, sufficient amounts of high quality data are required to produce accurate forecasts. Stochastic models for eruption onsets have developed from the homogeneous Poisson process (Wickman, 1966; De la Cruz-Reyna, 1991), through the nonhomogeneous Poisson process (Ho, 1991; Bebbington and Lai, 1996a) and renewal processes (e.g., Bebbington and Lai, 1996b; De la Cruz-Reyna and Carrasco-Nunez, 2002; Turner et al., 2008a; Garcia-Aristizabal et al., 2012), to a unifying treatment via the Trend Renewal Process (Bebbington, 2010) (see Section 2.3). All these approaches are purely temporal, in that they use only the information from the eruption onset times, whereas a parallel line of development has explored the utility of ancillary information such as the eruptive volumes (Burt et al., 1994; Sandri et al., 2005; Marzocchi and Zaccarelli, 2006; Bebbington, 2008; Passarelli et al., 2010a) or durations (Bebbington, 2007). Models incorporating volume have also been related to the petrology (origin, structure, and composition) (Hill et al., 1998) and stress-field (Wadge and Burt, 2011) of the volcano.

Although incorporating some measure of previous eruption volume into forecasts of future eruption timing can be shown (cf. Passarelli et al., 2010b), in many cases, to improve the model performance, it requires reliable, or at least consistent, volume estimates. These are rarely available, even for modern volcanoes. Only a very small number of well studied volcanoes and particular eruptions volumes have been estimated (as discussed in Section 2.4). Chapter 5 will explore methods of estimating eruption volumes, which could subsequently feed into probabilistic hazard models, but here ancillary information is considered in the form of titanomagnetite chemistry. Titanomagnetite is a mineral containing oxides of titanium and iron. The titanomagnetite chemistry records utilized here detail the percentage of certain elements (namely titanium, magnesium, and aluminium oxides) present in erupted material.

Mt Taranaki (New Zealand) is a typical stratovolcano (a conical volcano built up by many layers of volcanic deposits) that has been active for at least 100 kyr (Zernack et al., 2012), but not erupted since the mid-1800s. A catalog of Holocene eruptions from Mt Taranaki was assembled by combining tephras in lake sediment records with near-source tephrostratigraphy (Turner et al., 2008a, 2009, 2011b) (see Section 2.2.6 for a description of how the records were combined). From these records, underlying cycles in magma chemistry can be linked to the eruption rate. In particular, titanomagnetite has proved to be useful in elucidating long-period cycles in volcanic activity. There are phases of high to low magnesium oxide compositions which Turner et al. (2011a) interpreted as indicating a magma recharge from the lower depths of the volcanos reservoir. Erupted material may also be sourced directly

Figure 3.1: Map showing the location of Lake Umutekai and Lake Rotokare



from lower levels of the crust, or may gestate higher in the magma plumbing system (Turner et al., 2008b). Titanomagnetite chemistry is helpful in distinguishing these different magmatic processes. This chapter considers how such geochemical indices can be best incorporated into improved temporal volcanic hazard forecasts, using both event age and geochemical data from Mt Taranaki to illustrate the methodology.

The rest of the chapter is organized as follows. First is a description of the data utilized, a case study from Mt Taranaki. This is followed in Section 3.3 by a review of some of the renewal models introduced in Section 2.3.3, an outline of proportional-hazards models, and a discussion of how both model types will be fitted to the Mt Taranaki data. The results for both baseline (pure renewal) and geochemistry-enhanced models are presented in Section 3.4, and discussed in Section 3.5.

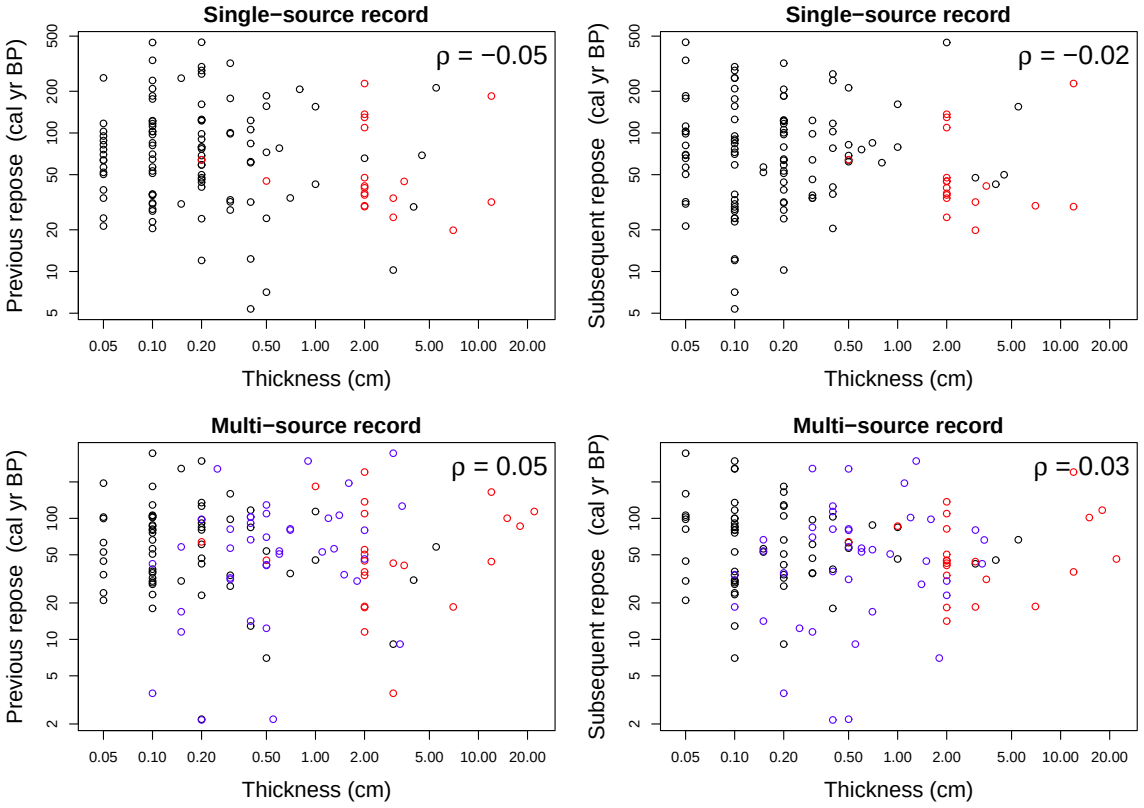
3.2 Data: Mt Taranaki eruption record

The eruption records used for this analysis are from Mt Taranaki, New Zealand. The first event record (Turner et al., 2008a), which will henceforth be termed 'single-source', was extracted from a sediment core from Lake Umutekai, NNE of the volcano. It contains 103 tephra units originating from Mt Taranaki, with ages ranging from 1400 cal yr BP to 11500 cal yr BP (see Section 2.2.4 for interpretation of 'BP'), interpolated via sediment thickness from a number of ^{14}C age determinations (see Figure 2.2, Section 2.2.4). This was extended to the present by using the most recent 19 eruptions in a stratigraphic record of 23 individually ^{14}C dated eruption deposits on the volcano upper flanks. The record used in this work differs from that in Turner et al. (2008a) in being presented in cal yr rather than ^{14}C yr.

The second record (Turner et al., 2009) was assembled by merging the data discussed above with an additional record of 42 tephras from Lake Rotokare, located SE of the volcano. Duplicates (correlated tephras) between sites were removed by statistical matching of ages and geochemistry, as detailed in Turner et al. (2009) (see Section 2.2.6). Again, the catalog used here has been calibrated in calendar years. Both lakes Umutekai and Rotokare are located <35 km from the volcano (Figure 3.1), within range to record both minor eruptions and those producing larger tephra falls.

Bebbington et al. (2008) estimated the volume of the eruptions corresponding to the tephra layers in the Umutekai core, by inverting a volume-distance-thickness relation from Rhoades et al. (2002) (see Section 2.4.5). Chapter 5 will revisit volume estimates for Mt Taranaki events. However, Bebbington et al. (2008) suggest that thicknesses of 0.5 mm indicate an expected minimum eruption volume of 10^7 m^3 (VEI 2-3), while thicknesses of 50 mm at Umutekai correspond to a likely eruption volume on the order of 1 km^3 (VEI 4-5). VEI refers to the volcanic explosivity index, which provides a relative measure of the explosiveness of volcanic eruptions. VEIs range from 0 to 8, depending primarily how much volcanic material is ejected during the eruption phase (see, Newhall and Self, 1982). However, other variables, such as the height of the eruption column and the eruption duration, are also considered in assigning a VEI level to an eruption. The VEI scale is logarithmic. Therefore, an increase of one on the scale, suggests an eruption ten times more powerful. No correlation was observed between the tephra thickness and either the preceding or subsequent repose length (see Figure 3.2). Therefore, occurrence rate can be modeled independently of eruption size (Bebbington et al., 2008).

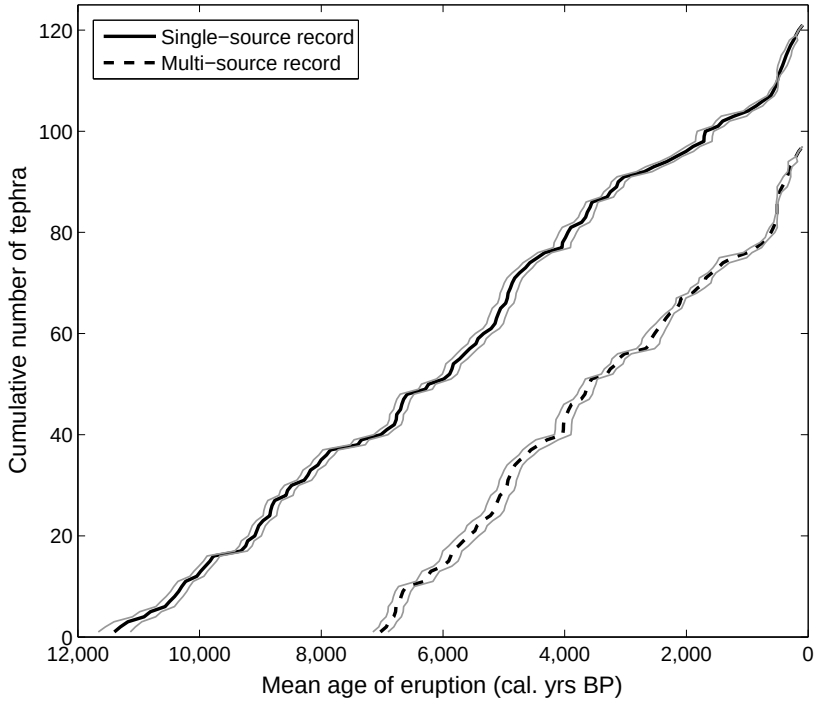
Figure 3.2: Observed tephra thicknesses versus previous and subsequent repose periods. Lake Umutekai (black), Lake Rotokare (blue), and Near-source (red). ρ denotes spearman rank correlations.



As the older part of the second record is sourced solely from the single site of lake Umutekai, a noticeable change in the observed event frequency occurs at around 7000 cal yr BP (Turner et al., 2009). Hence only the most recent 97 eruptions, which will be termed the 'multi-source record' can be considered to be time-homogeneous in nature, and will be analyzed here. The longer record of Turner et al. (2008a), which is derived from non-overlapping (in time) sources (excluding Rotokare), is also time-homogeneous and will be analyzed in parallel to provide a sensitivity analysis on the robustness of the methodology.

The observed mean event age record is shown in Figure 3.3 for both the single-source and multi-source records, which confirms that both records appear time-homogeneous. Each age is in fact a probability distribution, which is no longer normal after converting (McCormac et al., 2004) from ^{14}C to cal yr. Instead each distribution is transformed by the calibration procedure into a unique distribution; these may be multimodal (see, for example, Figure 2.1). These catalogs have been estimated to be approximately 79% (single-source) and 86% (multi-source) complete (Wang and Bebbington, 2012).

Figure 3.3: Event ages in the single-source and multi-source records. Bands indicate 2 standard error limits on the ages.



The geochemical data (Turner et al., 2008b, 2009, 2011a) presented in Appendix A, represents the percentage weight of TiO_2, Al_2O_3, and MgO present in the erupted material. Titanomagnetite composition is easy to measure via electron microprobe analysis and can be used to discriminate between eruptive sources and individual eruptions (Cronin et al., 1996; Turner et al., 2009). In addition, titanomagnetite compositions are very sensitive to details of magma storage times and conditions. If there is strong variability in the MgO and Al_2O_3 components of titanomagnetite crystals of a single eruption, it reflects a mixed or mingled magma. This means, particularly in larger-volume eruptions, that the magma erupted may have been assembled from two or more initially independent batches that mixed or mingled before eruption. Hence, variable MgO and Al_2O_3 could be a proxy for mid-crustal magma residence time before eruption, and thus be correlated to eruption intervals. Changes in temperature, or oxygen fugacity (or both) affect TiO_2 concentrations in andesitic magmas (Buddington and Lindsley, 1964; Devine et al., 2003). Hence, titanomagnetites have been used at Mt Taranaki to distinguish eruptions with differing magmatic histories. The variability of TiO_2 is related to how rapidly magma rises from mid-crustal magma storage areas (c. 7-10 km depth) to eruption, with rapid rise leading to more homogenous compositions (Turner et al., 2008b, 2011a).

In this study, electron microprobe derived spot concentrations of the major constituents substituting for iron in magnetite (TiO_2, Al_2O_3 and MgO) from 10 individual grains for each sample were used. Analytical conditions are reported in Turner et al. (2008b, 2009, 2011a). The precision (1σ) in the measurements is 0.08%, 0.06% and 0.07% for TiO_2, Al_2O_3 and MgO, respectively (Brenna, 2012). Turner et al. (2009) used the principal components of the titanomagnetite analysis to reduce the dimensions of compositional data and assist in correlating duplicate tephras between sites (see Section 2.2.6). However, incorporating the geochemical data in this form into a model makes interpreting the resulting parameters difficult in a volcanological context. As discussed above, the variation in compositions could be as important from a volcanological perspective as the means. Hence, the mean and standard deviation of the major element 'contaminants' measured in the titanomagnetites are used in this analysis. The geochemical data is shown in Figures 3.4 and 3.5, for the single-source and multi-source records respectively. There is a broad trend over time towards lower concentrations of mean TiO_2, but the other means, and the standard deviations, do not exhibit any trend.

There are two issues with the structure of the data. First, not all recorded eruptions have geochemical data attached to them, because in some cases, very thin ash units (0.5 mm thickness) did not provide enough material for analysis. In survival analysis there are three standard approaches for missing covariates (Collett, 2003, p. 255); to use the previous value, to use the closest value (in time) to the eruption, or to linearly interpolate between the two. Here the first approach will be taken, with the geochemistry imputed to be that of the previous eruption, applying this recursively as necessary. This is justified by the fact that the eruptions with missing geochemistry are almost certainly small, and thus are unlikely to result from a major change in the magmatic or volcanic system. Moreover, it is the only approach consistent with forecasting, as the estimated covariate is not available prior to eruption in the other two approaches and must then be modeled separately. As this imputation is not possible for the oldest eruption in the record (number 104 from Umutekai), this observation is omitted from all calculations. Secondly, in cases where observations from different sites are correlated from the same eruption, the geochemical data are combined before calculating the mean and standard deviation. For each eruption there are n (may vary up to a maximum of 10) multivariate measurements $X_1, \ldots, X_n$ where X_i = (wt% TiO_2, wt% Al_2O_3, wt% MgO). The Matlab code to combine and sort available geochemical data, and impute missing geochemical data, is given in Appendix B.1.

Figure 3.4: Evolution of geochemical covariates for the single-source record. The correlations can be found in Table 3.1.

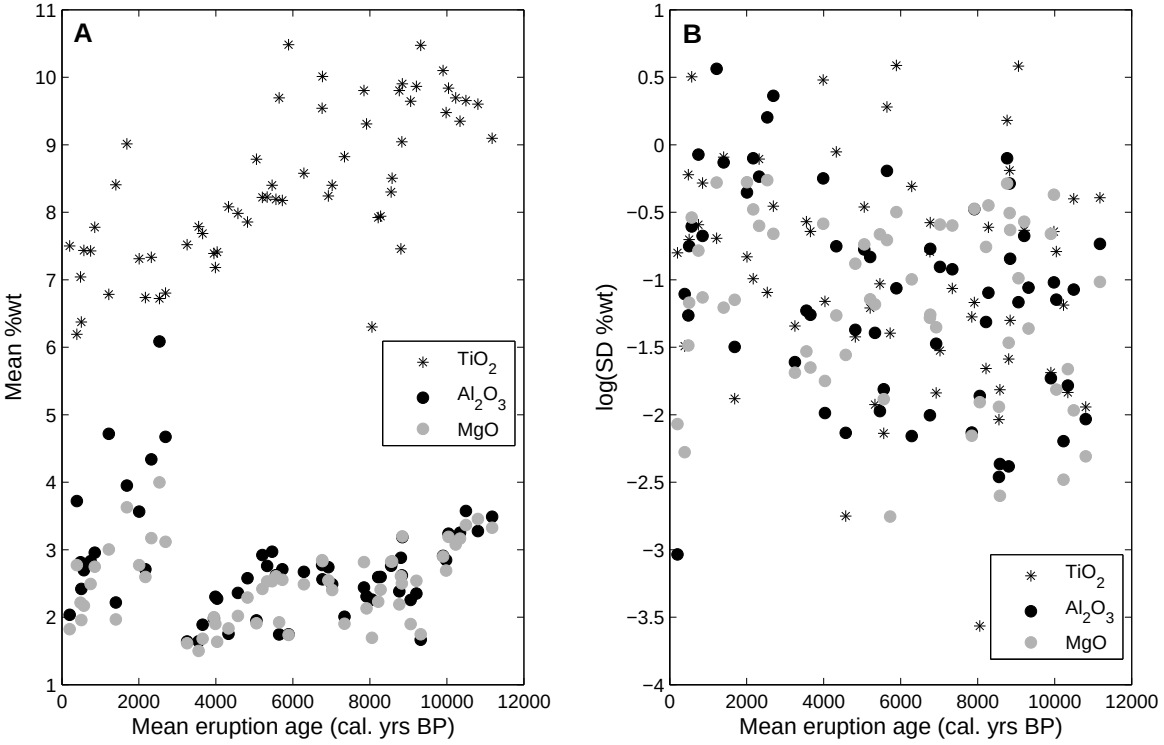


Figure 3.5: Evolution of geochemical covariates for the multi-source record. The correlations can be found in Table 3.1.

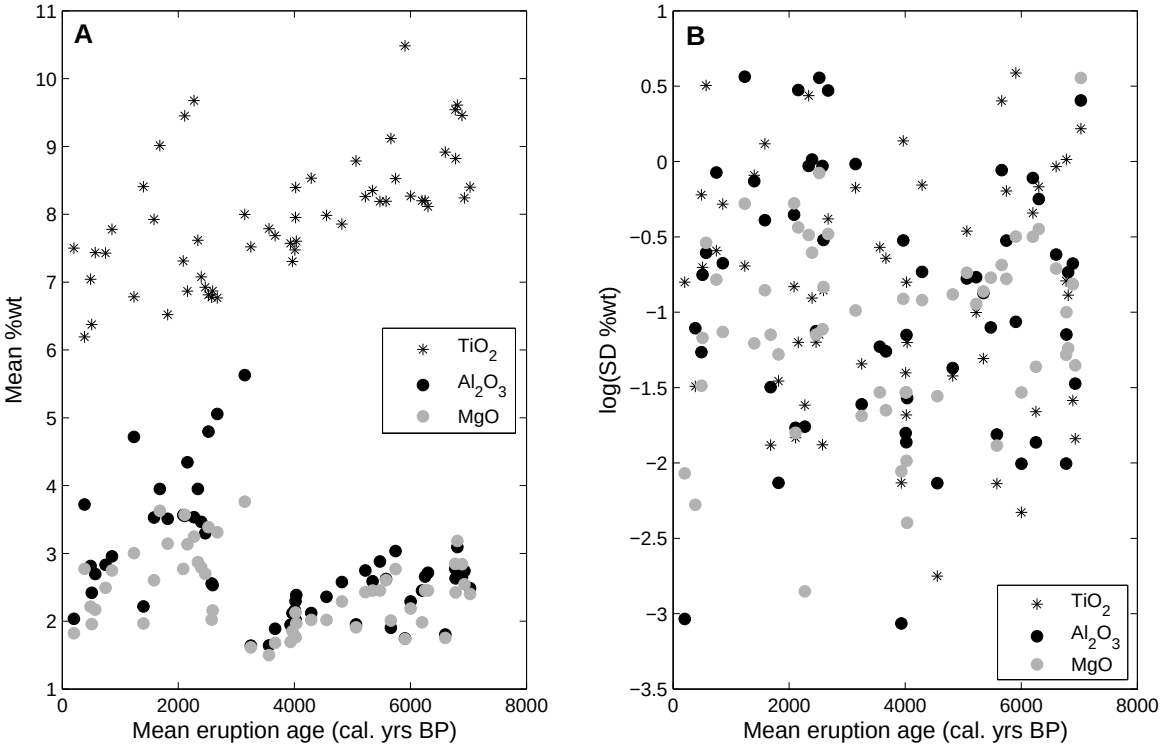


Table 3.1: Spearman rank correlations between repose lengths (τ) and geochemical covariates. Single-source on lower diagonal, multi-source on upper diagonal.

	τ	Mean(TiO_2)	Mean(Al_2O_3)	Mean(MgO)	SD(TiO_2)	SD(Al_2O_3)	SD(MgO)
τ		0.137	-0.001	0.069	-0.0125	0.119	0.229
Mean(TiO_2)	0.015		-0.240	-0.018	0.048	-0.233	-0.084
Mean(Al_2O_3)	0.126	-0.111		0.942	-0.079	0.363	0.304
Mean(MgO)	0.205	0.182	0.896		-0.182	0.262	0.250
SD(TiO_2)	-0.043	0.088	-0.183	-0.227		0.561	0.498
SD(Al_2O_3)	0.219	-0.116	0.076	0.012	0.627		0.798
SD(MgO)	0.093	-0.002	0.079	-0.001	0.462	0.737	

The Spearman rank correlations between the geochemical covariates and the repose times are shown in Table 3.1. None of the covariates are strongly correlated with the inter-event times when examined individually. However, Al_2O_3 and MgO are strongly correlated.

3.3 Methods: Probabilistic eruption forecasting

Recall from Section 2.3, that eruption occurrences can be modeled as a stochastic process that quantifies the randomness in the timings. This can be expressed as the hazard rate $\lambda(t)$ (elsewhere also known as the point process intensity), which specifies the instantaneous rate of event occurrence. If $N(s,t)$ denotes the random variable describing the number of events in the time interval (s,t), then

$$\lambda(t) = \lim_{dt \to 0} \frac{\Pr[N(t, t+dt) = 1]}{dt}. \tag{3.1}$$

We will suppose that events occur at times $t_1 < t_2 < \ldots < t_n$, with geochemical covariates $z_1, z_2, \ldots, z_n$, and that the time from t_1 to the present is T. The inter-event times (commonly known as reposes, although strictly reposes are the time from the end of one event to the start of the next) can then be defined as $\tau_1, \tau_2, \ldots, \tau_{n-1}$, where $\tau_i = t_{i+1} - t_i, i = 1, \ldots, n-1$, and $\tau^* = T - t_n$ is the current incomplete repose, which is a right-censored observation (i.e., $\tau^* < \tau_n$, the as-yet unknown repose).

3.3.1 Renewal models

As discussed in Section 2.3, there are many different possible parameterizations of $\lambda(t)$, depending on the data and situation. The simplest model is that of a Poisson process (see

Section 2.3.2), with a constant hazard function, $\lambda(t) = \lambda$. This means that events occur at a constant average rate, that the times between events are independent, and that the process is memoryless. That is, that the length of the current repose contains no information about the time remaining to the next eruption. Although this may sound unlikely, many volcanoes do exhibit such behavior at least some of the time (Marzocchi and Zaccarelli, 2006). A more general model can be developed by relaxing the memoryless tenet, resulting in a renewal process (see Section 2.3.3). In a renewal process the times between events are independent and identically distributed with probability density $f(\tau) = F'(\tau)$, and hence

$$\lambda(t) = \lambda(\tau) = \frac{f(\tau)}{1 - F(\tau)}, \tag{3.2}$$

where $\tau = t - s$, assuming that the previous eruption occurred at time s. Hence only the time since the previous eruption influences the timing of the next eruption.

As seen in Section 2.3.3, various distributions can be substituted into (3.2), each of which characterizes the time-varying hazard in different ways. Here we will follow the approach of Turner et al. (2008a) in considering, among others, the exponential density

$$f(\tau) = \beta \exp(-\beta\tau), \tag{3.3}$$

which is exactly the Poisson process. The Weibull renewal model (Bebbington and Lai, 1996b) with density

$$f(\tau) = \alpha\beta(\beta\tau)^{\alpha-1} \exp[-(\beta\tau)^{\alpha}], \tag{3.4}$$

has the capacity to model either under-dispersed (i.e. semi-periodic) or over-dispersed (i.e. clustered) inter-event times. When $\alpha = 1$ the Weibull density (3.4) reduces to (3.3), and hence the Poisson process is a special case of the Weibull renewal process. The lognormal renewal model (Bebbington and Lai, 1996a) with density

$$f(\tau) = \frac{1}{\tau\sigma\sqrt{2\pi}} \exp\left[-\frac{0.5(\log\tau - \mu)^2}{\sigma^2}\right], \tag{3.5}$$

only caters to under-dispersed data, but with data obtained from sediment records it is difficult to separate events that occur close together in time, and hence the lognormal is also a viable alternative. Further its relatively thick tail, compared to the Weibull, has the facility to model situations where there are a small number of relatively long reposes, as were observed in the single-source record (Turner et al., 2008a). These anomalously long reposes

led Turner et al. (2008a) to use the mixture of Weibulls density

$$f(\tau) = p\alpha_1\beta_1^{\alpha_1}\tau^{\alpha_1-1}\exp\left[-(\beta_1\tau)^{\alpha_1}\right] + (1-p)\alpha_2\beta_2^{\alpha_2}\tau^{\alpha_2-1}\exp\left[-(\beta_2\tau)^{\alpha_2}\right], \tag{3.6}$$

which allows for a bimodal distribution of repose lengths. Of course the mixture of Weibulls density (3.6) contains (3.4), and hence (3.3), as special cases.

Other models (such as those discussed in Section 2.3.6) generally require erupted volume information, which is not available here, or assume that the system, and hence the distribution of the reposes, changes over time. The latter does not appear to be the case from Figure 3.3, with the possible exception of the bimodality, or anomalously long reposes. As the objective here is to determine whether these can be tied to variations in the geochemistry, a steady state model (i.e., a renewal process) is required for reasons of identifiability; to ensure that long (or short) reposes are a feature of differences in geochemistry and not changes over time. Figures 3.4 and 3.5 show that the geochemistry evolves over time.

3.3.2 Proportional hazards

The natural way to incorporate extra covariates, such as geochemistry, into a renewal model is by using a proportional hazards (Cox, 1972) approach. This was an idea that was considered by Faenza et al. (2003) to characterize the spatio-temporal distribution of large earthquakes, and by Bebbington and Marzocchi (2011) to examine the triggering effect of large earthquakes on volcanic eruptions. A similar methodology can be applied here to model geochemical influences on the eruption process by attaching a vector of covariates to the inter-event hazard. Thus the hazard is a function of both elapsed time, and geochemistry:

$$\lambda(\tau, z) = \lambda_0(\tau)\exp\left(\delta^T z\right), \tag{3.7}$$

where $\lambda_0(\tau)$ is a baseline hazard function, and $z = \begin{pmatrix} z_1 & z_2 & \dots \end{pmatrix}^T$ is the vector of covariates, with corresponding vector of coefficients $\delta = \begin{pmatrix} \delta_1 & \delta_2 & \dots \end{pmatrix}^T$ as additional parameters to be fitted. The superscript T denotes the transpose. Note that each of the coefficients δ_j is a potential nuisance parameter, as the corresponding covariate need not be incorporated in the model.

In this model the covariates are acting multiplicatively on the hazard. Therefore, the model is an adaptation of Cox's proportional hazards (PH) model (Cox, 1972), which stipulates

that covariates have a constant proportional effect on the hazard. Although the Cox proportional hazards model assumes that the hazard is a function of the independent covariates, it doesn't assume any specific form for the baseline hazard $\lambda_0(\cdot)$ (Kalbfleisch and Prentice, 1980). Therefore, it is referred to as a semi-parametric model.

As this analysis deals with a small data set, in order to permit meaningful comparison, a parametric approach will be adopted to estimate the baseline hazard $\lambda_0(\cdot)$. In particular, it will be assumed that the baseline is one of the renewal hazard functions defined above.

The Weibull baseline distribution is particularly convenient in this regard because the hazard function (3.7) becomes

$$\lambda(\tau, z) = \alpha\beta^{\alpha}\tau^{\alpha-1}\exp\left(\delta^T z\right) = \alpha\left[\beta\exp\left(\delta^T z/\alpha\right)\right]^{\alpha}\tau^{\alpha-1}, \tag{3.8}$$

so the distribution with covariate values z is also Weibull, with the scale parameter β multiplied by the constant $\exp\left(\delta^T z/\alpha\right)$. The lognormal does not have this property, so if the baseline is taken to be lognormal then the addition of covariates leads to distributions that are not lognormal. For the mixture of Weibulls it is assumed that covariates act proportionally on the hazard for each separate component, so that β_1 and β_2 in (3.6) are adjusted for covariates as described above. Hence the model with a mixture of Weibulls baseline is not technically a proportional hazards model. However, it is a conditional proportional hazards model, i.e., the covariates act proportionally on the hazard when conditioning on the component of the mixture.

3.3.3 Model fitting and selection

The parameters in each model are estimated by maximizing the log-likelihood function

$$\log L(\tau_1, \ldots, \tau_{n-1}, \tau^*) = \sum_{i=1}^{n-1}\log f(\tau_i) + \log[1 - F(\tau^*)]. \tag{3.9}$$

If the model incorporates geochemical covariates, then the log-likelihood is given by

$$\begin{aligned}\log L(&\tau_1, \ldots, \tau_{n-1}, \tau^* \mid z_1, \ldots, z_n)\\ &= \sum_{i=1}^{n-1}\log f(\tau_i \mid z_i) + \log[1 - F(\tau^* \mid z_n)],\end{aligned} \tag{3.10}$$

where

$$F(\tau \mid z) = 1 - \exp\left(-\int_0^{\tau} \lambda(s, z)ds\right), \tag{3.11}$$

and $f(\tau \mid z) = (d/d\tau)F(\tau \mid z)$. Estimated standard errors for the parameters are obtained by inverting the Hessian matrix obtained from the numerical optimization solution (see, e.g., Press et al., 1986, pp 510-515). However, it must be noted that the error structure is highly correlated, and so the standard errors are not easily interpreted, and should not be used to infer confidence limits for parameters in isolation.

Instead, and as we have multiple models, with a number of geochemical covariates which may, or may not, be included, we need a way to select the best model. As different models can have different numbers of parameters, we need a means of compensating for the effect of additional parameters and thus avoiding over-fitting. As the models will be fitted using maximum likelihood, this can be done using the Akaike Information Criterion (Akaike, 1977),

$$\text{AIC} = -2\log L + 2k, \tag{3.12}$$

where k is the number of parameters in the model. Smaller AICs indicate better models. This also allows us to test whether parameters are justified; each is treated as a potential nuisance parameter by comparing the AICs of models including and excluding the parameter (see, e.g., Bebbington, 2010). Note that we cannot in general use the corrected AIC (Hurvich and Tsai, 1989). Unlike for linear regression and autoregressive models, there is no proof of its validity for point process models (Claeskens and Hjort, 2008).

Because the sequence of event ages has associated errors, possible realizations of the actual event ages can be obtained by Monte Carlo simulation. For each ^{14}C-dated tephra, a normal random variate with the given mean and standard deviation is generated, and used to determine a calendar year from the transformed distribution using the calibration curve ShCal04 (McCormac et al., 2004). The ShCal04 calibration curve was the latest curve available at the time this portion of work was completed.

Ages contradicting stratigraphy are rejected and resampled, and finally the remaining ages are then interpolated in calendar years by the spline-fitting method of Turner et al. (2008a). This is repeated to generate 100 sample catalogs, as given in Appendix B.2 (MultiSourceAges.csv and SingleSourceAges.csv). The same procedure was followed by Turner et al. (2008a), except that only ^{14}C ages were used. Hence the means and standard deviations of the ages in

Table A.1 (Appendix A) differ from those in Turner et al. (2008a). In the case of the multi-source catalog, tephras identified by Turner et al. (2009) as being from the same eruption were used to further constrain the spline fitting. The age statistics given in Appendix A, Tables A.1 and A.2, summarize the results. This gives $100n$ different sampled eruption ages and thus $100(n-1)$ sampled inter-event times, and 100 replications of τ^* on which to fit a model. For the single-source record $n = 121$, and for the multi-source record $n = 97$.

As the model is fitted to the aggregation of the 100 Monte Carlo sequences, the resulting log-likelihood, and hence AIC, needs to be adjusted to account for the total number of realizations (cf. Kagan and Knopoff, 1977). Hence (3.12) is replaced by

$$AIC = -2\widetilde{\log L} + 2k, \tag{3.13}$$

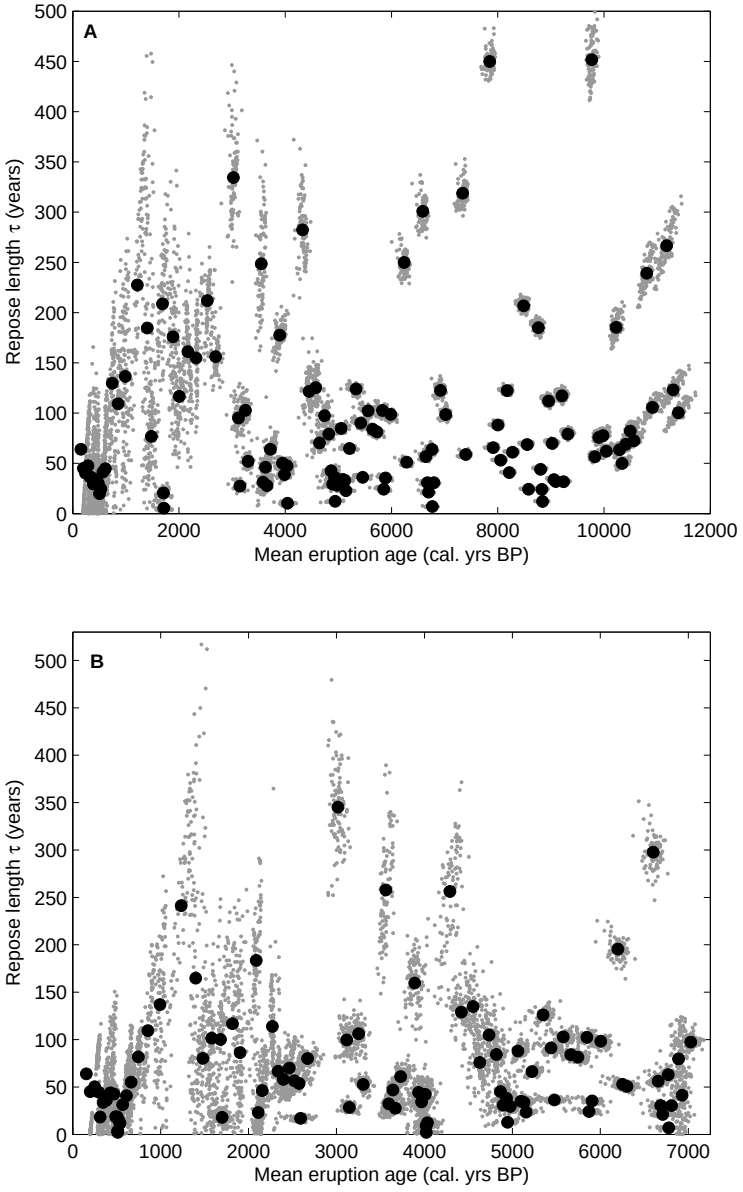
where

$$\widetilde{\log L} = \frac{\log L(\tau)}{100}. \tag{3.14}$$

3.4 Results

Figure 3.6 shows the simulated event ages (in cal yr BP) for the 'single-source' and 'multi-source' records, and their resulting repose lengths. There is more variability with some of the more recent eruptions, which were individually dated without a sedimentary sequence to further constrain them. The increase in the variability of the older part of the multi-source record derives from the additional events, correlations between sites, and age determinations, inserted in the record. The possible age realizations, consistent with stratigraphy and the sediment thicknesses, are conditional on these and are progressively tightened with each additional constraint. Hence there is less variation *within* a simulated sequence, but greater variation *between* sequences. For example, tephra units 57, 58 and 59 of the single-source record (Appendix A, Table A.1) have mean ages of 6646, 6676, and 6698 cal yr BP, and SDs of 56, 57 and 57 yr. These are three separate tephras, with interleaved peat thicknesses of 6.5 and 4.5 mm, respectively. Hence we can be sure that they are much more than three months apart in time, the criterion for distinct eruptions (Siebert and Simkin, 2002-). The large standard deviations reflect the variability in the simulated deposition curves.

Figure 3.6: Mean eruption age (black) and sampled event ages (grey). (A) Single-source data, (B) Multi-source data.



3.4.1 Baseline models

The results for fitting the baseline renewal models to the single-source record are shown in Table 3.2. Although the parameter estimates differ slightly from those in Turner et al. (2008a) (due to using cal yr rather than ^{14}C yr, as well as discarding the oldest event), the mixture of Weibulls distribution is still the best explanation of the record, repeating the conclusion of Turner et al. (2008a). The fitted density functions for the aggregated data set are shown in Figure 3.7, confirming the superior fit of the mixture of Weibulls, which accommodates the secondary mode of longer reposes between 200 and 300 yr. When considering the multi-source

Table 3.2: Fitted baseline models: single-source record.

Model	Exponential	Weibull	Lognormal	Mixture of Weibulls
Parameters	$\beta = 0.010 \pm 0.001$	$\alpha = 1.163 \pm 0.079$	$\mu = 4.151 \pm 0.089$	$p = 0.780 \pm 0.126$
		$\beta = 0.010 \pm 0.001$	$\sigma = 0.986 \pm 0.064$	$\alpha_1 = 1.578 \pm 0.176$
				$\beta_1 = 0.015 \pm 0.002$
				$\alpha_2 = 2.195 \pm 1.013$
				$\beta_2 = 0.004 \pm 0.001$
$\log L$	-667.1	-664.9	-666.4	-660.5
AIC	1336.2	1333.8	1336.8	**1330.9**

Table 3.3: Fitted baseline models: multi-source record.

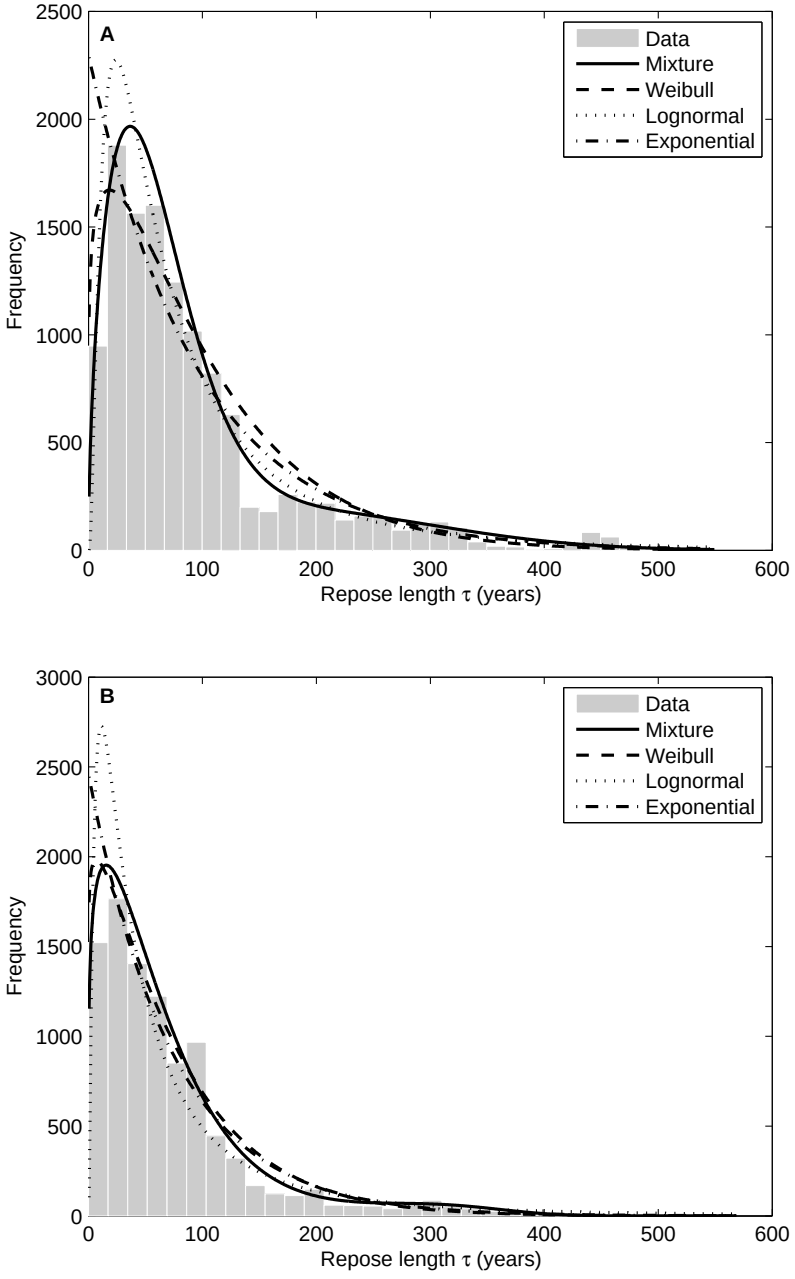
Model	Exponential	Weibull	Lognormal	Mixture of Weibulls
Parameters	$\beta = 0.013 \pm 0.001$	$\alpha = 1.079 \pm 0.084$	$\mu = 3.810 \pm 0.119$	$p = 0.964 \pm 0.048$
		$\beta = 0.013 \pm 0.001$	$\sigma = 1.173 \pm 0.085$	$\alpha_1 = 1.210 \pm 0.143$
				$\beta_1 = 0.015 \pm 0.002$
				$\alpha_2 = 5.148 \pm 4.913$
				$\beta_2 = 0.003 \pm 0.001$
$\log L$	-509.0	-508.5	-516.6	-507.0
AIC	**1020.0**	1021.0	1037.2	1024.0

record, the conversion to cal yr from ^{14}C yr results in a different preferred model (Table 3.3) to that found by Turner et al. (2009), i.e., the exponential renewal model (Poisson process) rather than the mixture of Weibulls renewal model. Examining the various fitted density functions for the aggregated multi-source data shown in Figure 3.7b, the secondary mode is still clearly seen, although it is reduced in extent. The problem arises because of the 'shoulder' at around 100 yr in the sampled repose lengths which makes all of the models fit poorly, and thus favors the simplest alternative.

3.4.2 Proportional hazard models

The exponential distribution (3.3), the Weibull distribution (3.4), and the mixture of Weibulls distribution (3.6) will be considered as possible baseline hazard functions in the PH model. The lognormal will be dispensed with as it fits poorly in Figure 3.7, and did not improve relative to the other alternatives with the addition of proportional hazards.

Figure 3.7: Density functions for a range of different distributions fitted to the sampled inter-event times of 100 Monte Carlo runs. (A) Single-source data, (B) Multi-source data.



Given that the geochemical signature has been reduced to six statistics: [1:Mean(TiO_2), 2:Mean(Al_2O_3), 3:Mean(MgO), 4:SD(TiO_2), 5:SD(Al_2O_3), 6:SD(MgO)], 64 possible combinations of geochemical precursor can be obtained by including or excluding each statistic. This includes the model without geochemistry, as given in Tables 3.2 and 3.3. The possibilities were investigated by fitting each baseline model to each catalog, with all possible combinations of geochemistry. For the mixture of Weibulls baseline the same covariates and coefficients were attached to each component of the mixture, as analysis showed that the

penalty from the additional parameters far outweighed any improvement in fit from allowing different covariate vectors or coefficients for each component (see Section 3.4.3). Matlab code to fit the various baseline PH models, for all 64 possible combinations of geochemical covariates, is provided in Appendix B.2.

Analyzing all the possible combinations of geochemistry data produced the preferred models in Tables 3.4 and 3.5. With the exception of the single-source mixture of Weibulls baseline, all the models for both catalogs preferred the same combination of covariates from z =(Mean(TiO_2) Mean(Al_2O_3) Mean(MgO) SD(TiO_2) SD(Al_2O_3) SD(MgO))^T, with $\delta = (\delta_1 \quad 0 \quad 0 \quad \delta_4 \quad \delta_5 \quad 0)^T$. The exclusion of MgO appears to be due to its strong correlation with Al_2O_3 (cf. Table 3.1). The signs and relative magnitudes of the coefficients are likewise very consistent across baseline models and catalogs. Where a different vector of coefficients was preferred in the case of the single-source mixture of Weibulls model, the estimated values for the (Mean(TiO_2) SD(TiO_2) SD(Al_2O_3)) model were $\delta_1 = -0.140, \delta_4 = 1.147, \delta_5 = -0.939$, again in line with the signs and magnitudes of the coefficients in the other models. Furthermore, according to the AIC, the Weibull baseline distribution outperforms both the exponential and mixture of Weibulls baseline distributions, for both the single-source and the multi-source records.

The geochemical precursor model is thus of the form:

$$\lambda(\tau, z) = \alpha\beta^{\alpha}(\beta\tau)^{\alpha-1}\exp(\delta_1 z_1 + \delta_4 z_4 + \delta_5 z_5) \tag{3.15}$$

where in the covariate vector z_1 =(Mean(TiO_2), z_4 = SD(TiO_2), and z_5 = SD(Al_2O_3).

An eruption that is high in Mean(TiO_2) or SD(Al_2O_3) will decrease the subsequent hazard function, and thus lengthen the next repose period. Conversely, a high value of SD(TiO_2) points to a higher subsequent hazard, and hence a shorter repose.

3.4.3 Sensitivity analyses

A sensitivity analysis was conducted to do with imputing missing covariates as the previous data. Using linear interpolation it was found that the z_1, z_4, z_5 covariate combination was not significantly worse than any other, although the mixture baseline model was significantly better for the single source data. This was influenced by the fact that the two most recent reposes had to be excluded, as it was not possible to interpolate the geochemistry.

Table 3.4: Fitted proportional hazard models: single-source record.

Model	Exponential	Weibull	Mixture of Weibulls
Parameters	$\beta = 0.064 \pm 0.052$	$\alpha = 1.227 \pm 0.084$	$p = 0.906 \pm 0.040$
		$\beta = 0.063 \pm 0.042$	$\alpha_1 = 1.497 \pm 0.129$
			$\beta_1 = 0.035 \pm 0.011$
			$\alpha_2 = 4.022 \pm 1.497$
			$\beta_2 = 0.004 \pm 0.001$
	δ_1 = -0.215 ± 0.096	δ_1 = -0.271 ± 0.099	δ_3 = -0.626 ± 0.200
	$\delta_4 = 0.556 \pm 0.237$	$\delta_4 = 0.684 \pm 0.244$	
	δ_5 = -0.654 ± 0.315	δ_5 = -0.795 ± 0.320	
$\log L$	-659.8	-655.9	-657.1
AIC	1327.7	**1321.7**	1326.3

Table 3.5: Fitted proportional hazard models: multi-source record.

Model	Exponential	Weibull	Mixture of Weibulls
Parameters	$\beta = 0.153 \pm 0.157$	$\alpha = 1.150 \pm 0.089$	$p = 0.930 \pm 0.051$
		$\beta = 0.138 \pm 0.127$	$\alpha_1 = 1.304 \pm 0.146$
			$\beta_1 = 0.076 \pm 0.073$
			$\alpha_2 = 4.293 \pm 3.206$
			$\beta_2 = 0.006 \pm 0.004$
	δ_1 = -0.283 ± 0.131	δ_1 = -0.317 ± 0.136	δ_1 = -0.271 ± 0.159
	$\delta_4 = 0.637 \pm 0.291$	$\delta_4 = 0.699 \pm 0.302$	$\delta_4 = 0.864 \pm 0.304$
	δ_5 = -0.922 ± 0.321	δ_5 = -1.045 ± 0.335	δ_5 = -0.759 ± 0.381
$\log L$	-499.6	-498.0	-502.7
AIC	1007.1	**1006.1**	1021.4

When fitting the PH model for the mixture of Weibulls baseline the same covariate vector (z) and coefficient vector (δ) was attached to each component of the mixture, giving the density function:

$$\begin{aligned} f(\tau) &= p\alpha_1\beta_1^{\alpha_1}\exp\left(\delta^T z\right)\tau^{\alpha_1-1}\exp\left[-(\beta_1\tau)^{\alpha_1}\exp\left(\delta^T z\right)\right] \\ &\quad +(1-p)\alpha_2\beta_2^{\alpha_2}\exp\left(\delta^T z\right)\tau^{\alpha_2-1}\exp\left[-(\beta_2\tau)^{\alpha_2}\exp\left(\delta^T z\right)\right]. \end{aligned} \tag{3.16}$$

As an alternative, the possibility of short and long repose periods being governed by different geochemical precursors was considered. This introduced a separate coefficient vector (γ) on the second component of the mixture. The density function in (3.16) was replaced by:

$$\begin{aligned} f(\tau) &= p\alpha_1\beta_1^{\alpha_1}\exp\left(\delta^T z\right)\tau^{\alpha_1-1}\exp\left[-(\beta_1\tau)^{\alpha_1}\exp\left(\delta^T z\right)\right] \\ &\quad +(1-p)\alpha_2\beta_2^{\alpha_2}\exp\left(\gamma^T z\right)\tau^{\alpha_2-1}\exp\left[-(\beta_2\tau)^{\alpha_2}\exp\left(\gamma^T z\right)\right] \end{aligned} \tag{3.17}$$

Table 3.6: Fitted proportional hazard models for the Mixture of Weibulls baseline where different covariate vectors and coefficients are imposed on each component of the mixture.

Record	Single-source	Multi-source
Parameters	$p = 0.810$	$p = 0.777$
	$\alpha_1 = 1.588$	$\alpha_1 = 1.347$
	$\beta_1 = 0.020$	$\beta_1 = 0.116$
	$\alpha_2 = 3.079$	$\alpha_2 = 1.617$
	$\beta_2 = 0.015$	$\beta_2 = 0.002$
	$\delta_1 = -0.083$	$\delta_1 = -0.332$
	$\delta_4 = 0.327$	$\delta_4 = 0.999$
	$\delta_5 = -0.030$	$\delta_5 = -0.830$
	$\gamma_1 = -0.499$	$\gamma_1 = 0.236$
		$\gamma_5 = -0.374$
$\log L$	-657.4	-503.0
AIC	1332.83	1025.9

where $z =$(Mean(TiO_2) Mean(Al_2O_3) Mean(MgO) SD(TiO_2) SD(Al_2O_3) SD(MgO))^T. This resulted in $64^2 = 4096$ possible combinations of geochemical precursor. Combinations range from the model that has no covariates attached to either component of the mixture ($\delta_i = 0$, $\gamma_i = 0$ for all $i = 1, \ldots, 6$), through to the model that has all six geochemical covariates attached to both components of the mixture ($\delta_i > 0$, $\gamma_i > 0$ for all $i = 1, \ldots, 6$).

The computing time required to find maximum likelihood estimates for all 4096 is excessive, requiring upwards of 11 weeks run time for each record. The resulting parameter estimates for both the single-source and multi-source records are shown in Table 3.6. Quite pleasingly, both catalogs prefer the same covariate vector for the first component of the mixture model. The signs and relative magnitudes of the covariate coefficients are similar to those in Tables 3.4 and 3.5. Although the second component of the mixture differs in preferred covariate combination, MgO is still noticeably absent from both covariate vectors of both records. This suggests that the longer repose periods (those in the second mode of the inter-event time distribution) are governed by changes in the Mean(TiO_2) of the previous eruption (and SD(Al_2O_3) for the case of the single-source record). Therefore, the geochemical precursors selected by allowing separate covariates and coefficient vectors for each component of the mixture, are at least amenable to that of the preferred models presented earlier. The added complexity however, does not suggest an improvement in fit. As mentioned on page 66, the penalty from the additional parameters far outweighs any improvement in fit from allowing different covariate vectors or coefficients for each component. The AIC values are substantially worse than those in Tables 3.4 and 3.5 for the single-source and multi-source records, respectively.

3.4.4 Goodness of fit

In order to check that the best fitting models are a suitable representation of the data, the point process residual theory of Ogata (1988) can be used (see Section 2.3.4). This requires calculating the 'compensator'

$$\Lambda(t) = \int_0^t \lambda(t)dt.$$

If the model $\lambda(t)$ satisfactorily accounts for the structure of the data, the transformed event times $\{\Lambda(t_i)\}$ will be from a Poisson process of rate one. As the preferred model has a renewal process (in the form of a baseline hazard function) with covariates that are invariant between events, it can be found that

$$\begin{aligned}\Lambda(t_{i+1}) &= \Lambda(t_i) + \exp\left[\delta^T z(t_i)\right] \int_{t_i}^{t_{i+1}} \alpha\beta^\alpha (t-t_i)^{\alpha-1} dt \\ &= \Lambda(t_i) + \exp\left[\delta^T z(t_i)\right] \beta^\alpha (t_{i+1}-t_i)^\alpha.\end{aligned}$$

Defining the inter-event times in the transformed process as $Y_i = \Lambda(t_{i+1}) - \Lambda(t_i)$, it is found that none of the $\{Y_{i+1}\}$ and $\{Y_i\}$ have significant correlations, that the cumulative counts $\{\Lambda(t)\}$ are not significantly different from a $U[0, N]$ distribution (Figures 3.8A and 3.9A), and that none of the $\{1 - \exp(-Y_i)\}$ are significantly different from a $U[0, 1]$ distribution (Figures 3.8B and 3.9B). Hence the transformed processes can be assumed to be independent, stationary, and have exponential inter-event times, respectively. Thus the transformed processes are not distinguishable from a Poisson process, and so the proposed models are a good representation of the data.

3.4.5 Hazard forecasts

The difference in the hazard forecast from the multi-source record using the baseline and PH Weibull models is shown in Figure 3.10A. This shows how the estimated hazard evolves over time following the last eruption. In this case the geochemical precursors are encouraging a shorter repose.

The probability of no eruption in the next t years, given that it has been s years since the last eruption, can also be calculated as

$$\Pr(\tau_n > s + t \mid \tau_n > s) = \frac{\Pr(\tau_n > s + t)}{\Pr(\tau_n > s)} = \frac{1 - F(s + t \mid z)}{1 - F(s \mid z)},$$

Figure 3.8: Stationarity (A) and exponential inter-event (B) diagnostic plots for the single-source record. The dotted lines are the piecewise 95% confidence limits; all 100 Monte Carlo sequences are contained within them.

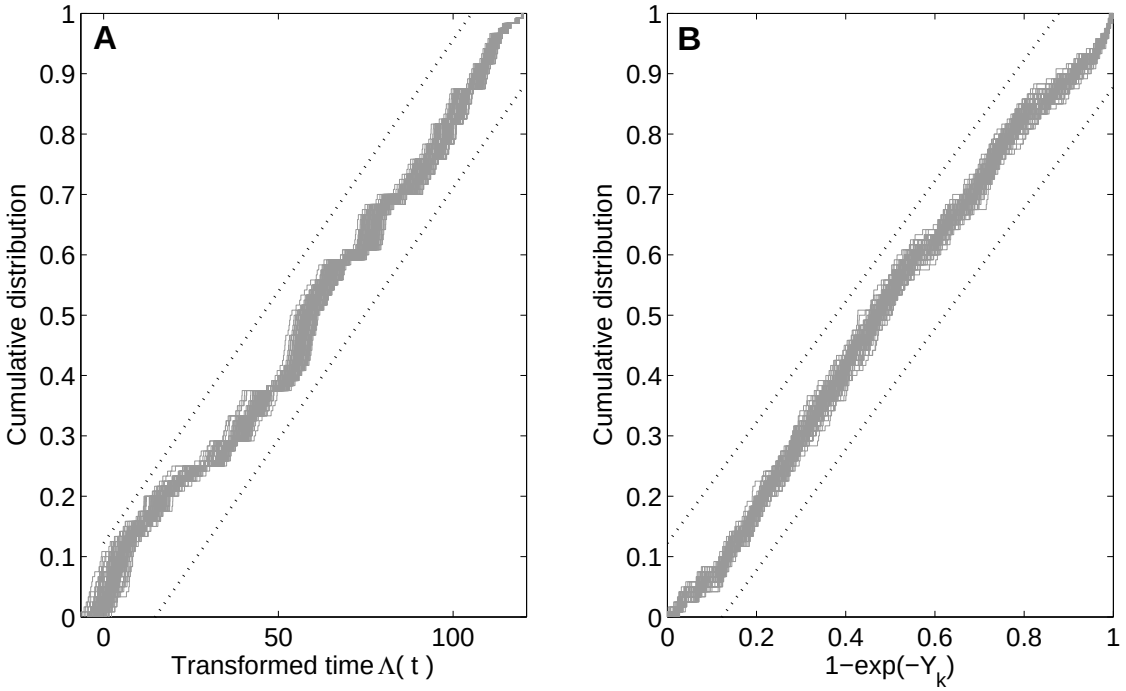


Figure 3.9: Stationarity (A) and exponential inter-event (B) diagnostic plots for the multi-source record. The dotted lines are the piecewise 95% confidence limits; all but two of the 100 Monte Carlo sequences (i.e., not a significant number) are contained within the stationarity limits, all 100 are within the exponential limits.

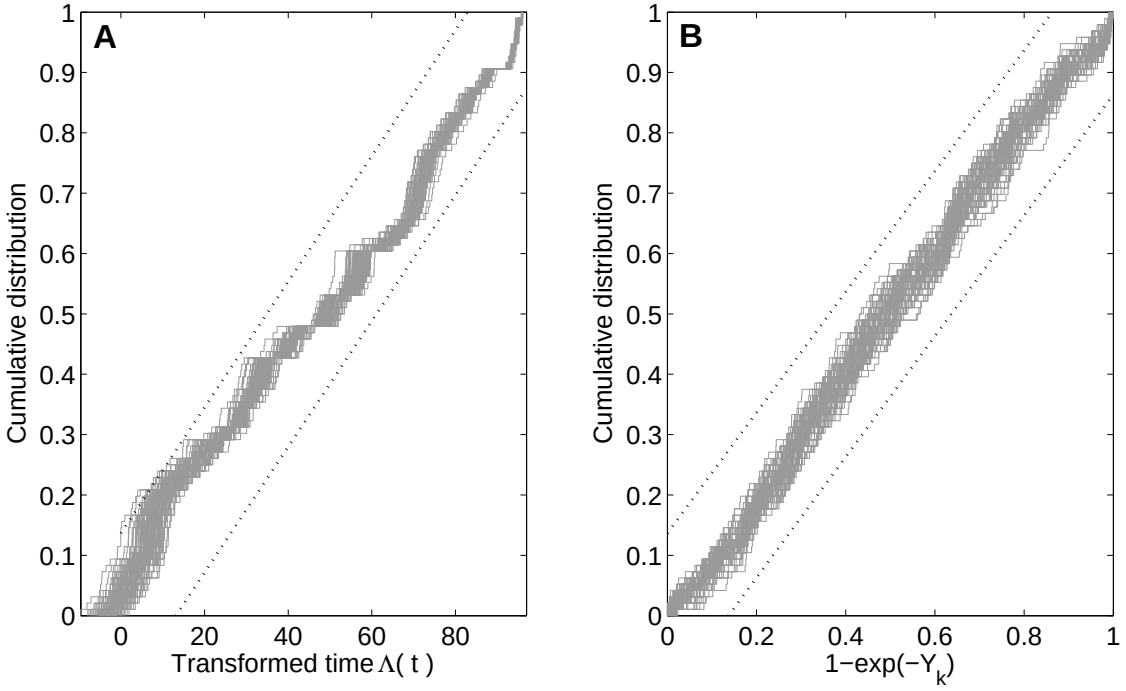
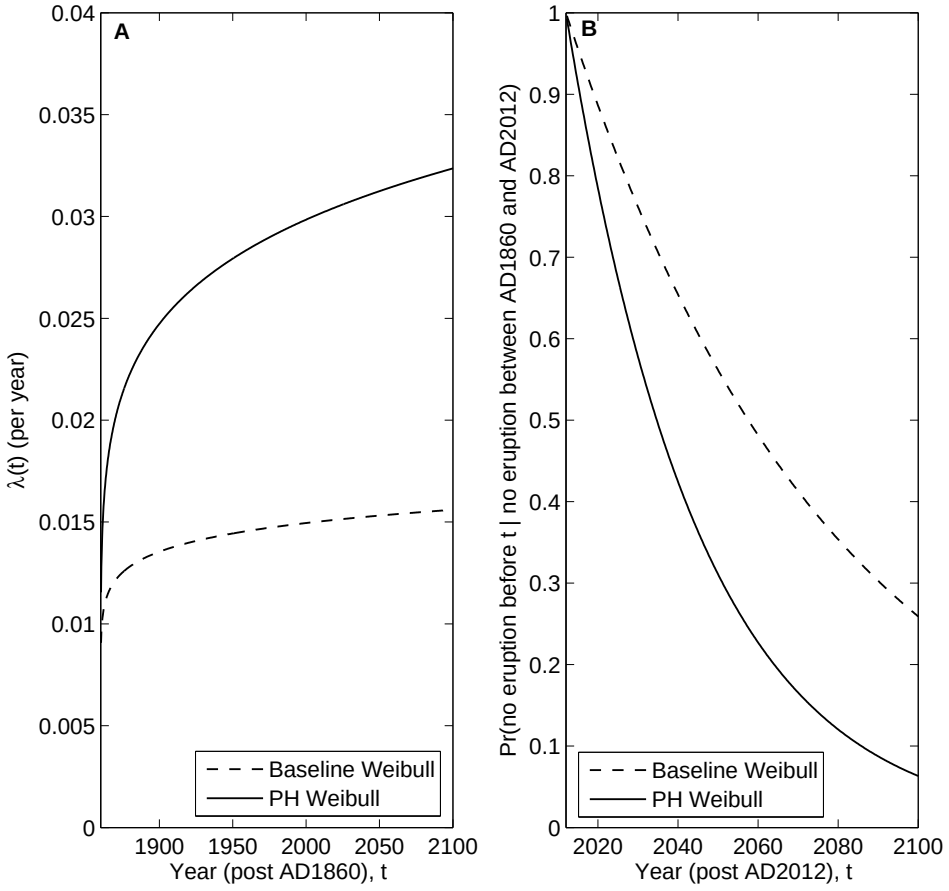


Figure 3.10: Current estimated hazard (A) and distribution of time to next eruption (B), based on the multi-source record.

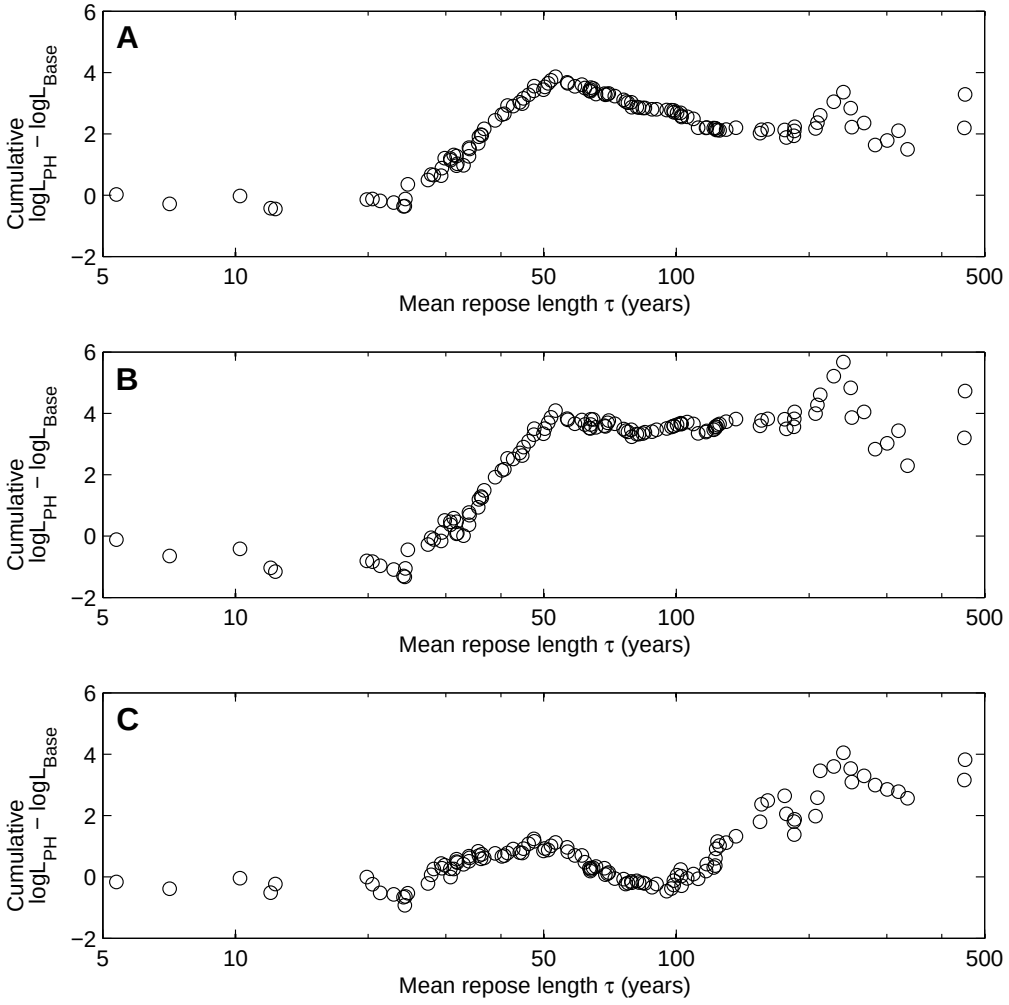


where $F(\cdot \mid z)$ is given by (3.11). The resulting estimated (as at the end of AD 2012) conditional distributions of the time to the next eruption are compared in Figure 3.10B. This can be recalculated at any time, and used to estimate the probability of an event in any finite time horizon.

3.5 Discussion

The proportional hazards models strongly improve the fit to repose periods over the baseline renewal models. For the single-source eruption record, the Weibull PH model has an AIC value of nine smaller than the initially preferred mixture of Weibulls renewal model, where a difference of approximately two indicates a statistically significant improvement (Utsu, 1999). Not only is the fit better, but the preferred baseline in the PH model has a unimodal density.

Figure 3.11: Difference in cumulative log-likelihood values for each repose (single source data) under the proportional hazards and baseline renewal models. (A) Exponential baseline, (B) Weibull baseline, (C) Mixture of Weibulls baseline.

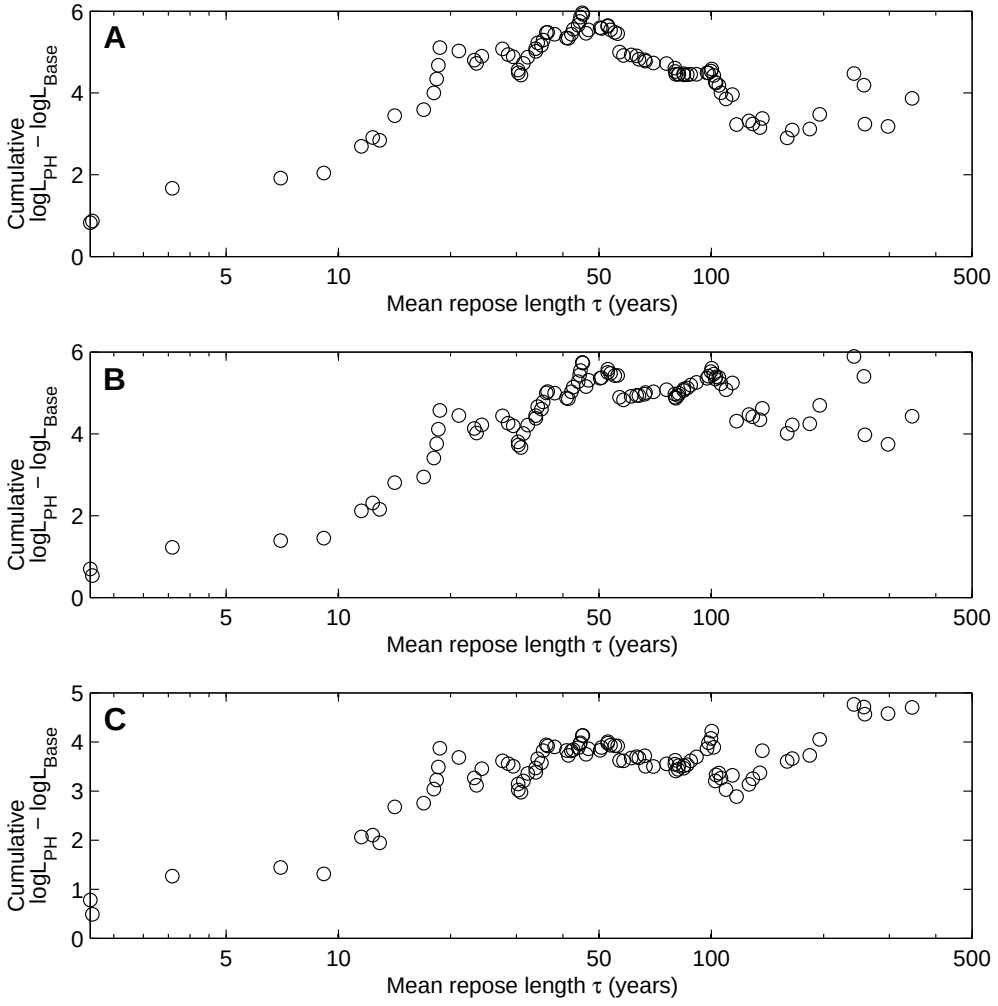


This raises a question as to how the obvious bimodal nature of the observed repose lengths are being modeled.

To examine what causes the significant improvement in fit, we can analyze the contribution each of the events makes to the overall log-likelihood. The cumulative difference between the log-likelihoods for the PH and renewal models are plotted against the inter-event time for each event in Figures 3.11 and 3.12. A positive difference in log-likelihood indicates that the PH model fits that event better than the renewal type model.

For the single-source record the most noticeable improvements in log-likelihood are exhibited by the positive slope for reposes of between 20 and 50 yr, and for some of the events with uncharacteristically long (> 200 yr) repose periods. For the multi-source record, many of the longer repose times are 'filled in' by the additional record(s), but again we see that the

Figure 3.12: Difference in cumulative log-likelihood values for each repose (multi-source data) under the proportional hazards and baseline renewal models. (A) Exponential baseline, (B) Weibull baseline, (C) Mixture of Weibulls baseline.



PH model is gaining much of its power from the shorter (< 50 yr) reposes. In all cases the fit to the intermediate repose lengths of 50-100 yr is similar with or without the additional geochemical parameters.

Wadge (1982) has noted that many polygenetic volcanoes alternate periods of activity with longer periods of quiescence. Mt Taranaki is almost certainly such a case, at least over the Holocene, with a noticeable second mode in the repose time distribution. Renewal models, which assume stationarity, for such behavior implicitly assume that the longer reposes occur at random. The analysis in this chapter has shown that by incorporating the geochemical parameters, the mixture distribution for the bimodal inter-event times is no longer needed, as the geochemical data is acting differentially on the repose time. This has great advantages in a forecasting setting, as this information is obtained (from the previous eruption) at the

beginning of the repose; we do not have to wait and see if we pass through the first mode as is the case under the mixture of Weibulls renewal model.

The way in which the geochemical predictors are operating is also consistent with petrological expectations as described above. First, the predictors of long following repose periods are high variability in Al_2O_3 and high mean TiO_2. Implicit in this is that the TiO_2 variability is low. As described above, variability in the Al and Mg components within the population of titanomagnetite crystals most likely represents mixing or mingling of slightly different magma batches. The main cause for the highest variability in TiO_2 in the Taranaki magmas (increasing the likelihood of a short following repose) is dwell time in the shallow conduit below the volcano (Turner et al., 2008b). Based on these considerations the following situations can be surmised to explain both the geochemical signals and future forecasts based upon them:

Case 1. Magmas erupting directly from mid-crustal (or lower-crustal) magma chambers (either 7-10 or c. 35 km deep in the Taranaki case; Turner et al. 2011b) that involve two or more mixed/mingled batches. These scenarios imply that the next eruption is likely to also be sourced from great depth, and that more time is required to assemble a large enough body of magma to buoyantly rise again to erupt.

Case 2. Magmas erupting from shallow-level conduits (within the edifice or upper 2-3 km, e.g., Platz et al., 2007b). These events imply that resident, highly crystalline shallow magma may remain, or be continuing to rise, to feed further eruptions in the short term.

Another implication of these two situations is that Case 1 eruptions should tend to be of larger volume, because they involve an accumulation of magma from more than one batch. By contrast, Case 2 eruptions may be small and non explosive. Platz et al. (2007a, 2012) have independently used geochemical techniques to conclude that Taranaki eruptions can be generated by either direct rise from mid-crustal reservoirs or also from shallow-stalled magmas.

3.6 Conclusions

The work in this chapter has shown that incorporating geochemical covariates in the estimated hazard function using a proportional hazards type approach leads to significant improvement

in fit. Moreover, as the geochemical precursors are produced by the previous eruption, the additional information is immediately available, and does not have to accumulate over time as under the mixture of Weibulls renewal model. For both catalogs, the Weibull baseline hazard was preferred, but the same covariate model was selected in almost all cases, and the coefficients indicated that the geochemistry influenced all the models in much the same way. Hence the conclusions can be considered to be robust. These findings represent a recipe that could applied to many similar 're-awakening' types of stratovolcano that exhibit cyclic variations in eruption frequency and magnitude. The findings of this work also open the door to exploring other relationships that could be made more tractable with geochemical precursory data, such as forecasting eruption magnitude/volume or style (see, for example, Section 2.3.6), which have a major bearing on hazard assessment.

Chapter 4

Automated statistical matching of multiple tephra records

Probabilistic hazard forecasts established from models such as the geochemical covariate model proposed in Chapter 3, rely on the accuracy and completeness of historical eruption records. This poses the question of how to establish a detailed record of past volcanic events. Multiple sites are needed to build the most accurate composite tephra record, but correctly merging them by recognizing events in common and site-specific gaps remains complex. In this chapter I present an automated statistical matching algorithm, developed using stochastic local optimization techniques, to correlate tephra records from multiple sites. Work discussed in this chapter is published in Green et al. (2014).

4.1 Introduction

Tephra studies play a pivotal role in reconstructing the history of Quaternary events throughout many parts of the world (Lowe, 2011). Since tephras provide isochronous markers in many depositional environments, they have been used to understand the age and development processes of landscapes, soil-forming and sediment-accumulation rates and the timing of regional, global paleoclimatic changes (Sandiford et al., 2003), and human evolution (Tyron et al., 2009; Brown and McDougall, 2011). The importance of establishing a comprehensive history of tephras and attaining higher resolution correlations between records has been emphasized in the New Zealand region through the development of projects such as INTIMATE (INTegration of Ice-core, MArine and TErrestrial records) (Lowe et al., 2008). Tephra records

have also proven to be one of the most reliable foundations for the probabilistic forecasting of future eruption potential from volcanoes (Turner et al., 2008a), as long records allow for a wider range of the volcanoes' behavior to be modeled (Marzocchi and Bebbington, 2012).

For all of these applications, compiling the most complete composite sequence of tephras in a region helps to improve the resolution and reliability of climate, geomorphic, or volcanic activity reconstructions. This means expanding from single-site records to comparing locations that may receive tephra deposits under subtly different wind conditions, or for eruptions of different sizes and styles. It also means combining records from sites (Turner et al., 2009) that have varying suitability for tephra preservation. Ideal sites, such as maar craters (closed depressions with a low catchment area) may need to be combined with records from upland soil sequences, where tephra preservation is only possible under forested conditions, and/or for only very thick falls. Inter-site variations in hydrology, along with the original depositional variability, make tephra correlation based on physical properties and appearance challenging. Chemical correlation may help, but is not a solution in volcanic areas where there are small chemical variations or where there are cyclical variations in chemistry that semi-regularly produce the same compositions (Pollard et al., 2006; Turner et al., 2011a).

Statistical techniques for correlating tephras between sites have included various types of discriminant function analysis, first used by Stokes and Lowe (1988) to classify the source of tephra deposits using glass-shard major element chemistry (see Section 2.2.6). This method has been successfully applied in a range of studies with mixed eruption sources, to either separate tephra from different volcanoes (Stokes et al., 1992; Cronin et al., 1996, 1997), or in some cases individual tephras (Cronin et al., 1997). When it comes to merging multiple eruption records, however, there is no standard means of obtaining and evaluating the most reasonable correlations of tephra across multiple sites, especially where large numbers of units are present (Charman and Grattan, 1999). Geochemical fingerprinting and dating of widespread marker tephras allow single stratigraphic records to be constructed, but the issue of how to identify and match other individual events to build robust composite stratigraphies remains.

Bursik and Rogova (2006) explored the use of a hybrid information processing system for correlating tephra layers between sites, but their analysis was restricted to major tephra layers within a single eruption sequence. The units were distinguished based on their physical characteristics and compositions (namely by the fraction of pumice, grading, zoning, thickness, and grain size) and although they were able to discriminate successfully among the tephra

layers, their data was sparse. Data were collected at five sites located no more than 7 km from the source vents and only one eruption sequence was analyzed.

When examining multiple eruptions from multiple, variously located source volcanoes, along with very long records (e.g. 70,000 yrs), the estimated ages of the tephras (combined with stratigraphic constraints) naturally become the focal point for matching, followed by lithology (general physical characteristics), petrology (origin, structure, and composition), and chemistry.

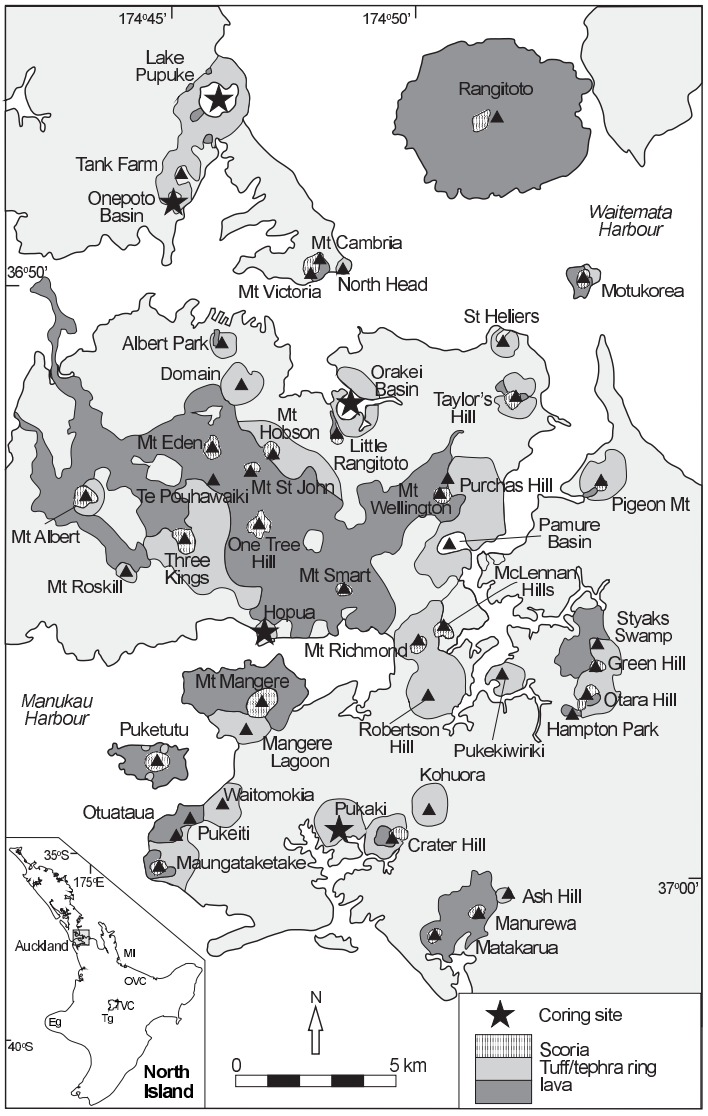
Given a large number of events observed at each location, merging records by recognizing common events and site-specific gaps becomes a statistical problem of some significance. Turner et al. (2009) suggested a method for correlating records sourced from two locations (see Section 2.2.6); however, extending this to the case of more than two locations becomes complex. The age and associated age error of each event means it is possible for events in a particular core to overlap multiple events in any of the other cores. This presents a large combinatorial problem, advocating the use of local search algorithms to find the most statistically likely arrangement of tephra across the cores.

The objective of this portion of research was to develop an automated algorithm for statistically matching tephra records from multiple locations. The methodology will be illustrated on the composite tephra record in the Auckland region of New Zealand, compiled from five individual maar lake sediment cores. The algorithm can, however be easily implemented on other volcanic tephra records, as will be illustrated in Chapter 5. The radiocarbon age and associated age error for each tephra form the basis of the matching algorithm, and the geochemical signature of each tephra is used to constrain the arrangement.

4.2 Auckland Volcanic Field data

The Auckland Volcanic Field (AVF) consists of ∼50 small basaltic volcanoes within an area of approximately 360 km^2 (Allen and Smith, 1994; Kereszturi et al., 2013). Auckland City is New Zealand's largest urban area, home to 1.4 million people and built almost entirely upon the AVF (Figure 4.1). Developing a hazard forecast (Magill and Blong, 2005a,b) for the AVF requires a robust understanding of the frequency of past eruptions. Unlike the temporal hazard models discussed in Section 2.3, or those developed in Chapter 3 for Mt Taranaki, probabilistic hazard models for the monogenetic Auckland Volcanic Field require additional

Figure 4.1: Volcanic geological map of the Auckland Volcanic field, North Island, New Zealand, adapted from Kermode (1992), showing named volcanic centers and the major lithologies (general physical characteristics) of eruptives. Coring sites for volcanic ash layers are located at: Lake Pupuke, Hopua, Orakei Basin, Onepoto Basin and Pukaki (indicated by stars). Reprinted from Bebbington and Cronin (2011) Spatio-temporal hazard estimation in the Auckland Volcanic Field, New Zealand, with a new event-order model. Bull Volcanol 73: 55-72 ©2010 Springer-Verlag, with kind permission from Springer Science and Business Media.



spatial elements to understand not only when, but *where* the next eruption is likely to occur. Bebbington and Cronin (2011) detail many of the problems associated with forming such hazard models for the AVF, such as the absence of obvious age:location trends, the lack of available chemical datasets appropriate for correlating tephras to their source vents, and difficulties establishing a robust stratigraphic ordering due to the often small eruption size of field volcanos. The over-riding problem in developing a hazard forecasts for the AVF is working out the spatio-temporal pattern (or at least the event order). Once this is established,

spatial, temporal and spatio-temporal models can be evaluated (Bebbington, 2013).

Tephra dispersal for the typically small basaltic eruptions of this field is highly variable and little detailed mapping is possible due to the rapid weathering in this humid environment. An intensive regimen of scientific coring has been carried out in the AVF not only to understand the volcanic history of the effects of local and far-field tephra (Sandiford et al., 2001; Shane and Hoverd, 2002; Shane, 2005; Molloy et al., 2009), but also to examine paleo-climatic and paleo-geomorphological histories of the region (Sandiford et al., 2002; Stephens et al., 2012). Tephras were recovered from five old maar volcanic structures with good stratigraphic records: Lake Pupuke, Onepoto Basin, Orakei Basin, Hopua Crater, and Pukaki Crater (Figure 4.1).

Thirty-seven tephras were recovered in the sediment core from Lake Pupuke maar (Molloy et al., 2009), 48 from the Onepoto Basin (Shane and Hoverd, 2002), 41 from the Orakei Basin (Molloy et al., 2009), 29 from Hopua Crater (Molloy et al., 2009), and 62 from Pukaki Crater (Sandiford et al., 2001; Shane, 2005). In addition to a series of basaltic tephras from the Auckland Volcanic Field (AVF), tephras from three distal rhyolitic volcanoes were found: Taupo (TVC), Okataina (OVC), and Mayor Island (MI), as well as tephra from two andesitic volcanic centers (Taranaki/Egmont (Eg), and Tongariro/Ruapehu (Tg)). These distal volcanoes are located between 285 km SSW (Taranaki), 290 km South (Tongariro/Ruapehu) and 150 km SE (Mayor Island) of Auckland (Figure 4.1).

The source of these tephras was determined on the basis of glass chemistry composition using the total alkali silica diagram of Le Maitre (1984). Individual marker tephras were separated primarily on the basis of K_2O vs CaO composition. For further details see Sandiford et al. (2001); Shane and Hoverd (2002); Shane (2005) and Molloy et al. (2009).

4.2.1 Reconciliation with original data

Being the compilation of several datasets developed over a number of years, there are some inconsistencies in the data from Molloy et al. (2009). The first task was to reconcile the data presented in the most recent analysis (Molloy et al., 2009) with that presented in earlier publications (Shane, 2005; Shane and Hoverd, 2002; Sandiford et al., 2001). The following details some of the disparities and how they will be dealt with in this analysis.

Shane (2005) describes two AVF events in the Pukaki crater at depths of 56.35 m and 56.53 m with thicknesses of 310 mm, and 180 mm respectively. Molloy et al. (2009) combined these

events because they were not separated by non-volcanic sediments (Phil Shane, pers. comm.); this approach was also adopted for the following analysis. Other differences include:

- Shane and Hoverd (2002) report the Rotoehu tephra in Onepoto Basin at a depth of 41.42 m with a thickness of 630 mm. However, the occurrence of two Egmont tephras at 40.87 m and 40.82 m with respective thicknesses of 2 mm and 1 mm would restrict the Rotoehu tephra to < 550 mm. Figure 3 of Molloy et al. (2009) suggests that the thickness of Rotoehu is c. 530 mm.
- Shane and Hoverd (2002) mention that the composition of the Te Mahoe tephra found at a depth of 39.9 m overlaps that of the Maketu tephra. Molloy et al. (2009) prefer correlating this event to Maketu because of its thickness. This reassignment will be adopted in the following analysis.
- Finally, Shane and Hoverd (2002) document 80 tephra beds in Onepoto Basin to a depth of approximately 61 m but the analysis here will be restricted to the 48 tephras above 45 m. The age record is only constrained above the Rotoehu tephra at a depth of 41.42 m, and as a result there is considerable age uncertainty in tephras observed below the Rotoehu tephra. In addition, tephras below 45 m are much older than any of the tephras found in the other four cores so including them would not affect the matching procedure.

Figures 4.4 and 4.5 show the stratigraphic relationship of tephra across the five cores, along with the Molloy et al. (2009) correlations.

4.2.2 Sedimentation rates and age estimation

Several marker rhyolitic tephras derived from the Taupo and Okataina volcanic centers were identified at various sites. The distinctive petrological and chemical characteristics of these key marker tephras and the fact that their ages are mostly well constrained allow a stratigraphic record to be established. The ages of the remaining (non-marker) tephras are subsequently obtained by linear interpolation using the stratigraphic marker horizons as published in Molloy et al. (2009), Danišík et al. (2012), and Lowe et al. (2013). The ages of these stratigraphic marker horizons are given in Table 4.1.

The error associated with these age determinations allows possible realizations for the actual ages to be obtained. Each tephra age in the cores was obtained through repeated (100

Table 4.1: Tephras used for age control

Tephra name	Age (cal ka BP)	$\pm 2\sigma$	Reference
Taupo (Tp)	1.72	0.01	Lowe et al. (2013)
Tuhua (Tu)	6.58	0.55	Lowe et al. (2013)
Mamaku (Ma)	7.94	0.26	Lowe et al. (2013)
Rotoma (Rm)	9.42	0.12	Lowe et al. (2013)
Opepe (Op)	9.99	0.16	Lowe et al. (2013)
Waiohau (Wh)	14.01	0.16	Lowe et al. (2013)
Rotorua (Rr)	15.64	0.41	Lowe et al. (2013)
Rerewhakaaitu (Rw)	17.50	0.46	Lowe et al. (2013)
Okareka (Ok)	21.86	0.29	Lowe et al. (2013)
Te Rere (Te)	25.17	0.96	Lowe et al. (2013)
Oruanui (Kk)	25.36	0.16	Lowe et al. (2013)
Poihipi (Po)	28.45	0.67	Lowe et al. (2013)
Okaia (O)	28.62	1.43	Lowe et al. (2013)
Maketu (Mk)	36.32	1.15	Molloy et al. (2009)
Tahuna (Ta)	39.27	2.39	Molloy et al. (2009)
Rotoehu (Re)	45.17	3.30	Danišík et al. (2012)

times) simulation from a Gaussian distribution, using the mean and standard deviation of the age estimates as given in Table 4.1, and respecting stratigraphy in the cores. After subtracting the thicknesses of the tephra itself, the ages of the remaining non-marker tephras were estimated through linear interpolation, and in some cases extrapolation (Figure 4.2) to obtain a statistical distribution for the age of each tephra.

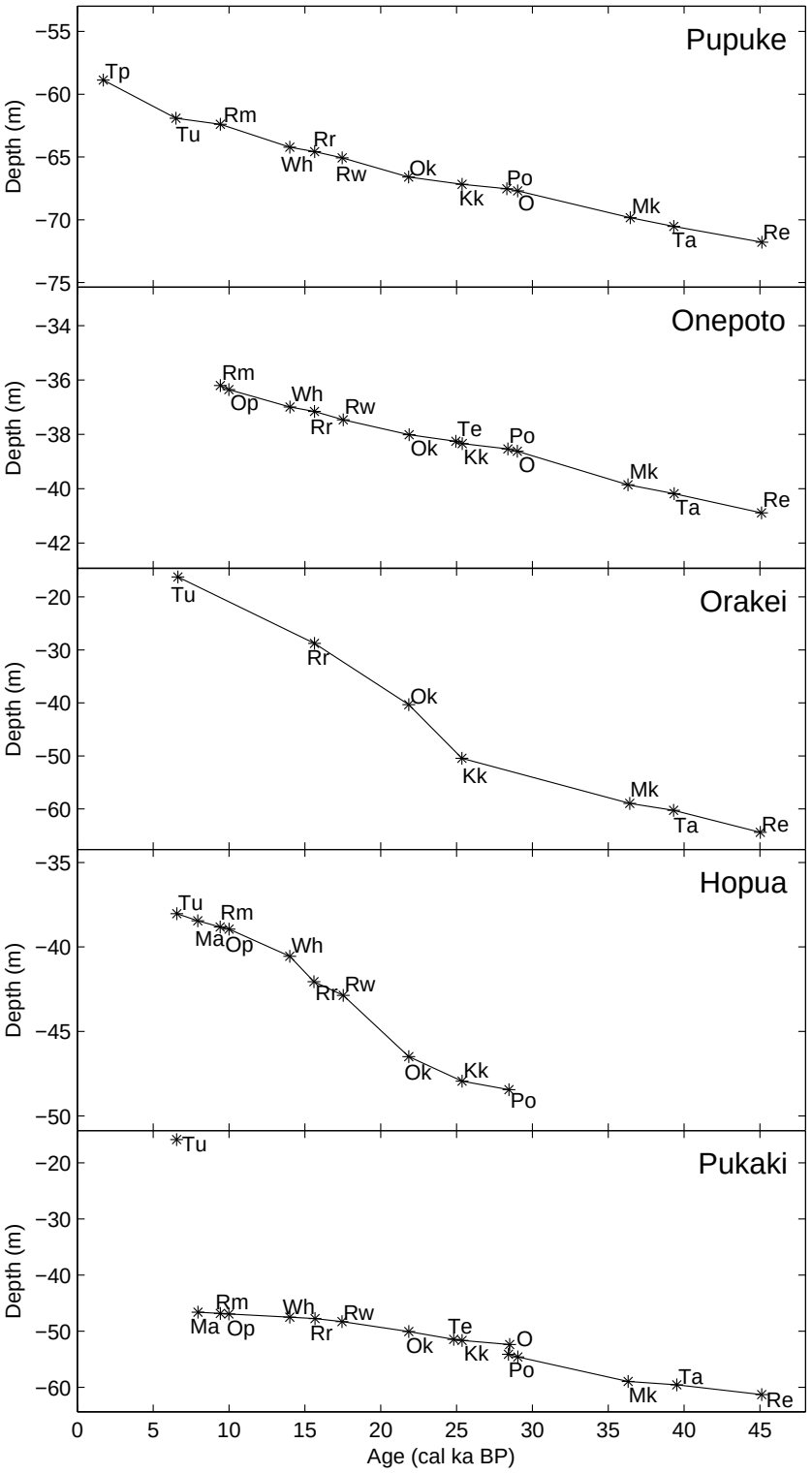
Given the adjusted depth (after subtracting the thickness) x of any non-marker tephra, the linear interpolated age y can be calculated by

$$y = y_a + (y_b - y_a)\frac{x - x_a}{x_b - x_a}, \tag{4.1}$$

where (x_a, y_a) and (x_b, y_b) are the adjusted depth and simulated age of the nearest marker tephra found above and below the given tephra, respectively. This was repeated a number of times (100) to obtain mean ages and standard deviations for the tephras. Example 4.1 has used a single pair of marker tephras from a single core. The full problem involves many marker tephras, five cores, and many of the tephras are missing. Matlab code is provided in Appendix C.1 to estimate the ages. The code is based on the idea illustrated in Example 4.1, but uses the full suite of marker ages in Table 4.1 and all the cores.

The sedimentation rate between the Tuhua and Mamaku tephras in the Pukaki core is unusual, apparently because of the breach of the crater rim at approximately 7600 cal yr BP

Figure 4.2: Sedimentation rates after subtracting tephra thicknesses. Abbreviations for the tephra names are given in Table 4.1.



Example 4.1. *Consider the three most recent events observed in the Onepoto sequence. The first tephra, found at an adjusted depth of 36.207 m (after subtracting the thickness), corresponds to the Rotoma tephra which has a mean age (and standard deviation) of 9.42 (0.06) cal ka BP (Table 4.1). The third tephra, found at an adjusted depth of 36.358 m, corresponds to the Opepe tephra which has a mean age (and standard deviation) of 9.99 (0.08) cal ka BP. In order to calculate the linear interpolated age of the second tephra, found at an adjusted depth of 36.308 m, realizations of the actual ages of both the Rotoma and Opepe tephras must first be obtained. This is done by simulating from a $N(\mu, \sigma^2)$ distribution given the mean age (μ) and associated error (σ) of each marker tephra (above). As the Opepe tephra lies stratigraphically below the Rotoma tephra, it must be older. If this constraint is not met when simulating the ages, they must be re-sampled. Using this procedure possible age realizations of 9.4601 and 10.025 cal ka BP are obained for the Rotoma and Opepe tephras, respectively. The linear interpolated age of the intermediate tephra is calculated as*

$$y = 9.4601 + (10.025 - 9.4601)\frac{36.308 - 36.207}{36.358-36.207} = 9.8379. \quad (4.2)$$

(Sandiford et al., 2001), but no tephras lie in this age range.

Most of the cores are underpinned by the widespread Rotoehu tephra. Molloy et al. (2009) chose not to use the Rotoehu tephra as a marker horizon because of controversy over the age of this event at the time, but this resulted in a larger age uncertainty in the lower part of the core since the sedimentation rate was extrapolated rather than interpolated. To improve the precision in age estimates for the older tephras, the analysis here was applied using the new constrained age for the Rotoehu tephra given by Danišík et al. (2012) (see Table 4.1). The compositionally distinctive Poihipi tephra, identified in all of the cores except Orakei, was not used for age control in Molloy et al. (2009). In contrast in this analysis, the age estimate given by Lowe et al. (2013) is applied to further constrain the linear interpolation.

The rhyolitic Hauparu tephra, which is stratigraphically slightly younger than the Maketu tephra, has an age estimate of 35.87 ± 12.7 ^{14}C ka BP (Froggatt and Lowe, 1990). Because of the large error on this age estimate it is not used for age control, but instead it is ensured that the algorithm correctly correlates the Hauparu tephra across the four cores in which it was observed.

The linear interpolation of the tephra ages assumes that the sedimentation rate is constant between the marker horizons. The error inherent in this assumption is allowed for by following the idea of Bebbington and Cronin (2011), whereby a small constant is combined with the

precision error for each unknown age. The error associated with the ages is obtained using

$$\sigma = \sqrt{s^2 + c^2}, \tag{4.3}$$

where s is the standard deviation of the ages obtained through repeated sampling from the linear interpolation and c denotes the non-linearity constant. Various non-linearity constants were considered, ranging from 0.3 ka to 0.7 ka. The resulting sensitivity in the optimal arrangement returned by the procedure will be discussed later.

4.3 Stochastic local search

Once the marker horizons are aligned, the most likely arrangement of the other tephras across the cores can be obtained. An eruption may have left a deposit at all of the sites, or any subset. Additionally, because age estimates have associated age errors, it is possible for events in a particular core to overlap multiple events in any of the other cores. Thus, there are a large number of possible combinations. This was treated as a *combinatorial problem* as we wish to identify the optimal arrangement, from a large set of possible arrangements, according to a set of well defined rules.

A popular approach towards solving challenging combinatorial problems is to utilize local search algorithms to look for an optimal solution from all possible pairwise candidate solutions. Candidate solutions are discovered during an attempt to solve a problem. In the context of tephra matching, candidate solutions are the alternative arrangements of matches or correlations encountered while looking for the 'optimal arrangement'. There is usually a large set of possible alternatives, hence finding an optimal solution can be computationally intensive.

Local search algorithms start at some position of the given search space and iteratively move to a neighboring position. Here the search space consists of all possible arrangements of the tephras across the five cores. An iterative move slightly alters the arrangement of a particular segment of the tephra record, generating a new candidate solution. Each new candidate solution, or arrangement of tephras, is evaluated on the basis of how statistically likely it is, as measured by an objective function, the calculation of which is discussed below. The search process looks for the arrangement that maximizes this objective function.

4.3.1 Feasible candidate matches

To specify a method for randomly generating particular 'moves' or rearrangements of the observations within each core, we consider the method proposed by Turner et al. (2009). Given age determinations which are normally distributed with means (μ_1 and μ_2) and errors (σ_1 and σ_2) (see Figure 2.1) for two tephras from different cores, candidate matches can be considered by calculating the age distance between them. The age distance (D) is a standard normal random variable calculated by

$$D_{1,2} = \frac{\mu_1 - \mu_2}{\sqrt{\sigma_1^2 + \sigma_2^2}}. \tag{4.4}$$

Thus absolute distances greater than three would be considered significantly different at the 0.3% significance level, and so any pair of events that has an absolute age distance less than three could be from the same eruption.

With just two cores the most likely arrangement of the deposits can be easily determined (Turner et al., 2009). Extending this to the case of five cores is more complicated. A particular eruption may be widespread enough to be observed in all five cores, or it could be seen in only a few (or one) of the cores. Equation (4.4) is used to compile a list of candidate matches with an absolute age distance $|D_{1,2}| < 3$. Henceforth this list will be referred to as the *list of possible moves*. This list is used to randomly select segments of the formation for possible rearrangement. These segments consist of tephras across the five cores that have mean ages between μ_1 and μ_2. The rearrangement of a particular segment provides the neighboring state s' of the current state or arrangement s.

Stochastic local search procedures explore the solution space by making minimal alterations to candidate solutions in the hope of finding an optimal solution. The optimal solution is returned when no improving moves can be made and some stopping criterion is satisfied.

4.3.2 Search constraints

With each possible rearrangement of a selected segment, the algorithm ensures that the stratigraphy of events is maintained. Additionally the procedure can force matches between certain tephra units based on other criteria (e.g., lithology – general physical characteristics). It can also be adapted to prevent certain matches, although only if there is sufficient evidence to suggest that the tephras are not compatible. The marker horizons described above (see

Table 4.1) were thus forced to be matched by the algorithm and other tephras were prohibited from merging with them.

Incompatible matches may be identified using tephra thicknesses, grain size, glass or titanomagnetite chemistry (Turner et al., 2008b), or any other meaningful variables. As discussed on page 81, the source of the tephra deposits found in the five cores was identified using glass-shard major element chemistry (Sandiford et al., 2001; Shane and Hoverd, 2002; Shane, 2005; Molloy et al., 2009). Taking the classifications as given, the following analysis only permits matches among tephras that have similar geochemical compositions. For instance, the process will not allow tephras from Taranaki/Egmont to match with the AVF tephras. Furthermore, tephras observed at different ages at the same location cannot be pooled to form one continuous event.

4.3.3 Initial configuration

To begin the search procedure an initial solution, or formation for the tephras across the five cores, must be specified. The search procedure begins by forcing events in the marker horizons to be matched, while considering each unknown tephra to be a separate eruption. Alternatively, one could start with some reasonable predetermined arrangement for the unknown tephras. This may decrease the time the algorithm takes to find the optimal arrangement, but the downside is that the algorithm may be less likely to explore the entire solution space.

4.3.4 Simulated Annealing

Local search methods (Hoos and Stützle, 2005) can have a tendency to become stuck in suboptimal regions of the search space, or plateaus where many solutions are equally valid, so the local optimum obtained may often not be the global one. An additional technique to help the process escape from local minima is to allow the search process to perform worsening steps. Simulated annealing (Kirkpatrick et al., 1983; Černý, 1985) is a special case of a metaheuristic search procedure that can accept worse solutions to allow for a more extensive search for the optimal solution. It involves optimizing combinatorial problems by a process analogous to the physical annealing process, in which a substance is melted and then cooled slowly until a perfect crystalline structure is obtained. In this analogy, the candidate solutions are the states of the physical system, the objective function is the energy of the physical system, and

the optimal solution is found when the system reaches equilibrium and no further changes occur.

The probability of accepting a step that leads to a worse candidate solution should be dependent on the worsening of the objective function. Significantly worse steps should be given a smaller probability of acceptance. The probability of accepting a worsening step is controlled by an annealing constant k. This annealing constant can be adjusted throughout the duration of the search process according to some schedule, so that the probability of accepting worse solutions decreases slowly. In many cases a schedule in which the annealing constant is updated proportionally ($k \rightarrow \alpha k$) is the most efficient (Kirkpatrick et al., 1983).

Simulated annealing performed for a sufficiently long period is guaranteed to converge to an optimal solution (Geman and Geman, 1984; Hajek, 1988). In practice however, the time to find an optimal solution is often limited by a variety of different termination conditions. This could be when the system reaches a suitable state, or when no improving candidate solutions have been found after a given number of iterations. In order to apply the simulated annealing method to tephra matching we need: a quantitative objective function to optimize; an acceptance probability function; an annealing schedule; and a specified length of time for which the system is to be evolved.

4.3.5 Tabu Search

One of the difficulties with local search methods is that the procedure can visit the same candidate solution multiple times. *Tabu search* (Glover, 1989, 1990) was used to prevent the search from immediately returning to previously visited candidate solutions. The tabu search constrains the search by using memory structures, in the form of a *tabu list*, to store previously visited solutions and classify certain moves as forbidden (or tabu). Each randomly selected segment is added to the tabu list where it remains until a number of iterations (the tabu tenure) have passed. If the tabu tenure chosen is too large then the search path may be overly restricted and the algorithm may fail to locate the optimal solution. If it is too small, the search may take too long.

Tabu search enhances the performance of local search algorithms by forcing the exploration of regions of the solution space that might have otherwise remained unexplored. Cycling is avoided by preventing the search from immediately returning to a particular segment of the tephra record.

4.3.6 Objective function

Whether the new rearrangement is accepted, moving the system to state s' from the current arrangement s, depends on the value of the objective function. Ward and Wilson (1978) describe how, given a series of ^{14}C age estimates with mean ages $(\mu_i, i = 1, \ldots, n)$ and associated normally distributed errors $(\sigma_i, i = 1, \ldots, n)$ one can test the hypothesis that the age estimates correspond to the same eruption. This involves first calculating the pooled age given by

$$\mu_p = \frac{\sum_{i=1}^{n} \mu_i/\sigma_i^2}{\sum_{i=1}^{n} 1/\sigma_i^2}, \tag{4.5}$$

which has a variance

$$\sigma_p^2 = \left(\sum_{i=1}^{n} \frac{1}{\sigma_i^2} \right)^{-1}. \tag{4.6}$$

The tephras can be considered to be from the same eruption if the age estimates are judged not to be significantly different. This is tested using the statistic

$$T = \sum_{i=1}^{n} \frac{(\mu_i - \mu_p)^2}{\sigma_i^2} \sim \chi^2_{n-1}. \tag{4.7}$$

This is equivalent to (4.4) when $n = 2$. Such a T statistic is obtained for each composite event in the record. If there are m distinct eruptions in the current arrangement, then summing (4.7) over these gives a likelihood ratio test statistic

$$\sum_{i=1}^{m} T = -2 \log L. \tag{4.8}$$

Although arrangements resulting in a smaller statistic are more likely, different arrangements will have a different number of events depending on how many tephras are matched across the cores. As unmatched tephras contribute zero to the likelihood ratio statistic (4.8) in spite of representing an event, we need a means of compensating for the effect of allowing additional events. For this the Akaike Information Criterion (Akaike, 1977) is used:

$$\text{AIC} = -2 \log L + 2m. \tag{4.9}$$

AIC is used to measure the relative quality of statistical models, where m denotes the number of parameters in a statistical model. In this context the number of parameters m is the number of distinct mean ages/events in a given tephra arrangement. Since AIC is dependent on σ_i

through (4.7) and (4.8), it is affected by the choice of non-linearity constant c in (4.3). AIC rewards more likely (higher $\log L$) tephra arrangements but it includes a penalty that increases with the number of events in any arrangement. Smaller AICs indicate better arrangements, so the objective function is to minimize the AIC for the arrangement of tephras across the five cores.

After randomly selecting a segment of the current arrangement from the list, all possible rearrangements of that portion of the record are considered. These combinations range from all ages corresponding to separate events in the record, through to arrangements with multiple events represented in all five cores. Note that the possible *re*arrangements *exclude* the current arrangement.

The 'best' rearrangement of the selected segment is the one which results in the smallest AIC. In some cases the 'best' rearrangement may be worse than the current arrangement, in which case a decision is made, probabilistically, whether or not to accept it. The probability of accepting a new arrangement is given by

$$p(k, s, s') = \begin{cases} 1, & \text{if } \mathrm{AIC}_{s'} \leq \mathrm{AIC}_s \\ \exp\left[-k(\mathrm{AIC}_{s'} - \mathrm{AIC}_s)\right], & \text{otherwise} \end{cases} \tag{4.10}$$

where k is an arbitrary *annealing constant* that determines how willing we are to accept 'worse' rearrangements. At small values of k, the algorithm can accept considerably worse steps with a reasonably high probability. For large values of k the probability of accepting a worse rearrangement is low. Eventually as k increases the algorithm will not accept worsening moves. Any rearrangement that results in an improvement in AIC is automatically accepted. This acceptance probability is an adaptation of the Metropolis condition introduced by Metropolis et al. (1953), and permits a more extensive search for the optimal solution arrangement.

Throughout the search process, the annealing constant k is adjusted according to an *annealing schedule*. The algorithm starts initially with k set to an arbitrarily small positive value, and after a certain number of iterations it is increased proportionally ($k \to \alpha k$, where $\alpha > 1$). Initially the procedure will consider many different arrangements, before drifting towards solutions that only differ slightly in AIC values. Eventually it will stabilize around a particular arrangement, choosing to not accept any alternatives. The procedure is terminated once it reaches a value for k that does not accept any worse rearrangements after a certain number

Figure 4.3: Description of the local search algorithm.

```
k = 0.05;                                          \\ initialize annealing constant k
tabuList = [];                                     \\ initialize tabuList as empty

while ContinueSearch~=0                            \\ continue until stopping condition met
  ContinueSearch = 0;                              \\ initialize stopping condition
  Counter = 0;                                     \\ initialize run count since updating k

  while Counter < CounterMax                       \\ if the count is less than our
                                                   \\ prespecified max count
    List = setdiff(1:length(ListOfMoves),tabuList);\\ find rows from ListOfMoves not in tabuList
    MoveNo = randsample(List,1);                   \\ select random MoveNo from those
                                                   \\ not in tabuList

    tabuList(end+1) = MoveNo;                      \\ add selected MoveNo to the tabuList
    if numel(tabuList) > tabuTenure                \\ if the tabuList is full
      tabuList = tabuList(2:end);                  \\ let the top row expire
    end

    [AICnew,MatchNew] = Rearrange(MoveNo,...       \\ find the best rearrangement
                                MatchOld,AICold);  \\ of the selected segment

    if AICnew > AICold                             \\ if the new AIC is worse than the old AIC
      if exp(-k*(AICnew-AICold)) > rand()          \\ should we accept the worse match?
        AICold = AICnew; MatchOld = MatchNew;      \\ yes, update the old combination and AIC
        ContinueSearch = 1;                        \\ ammend stopping condition
      end
    else                                           \\ if the new AIC is better than the old AIC
      AICold = AICnew;  MatchOld = MatchNew;       \\ update the stored old combination and AIC
      if AICnew < AICbest                          \\ if the new AIC is better than the best AIC
        AICbest = AICnew; MatchBest = MatchNew;    \\ update the best combination and AIC
      end
    end
    Counter = Counter + 1;                         \\ add 1 to the count since updating k
  end
  k=alpha*k;                                       \\ update the annealing constant.
end

MatchBest                                          \\ return the optimal solution
```

of iterations. At this point the optimal solution is returned. Unfortunately, there is no set number of iterations to perform before updating the annealing constant k, or a value for the multiplier α, that will be suitable for all problems, hence some trial and error is required to develop an appropriate annealing schedule.

The algorithm for the local search is described in Figure 4.3. Matlab code to perform the automated statistical matching algorithm is presented in Appendix C. The associated data used in this analysis can also be found in Appendix C or in the electronic supplementary material accompanying Green et al. (2014).

4.4 Results

Application of the local search algorithm to the Auckland Volcanic Field data resulted in the arrangement seen in Table 4.2, where the non-linearity constant in (4.3) is $c = 0.5$ ka. Comparison between this arrangement and the previously published arrangement from Molloy et al. (2009) is shown in Figures 4.4 and 4.5. The following results all assume a non-linearity constant of 0.5 ka except where specifically stated otherwise.

The search procedure began with an initial formation that had all tephras corresponding to separate eruptions. The specified marker horizons were then aligned across the cores, resulting in an arrangement that had an objective function value, or AIC, of 344.95.

Table 4.2: Resulting arrangement of tephra thicknesses and estimated ages across the cores when the non-linearity constant added equals 0.5 ka. Marker tephras used to constrain the ages are denoted with an asterisk with abbreviations given in Table 4.1.

Source	Thickness (mm)					Age (ka)					Combined
	Pupuke	Onepoto	Orakei	Hopua	Pukaki	Pupuke	Onepoto	Orakei	Hopua	Pukaki	age (ka)
AVF	22					0.80 ± 0.50					0.80 ± 0.50
Tp*	3					1.72 ± 0.01					1.72 ± 0.01
Tu*	15		18	20	70	6.51 ± 0.27		6.63 ± 0.26	6.56 ± 0.29	6.55 ± 0.28	6.56 ± 0.14
Ma*				1	2				7.95 ± 0.13	7.96 ± 0.11	7.95 ± 0.08
Rm*	45	43		10	7	9.43 ± 0.06	9.43 ± 0.06		9.42 ± 0.06	9.43 ± 0.07	9.43 ± 0.03
Eg		2					9.81 ± 0.50				9.81 ± 0.50
AVF				3					9.97 ± 0.50		9.97 ± 0.50
Op*		2		2	4		10.00 ± 0.07		10.00 ± 0.07	9.99 ± 0.09	10.00 ± 0.05
Tg					1					12.19 ± 0.50	12.20 ± 0.50
Eg	1	0.5			4	12.24 ± 0.50	11.64 ± 0.50			12.24 ± 0.50	12.04 ± 0.29
Eg		2			1		12.32 ± 0.50			12.76 ± 0.50	12.54 ± 0.36
Eg	1	2		3	1	13.17 ± 0.50	12.76 ± 0.50		13.59 ± 0.51	12.90 ± 0.50	13.10 ± 0.25
Tg		0.5					13.21 ± 0.51				13.21 ± 0.51
Eg				1					13.82 ± 0.51		13.82 ± 0.51
Wh*	2	2		3	3	14.01 ± 0.08	14.01 ± 0.09		14.01 ± 0.08	14.01 ± 0.07	14.01 ± 0.04
Eg		0.5		1			14.43 ± 0.51		14.13 ± 0.51		14.28 ± 0.36
Tg				1	1				14.24 ± 0.50	14.10 ± 0.50	14.17 ± 0.36
Tg				1	1				14.49 ± 0.51	14.52 ± 0.51	14.51 ± 0.36
Tg		2			1		14.91 ± 0.52			14.67 ± 0.51	14.79 ± 0.36
Tg	1				1.5	15.43 ± 0.53				15.30 ± 0.52	15.37 ± 0.37
AVF					1					15.45 ± 0.53	15.45 ± 0.53
Rr*	45	20	8	25	20	15.64 ± 0.20	15.64 ± 0.23	15.64 ± 0.19	15.60 ± 0.20	15.67 ± 0.20	15.64 ± 0.09
Eg					1					16.01 ± 0.53	16.02 ± 0.53
Eg					1					16.14 ± 0.53	16.14 ± 0.53
Eg	2	4	3	2	2	17.41 ± 0.54	17.46 ± 0.55	17.50 ± 0.52	17.48 ± 0.55	17.43 ± 0.55	17.46 ± 0.24
Rw*	1.5	3		2	1.5	17.46 ± 0.22	17.53 ± 0.24		17.53 ± 0.25	17.45 ± 0.24	17.49 ± 0.12
Eg		2	2	5	8		19.04 ± 0.53	18.94 ± 0.51	19.37 ± 0.52	18.80 ± 0.53	19.04 ± 0.26
AVF				290	3				19.94 ± 0.52	19.54 ± 0.52	19.74 ± 0.37
AVF				235	3				20.40 ± 0.52	19.59 ± 0.52	20.00 ± 0.37
Eg			2	2				20.68 ± 0.51	21.01 ± 0.51		20.84 ± 0.36
Eg			2	2	1			21.68 ± 0.51	21.04 ± 0.51	21.49 ± 0.52	21.40 ± 0.30
Ok*	2	2	3	15	4	21.84 ± 0.14	21.87 ± 0.15	21.85 ± 0.13	21.86 ± 0.13	21.86 ± 0.14	21.86 ± 0.06
Eg			2		2			22.52 ± 0.51		22.98 ± 0.53	22.74 ± 0.37
Eg		1	10	2	1		23.80 ± 0.55	23.21 ± 0.51	23.67 ± 0.51	23.95 ± 0.57	23.64 ± 0.27
AVF			8					23.25 ± 0.51			23.25 ± 0.51
AVF			5		1			23.39 ± 0.51		23.95 ± 0.57	23.64 ± 0.38
Tg				1					24.04 ± 0.51		24.04 ± 0.51

Continued on next page

Table 4.2 – *Continued from previous page*

Source	Thickness (mm)					Age (ka)					Combined
	Pupuke	Onepoto	Orakei	Hopua	Pukaki	Pupuke	Onepoto	Orakei	Hopua	Pukaki	age (ka)
Eg	1		5	1	1	24.31 ± 0.51		24.32 ± 0.50	24.54 ± 0.51	24.11 ± 0.58	24.33 ± 0.26
AVF			12		0.5			24.42 ± 0.50		24.26 ± 0.59	24.35 ± 0.38
Te*		0.5			0.5		24.95 ± 0.34			24.82 ± 0.38	24.89 ± 0.26
AVF			340		50			24.53 ± 0.50		24.88 ± 0.61	24.67 ± 0.39
Eg				2	1				24.55 ± 0.51	25.17 ± 0.53	24.85 ± 0.36
AVF			740	40				25.01 ± 0.50	24.94 ± 0.51		24.97 ± 0.36
Kk*	25	27	10	60	22	25.36 ± 0.08	25.37 ± 0.09	25.35 ± 0.08	25.36 ± 0.08	25.37 ± 0.08	25.36 ± 0.04
Eg		0.5			0.5		27.31 ± 0.54			28.17 ± 0.58	27.71 ± 0.40
Eg		2		2	0.5		28.06 ± 0.57		28.11 ± 0.61	28.29 ± 0.59	28.15 ± 0.34
Po*	1.5	1		1	1	28.31 ± 0.31	28.38 ± 0.30		28.46 ± 0.39	28.50 ± 0.33	28.41 ± 0.16
Eg	2					28.33 ± 0.59					28.33 ± 0.59
AVF	7	12	410	335	70	28.56 ± 0.59	28.45 ± 0.58	28.18 ± 0.53	28.59 ± 0.64	28.63 ± 0.61	28.46 ± 0.26
Eg	2	2	8		4	29.01 ± 0.73	28.97 ± 0.69	29.50 ± 0.55		28.96 ± 0.68	29.16 ± 0.33
O*	6	7			15	29.03 ± 0.55	29.01 ± 0.50			29.04 ± 0.51	29.03 ± 0.30
Eg	4				2	30.33 ± 0.69				30.14 ± 0.66	30.23 ± 0.48
AVF	3		400		1	30.40 ± 0.69		29.76 ± 0.56		30.26 ± 0.65	30.09 ± 0.36
AVF	6	2	45			30.60 ± 0.68	30.75 ± 0.64	30.28 ± 0.57			30.52 ± 0.36
AVF	20	4	20		490	30.86 ± 0.67	30.80 ± 0.64	30.39 ± 0.57		31.43 ± 0.62	30.85 ± 0.31
AVF	2	8			200	31.77 ± 0.65	31.77 ± 0.63			32.48 ± 0.62	32.02 ± 0.36
Eg	0.5					31.94 ± 0.65					31.94 ± 0.65
Eg	1					32.23 ± 0.64					32.23 ± 0.64
AVF					5					32.89 ± 0.62	32.89 ± 0.62
AVF					500					33.05 ± 0.62	33.05 ± 0.62
Eg			1		2			33.71 ± 0.68		33.93 ± 0.64	33.83 ± 0.47
Eg					1					33.96 ± 0.64	33.96 ± 0.64
AVF			110					34.31 ± 0.70			34.31 ± 0.70
Eg	1		1		1	34.44 ± 0.65		34.90 ± 0.72		34.08 ± 0.65	34.44 ± 0.39
AVF	15	12	35			34.61 ± 0.66	34.87 ± 0.70	35.27 ± 0.74			34.90 ± 0.40
OVC		0.5					35.71 ± 0.74				35.71 ± 0.74
Hauparu	13	6	5		5	35.63 ± 0.69	36.09 ± 0.76	36.01 ± 0.77		35.95 ± 0.74	35.90 ± 0.37
Mk*	60	40	40		25	36.45 ± 0.52	36.30 ± 0.60	36.41 ± 0.60		36.31 ± 0.58	36.37 ± 0.29
Eg			0.5		1			37.32 ± 0.77		37.39 ± 0.75	37.36 ± 0.54
OVC		1					37.88 ± 0.90				37.88 ± 0.90
Eg	1		1			37.88 ± 0.79		38.59 ± 1.07			38.13 ± 0.63
OVC		0.5			1		38.16 ± 0.96			38.86 ± 1.06	38.48 ± 0.71
Ta*	8	4	5		4	39.32 ± 1.03	39.34 ± 1.22	39.31 ± 1.22		39.52 ± 1.16	39.37 ± 0.57
Eg	0.5				1	40.67 ± 0.98				40.28 ± 1.17	40.51 ± 0.75
OVC		0.5					40.76 ± 1.09				40.76 ± 1.09
Eg		1	2		1		43.38 ± 1.23	43.12 ± 1.30		43.21 ± 1.25	43.24 ± 0.73
OVC		1					43.62 ± 1.28				43.62 ± 1.28
Eg		0.5			1		43.71 ± 1.30			43.61 ± 1.32	43.66 ± 0.92
Eg	1	2				44.00 ± 1.33	43.78 ± 1.31				43.89 ± 0.93
Eg	3	1				44.70 ± 1.50	44.52 ± 1.48				44.61 ± 1.05
Eg		2			3		44.92 ± 1.59			44.92 ± 1.59	44.92 ± 1.12
Re*	630	530	70		430	45.13 ± 1.53	45.10 ± 1.55	45.02 ± 1.73		45.11 ± 1.56	45.10 ± 0.79
AVF			120					48.61 ± 3.00			48.61 ± 3.00
Eg	5	5	1		5	50.86 ± 3.30	52.68 ± 4.06	53.37 ± 4.73		49.81 ± 2.96	51.19 ± 1.79
AVF			510					56.84 ± 6.03			56.84 ± 6.03
Eg			6		0.5			57.74 ± 6.36		55.08 ± 4.61	56.00 ± 3.73
Eg		4	2		3		66.27 ± 8.79	58.13 ± 6.51		56.26 ± 4.99	58.52 ± 3.61
Eg		1	2		2		68.67 ± 9.63	60.03 ± 7.22		56.39 ± 5.03	59.30 ± 3.79
MI			2					64.27 ± 8.82			64.27 ± 8.82
Eg		3	3				71.19 ± 10.51	64.31 ± 8.84			67.16 ± 6.77
AVF		4	40				73.72 ± 11.40	66.39 ± 9.63			69.45 ± 7.36
Eg		1					75.47 ± 12.02				75.47 ± 12.02

The annealing constant k was initialized with a value of 0.05. This is small enough that even significantly worse arrangements (where the change in AIC exceeds 50) could be accepted approximately 10% of the time. After each 3000 iterations, the annealing constant was doubled. Slower annealing schedules, where the annealing constant was only increased by 20-50% each time, resulted in the same optimal arrangement but required a significantly longer run time. In each iteration, an element from the list of possible moves was randomly selected and the associated segment was rearranged, the decision to accept or reject the new formation being made according to (4.10), and the current AIC checked against the best AIC found so far to see if an improved arrangement was found.

The list of possible moves consists of 752 different two-way matches; each of which selects a segment of the record to rearrange. In order to prevent the search procedure from returning to a particular segment of the record multiple times, a tabu list of length 200 was created. Once an element from the list of possible moves had been selected, and the rearrangement of the associated segment was performed, the element was added to the tabu list. This element of the list could not be reselected within the next 200 iterations of the search procedure. Various different tabu tenures were considered, ranging in length from 50 to 400. However, the smaller tabu lists did not explore the solution space enough for the optimal arrangement observed in Table 4.2 to be found.

The search algorithm ran for a total of 48000 iterations at which point there had been no improvements to the best AIC value for some time, and the procedure had not accepted any worse matches for over 3000 iterations (see Figure 4.6). The optimal arrangement of tephras across the five cores had an AIC value of 205.69.

Using the estimated ages and associated errors obtained through linear interpolation, the AIC associated with the previously published arrangement of tephras given in Molloy et al. (2009) was calculated. Assuming that the inherent error associated with the linear interpolation of ages is 0.5 ka, the AIC obtained under the previously published arrangement is 278.13. The proposed automated procedure resulted in an arrangement with a better AIC after only 16 iterations. The AIC improvement and the optimal arrangement, are dependent on the value of the nonlinearity constant defined in (4.3), as shown in Table 4.3.

The number of marker tephras is invariant across all arrangements because these matches were forced across all cores. The large age separation between individual Okataina, Mayor Island, Tongariro, and Auckland Volcanic Field events leaves little room for changes in the

Figure 4.4: Stratigraphy of post-29 cal ka BP tephra layers. *Dotted lines* show the tephra correlations from Molloy et al. (2009). *Dashed lines* show the correlations returned by the automated procedure when the non-linearity constant defined in (4.3) equals 0.5 ka. *Solid lines* show correlations consistent between the arrangement obtained using the automated procedure and that of Molloy et al. (2009).

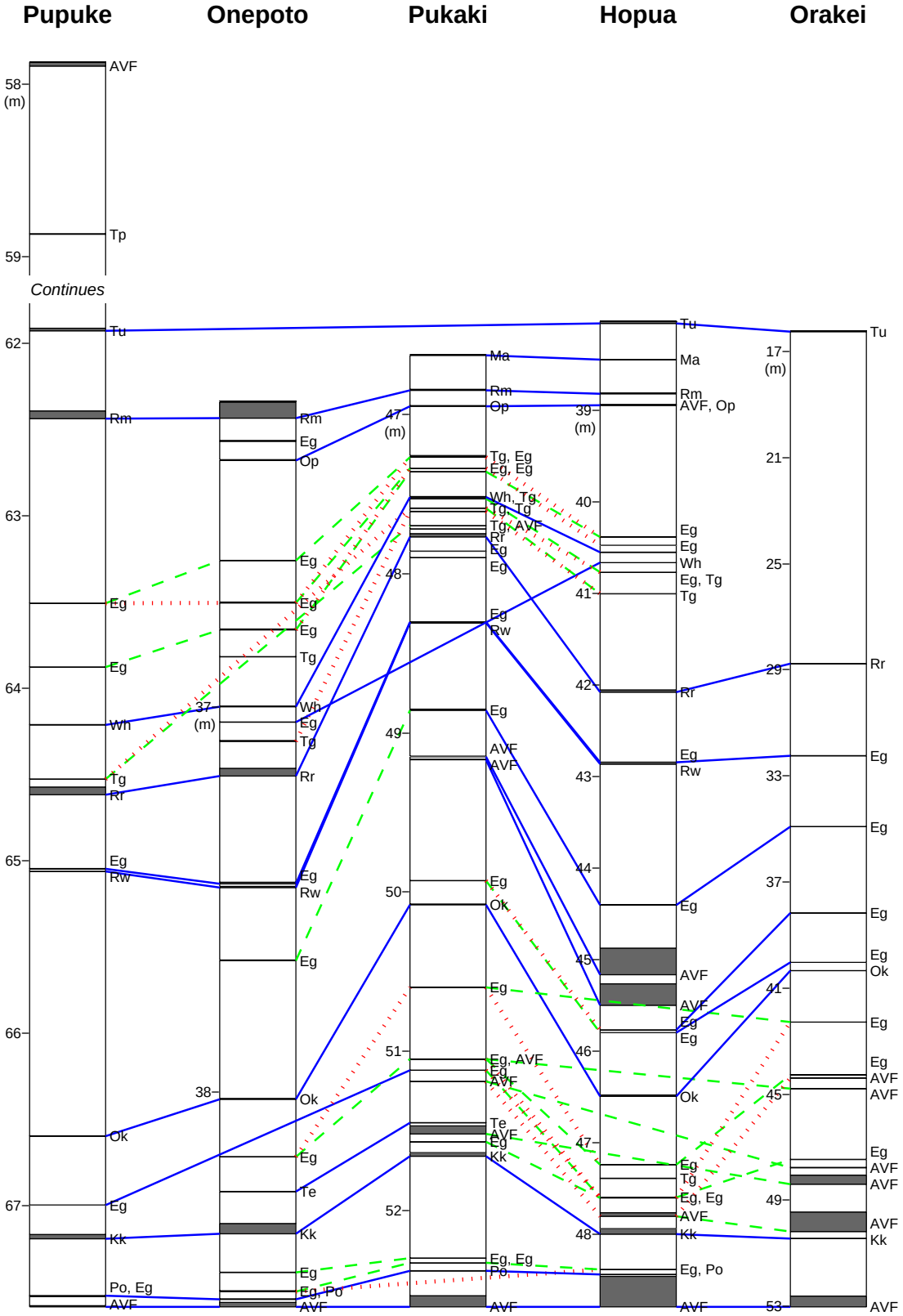


Figure 4.5: Stratigraphy of pre-29 cal ka BP tephra layers. *Dotted lines* show the tephra correlations from Molloy et al. (2009). *Dashed lines* show the correlations returned by the automated procedure when the non-linearity constant defined in (4.3) equals 0.5 ka. *Solid lines* show correlations consistent between the arrangement obtained using the automated procedure and that of Molloy et al. (2009).

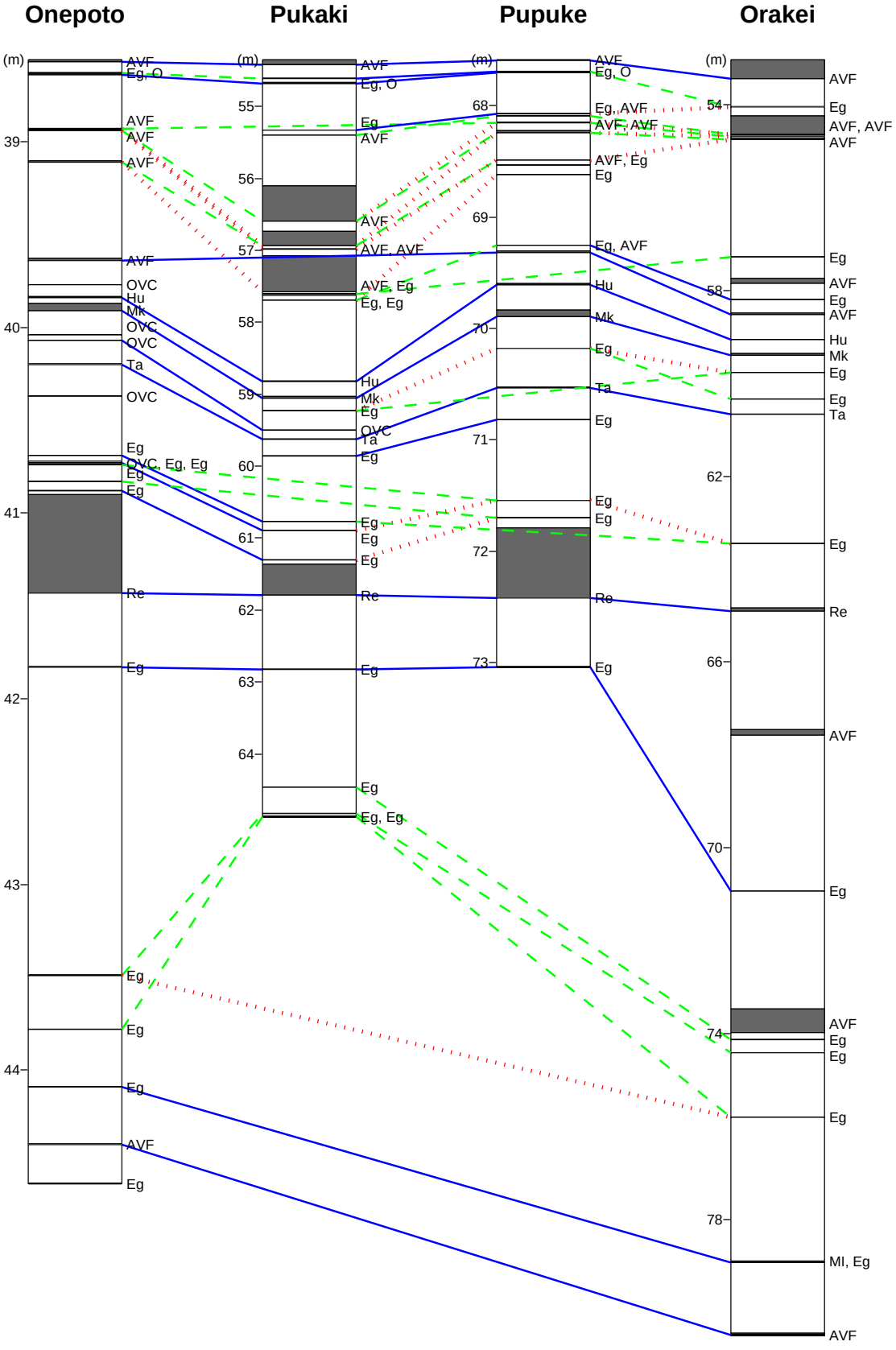


Figure 4.6: Plot of the changes in AIC over the duration of the heuristic optimization.

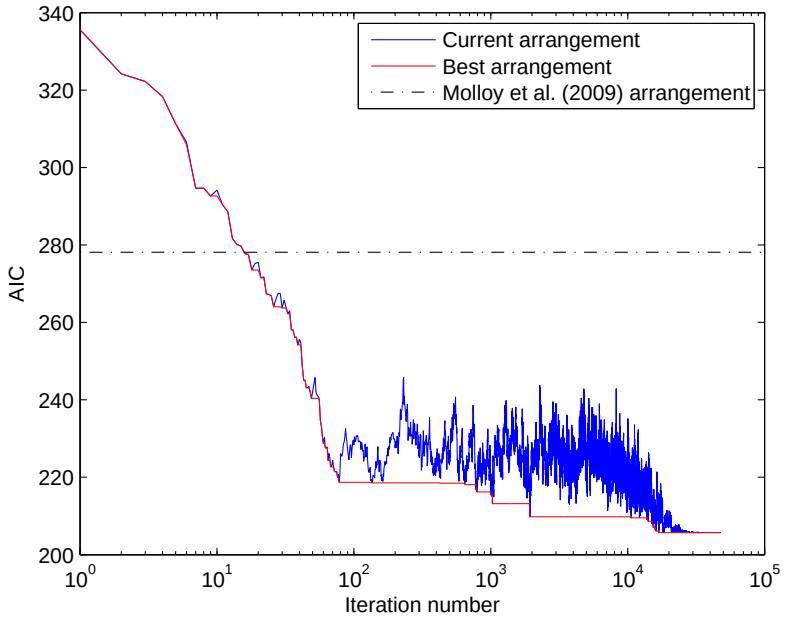


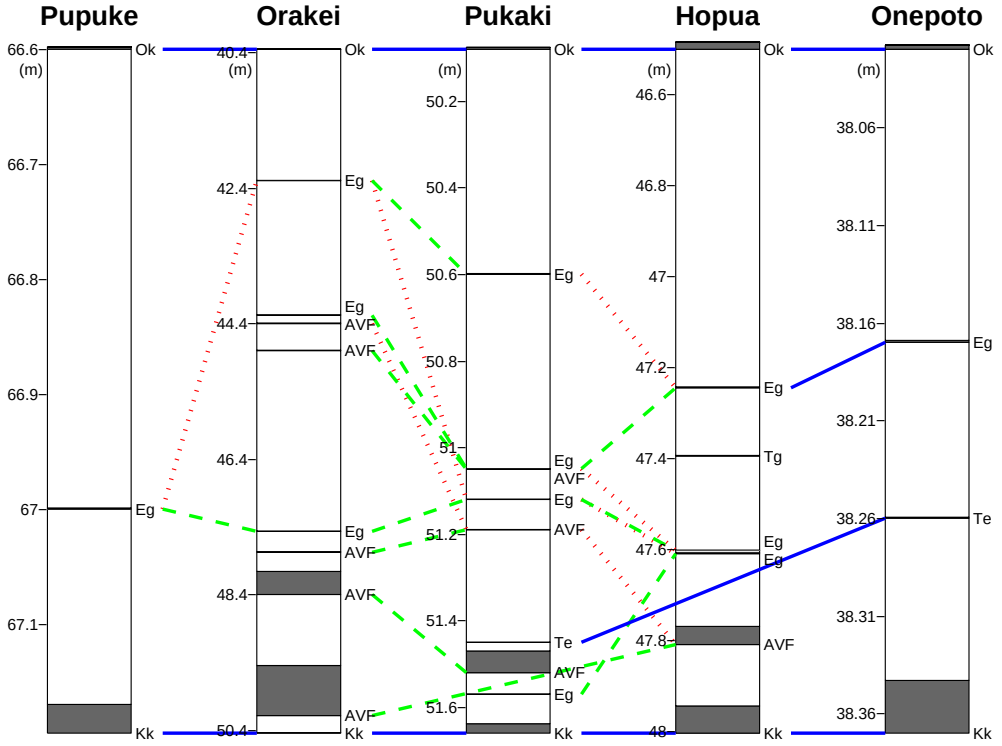
Table 4.3: A comparison of the number of distinct tephra from each volcanic center observed across all five cores when the non-linearity constant, defined in Equation (4.3), varies. ΔAIC represents the difference in AIC between the arrangement of Molloy et al. (2009) and that returned by the automated procedure. ΔAIC > 2 indicates a statistically significant improvement (Utsu, 1999).

	Non-linearity constant			
Source	0.3	0.5	0.7	Molloy et al. (2009)
Eg	45	40	40	52
AVF	24	22	22	24
OVC	6	6	6	6
Tg	7	7	7	7
MI	1	1	1	1
Marker tephras	16	16	16	16
Total number	99	92	92	106
ΔAIC	135.09	72.43	52.23	-

tephra alignments. Of particular interest however, is that by increasing the value of the non-linearity constant, the number of inferred Taranaki tephras decreases by up to 23%. This reflects an increasing likelihood of events overlapping among cores, permitting tephras that may have previously been left unmatched to merge.

The difference in AIC (ΔAIC) between the Molloy et al. (2009) arrangement and the arrangement returned by the automated procedure can be seen in Table 4.3. A difference of approximately two indicates a statistically significant improvement (Utsu, 1999). Hence, each

Figure 4.7: Stratigraphy of tephra layers found between the Okareka and Oruanui marker horizons. *Dotted lines* show the tephra correlations from Molloy et al. (2009). *Dashed lines* show the correlations returned by the automated procedure when the non-linearity constant defined in (4.3) equals 0.5 or 0.7 ka. *Solid lines* show correlations consistent between the arrangement obtained using the automated procedure and that of Molloy et al. (2009).



amalgamated record obtained using the automated procedure is, to a great extent, statistically more likely than the previously published arrangement. Although the total number of events does not vary greatly between arrangements, the alignment of tephras across the cores does. This is illustrated by a segment of the record that differs the most among the three arrangements obtained using the automated procedure and that identified by Molloy et al. (2009). The segment of the record consists of all tephras lying between the Okareka and Oruanui marker tephras. The core stratigraphy and likely correlations associated with this segment of the record are given in Table 4.4 and Figure 4.7.

The arrangement of Molloy et al. (2009) shown in Table 4.4(a) contains six Taranaki/Egmont tephras, one Tongariro/Ruapehu tephra, and seven AVF tephras. The single Tongariro unit is observed in all arrangements of Table 4.4 but the total number of Egmont and AVF events changes depending on the value of the non-linearity constant. In the analysis of Molloy et al. (2009) the Oruanui marker horizon was constrained at an age of 27.10 ± 0.96 cal ka BP (Lowe et al., 2008). The revised age of 25.36 ± 0.16 cal ka BP adopted in this analysis from Lowe et al. (2013) is much younger; as a result there are far fewer tephras observed

Table 4.4: Sensitivity in arrangements of tephras between the Okareka and Oruanui marker horizons to variability in sediment deposition rates.

(a) Arrangement from Molloy et al. (2009).

Source	Thickness (mm)				
	Pupuke	Onepoto	Orakei	Hopua	Pukaki
Okareka	2	2	3	15	4
Eg38		1		2	2
Tg1				1	
Eg37				1	1
AVF19					1
Eg36	1		2	2	1
Eg35			10		
AVF18			8	40	0.5
AVF17			5		
Te Rere		0.5			0.5
AVF16					50
Eg34					1
Eg33			5		
AVF15			12		
AVF14			340		
AVF13			740		
Oruanui	25	27	10	60	22

(b) Non-linearity constant = 0.3

Source	Thickness (mm)				
	Pupuke	Onepoto	Orakei	Hopua	Pukaki
Okareka	2	2	3	15	4
Eg			2		
Eg			10		2
AVF			8		
Eg		1		2	1
AVF			5		1
Tg				1	
Eg	1		5	1	1
AVF			12		0.5
AVF			340		
Eg				2	
Te Rere		0.5			0.5
AVF			740	40	50
Eg					1
Oruanui	25	27	10	60	22

(c) Non-linearity constant = 0.5, 0.7

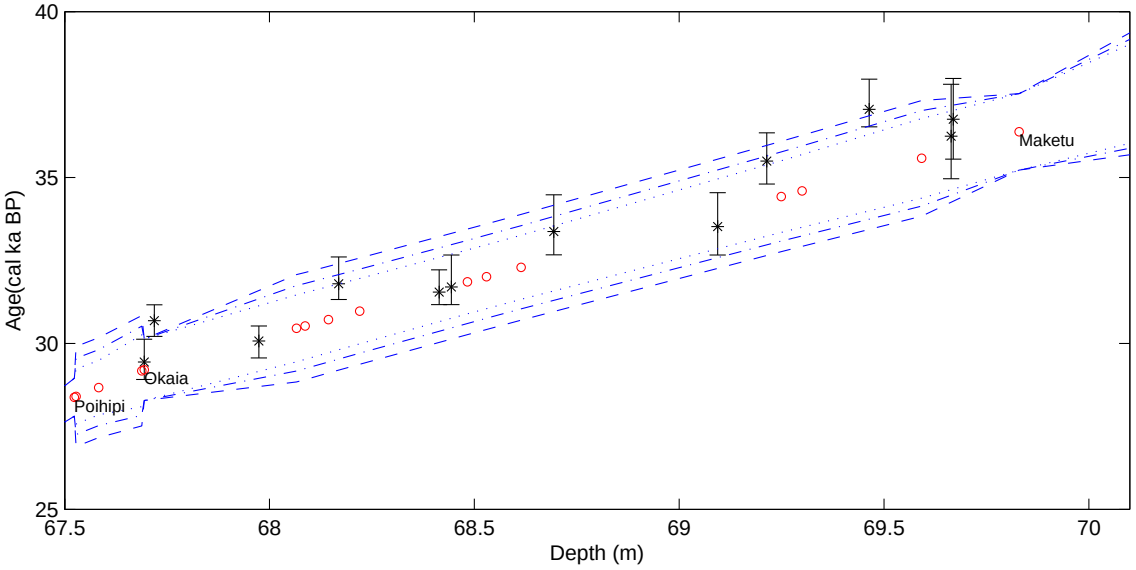
Source	Thickness (mm)				
	Pupuke	Onepoto	Orakei	Hopua	Pukaki
Okareka	2	2	3	15	4
Eg			2		2
Eg		1	10	2	1
AVF			8		
AVF			5		1
Tg				1	
Eg	1		5	1	1
AVF			12		0.5
Te Rere		0.5			0.5
AVF			340		50
Eg				2	1
AVF			740	40	
Oruanui	25	27	10	60	22

between the Te Rere and Oruanui marker horizons irrespective of the non-linearity constant. The tephras in this segment that were recovered from Hopua Crater and Pukaki Crater are likely to be much older and thus match with the older tephras recovered from Orakei Basin. In the arrangement of Molloy et al. (2009) (Table 4.4(a)) there are a large number of units that are observed in only a single core. When the non-linearity constant $c = 0.3$ ka, two fewer AVF units are identified, and the arrangement of many of the Taranaki/Egmont events changes (Table 4.4(b)). Increasing the non-linearity constant c to 0.5 or 0.7 ka results in more tephras being able to match across the cores (Table 4.4(c)). The total number of events in the segment is compressed with two fewer Taranaki/Egmont units due to the older tephras in the Pukaki core being shifted up relative to the other cores, resulting in only four Egmont units and five AVF units, with all but one of them being observed in more than one of the locations.

Table 4.4 highlights a section of the record where the arrangement changes the most depending on the choice of the non-linearity constant. For the remainder of the record, the arrangements returned by the automated procedure are not as sensitive to the choice of the non-linearity constant in (4.3) added during the linear interpolation step. Using $c = 0.5$ ka as the baseline, the arrangements returned when $c = 0.3$ or 0.7 ka differs only slightly:

- Two more Taranaki/Egmont units are identified between the Opepe and Waiohau marker horizons when $c = 0.3$ ka.
- When $c = 0.3$ ka the 235 mm thick AVF event recovered from Hopua crater between the Rerewhakaaitu and Okareka marker horizons are separated from the rest, resulting in one additional unit in this segment.
- There is one additional Egmont/Taranaki unit identified between the Oruanui and Poihipi marker horizons when $c = 0.3$ ka. The Egmont/Taranaki tephras recovered from Pukaki crater are considered to be slightly older than those found in the other cores; as a result the oldest one corresponds to a separate event.
- Finally, when $c = 0.3$ ka there is one additional AVF unit identified between the Okaia and Maketu tephras; the 200 mm thick AVF tephra recovered from Pukaki crater is considered to be from an older separate eruption.

Figure 4.8: Estimated ages for the tephras recovered between the Poihipi and Maketu marker horizons in the Lake Pupuke sediment core. *Circles* denote the estimated mean ages obtained through linear interpolation. The *dotted lines, dot-dashed lines,* and *dashed lines* shows the 2σ bands for these estimates obtained using (4.3) where $c = 0.3, 0.5$ and 0.7 ka respectively. The error bars show the age ranges from Nilsson et al. (2011) associated with this segment of the record.



4.5 Discussion

4.5.1 Validation of the non-linearity constant

Nilsson et al. (2011) provide a detailed record of a portion of the Lake Pupuke sediment core, where a large number of individual ^{14}C age determinations were made. The age determinations between the Poihipi and Maketu marker horizons are compared, in Figure 4.8, with the estimates obtained through the local search algorithm. While the age determinations appear to agree within the given margins of error, it is possible to derive an approximate estimate of c from the Nilsson et al. (2011) ages. This is done by first using Monte Carlo simulation from the ^{14}C ages, converting them to calendar years and discarding any age sequence that contradicts the stratigraphy. Using the medians of the resulting age distributions, a linear regression analysis in conducted. The residual standard error, $s = 0.485$ ka, of the median ages around the linear fit is then the estimate of non-linearity. This validates the choice of non-linearity constants $c = 0.5 \pm 0.2$ ka. Hence the remainder of the discussion focuses on the arrangement with non-linearity constant $c = 0.5$ ka.

Table 4.5: Probability of observing distal tephras at a given maar, assuming a non-linearity constant of $c = 0.5$ ka. AVF events are not included. Max age (cal ka BP) denotes the mean age for the oldest event observed in each core.

	Pupuke	Onepoto	Orakei	Hopua	Pukaki
Max age (cal ka BP)	51.2	75.5	69.5	28.5	59.3
Observed #	30	42	27	24	50
Expected #	64	67	68	37	66
Probability	0.47	0.63	0.40	0.65	0.76

4.5.2 Implications for detecting tephras in maars

Using the arrangement given in Table 4.2 the probability of observing a given event at each of the five locations (Table 4.5) can be examined. Other eruptions may have occurred during this time without being observed in any of the cores so the expected number in Table 4.5 represents the minimum possible number of tephras observable at a location, i.e., given that the unit was observed in at least one of the five locations. The probability of observing an AVF event at each of these locations would depend largely on the proximity to the vent, therefore only the probability of observing the distal Eg, Tg, OVC and MI events is considered. These should blanket the Auckland area approximately uniformly, so the presence or absence in the record should be a function of the given maar alone.

The southernmost maar, Pukaki crater, has the greatest probability of recording distal tephras. The probability of observing a distal tephra decreases farther north, with Lake Pupuke, 24 km north, preserving only 47% of the distal tephras found in the Auckland region. Orakei Basin has an unusually small observation probability, lying in the middle of the Auckland region but preserving only 27 distal tephras out of an expected 68 (40%). Many of the widespread and easily distinguishable, marker tephras such as the Rotoma, Waiohau, Rerewhakaaitu, Poihipi and Okaia are missing in the Orakei core, but were observed at all of the other locations. Although the Rotorua, Oruanui, and Rotoehu marker tephras appear across all five cores, the thicknesses observed in Orakei Basin are noticeably smaller than at the other locations. The Rotoehu tephra was between 430 and 630 mm thick at the other locations but only 70 mm thick in Orakei Basin. However, although the probability of observing distal tephras is particularly small at Orakei Basin, the thicknesses of many of the Egmont and Tongariro tephras that were observed are much greater.

Figure 4.9: Boxplots for the andesitic tephra thicknesses.

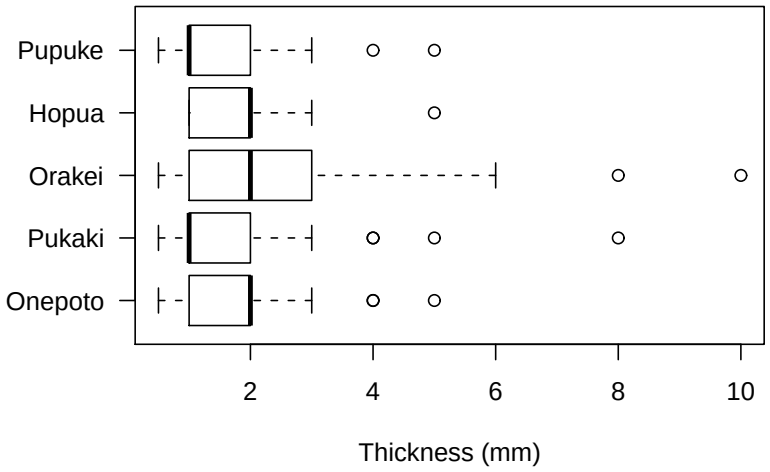


Figure 4.9 shows boxplots for the thicknesses of the andesitic tephras. An analysis of variance returns a P-value of 0.057 against the null hypothesis of equal means; Orakei Basin being the obvious cause. Orakei Basin erupted in an active river system (Németh et al., 2012), and while it cuts off the base of the stream system, a large catchment feeds it and the stream recuts through the tuff ring. This basin is also surrounded by steep and highly unstable soils, with active soil slip on the northern inner face. This basin is thus less likely to preserve records of tephras undisturbed and may in some cases encourage erosion (when stream flows into and through the basin are high) or over-thickening, when tephras are washed into the basin from its large surrounding catchment.

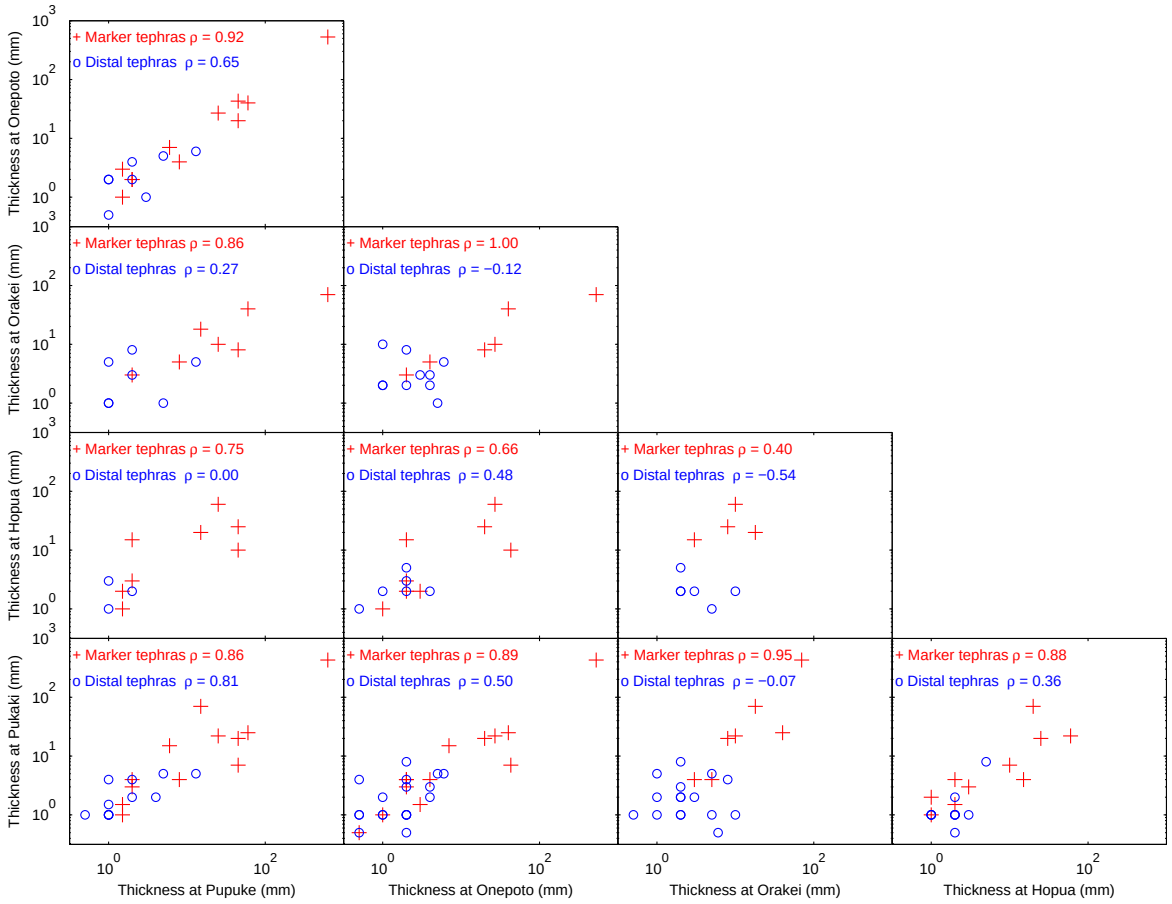
4.5.3 Quantifying the variability in observance probabilities

Figure 4.10 shows the correlations in thickness for identified common tephras among the five maars. While there are generally strong correlations for the marker tephras (the Spearman rank correlation is used to avoid leverage from the larger thicknesses of these tephras), correlations between distal andesitic tephra thicknesses are often negligible, even negative in the case of Orakei/Hopua, Orakei/Onepoto, and Orakei/Pukaki.

In order to quantify the variability in the probability of observing a distal tephra within a maar a statistical model is formulated with likelihood function

$$L(p_j, q, r_j | x_{i,j}, t_i) = \prod_{j=1}^{5} \prod_{i=1}^{n} (1-p_j)^{1-\mathbf{I}_{(0,\infty)}(x_{i,j})} \times \left(\frac{p_j f(x_{i,j})}{1 - \Phi\left(\frac{-r_j}{q}\right)} \right)^{\mathbf{I}_{(0,\infty)}(x_{i,j})} . \qquad (4.11)$$

Figure 4.10: Tephra thicknesses observed at each of the locations; ρ denotes the Spearman rank correlations between the observed thicknesses.



The second term represents a truncated (> 0) normal distribution where

$$f(x_{i,j}) = \frac{1}{qt_i\sqrt{2\pi}} \exp\left(-\frac{(x_{i,j} - r_j t_i)^2}{2(qt_i)^2}\right), \tag{4.12}$$

$$\Phi\left(\frac{-r_j}{q}\right) = \frac{1}{2}\left[1 + \text{erf}\left(\frac{-r_j}{q\sqrt{2}}\right)\right], \tag{4.13}$$

$x_{i,j}$ denotes the tephra thickness from event i that was observed at site j, and $\mathbf{I}_{(0,\infty)}(x_{i,j})$ is an indicator function that equals one if event i was observed at site j (i.e. if $x_{i,j} > 0$) and zero otherwise. The observed thicknesses have means (across the maars where the tephra was observed) of $t_i = \overline{x_{i,j}} : x_{i,j} > 0$. The site numbers $j = 1, \ldots, 5$ correspond to Lake Pupuke, Onepoto Basin, Orakei Basin, Hopua Crater, and Pukaki Crater respectively. The parameters to be estimated are the probability of observing any given tephra in maar j, $p_j, j = 1, \ldots, 5$, the site-specific thickness ratios $r_j, j = 1, \ldots, 5$, and the unknown natural thickness variability q. Hence the thickness at which each tephra is observed is dependent on a) it being present, and b) the over- or under-thickening characteristic of that location. The

Table 4.6: Estimated parameters in (4.11) for both the marker tephras and the non-marker distal tephras. Subscripts $1, \ldots, 5$ correspond to Lake Pupuke, Onepoto Basin, Orakei Basin, Hopua Crater, and Pukaki Crater respectively.

Marker tephras	Distal tephras
$p_1 = 0.82, \quad r_1 = 1.16$	$p_1 = 0.34, \quad r_1 = 0.88$
$p_2 = 0.93, \quad r_2 = 0.95$	$p_2 = 0.53, \quad r_2 = 0.92$
$p_3 = 0.50, \quad r_3 = 0.56$	$p_3 = 0.37, \quad r_3 = 1.21$
$p_4 = 0.91, \quad r_4 = 1.16$	$p_4 = 0.54, \quad r_4 = 1.07$
$p_5 = 1.00, \quad r_5 = 1.02$	$p_5 = 0.68, \quad r_5 = 0.98$
$q = 0.48$	$q = 0.42$

natural variability is assumed to be proportional to the mean tephra thickness t_i (Rhoades et al., 2002; Kawabata et al., 2013).

Using the arrangement given in Table 4.2 we can investigate the probability of observing both distal tephras and marker tephras at each of the sampling locations. The parameter estimates p_j, r_j and q are obtained by numerical maximization of the likelihood function given in (4.11) and are shown in Table 4.6. The variability in the tephras is modeled through the denominator $(qt_i)^2$. For both the larger sample of distal tephras and the smaller sample of marker tephras the value of q is similar indicating that the inherent variability in the process is similar, and validating the proportionality assumption.

The observance probabilities are higher for marker tephras than distal (andesitic) tephras, consistent with the former being both larger and more easily distinguished. For both marker and andesitic tephras, the same south-north pattern is seen as in Table 4.5, with Orakei again as an outlier. More interestingly, the relative thicknesses at which Orakei tephras in particular are observed differ depending on whether the tephra corresponds to a marker horizon or a distal event. For the marker events, the relative thickness of tephras observed in Orakei Basin, $r_3 = 0.56$, is much smaller than that of the other locations. However, for the non-marker distal tephras the estimated thicknesses of tephras observed in Orakei Basin are much greater in comparison to the other locations.

Given an average thickness at a given location, which can be determined using an empirical model for polygenetic (Rhoades et al., 2002) or monogenetic (Kawabata et al., 2013) volcanoes, (4.11) provides an estimated distribution of observed thicknesses.

4.5.4 Implications for hazard from distal volcanic centers

The most frequently observed tephras in the Auckland region are from the widespread distal eruptions. Because of the frequency and large dispersal, these distal eruptions pose a large risk to the Auckland region (Newnham et al., 1999). When considering loss of life and infrastructure combined, Mt Taranaki poses the biggest risk to the Auckland region (Magill and Blong, 2005a,b).

The arrangement of Molloy et al. (2009) identified 52 distinct Taranaki/Egmont tephras; 55% of which are observed in only a single core. The arrangement resulting from the automated procedure reduces the number of distinct Taranaki/Egmont tephras to 40 with only 15% of those being observed in a single core. The greatest reduction in Taranaki tephras is between the Okareka and Okaia marker horizons (approximately 21.8-29 cal ka BP), and below the Rotoehu marker horizon ($>$ 45 cal ka BP) (Figures 4.4 and 4.5). These reductions are partly due to the fact that this analysis used the new constrained ages for the Oruanui tephra (Lowe et al., 2013) and the Rotoehu (Danišík et al., 2012) tephra, respectively.

The reduction in the number of Taranaki events has implications for hazard assessment, suggesting that the risk from distal volcanic centers may not be as great as that suggested by Molloy et al. (2009). However, the fact is that there are fewer Taranaki/Egmont tephras observed in only a single site which argues perhaps for a wider impact from each event.

4.5.5 Implications for hazard from the AVF

Bebbington and Cronin (2011) used the arrangement of Molloy et al. (2009) to construct a *feasible* matching between AVF tephras recovered from the maars and the eruptive centers, constrained by stratigraphy, volume and other available age determinations. The result has been the basis of spatio-temporal hazard estimates (Bebbington and Cronin, 2011; Bebbington, 2013), and investigations of a possible geochemical control on the evolution of the field (Le Corvec et al., 2013).

In the arrangement of Molloy et al. (2009) the sole AVF tephra (AVF12 in Molloy et al., 2009) between the Poihipi and Okaia marker horizons appeared in all five maars. Hence this was the foundation in the Bebbington and Cronin (2011) analysis, where it was correlated with the Three Kings eruption. This had the largest tephra volume (Allen and Smith, 1994) from the field, and was of approximately the right age. This alignment is unchanged for 0.3

ka $\leq c \leq$ 0.7 ka, although a small enough value of c can result in the Orakei Basin tephra no longer correlating this with those from the other four maars. Although the top five and bottom five AVF tephras are also the same as in Molloy et al. (2009), the intervening units are assigned and correlated differently.

Overall, the automated procedure results in two fewer AVF events above 'Three Kings', although as discovered by Bebbington and Cronin (2011), the process of successively allocating tephras to sources can result in tephras of the same age at different sites being correlated to separate sources, due to the physical implausibility of a particular spatial distribution. Between the top five AVF tephras and 'Three Kings', the Hopua and Pukaki cores are dragged downwards. This makes one of the Bebbington and Cronin (2011) correlations more likely, as the thick (40 mm) tephra in Hopua is now aligned with the very thick (740 mm) tephra in nearby Orakei Basin, with a putative source (Mt Eden, see Figure 4.1) between them. However, the 340 mm tephra in Orakei Basin is now aligned with a 50 mm tephra in Pukaki, making Mt Hobson a very unlikely source. Below 'Three Kings', the 400 mm tephra in Orakei Basin corresponding to Panmure Basin becomes slightly younger, in accordance with Lindsay et al. (2011). Overall, the new arrangement is at least as amenable as that of Molloy et al. (2009) to the methodology of Bebbington and Cronin (2011), although some correlations will change.

4.5.6 Limitations and extensions of the methodology

This chapter has successfully demonstrated the application of the methodology to five sediment cores from the Auckland volcanic field. The algorithm is easily adapted to merge records obtained from any number of sites. However, caution should be taken when correlating a larger number of sites, as finding the optimal arrangement of tephra is computationally intensive and would require the maximum feasible size of the candidate segments to be reduced. If this falls below the three-sigma limit, the results should be considered carefully. Similarly, finding optimal matches among a very large number of tephras deposited across a relatively (to the age errors) short amount of time would be difficult because of the huge number of feasible combinations.

The method in this chapter has the advantage that it is not restricted to tephra deposits dated using radiocarbon dating. Any alternative dating technique that provides an estimated tephra age and approximately normally distributed associated error would be suitable.

An important capability of the method in this chapter is that of ruling out certain matches by the use of a list of incompatible tephras. In particular, should numerical limitations such as those above impinge on the execution, appropriate analyses can be used to thin out the list of feasible assignments. The more matches that can be ruled out the more efficient the procedure will be at finding a solution. In the case of this Auckland example incompatible matches were determined on the basis of glass chemistry composition. However, the matches considered could be further constrained by considering for example tephra thicknesses, grain size distribution, and/or titanomagnetite chemistry (as will be illustrated in Chapter 5 for Mt Taranaki records). The restrictions applied would depend on the problem at hand and the amount of auxiliary information available.

Besides incompatible matches, computational efficiency can be increased (or equivalently the solution set can be made smaller) by decreasing the age errors. Thus selective dating may be a useful ancillary approach.

4.6 Conclusions

I have developed an algorithm that is capable of finding the statistically most likely matches between multiple tephras at multiple locations. By considering the estimated age and associated error of each observed tephra, and ensuring that the stratigraphy of the events is maintained, I was able to compile the most plausible amalgamated event record. The local search algorithm has the capability to force or eliminate specified tephra matches. The restriction implemented in this example, in addition to forcing matching among tephras in the marker horizons, was that the glass chemistry composition of unknown tephras must be consistent for a match to be allowed, expressed as a list of implausible matches. The methodology can easily accommodate further constraints by adapting the list of matches to exclude tephras with incompatible geochemistry, thickness, grain size distributions, or any other auxiliary information. The new amalgamated record produced by this method is statistically more likely than the previously published arrangement from the Auckland area. The number of tephras, and hence the hazard, from the distal center of Mt Taranaki is reduced by approximately 23%. The new arrangement also alters the matches among the AVF-sourced tephras across cores, and thus the results may have consequences for modeling the spatial-temporal behavior of the Auckland Volcanic Field.

While I have illustrated the methodology through application to multiple tephra records obtained from the Auckland region, the algorithm can be easily implemented on other volcanic tephra records, as will be illustrated in Chapter 5. The procedure automates the task of matching tephra records and significantly improves on the alternative time consuming and arguably subjective ways of matching 'by hand'.

Chapter 5

Bayesian estimation of eruptive volumes

While temporal forecasts for eruption occurrence have been widely studied (see Chapter 2.3), models for forecasting the size of future events are rare, and not particularly informative. The two main models are that of Wadge and Guest (1981), and the size-predictable model discussed in Section 2.3.6. The former is based on a Kolmogorov-Smirnov type envelope defining steady-state, while the latter is a regression model. Probabilistic forecasting of eruption size remains a major volcanological challenge (Bebbington and Marzocchi, 2011). Establishing probabilistic hazard forecasts relies on some knowledge of past events. Chronological records can be compiled to detail the timing of previous events (see Chapter 4), but constructing a record of the size/volume of previous events is more challenging. In this chapter I formulate a model, in a Bayesian framework, to estimate eruptive volumes of Mt Taranaki events.

5.1 Introduction

Eruptive volumes are commonly estimated from tephra thicknesses measured at various exposed locations on and around the volcano. This involves hand-contouring thickness measurements to create isopach maps, from which two dimensional plots of the logarithm of thickness versus the square root of isopach area are constructed (Section 2.4). Various empirical functions are fitted to the data then integrated, with respect to area, to obtain estimates of eruptive volumes.

Pyle (1989) and Fierstein and Nathenson (1992) proposed fitting one or two straight-line segments to the logarithm of thickness, versus the square root of area. This implies an exponential thinning relationship of tephra with distance from the vent. Bonadonna and Houghton (2005) extended this idea to the case of three or more piecewise exponential functions. These methods introduce extra uncertainty into volume estimates, due to the subjective choices of the number of line-segments and the break-in-slope points. To overcome these problems, Bonadonna and Houghton (2005) proposed a power-law relationship to better describe the thinning of well-preserved deposits.

Continuous thickness decay functions should be integrated over infinite area. Since deposit thickness measurements are usually sparse and rarely available in the immediate vicinity of the vent, or at extreme distal locations, extrapolation is required to infer missing data. The power-law function proposed by Bonadonna and Houghton (2005), is not integrable over $(0, \infty)$. Arbitrary integration limits must be specified to obtain total volume estimates. As an alternative, Bonadonna and Costa (2012) proposed a three-parameter Weibull function to model the relationship between thickness and the square root of isopach area, which doesn't require manual selection of integration limits.

These empirical approaches all ignore the idea of sampling errors and variations in thickness measurements. The construction of isopach maps is a subjective process involving the interpretation of field measurements. Contours constructed can be influenced by geologists experience in working with tephra deposits, and the locations and spatial density of measurements. Different individuals may take different approaches when contouring tephra measurements. This can lead to different volume estimates when fitted with different continuous functions (Klawonn et al., 2014b). Despite this, such empirical methods of estimating eruptive volumes remain widely used. A more detailed review of these approaches is given in Section 2.4.

Rhoades et al. (2002) suggested a statistical approach (informed by physics) for modeling tephra thicknesses to allow for uncertainty in field measurements. They proposed a linear relationship between the logarithm of tephra thickness and eruptive volume, modulated by a non-circular wind term. Volume is explicitly incorporated as a variable. Thickness is modeled as a function of distance and angular direction from the center of the deposit. The model of Rhoades et al. (2002) can be easily inverted to estimate eruptive volumes, as illustrated by Bebbington et al. (2008), if the thicknesses at a single location are treated as expected thicknesses (i.e., as without sampling error).

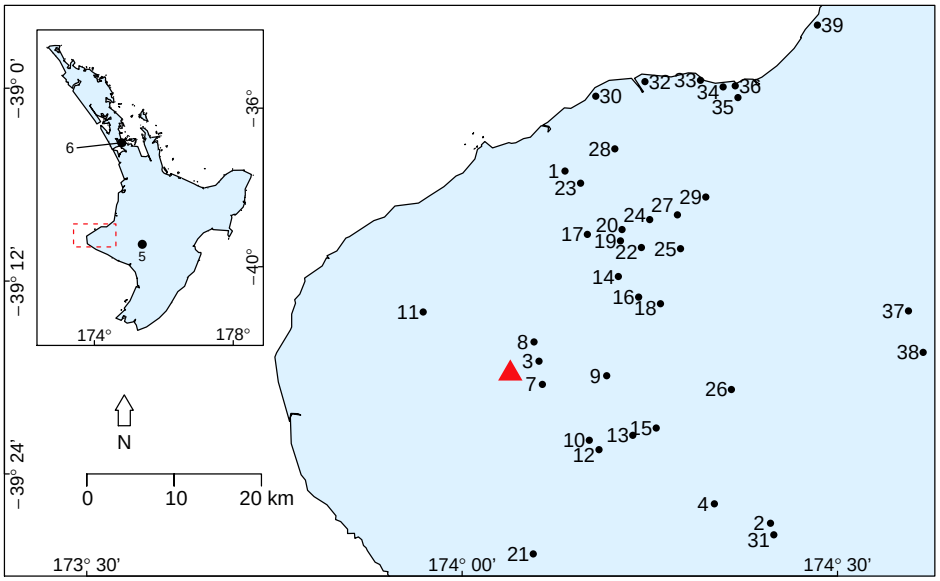
Rhoades et al. (2002) model the *average* dispersal from multiple events, as opposed to modeling individual eruptions. Their analysis utilizes thickness measurements for eruptions from Taupo Volcano (New Zealand) collected from a series of exposed locations (outcrops). Since post-depositional effects and weathering can erode tephra deposits, it is common for only large events to be recognized in exposed locations. In this chapter, I propose an adaptation of the model of Rhoades et al. (2002): eruptive volumes from multiple events at Taranaki (New Zealand) are estimated, based on sparse and incompletely observed thickness measurements from a number of both exposed and unexposed locations.

Unexposed locations, such as lakes and swamps, provide ideal sites for observing very fine tephras in addition to the large tephras observed in exposed locations. Records from unexposed locations can commonly exhibit over- or under-thickening of tephra, depending on the size and shape of the catchment area and surrounding topography. To avoid over- or under-inflating the size of previous events, site-specific thickness variations should be considered when estimating eruptive volumes. This chapter pays particular attention to estimating relative site-specific effects for each location.

Other studies of eruption size focus largely on single, well documented eruptions, such as Mount St Helens 18 May 1980 (Pyle, 1989; Fierstein and Nathenson, 1992; Bonadonna and Houghton, 2005; Bonadonna and Costa, 2012), whereas here the focus is also on estimating the volume of smaller events ($\ll 1$ km^3). 'Small' events may not be catastrophic in terms of loss of human life. However, they can still pose a serious risk to essential communication, water, and power supplies, industrial and agricultural processes, and infrastructure, which ultimately leads to substantial economic costs (Bebbington et al., 2008).

Dispersal from large Mt Taranaki eruptions has been well documented (Alloway et al., 1995), but available data for smaller Mt Taranaki events is sparse. The latter consists of a large number of thickness measurements from a very small number of unexposed locations. There are too few thickness measurements, often only one per eruption, to construct isopach maps for the smaller Mt Taranaki events. Direct estimation from point thickness data is preferable (Kawabata et al., 2013). An adaptation of the attenuation model of Rhoades et al. (2002) will be proposed in a Bayesian framework. Larger well estimated events are used as leverage for understanding the smaller unknown events, and the uncertainty in thickness measurements can be properly accounted for.

Figure 5.1: Location map of Taranaki. Locations mentioned in the text are indicated by numbers. The triangle denotes the position of the vent.



The rest of this chapter is organized as follows. First the eruption records from Mt Taranaki are introduced. This is followed in Section 5.3 by an outline of the model formulated to estimate eruptive volumes. Section 5.4 explains the model implementation in a Bayesian framework. The results and sensitivity analyses are presented in Section 5.5 and Section 5.6, followed by discussion and conclusions.

5.2 Mt Taranaki event records

The event records used in this analysis are from Mt Taranaki, New Zealand. Records contain observations of tephra thicknesses, measured from sediment cores extracted from unexposed locations that are both proximal and distal to the vent. Samples are also included from exposed locations around the volcano.

Thirty-nine records are utilized, which were obtained from different locations (Figure 5.1). The first six are from sediment cores extracted from within lakes and swamps. These will be referred to as *unexposed* locations. The remaining 33 locations, collectively referred to as *exposed* locations, provide individual point observations of tephra thicknesses sampled from within deep drains and road-side outcrops (Alloway, 1989). Prevailing high level westerly winds deposit tephra east of the vent. Therefore the deposition sites where tephras were sampled are predominantly east of the vent.

Figure 5.2: Comparison of exposed and unexposed locations.

(a) Exposed locations	**(b)** Unexposed locations
	
Tephra section from the Kaupokonui Valley on the flanks of Mt Taranaki, at approximately 1400 m elevation. Photo courtesy of Shane Cronin	*Sample core extracted from Lake Umutekai. Photo courtesy of Shane Cronin*
- 33 exposed locations - Vulnerable to erosion - Preserve tephra from 22 named events, 14 of which have known isopach diagrams	- 6 unexposed locations - Exhibit site-specific effects - Preserve tephra from 22 named events in addition to 248 unnamed events

The different sites have varying suitability for tephra preservation. The exposed locations are vulnerable to erosion and interference from reforestation and construction work. They only preserve thick tephra falls, henceforth referred to as *named* tephras. The unexposed locations, on the other hand, provide ideal sites for observing very *fine* ash in addition to the thick (named) tephra falls.

Depending on the surrounding topography and variability in size, shape, and depth of the catchment areas, different unexposed locations can exhibit large variations in tephra thicknesses. Over- or under-thickening of tephra can have consequences for volume estimation. Volume estimates are obtained on the basis of observed tephra thicknesses, and can be subsequently fed into probabilistic models for forecasting volcanic hazard. It is these site-specific effects that we are interested in understanding and quantifying, so that reliable volume estimates can obtained and effective disaster mitigation procedures can be developed. Figure 5.2 shows an exposed location and a sample portion of a record from an unexposed location along with a summary of their typical tephra preservation characteristics.

5.2.1 Tephra records from unexposed locations

The locations and details of the six records obtained from unexposed locations, used in this analysis, are given in Figure 5.1 and Table 5.1. The first four locations are near enough to the

Table 5.1: Details of unexposed locations ($j = 1 \dots 6$). The variable r_j denotes the distance from the vent (km), θ_j denotes the angular direction from the vent (radians anticlockwise from east), n denotes the number of tephras recognized in the location.

ID j	Location	Latitude	Longitude	r_j	θ_j	n	Age range (cal. yr BP)	Source
1	Lake Umutekai	39°05.27'S	174°08.22'E	24.15	1.34	104	1400 - 11500	(Turner et al., 2008a)
2	Lake Rotokare	39°27.10'S	174°24.64'E	34.49	-0.49	42	490 - 7050	(Turner et al., 2009)
3	Near Source	39°16.59'S	174°06.09'E	3.59	0.44	23	94 - 2200	(Turner et al., 2009)
4	Eltham Swamp	39°25.52'S	174°20.09'E	27.83	-0.54	177	3000 - 36900	(Tinkler, 2013)
5	Lake Rangatauanui	39°26.04'S	175°22.45'E	114.39	-0.11	37	2100 - 24500	(Moebis, 2010)
6	Auckland			276.29	1.37	40	9800 - 75000	(Green et al., 2014)
	- *Lake Pupuke*	36°47.25'S	175°46.25'E	286.00	1.38	15		(Molloy et al., 2009)
	- *Onepoto Basin*	36°48.29'S	174°45.02'E	282.72	1.38	21		(Shane and Hoverd, 2002)
	- *Orakei Basin*	36°52.00'S	174°48.40'E	277.62	1.36	18		(Molloy et al., 2009)
	- *Hopua Crater*	36°55.46'S	174°47.05'E	270.24	1.36	11		(Molloy et al., 2009)
	- *Pukaki Crater*	36°58.58'S	174°48.37'E	265.01	1.35	28		(Sandiford et al., 2001; Shane, 2005)

volcano (< 35 km) to potentially record both minor unnamed eruptions and those producing larger tephra falls. The remaining two unexposed sites are located much further from the volcano (> 100 km). Although the distal sites include a wide age range of tephras, due to their proximity, they are less likely to contain a given tephra. The thinnest tephra observable in a core is 0.5 mm.

The Auckland record (the sixth record in the analysis) arises from the combination of multiple tephra records collected from within *maar* depressions in Auckland, New Zealand. A maar is a crater formed by a volcanic eruption that is (typically) subsequently filled with water and sediment to become a lake or a swamp. This analysis uses the detailed tephra records established in Molloy et al. (2009) from Lake Pupuke, Hopua Basin, and Orakei Basin. In Chapter 4 these records were combined with tephra records from Onepoto Basin (Shane and Hoverd, 2002) and Pukaki crater (Sandiford et al., 2001; Shane, 2005) to identify 40 distinct Mt Taranaki events. The Auckland sites are located more than 260 km from Mt Taranaki and only contain very fine tephra (~ 2 mm thick). To avoid letting the distal data dominate the analysis, the records from the five Auckland sites are aggregated. The average tephra thickness observed, the average distance from the vent, and the average azimuth of the five Auckland sites in Chapter 4, provide the sixth record for this analysis.

The approximate age ranges in Table 5.1 indicate an age overlap between many of the cores. A particular recorded eruption may have left a deposit at all of the sites, or any subset. The local search algorithm presented in Chapter 4 is utilized to find the most likely arrangement of tephras observed across the six unexposed locations. In addition to thickness measurements, the records contain geochemical data (where available), and estimated ages and associated

age errors for each tephra. The age and associated age errors form the basis of the matching algorithm, and arrangements are constrained by considering the geochemical compositions and stratigraphic relationship among tephras.

In Chapter 4 the matching algorithm was illustrated through application to the five Auckland records discussed above. The only geochemical constraint applied there was that tephras must be sourced from the same volcano. In the case of the records used here however, all of the tephras are sourced from Mt Taranaki so that constraint is no longer applicable. There is, however, more detailed geochemistry data available in the form of titanomagnetite chemistry.

Employing the approach of Turner et al. (2009), matches among tephra are permitted or ruled-out through principal component analysis on the major elements (TiO_2, Al_2O_3, MgO) of the titanomagnetite compositions. The following principal components (PCs) are computed for each sample from each tephra:

$$\text{PC1} = 0.18\ \text{TiO}_2 - 0.71\ \text{Al}_2\text{O}_3 - 0.68\ \text{MgO}, \tag{5.1}$$

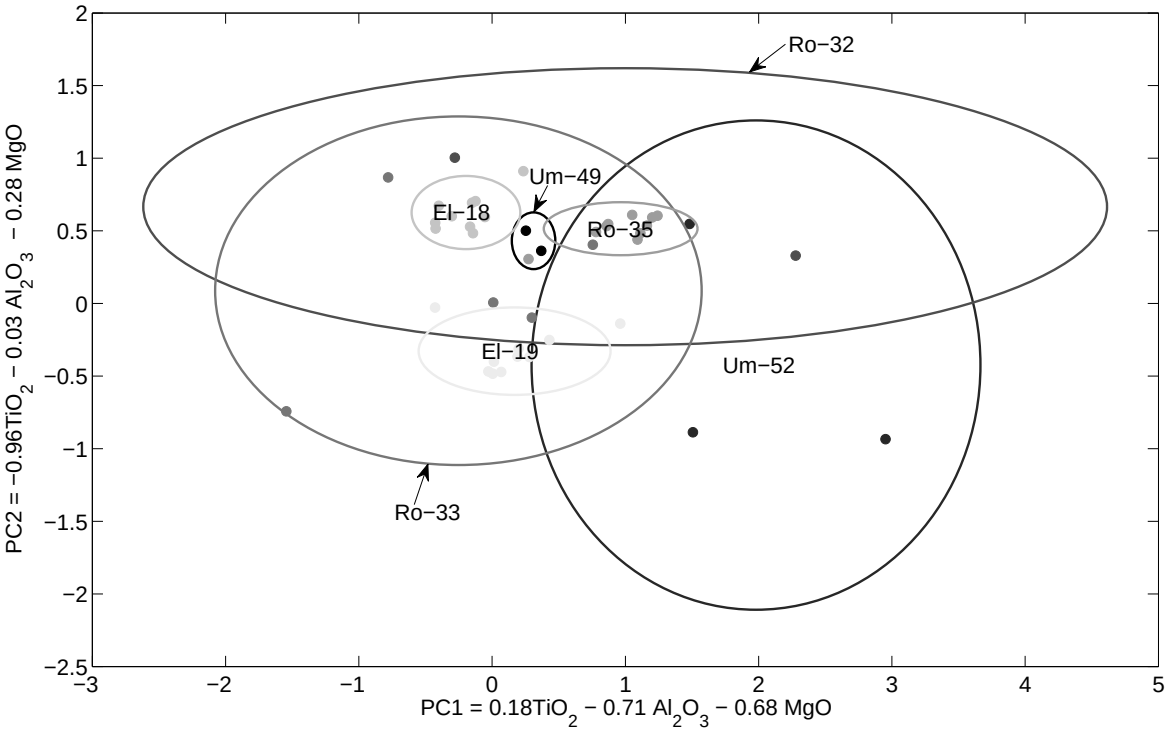
$$\text{PC2} = -0.96\ \text{TiO}_2 + 0.03\ \text{Al}_2\text{O}_3 - 0.28\ \text{MgO}. \tag{5.2}$$

These PCs differ from those in Turner et al. (2009) because of the inclusion of the Eltham Swamp and Lake Rangatauanui tephras. Not all of the tephras have geochemistry data available. In particular *none* of the tephras sourced from the Auckland region have titanomagnetite geochemistry available. While the Auckland tephras do have geochemistry data available, it is in the form of glass chemistry so is not comparable to the other records.

An eruption consists of many geochemistries; different geochemistries occur at different phases of eruption, so can be found in different combinations at different sites. Ellipses, centered at the mean principal component values, with axis length equal to four times the standard deviation, are constructed for each tephra. Only candidate age matches among tephras with overlapping ellipses are permitted.

Figure 5.3 shows an example for a section of the record aged 5500 - 6000 cal yr BP. Candidate matches among three or more tephras are allowed, provided there is at least one tephra with geochemistry consistent with all other tephras. For example, a match between Rotokare-35 and Eltham-19 would be permitted only with the inclusion of Umutekai-52. This is keeping with the nature of a magma as an aggregate of individual magmas. As deposits in different cores tend to reflect different wind conditions, they are often from different phases of an eruption. The requirement of at least one overlapping tephra ensures that there is a temporal

Figure 5.3: Geochemistry of tephras aged 5500 - 6000 cal yr BP. Labels indicate the source (Lake Umutekai (Um), Lake Rotokare (Ro), and Eltham swamp (El)) and the number for each of the tephras.



magma evolution path through the eruption. It also gives the matching algorithm greater flexibility and ensures that the search for the optimal arrangement is not over-constrained. In any case, the estimated ages for the tephras provide the true control over the arrangement. The geochemistry constraints serve to rule out infeasible matches to which we take a conservative approach.

Checking the geochemical correlations among tephra is done automatically through adaptation of the automated matching procedure in Chapter 4. There are a total of 270 distinct events identified by the matching algorithm. The final arrangement is given in Appendix D.

5.2.2 Tephra records from exposed locations

Alloway (1989) gives point observations of tephra thicknesses from 35 exposed locations within deep drains, farm cuttings, and prominent road sections. Although Alloway (1989) details 35 exposed sections, only 33 are used in this analysis. Section 35 is excluded because it represents

tephra much older than that found in the unexposed locations. Section 17 is excluded because it does not include Mt Taranaki sourced tephra.

The preservation environment of the exposed locations means fine ash layers cannot be macroscopically identified. Therefore, the exposed locations only contain records from 22 widespread easily distinguishable tephra formations. These will henceforth be referred to as being from *named events*. The names and approximate ages of these tephra formations, as identified by Alloway (1989) and Alloway et al. (1995), are as follows: Manganui (c. 3.1 ka), Inglewood (c. 3.6 ka), Korito (c. 4.1 ka), Mangatoki (c. 4.4 ka), Tariki (c. 4.6-4.7 ka), Waipuku (c. 5.2 ka), Kaponga (c. 8.0-10.0 ka), Konini (c. 10.1 ka), Mahoe (c. 11.0-11.4 ka), Kaihouri (c. 12.9-18.8 ka), Paetahi (c. 19.4-20.2 ka), Poto (c. 20.9-22.7 ka), Tuikonga (c. 23.4-24.0 ka), Koru (c. 24.8-25.2 ka), Pukeiti (c. 26.2 ka), Waitepuku (c. 27.5-28.0 ka), Mangapotoa (c. 28.0 - 50.0 ka), Waitui (c. 55.0 ka), Araheke (c. 55.0-75.0 ka), and Te Arei (c. 75.0 ka).

These named tephras correspond to large, widespread eruptions. Therefore, they are also recognized in the more complete records obtained from the unexposed locations. Approximate age ranges for the named tephras help to identify sections in the unexposed records where the named tephras may be found. The named tephras are reported in radiocarbon C^{14} ages in Table 1 of Alloway et al. (1995). For comparison with the unexposed records, radiocarbon C^{14} ages reported in Alloway et al. (1995) are converted into calendar ages using using CalPal Online (`http://www.calpal-online.de/`). The named tephras are identified in the unexposed records based on the approximate age ranges and large observed thicknesses in the appropriate direction(s).

Many of the named tephra formations have multiple beds corresponding to sub-phases from the same eruption. Here a conservative approach is taken again, using the thicknesses from the dominant phases(s) of each event. Phases may have occurred hundreds of years apart. For example the Kaihouri tephra formation has at least eight tephra beds aged 12.9 - 18.8 ka (Alloway, 1989). While tephra from each phase *may* be present in the records from unexposed locations, only the dominant phase(s) is assumed to have deposited tephra that is clearly distinguishable (i.e has a much larger observed thickness).

The thicknesses in the exposed records are obtained from descriptions of the reference sections in Alloway (1989) or, where not specifically stated, have been measured from scale figures showing the section stratigraphy. In situations where the figures in Alloway (1989) do not

separate the different beds, the total thickness is apportioned so that the thickness of the dominant phase(s) is proportional to that observed in the reference sections. Table 5.2 gives the final record of tephra thicknesses at each of the exposed locations. Larger observed tephras in the unexposed locations are matched to dominant phase tephras (of similar age) in the exposed locations. The alignment of named tephras across the unexposed locations is given in Appendix D.

5.3 Model development

Based on a data set of eruptions from Taupo Volcano (New Zealand), Rhoades et al. (2002) proposed the following relationship between the logarithm of tephra thickness (T in cm) and eruptive volume (V in km^3):

$$\log T = a + b \log V - c \log(r + dV^{1/3}) + f(\theta) + \epsilon, \tag{5.3}$$

where $\epsilon \sim N(0, \sigma^2)$, and r and θ denote the distance (in km) and angular direction (in radians) from the center of the deposit (see Figure 5.4). Tephra thickness and the column height of the eruptive plume must be finite and positive at the vent, which requires $b = \frac{c+1}{3}$ and $c > 2$ in (5.3) (cf. Rhoades et al., 2002).

The directional function $f(\theta)$ is of the form:

$$f(\theta) = \sum_{k=1}^{n} \alpha_k \sin(k\theta) + \beta_k \cos(k\theta), \tag{5.4}$$

where only α_k and β_k values that are significantly different from zero are allowed to take non-zero values. The function $f(\theta)$ is a semi-elliptical wind term included to modify the otherwise circular distribution of tephra deposits, by stretching and orienting it. Effectively the α and β terms become proxies for the wind strength, direction, and variability.

Rhoades et al. (2002) fitted their model via linear regression techniques, assuming log-normally distributed thicknesses. The estimated parameter values obtained by Rhoades et al. (2002) indicate an average relationship between tephra thickness and erupted volume. More importantly, the assumed identical parameters, including wind effects $f(\theta)$, for all events, and hence the residual error in (5.3) is much larger than would be expected for a single event. Due to insufficient data, and their linear regression approach, Rhoades et al. (2002) were not

Table 5.2: Thicknesses observed at exposed locations ($j = 7 \ldots 39$). r_j denotes the distance from the vent (km), θ_j denotes the angular direction from the vent (radians anticlockwise from east).

	Name																						
	ID i																						
ID j	r_j (km)	θ_j (radians)	Thickness (cm)																				
7	3.89	-0.3	11	15	9.3																		
8	4.54	0.96	6.9	12	8																		
9	11.08	0.01	7																				
10	11.89	-0.67	20	10	8.7	1	18	15						13									
11	12.27	-0.58							3.5	11.3	1.6	7	17	6	7	8.8							
12	13.45	-0.68	5.3	7	6.7	3	15	14	5	15	6					6.7	13		10				
13	15.77	-0.44						7	7	23	6	17	8	27			21	7	7				
14	16.71	0.76							3	2.7	2						3	4					
15	17.9	-0.33				1.1	8.4	8		12	5	18							9				
16	17.18	0.57	24	5	10	2	3	10	14	13	5.6								7				
17	18.31	1.1																		5	7	14	9
18	19.05	0.47	10	5	4.7	3		2	13	15	1.2												
19	19.85	0.91			6	2.1	3.2	9	8	4	0.4	14	15	19	9	13			1.2				
20	20.98	0.94											17	7	6	6							
21	20.95	-1.41				0.3	2.4		4.1	9.3	2.4	8					0.9						
22	20.93	0.8											24	6	4	11							
23	23.36	1.25											8	7	6	4			4		9	5	6	15
24	23.9	0.87										11	5	8	5				5	3	5	16		
25	24.29	0.66	10.7	8	5.3	3	8	4															
26	25.48	-0.04							4	7.3	2	11											
27	26.53	0.79			10.7	0.5	4	10	9			7	17	6	6	3							
28	28.53	1.17										13	5			5						12	
29	30.32	0.77																			6	6	18
30	34.48	1.31											2							3	2	1	8
31	35.5	-0.52		6	1.3																		
32	38.03	1.18											1.3								2		6
33	40.72	1.03											4.7										4
34	41.56	0.97	5.3	3.5	2	2	3	3														10	3
35	42.06	0.93										17	11.3										
36	42.49	0.95		2.5	2	0.2	1.2	4															
37	46.35	0.19										29		13									
38	47.52	0.08												4									
39	54.34	0.89			0.7	0.2	1.6																

able to consider changes in tephra isopachs, such as the tendency for larger events to exhibit multiple lobes. In addition, Rhoades et al. (2002) only modeled data from exposed locations. Extending the model in Rhoades et al. (2002), for application to both exposed and unexposed records, requires some consideration of site-specific effects for over- or under-thickening.

The following adaptation of (5.3) allows parameter values to differ according to the location at which thicknesses were observed and what event the thicknesses correspond to:

$$\log T_{i,j} = a_j + b_i \log V_i - c_i \log(r_j + d_i V_i^{1/3}) + f(\theta_j - \phi_i). \tag{5.5}$$

Here $T_{i,j}$ denotes the tephra thickness (in cm) of the ith event observed at the jth location. Locations $j = 1, \ldots, 6$ represent the unexposed locations: Lake Umutekai, Lake Rotokare, Near-Source, Eltham Swamp, Lake Rangatauanui, and Auckland respectively. Locations $j = 7, \ldots, 39$ are the exposed locations from Alloway (1989). The distance (in km) from the vent and angular direction (in radians, anticlockwise from east) of location j are denoted by r_j and θ_j respectively. Following the reasoning of Rhoades et al. (2002), the replacement $b_i = \frac{c_i+1}{3}$ is made so that only c_i is estimated.

Rhoades et al. (2002) assume a common wind direction for each event. They specify θ in (5.3) as the direction from the center of the deposit. Here θ is replaced with $\theta_j - \phi_i$, where ϕ_i denotes the wind direction of event i (to be estimated) and θ_j is the angular direction of location j (see Figure 5.4). The site-specific effects a_j, to be estimated, are assumed to act multiplicatively on observed thicknesses. They are included to explain over- or under-thickening characteristics of each location. The variable V_i denotes the volume (in km^3) of the ith event (to be estimated). The additional event dependent parameters to be estimated are c_i, and d_i, along with those involved in the directional function f. The parameter c_i represents the rate of thickness decay with distance from the vent, and d_i is included to ensure that thicknesses are finite at the vent (i.e. when $r_j = 0$).

Finding the correct functional form of the semi-elliptical wind dependent term $f(\theta_j - \phi_i)$ is complicated by the fact that the distribution of tephra from Mt Taranaki events is commonly multi-lobed. Predominant winds blowing from the south-west commonly result in a large lobe north-east of the vent. North-westerly winds can contribute to a smaller less prominent lobe south-east of the vent. Figure 5.5 shows an isopach map for the Inglewood tephra illustrating the double-lobed tephra dispersal. The Inglewood isopach map was inferred by Alloway et al. (1995), from point thickness measurements at the unexposed locations.

The parameterization given in (5.4) does not capture the double lobed nature of the tephra dispersal, so a mixture model is formulated to combine tephra thicknesses deposited north-east and south-east of the vent. The usefulness of mixtures of elliptical lobes for modeling the dispersal of tephra has been successfully illustrated by Kawabata et al. (2013). Equation (5.5) is equivalent to

$$T_{i,j} = \frac{\exp(a_j)V_i^{(c_i+1)/3}}{\left(r_j + d_i V_i^{1/3}\right)^{c_i}} \exp(f(\theta_j - \phi_i)), \tag{5.6}$$

which can be formulated as a mixture

$$T_{i,j} = \frac{\exp(a_j)V_i^{(c_i+1)/3}}{(r_j + d_i V_i^{1/3})^{c_i}} \left[p \exp(f_1(\theta_j - \phi_{1i})) + (1-p)\exp(f_2(\theta_j - \phi_{2i}))\right], \tag{5.7}$$

where $f_1(\theta_j - \phi_{1i})$ and $f_2(\theta_j - \phi_{2i})$ are given explicit functions as follows

$$f_1(\theta_j - \phi_{1i}) = \alpha_{1i}\cos(\theta_j - \phi_{1i}) + \beta_{1i}\cos(2\theta_j - 2\phi_{1i}), \tag{5.8}$$

$$f_2(\theta_j - \phi_{2i}) = \alpha_{2i}\cos(\theta_j - \phi_{2i}) + \beta_{2i}\cos(2\theta_j - 2\phi_{2i}). \tag{5.9}$$

Rhoades et al. (2002) use an east dispersal axis based on present day wind records, with sine terms included in the directional function (5.4) to provide an off-axis dispersal. This follows from their treatment of an aggregation of multiple eruptions. Rhoades et al. (2002) model the sum of probabilistic wind directions for multiple events, rather than the wind direction during a single eruption. Here we want to model eruptions individually. Therefore, a wind direction ϕ_i is explicitly included, and off-axis terms are not required.

Equation (5.7) introduces a mixing parameter p, representing the proportion of total thickness

Figure 5.4: Diagrammatic representation of the angular directions. Rhoades et al. (2002) specify θ as the angular direction from the center of the deposit. The parameterization proposed makes the replacement $\theta = \theta_j - \phi_i$, see text for details.

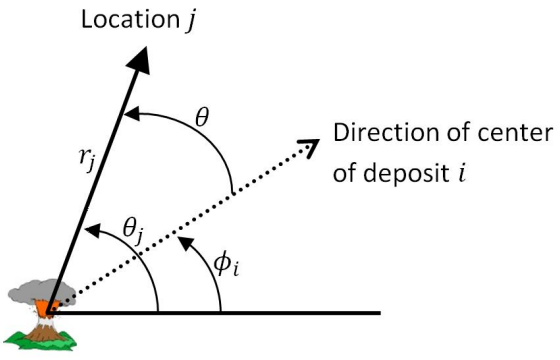
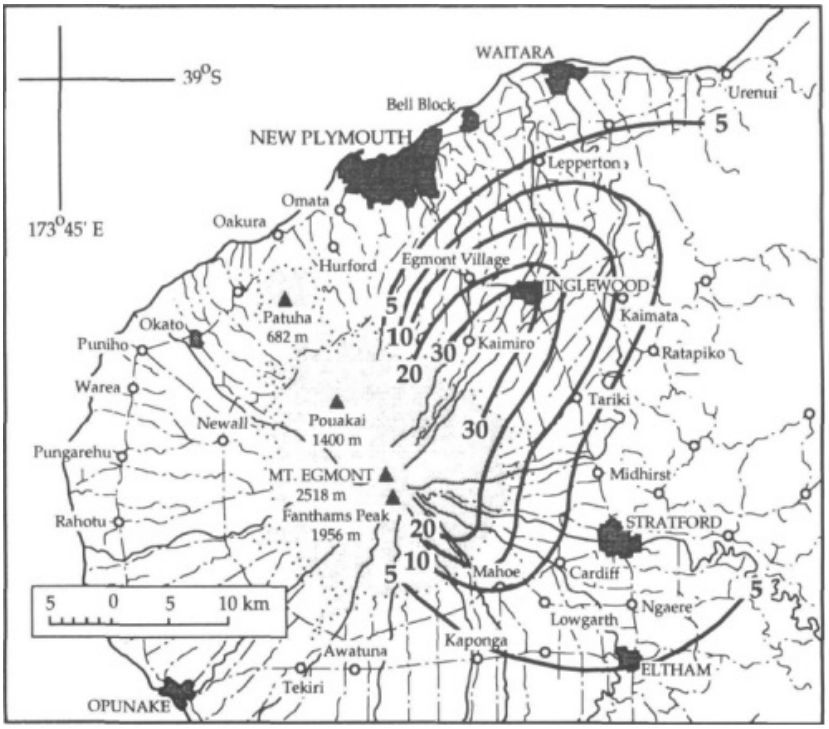


Figure 5.5: Isopach map for the Inglewood Tephra. Isopachs in centimeters. Reprinted from Alloway et al. (1995), with kind permission from Taylor and Francis.



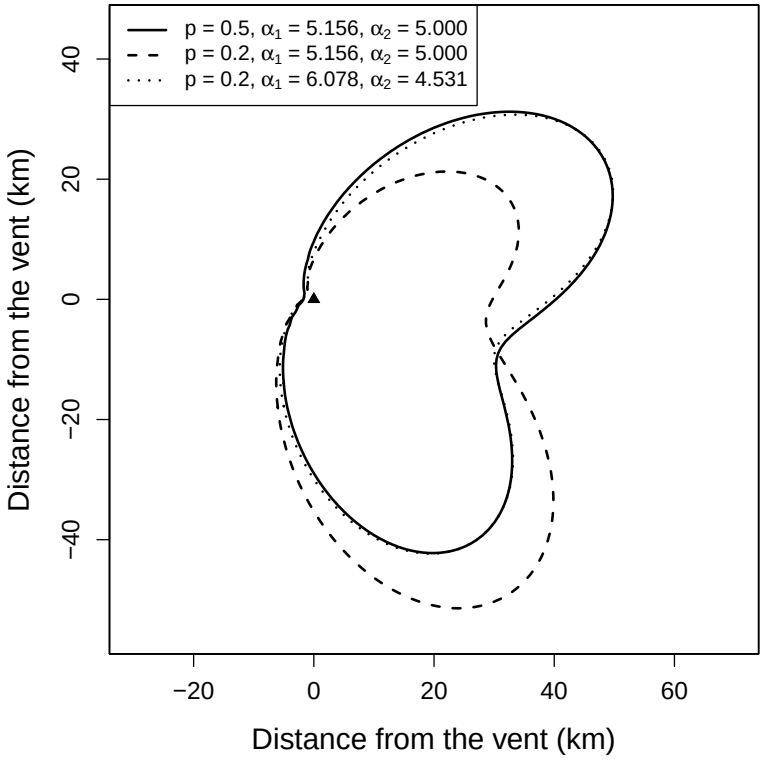
deposited in each lobe. In theory when $p > 0.5$ the majority of tephra should be deposited in the direction ϕ_{1i}. This is true if the parameter values involved in (5.8) and (5.9) are held constant. However, allowing these values to change absorbs the effect of changing p (for $p \in [0, 1]$). Figure 5.6 gives one such example; after adjusting α_1 and α_2 the contour lines for $p = 0.5$ and $p = 0.2$ are virtually indistinguishable. This identifiability problem is resolved by setting $p = 0.5$, for those tephra that are clearly multi-lobed. This reduces the number of parameters to estimate and allows the shape to be determined by changes in $\alpha_{1i}, \alpha_{2i}, \beta_{1i}$, and β_{2i}. In this parameterization, wind direction is controlled by changes in ϕ while α and β are proxies for the wind strength and variability.

Restrictions are imposed on the values of the parameters. To ensure thickness and column height are finite and positive at the vent, it is necessary that $d_i > 0$ and $c_i > 2$ (Rhoades et al., 2002). In addition to these constraints, are the physical constraints on the parameters α_{1i}, β_{1i}, α_{2i}, and β_{2i} which describe the shape of tephra dispersal.

To find the angular direction of maximum and minimum thickness, the semi-elliptical functions (5.8) and (5.9) can be differentiated for the case when $\phi = 0$ (westerly winds). Ignoring the subscripts,

$$f'(\theta) = -\alpha \sin(\theta) - 2\beta \sin(2\theta).$$

Figure 5.6: Example contours for a 2 cm thick tephra deposit. $V = 1, a = 5, c = 2.38, d = 0, \beta_1 = 0.703, \beta_2 = 0.613, \phi_1 = 0.442$, and $\phi_2 = -0.983$. The triangle denotes the position of the vent.



At a stationary point $f'(\theta) = 0$. Therefore $\theta = 0$ or π, and

$$f''(\theta) = -\alpha\cos(\theta) - 4\beta\cos(2\theta).$$

Therefore, in order to have a maximum in the downwind direction at $\theta = 0$ ($f''(0) < 0$) and a minimum in the upwind direction at $\theta = \pi$ ($f''(\pi) > 0$), we must restrict $\alpha > 4\beta$. This requires $\alpha_{1i} > 4\beta_{1i}$ and $\alpha_{2i} > 4\beta_{2i}$ for all i, where α_{1i}, β_{1i}, α_{2i}, and β_{2i} are ≥ 0 . If this constraint is violated, the dispersal shape would no longer be elliptical. Tephra would be forced west ($\theta = \pi$) of the vent (in addition to east $\theta = 0$) with the minimum thicknesses north ($\theta = \pi/2$) and south ($\theta = -\pi/2$) of the vent.

5.4 Bayesian implementation

The model in (5.5) is a highly complex statistical model for tephra attenuation, it is the simplest model consistent with the physical reality. The reparameterization introduces a much larger number of parameter values than proposed by Rhoades et al. (2002) (5.3). This

is because parameter values differ according to the location at which thicknesses were observed and what event the thicknesses correspond to.

Point thickness data available for Mt Taranaki events is extremely unbalanced. Not all events are observed at each location. Nonetheless, there is a lot of structure to the data. Despite being deposited under common wind conditions, tephra from a given event can be observed differently at each location. This is due to site-specific thickening effects and preservation environments. Eruption volumes, wind, and site-specific effects all act jointly on the observed thicknesses. Therefore, model parameters cannot be estimated independently. The natural way to manage high-dimensional complex models, such as the one proposed here, is in a Bayesian framework.

A lognormal distribution for tephra thickness is assumed, which follows from the model proposed by Rhoades et al. (2002) in (5.3)

$$T_{i,j} \sim LN(\mu_{i,j}, \sigma^2), \tag{5.10}$$

where $\mu_{i,j}$ is given by the logarithm of the right hand side of Equation (5.7)

$$\mu_{i,j} = \log\left(\frac{\exp(a_j)V_i^{(c_i+1)/3}}{\left(r_j + d_i V_i^{1/3}\right)^{c_i}}\left[p\exp(f_1(\theta_j - \phi_{1i})) + (1-p)\exp(f_2(\theta_j - \phi_{2i}))\right]\right), \tag{5.11}$$

where σ is the residual standard error (1.78) from Rhoades et al. (2002). A sensitivity analysis on the choice of σ is given in Section 5.6.1; instead of fixing $\sigma = 1.78$, which is likely an overestimate for a single event (see discussion on page 120), a prior distribution is imposed on the precision $\tau = 1/\sigma^2$.

The detection limit for observing tephra at the unexposed locations is 0.05 cm, thus thicknesses are left censored. If tephra from a particular eruption is observed in at least one unexposed location, then unexposed locations that did not observe the eruption must have less than 0.05 cm of tephra.

Implementing the model within a Bayesian framework requires appropriate prior distributions to be assigned to the model parameters V_i, a_j, c_i, d_i, ϕ_{1i}, ϕ_{2i}, α_{1i}, α_{2i}, β_{1i}, and β_{2i} which will be considered next.

5.4.1 Construction of priors

Of the 270 distinct events recognized in the combined record, 22 correspond to widespread and easily distinguishable named events, 14 of which have known isopach diagrams. The isopach maps of major post 28,000 yr BP eruptions from Mt Taranaki were constructed by Alloway et al. (1995), from the thicknesses observed in the exposed locations mentioned above.

The construction of isopach maps is a subjective process. It involves interpreting raw thickness measurements and creating hand-drawn contour lines for expected thickness boundaries. Among other things, one must decide where to place the contour lines, how many isopachs to include, what the largest and smallest isopach levels should be, and how smooth the contours should be drawn. Of course the outcome is dependent on who created the isopach map and their understanding of tephra dispersal and field data. Isopachs are useful as an indication of the dispersal of tephra. However, deriving estimates of previous eruption volumes from the isopach maps is risky, especially without proper consideration of the errors in thickness measurements and the subjectivity involved in constructing the isopachs (Klawonn et al., 2014b; Burden et al., 2013).

The isopach maps in Alloway et al. (1995) provide 'expert opinion' of the dispersal pattern for 14 of the named events. They can be used to elicit prior distributions for the model parameters. The tephra dispersal and volume of the large named events have already been well described (cf. Neall, 1972; Franks, 1984; Latter, 1990; Alloway et al., 1995; Turner et al., 2011a). Therefore the objective of this study is not to estimate anything concerning the large events, it is rather to estimate the size of the smaller eruptions. Large named events deposit tephra under the same wind conditions and site specific effects as smaller events. Knowledge of the dispersal of larger events can, therefore, be used as leverage for understanding the tephra dispersal from smaller events. A semi-empirical Bayes type approach is taken: first the parameter values in (5.11) are estimated for the named events, via maximum likelihood techniques. Then prior distributions are specified for all event parameters based on the resulting estimates.

5.4.2 Estimating parameter values for named tephras

Ten to thirty observations of $(T_{i,j},\, r_j,\, \theta_j)$ are taken from each isopach line in Alloway et al. (1995), associated with each event. Parameter values for the well known named events ($i =$

$1 \ldots, 14$) are estimated by numerical maximization of the log-likelihood function

$$\log L = \sum_{i=1}^{14} \left[-\sum_{j=1}^{n} \log T_{i,j} - \frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{j=1}^{n} (\log T_{i,j} - \mu_{i,j})^2 \right]. \tag{5.12}$$

This follows from (5.10) where thickness ($T_{i,j}$) is assumed to be lognormally distributed with $\mu_{i,j}$ given by (5.11), and σ is taken as the residual standard error (1.78) from Rhoades et al. (2002). As these events are all large and hence very irregularly shaped, this residual standard error is considered appropriate.

The purpose of estimating the parameter values for the named tephras is to elicit prior distributions for the model parameters. Mean and standard deviation volume estimates have been independently derived, by Turner et al. (2011a), for the large named tephras. In order to obtain a suitable prior, a penalty term, for deviations from these estimated volumes, is added to the log-likelihood function before maximization (Good and Gaskins, 1971).

The resulting parameter estimates and standard errors, the latter obtained by inverting the Hessian matrix obtained from the numerical optimization solution, are given in Table 5.3. Only standard errors that are explicitly used in the formulation of prior distributions are reported.

Only one site-specific effect a_j is estimated during the maximum likelihood approach, as exposures are not coupled with catchments of unknown size, or affected by unknown water inflow and outflow. The resulting value is equal to 5.812, with a standard error of 0.057 (obtained by inverting the Hessian matrix). The larger errors associated with the ϕ_{2i} estimates are due to the exposed locations being predominantly located in the north-east quadrant from the vent, rather than the south-east (Figure 5.1). The zero values for α_{2i} and β_{2i} for the Tuikonga, Pukeiti, and Waitepuku tephras are consistent with documented single-lobed isopach diagrams.

Figure 5.7 gives a comparison between the isopach diagrams from Alloway et al. (1995), and those constructed using the parameter estimates given in Table 5.3. The model is able to cope with the varying shapes in the tephra dispersal, and can still approximate the shape of the more vague isopachs in Alloway et al. (1995) (namely the Tuikonga, Koru.a, Pukeiti and Waitepuku tephras). It is important to note that the isopach maps from Alloway et al. (1995) are hand-drawn, and in all cases the isopachs are not closed. It is not possible, under

Table 5.3: Estimated parameter values and standard errors (in parentheses).

ID i	Name	$\hat{V}_i$	$\hat{c}_i$	$\hat{d}_i$	$\hat{\alpha_{1i}}$	$\hat{\alpha_{2i}}$	$\hat{\beta_{1i}}$	$\hat{\beta_{2i}}$	$\hat{\phi_{1i}}$	$\hat{\phi_{2i}}$
1	Manganui.a	1.328 (0.196)	2.189	1.452	2.927	2.892	0.376	0.481	0.470 (0.108)	-0.593 (0.218)
2	Inglewood.b	1.734 (0.193)	2.159	0.840	3.008	2.463	0.679	0.612	0.729 (0.093)	-0.551 (0.233)
3	Korito.b	1.444 (0.149)	2.260	1.381	3.123	3.259	0.538	0.808	0.550 (0.103)	-0.436 (0.191)
4	Tareki.f	0.268 (0.037)	2.213	0.913	2.808	1.885	0.610	0.117	0.174 (0.132)	-0.142 (0.336)
5	Tareki.e	1.038 (0.153)	2.183	0.527	3.335	2.726	0.811	0.228	0.290 (0.081)	-0.415 (0.257)
6	Waipuku	1.113 (0.123)	2.233	0.515	2.903	2.825	0.374	0.171	0.645 (0.160)	-0.636 (0.211)
7	Kaponga.f	0.848 (0.108)	2.176	0.613	4.440	1.266	0.000	0.025	0.437 (0.084)	-1.170 (0.677)
8	Konini.b	0.820 (0.111)	2.182	0.549	4.171	4.120	0.652	1.027	0.000 (0.102)	-1.143 (0.165)
9	Mahoe.a	1.347 (0.225)	2.230	1.003	2.536	2.607	0.634	0.652	0.000 (0.170)	-0.663 (0.189)
10	Paetahi.a	1.534 (0.301)	2.192	0.996	3.081	3.828	0.316	0.590	0.286 (0.152)	-0.697 (0.134)
11	Tuikonga.d	1.434 (0.258)	2.186	0.749	3.882	0.000	0.427	0.000	0.757 (0.138)	-1.463 (0.298)
12	Koru.a	1.074 (0.133)	2.195	0.389	3.633	4.433	0.227	0.072	0.955 (0.409)	-0.597 (0.321)
13	Pukeiti	0.507 (0.108)	2.134	0.405	3.163	0.000	0.469	0.000	1.033 (0.216)	-1.571 (0.298)
14	Waitepuku.a	0.532 (0.118)	2.180	0.594	3.430	0.000	0.420	0.000	0.632 (0.170)	-1.178 (0.298)

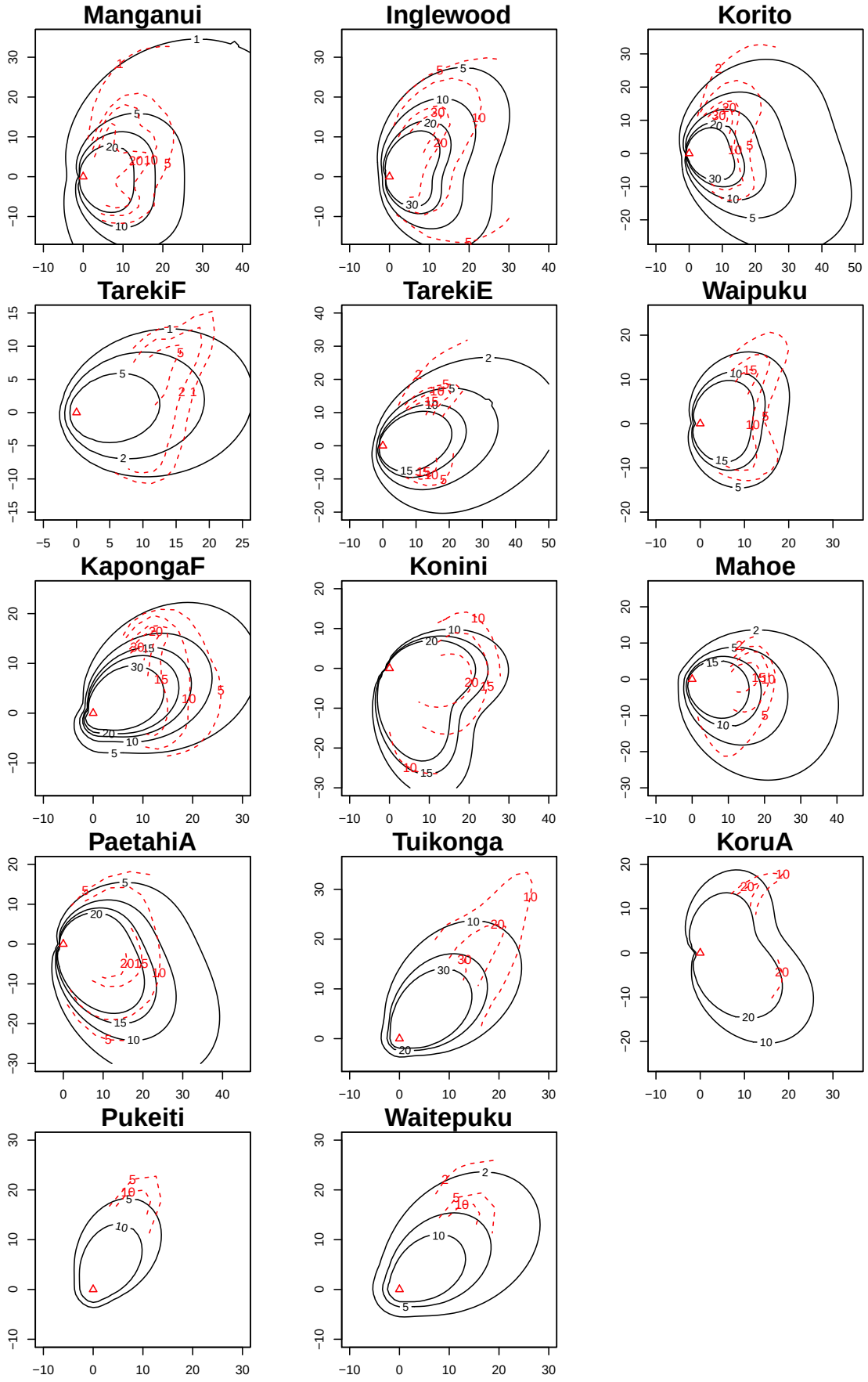
the physical constraints that $\alpha_{1i} > 4\beta_{1i}$ and $\alpha_{2i} > 4\beta_{2i}$ for all i, to obtain isopachs more consistent with those in Alloway et al. (1995) using maximum likelihood estimation.

The estimated values in Table 5.3 were obtained through maximum likelihood estimation using measurements taken from the hand-drawn contours in Alloway et al. (1995). They only provide empirical *a priori* estimates of the parameter values from the eruption records obtained from both exposed and unexposed locations. Maximum likelihood estimation is useful when considering only the 14 named events. However due to the large number of parameters, estimating the eruptive volumes of all 270 events recognized in both the unexposed and exposed locations is not possible using simple maximum likelihood techniques. To handle the complex nature of the data, the model will be implemented in a Bayesian framework. The estimated parameter values for the named events in Table 5.3 are used to formulate the prior distributions for the model parameters. This provides the facility to leverage the well recorded events into site-specific effects and volumes, while properly considering sampling error.

5.4.3 Specification of priors for the site-specific effects a_j

For the unexposed locations ($j = 1, \ldots, 6$), very little is known about the relative over- and under-thickening of tephra. Therefore a reference prior, $a_j \sim N(0, 10^4)$, is assumed for the site-specific effects.

Figure 5.7: Distribution of tephra thicknesses for named events. The triangle denotes the position of the vent, axis units are in kilometers from the vent (0,0), and isopachs are in centimeters. Solid lines represent isopachs constructed using parameter estimates in Table 5.3. Dotted lines represent isopachs from Alloway et al. (1995).



For the exposed locations ($j = 7, \ldots, 39$), a normal prior distribution $a_j \sim N(5.812, 0.057^2)$ is assumed. The values 5.812 and 0.057 are those obtained from the maximum likelihood approach in Section 5.4.2. A common prior distribution for a_j is used for all exposed locations as they are not subject to local catchment and stream-flow conditions.

5.4.4 Specification of priors for the eruptive volumes V_i

For each named event ($i = 1, \ldots, 14$) a lognormal prior is assumed for the eruptive volume, $V_i \sim LN(\mu_i, \sigma_i^2)$. This satisfies the non-negativity constraint on V_i. The parameters for the lognormal distribution are:

$$\mu_i = \log\left(m_i^2/\sqrt{s_i^2 + m_i^2}\right) \tag{5.13}$$

$$\sigma_i^2 = \log\left(s_i^2/m_i^2 + 1\right) \tag{5.14}$$

where m_i and s_i are the estimate and associated error (obtained by inverting the Hessian matrix), respectively, for each of of the 14 $\hat{V}_i$ given in Table 5.3.

Event numbers $i = 15, \ldots, 22$ do not have known isopach diagrams. Therefore, they are treated as unnamed events with regards to the specification of the event dependent parameters $\alpha_{1i}, \alpha_{2i}, \beta_{1i}, \beta_{2i}, \phi_i, c_i$, and d_i. However, the observed tephra thicknesses of these events are large enough to require a different prior distribution for the eruptive volume:

$$V_i \sim 2 \times \text{Beta}(1.65, 2.33). \tag{5.15}$$

This was formulated from discussions with an expert volcanologist (Shane Cronin, pers. comm.) experienced in working with Mt Taranaki tephras. The prior distribution assumes 50% of the events will have a volume less than 0.8 km^3 with a mode at 0.66 km^3. The inclusion of the multiplier allows for a maximum volume of 2 km^3.

From further discussion with our volcanological expert (Shane Cronin, pers. comm.), we assume for the remaining unnamed events ($i = 23, \ldots, 270$) a maximum eruption volume of 1 km^3. The elicited prior information has 90% of the events with a volume less than 0.1 km^3, and a mode at 0.01 km^3. This induces the prior

$$V_i \sim \text{Beta}(1.25, 25.23). \tag{5.16}$$

A sensitivity analysis of the choice of distribution in (5.16) is conducted in Section 5.6.4.

5.4.5 Specification of priors for the wind directional parameters ϕ_{1i}, ϕ_{2i}

The possibility of only one lobe is introduced through the the Bernoulli random variable

$$X_i \sim \text{Bernoulli}(q_i), \tag{5.17}$$

where q_i denotes the probability of observing two lobes. If $X_i = 0$ only one lobe is observed, whereas $X_i = 1$ under the presence of two lobes. Tephra from the majority of the named events ($i = 1, \ldots, 14$) was distributed both north-east and south-east of the vent, therefore $q_i = 1$. However, the Kaponga.f, Tuikonga, Pukeiti, and Waitepuku tephras ($i = 7, 11, 13, 14$ respectively) were only distributed north-east of the vent (Figure 5.7). Therefore $q_i = 0$ for these events. For the remaining events ($i = 15, \ldots, 270$), we use the prior $q_i = 0.1$. These eruptions are assumed to be much smaller due to their absence from the exposed records. Their tephra producing phase thus is more likely to have occurred under a single wind condition. Setting $q_i = 0.1$ assigns a small probability of observing two lobes, yet allows enough flexibility to cope with events that deposited thicker tephras in both Lake Umutekai (NNE of the vent) and Eltham Swamp/Lake Rotokare (SSE of the vent). A sensitivity analysis for the choice of q_i for the unnamed events is performed in Section 5.6.2.

Beta distributions are assigned to the wind direction, modulated to allow values in the range $(-\pi/2, \pi/2)$:

$$\xi_{1i} \sim \pi\text{Beta}(\upsilon_{1i}, \eta_{1i}) - \pi/2 \tag{5.18}$$

$$\xi_{2i} \sim \pi\text{Beta}(\upsilon_{2i}, \eta_{2i}) - \pi/2. \tag{5.19}$$

All of the unexposed sites are east of the vent (see Figure 5.1), and are far enough away that any easterly winds would not provide a deposit in the locations. Therefore, wind directions must be restricted to the range $(-\pi/2, \pi/2)$.

If $X_i = 0$, only one lobe is observed. Therefore, ϕ_{2i}, α_{2i}, and β_{2i} are set equal to ϕ_{1i}, α_{1i}, and β_{1i}, respectively. However if $X_i = 1$, under the presence of two lobes, then ϕ_{2i} is located

some angle (Y_i) away from ϕ_{1i} :

$$\phi_{1i} = \xi_{1i} \tag{5.20}$$

$$\phi_{2i} = \phi_{1i} + \text{sign}(\phi_{1i}) Y_i X_i \tag{5.21}$$

where Y_i is the absolute difference between ξ_{1i} and ξ_{2i}. The term sign(ϕ_{1i}) is included to indicate whether ϕ_{2i} lies north or south of ϕ_{1i}. If $\phi_{1i} > 0$ then ϕ_{2i} will be located Y radians *clockwise* of ϕ_{1i}, whereas if $\phi_{1i} < 0$, ϕ_{2i} will be located Y radians *anticlockwise* of ϕ_{1i}. In the case of two lobes, the labels ϕ_{1i} and ϕ_{2i} are interchangeable.

The values of υ_{1i}, η_{1i}, υ_{2i}, and η_{2i} in (5.18) and (5.19) depend on how the event is classified. For the named events ($i = 1, \ldots, 14$) values are obtained by method of moments. The mean and variance of a Beta(υ_{ki}, η_{ki}) distribution are:

$$\mu_{ki} = \frac{\upsilon_{1i}}{\upsilon_{1i} + \eta_{ki}} \tag{5.22}$$

$$\sigma^2_{ki} = \frac{\upsilon_{1i}\eta_{ki}}{(\upsilon_{1i} + \eta_{ki})^2(\upsilon_{1i} + \eta_{ki} + 1)}. \tag{5.23}$$

When rearranged this gives:

$$\upsilon_{ki} = \left(\frac{1 - \mu_{ki}}{\sigma^2_{ki}} - \frac{1}{\mu_{ki}}\right)\mu^2_{ki} \tag{5.24}$$

$$\eta_{ki} = \upsilon_{ki}\left(\frac{1}{\mu_{ki}} - 1\right). \tag{5.25}$$

Given the estimated wind directions $\hat{\phi_{ki}}$ ($k = 1, 2$) and associated errors (s_{ki}, obtained by inverting the Hessian matrix) for each event in Table 5.3, transformation from the interval $(-\pi/2, \pi/2)$ to (0,1) gives $\mu_{ki} = (\hat{\phi_{ki}} + \pi/2)/\pi$, and $\sigma^2_{ki} = (s_{ki}/\pi)^2$. These can then be substituted into (5.24) and (5.25) to give υ_{ki} and η_{ki} for all ($k = 1, 2$) and ($i = 1, \ldots, 14$).

Since the remaining events ($i = 15, \ldots, 270$) have a small probability of exhibiting two lobes, a common value of υ_{ki} and η_{ki} are used for both $k = 1, 2$. Rhoades et al. (2002) compared the wind velocity at various different altitudes for two North Island sites. These sites were Auckland Airport (37°00'S, 174°47'E) and Ohakea Airforce Base (40°12'S, 175°23'E), located approximately 260 km NNE and 150 km SE of Mt Taranaki, respectively. Assuming no difference from present day wind conditions, the similarities between wind velocities at both locations suggests common prevailing winds at other North Island locations (Rhoades et al., 2002). Assuming a plume altitude of approximately 10,000 m, Figure 12 of Rhoades et al.

(2002) suggests a northern wind velocity $Z_1 \sim N(0.5, 10^2)$ m/s, and an eastern wind velocity $Z_2 \sim N(9, 12^2)$ m/s. The dispersal axis of the deposit (ϕ) depends on these velocities through the relationship $\phi = \tan^{-1}(Z_1/Z_2)$. The near-zero northern mean velocity suggests an easterly prevailing wind ($\phi = 0$), but the variability allows for deposits in the NE and SE directions. Through Monte Carlo simulation of Z_1 and Z_2, υ_{ki} and η_{ki} were found equal to 2.09 and 2.08.

5.4.6 Specification of priors for the shape parameters $\alpha_{1i}, \alpha_{2i}, \beta_{1i}$ and β_{2i}

The parameters α_{1i},α_{2i}, β_{1i}, and β_{2i} control the semi-elliptical shape of the deposit. A gamma prior distribution is employed for each shape parameter, which satisfies the non-negativity constraints. In order to satisfy the physical constraints that $\alpha_{1i} > 4\beta_{1i}$ and $\alpha_{2i} > 4\beta_{2i}$ for all i, the following prior distributions are constructed:

$$\beta_{1i} \sim \text{Gamma}(\gamma_1, \gamma_2) \tag{5.26}$$

$$\alpha_{1i} \sim 4\beta_{1i} + \text{Gamma}(\gamma_3, \gamma_4). \tag{5.27}$$

The parameters $\gamma_1 = m^2/s^2 = 3.476$ and $\gamma_2 = m/s^2 = 7.374$, where m and s are the mean (0.471) and standard deviation (0.253) of all nonzero $\hat{\beta_{1i}}$ and $\hat{\beta_{2i}}$ values from Table 5.3. Likewise, $\gamma_3 = m^2/s^2 = 0.990$ and $\gamma_4 = m/s^2 = 0.828$, where m and s are the mean (1.196) and standard deviation (1.202) of all $\hat{\alpha_{1i}} - 4\hat{\beta_{1i}}$ and $\hat{\alpha_{2i}} - 4\hat{\beta_{2i}}$ from Table 5.3. If there is only one lobe (i.e. $X_i = 0$) then $\alpha_{2i} = \alpha_{1i}$ and $\beta_{2i} = \beta_{1i}$, whereas for two lobes (i.e. $X_i = 1$), β_{2i} and α_{2i} have the same prior distributions as β_{1i} and α_{1i} above. The prior distributions for α_{1i},α_{2i}, β_{1i}, and β_{2i} are common across all 270 events.

The directional functions $f_1(.)$ and $f_2(.)$ in (5.11) act multiplicatively on the observed tephra thicknesses. Both the unnamed and named tephras are assumed to be deposited under the same wind conditions. Therefore the prior distributions for α_{1i},α_{2i}, β_{1i}, and β_{2i} are common across all 270 events.

5.4.7 Specification of priors for the remaining event dependent parameters c_i and d_i

The parameter c_i determines the rate at which the deposit thickness decays with distance. The restriction that c_i must be greater than two, suggests a lognormal distribution for $c_i - 2$. The term $d_i V_i^{1/3}$ is included in (5.7) to ensure a finite tephra thickness at the vent (when

$r_j = 0$). This imposes a non-negativity constraint on d_i, hence a lognormal distribution is employed for d_i.

Rhoades et al. (2002) give an estimated mean (2.38) and error (0.26) for c_i, and an estimated mean (1.85) and error (2.18) for d_i. By matching moments, the estimates given by Rhoades et al. (2002) supply the prior distributions for both the unnamed events and the named events that do not have known isopach diagrams ($i = 15, \ldots, 270$):

$$c_i - 2 \sim LN(-1.160, 0.384) \tag{5.28}$$

$$d_i \sim LN(0.180, 0.871). \tag{5.29}$$

For the remaining named events ($i = 1, \ldots, 14$) parameters of the lognormal distributions are chosen such that c_i and d_i have means (2.194 and 0.780 respectively) and variances (0.001 and 0.114 respectively) equal to that of the 14 estimates given in Table 5.3. The prior distributions for c_i and d_i are common across all 14 named events:

$$c_i - 2 \sim LN(-1.655, 0.027) \tag{5.30}$$

$$d_i \sim LN(-0.333, 0.171) \tag{5.31}$$

Different priors are specified here for the named and unnamed events, as the objective is to leverage the well described named tephras into understanding the smaller unnamed events. The estimated parameter values from Rhoades et al. (2002) apply to events of many orders of magnitude, whereas the volumes of large named events are much more consistent.

5.5 Results

Calculations were performed using WinBUGS (Lunn et al., 2000). The program code is given in Appendix E. The model was run for 50,000 iterations, with a burn-in period of 5000. A thinning factor of 50 was applied to remove autocorrelations and provide an approximate random sample. Site-specific parameters a_j were initialized at 0 for the unexposed locations ($j = 1, \ldots, 6$), and 8 for the exposed locations ($j = 7, \ldots, 39$).

Goodness-of-fit is assessed using the Deviance Information Criterion (Spiegelhalter et al., 2002)

$$\mathrm{DIC} = \bar{D} + p_D, \tag{5.32}$$

where $\bar{D}$ is the posterior mean deviance which measures model fit or adequacy, and p_D is a complexity measure for the effective number of model parameters. DIC is a Bayesian generalization for the frequentist Akaike Information Criterion (Akaike, 1977). As such, smaller DICs indicate better models. The baseline model produced a DIC of 3408.15. In Section 5.6 DIC is used for goodness-of-fit comparison of various sensitivity analyses.

Three chains were used to calculate the posterior distributions. Convergence was assessed using the Gelman-Rubin diagnostic (Gelman and Rubin, 1992), calculated using the coda package in R (Plummer et al., 2006; R Core Team, 2014). The Gelman-Rubin diagnostic compares the within-chain variance to the between-chain variance. The estimated variance of each parameter is calculated as a weighted sum of the within-chain and between-chain variance, from which the potential scale reduction factor (PSRF) is calculated. Convergence is deduced when all PSRFs are near one. Trace plots and Gelman-Rubin diagnostic plots, for the site-specific effect parameters for the unexposed locations, are shown in Figure 5.8. Similar plots were obtained for all other model parameters, all of which showed convergence.

The posterior distributions of the site-specific effects are given in Figure 5.9. The small variability in aggregated site-specific effects, across all exposed locations, is the imposed constraint, due to these locations not being prone to over- or under-thickening. Treating the site-specific effects for the exposed locations as a benchmark, there is a clear difference in expected thicknesses for the unexposed locations. Proximal (near-source) tephras have a considerably smaller site-specific effect, suggesting observed tephras are thinner than expected. There is a tendency for distal tephras to be over-thickened. Auckland has a notably larger site-specific effect. Absolute thicknesses of observed tephras in the Auckland record (average thickness 0.18 cm) are much smaller in comparison to that of near-source tephras (average thickness 5.18 cm). Small measurement errors can, therefore, have a larger effect on the thickness of tephras observed in the Auckland record. Of particular interest is the relative size of the Eltham Swamp site-specific effect. Despite Eltham Swamp lying only $\sim$7 km from Lake Rotokare, and on practically the same dispersal axis as both Lake Rotokare and Lake Rangatauanui, tephras found in Eltham Swamp are typically thicker. It is possible that these tephras may have fallen into a vegetated swamp. Therefore, measured thicknesses could have been expanded by the presence of vegetation fragments throughout the tephra layers.

Figure 5.8: Gelman-Rubin diagnostic plots (left) and trace plots (right) for the site-specific effects a_j for unexposed locations ($j = 1, \ldots, 6$).

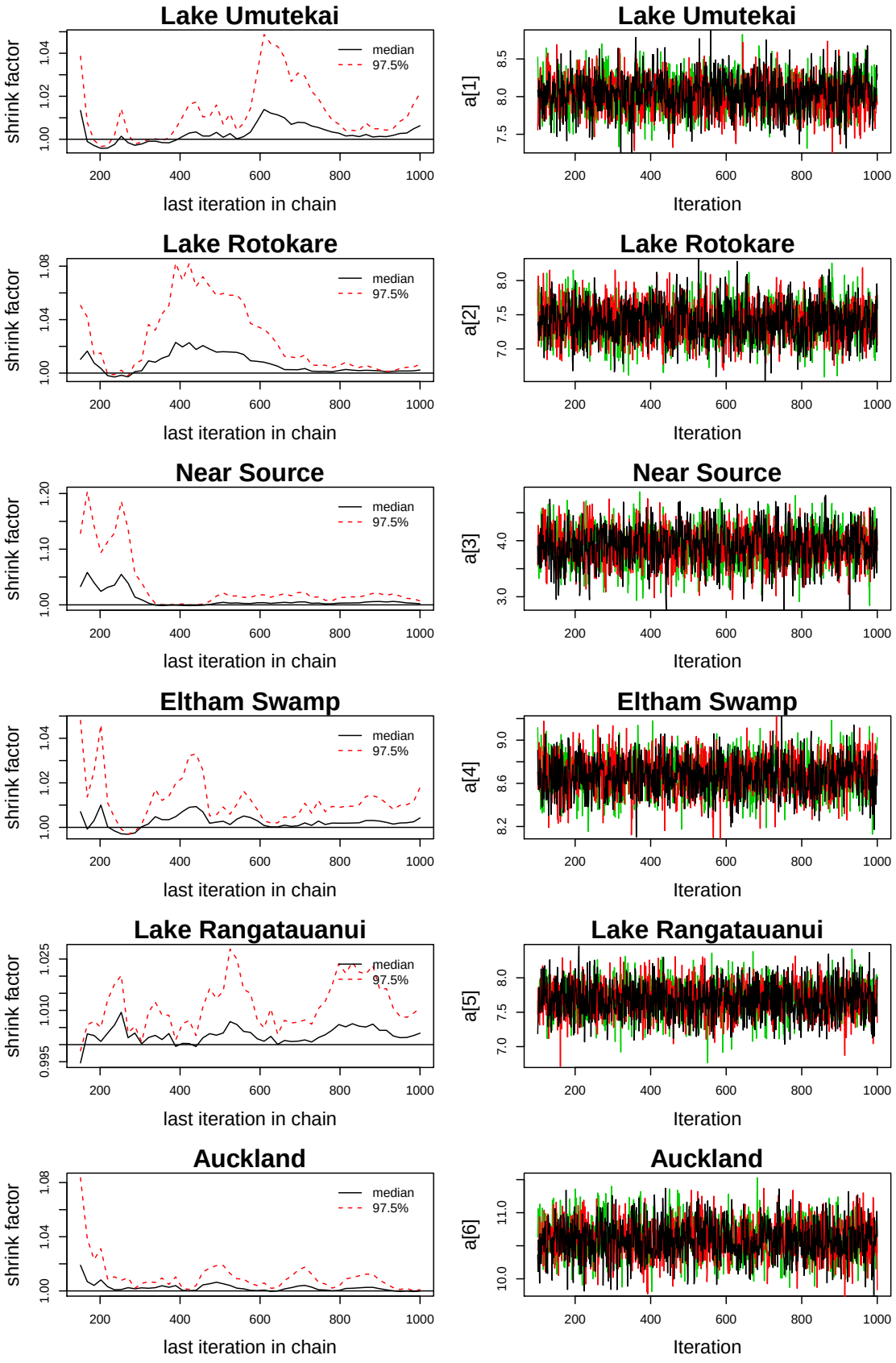
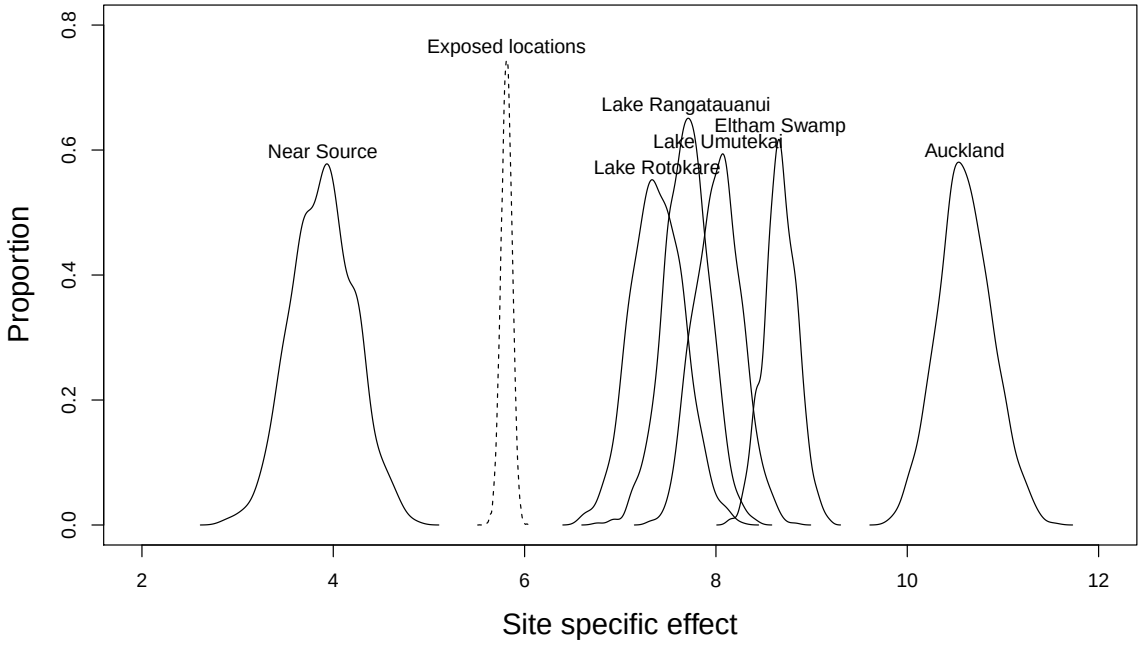


Figure 5.9: Posterior distributions for the site-specific effects a_j. The site-specific effects for the exposed locations ($j = 7, \ldots, 39$) are shown as one posterior distribution.



Posterior distributions for the volumes of the well-known named events are given Figure 5.10. There is more variability around the larger ($\sim$ 1.5 km^3) named events (Inglewood, Mahoe, Paetahi A and Tuikonga). What is particularly attractive is the relatively small variability observed for the older named tephras (Koru A, Pukekiti and Waitepuku). These tephras had poorly constrained isopach maps in Alloway et al. (1995) (see Figure 5.7). The volume distribution of the smallest named event (Tareki F) suggests tephras corresponding to events greater than $\sim$ 0.3 km^3, would be widespread and easily recognizable in the geological record. This implies a small probability of unnamed events greater than $\sim$ 0.3 km^3 going unmissed.

Figure 5.11 compares the posterior mean parameter values for the unnamed events. There is a strong negative relationship between eruptive volume and the parameter c, which describes the rate at which thickness decays with distance from the vent. According to Gonzalez-Mellado and De la Cruz-Reyna (2010), the rate of decay is inversely related to the height of the eruptive column. Larger events distribute ash further into the atmosphere, producing eruptive columns that reach higher altitudes. Therefore, a smaller rate of decay (c) is as expected for larger volume eruptions.

The nuisance parameter d was included by Rhoades et al. (2002) to ensure thicknesses are

Figure 5.10: Posterior distributions for the eruptive volumes V_i of named events ($i = 1, \ldots, 14$). The dashed lines indicate the estimated values from Table 5.3.

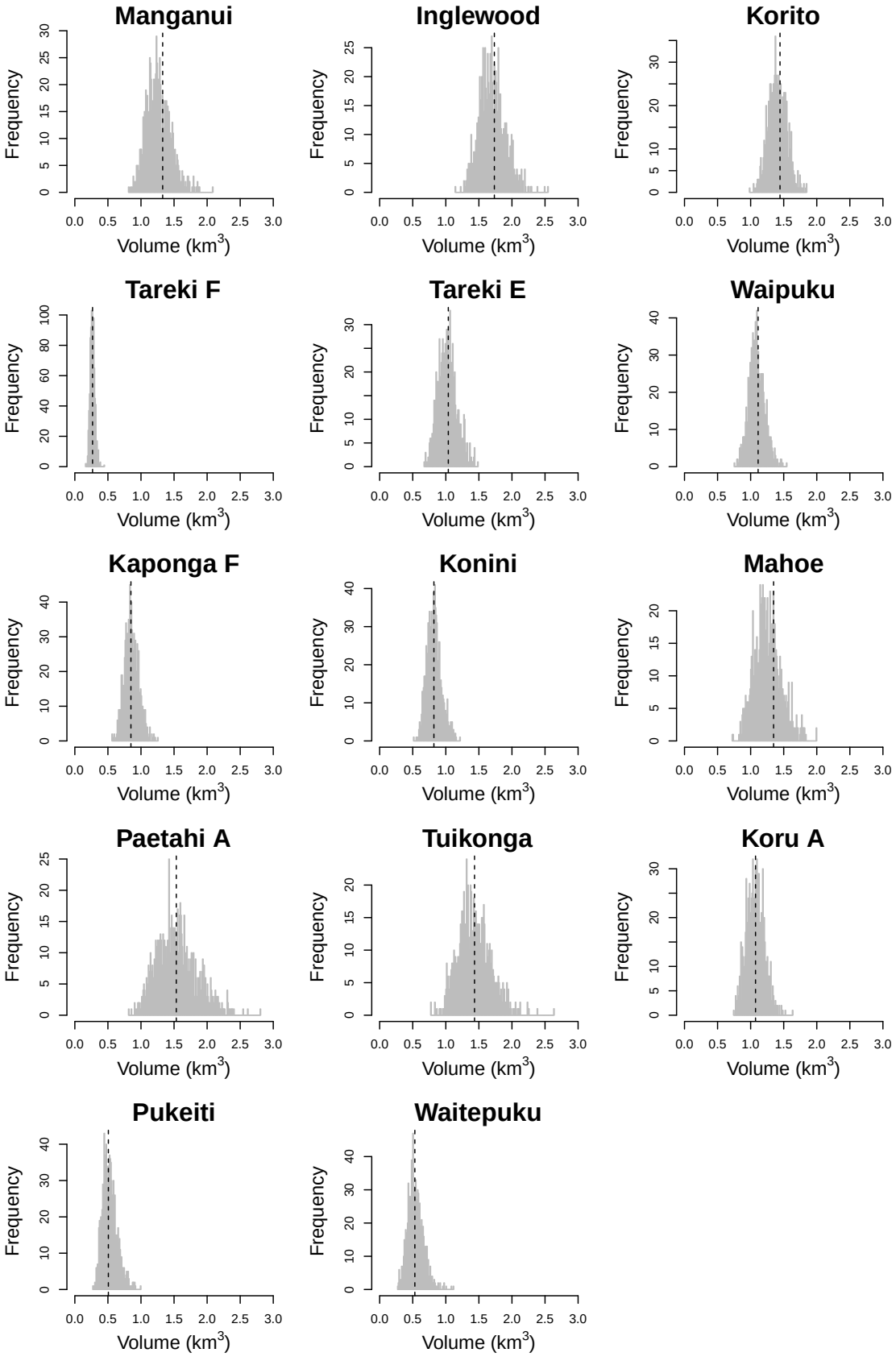
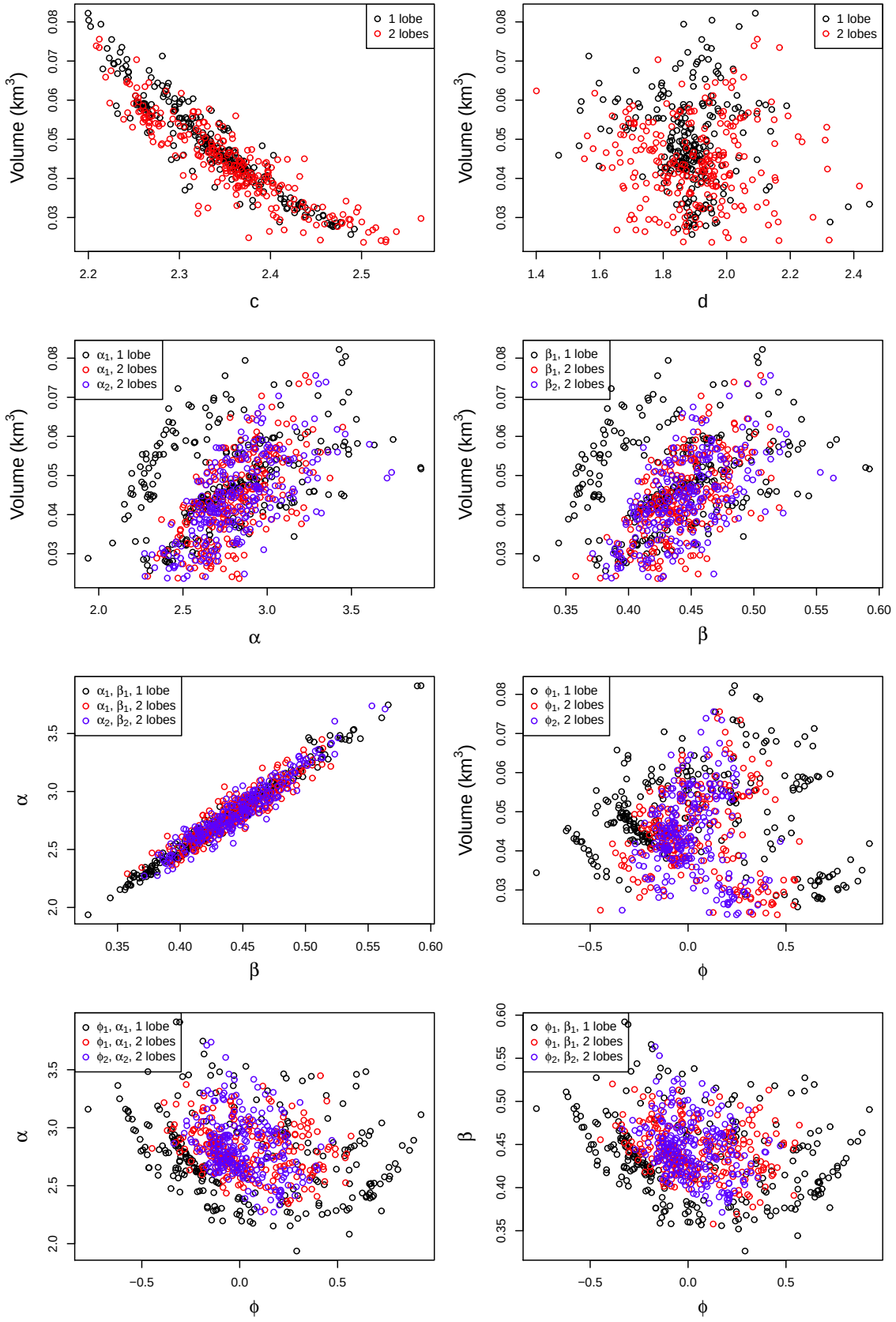


Figure 5.11: Posterior mean parameter values of unnamed events ($i = 23, \ldots, 270$) separated according to the number of dispersal lobes.



finite at the vent. Unsurprisingly, there doesn't appear to be any relationship with eruptive volumes.

The α and β parameters describe the elliptical shape of each dispersal lobe, and are considered proxies for the wind strength and variability. There is a strong linear relationship between these two shape parameters. This is a result of the physical constraint that α must be greater than 4β (see page 125). As α and β increase, the dispersal ellipse gets longer and wider to varying degrees. β lengthens the ellipse at a faster rate. Large α values characterize stronger more variable winds, whereas large β values portray stronger more consistent winds.

There is a slight tendency for larger volume eruptions to be associated with larger α and β values. Given the model equation in (5.5); for a fixed volume V, if any of the α or β terms increase the size of the ellipse increases for every isopach. Hence the thickness must increase at every point for the same volume. However, since the thicknesses are fixed data points, this inconsistency can only be resolved by a positive correlation between α, β and V. Intuitively, one would expect wind speed to be independent of eruption volume. However, in this case the model parameters do not appear to be separable.

Almost all of the deposit locations are east of the volcano (Figure 5.1). The wind parameter ϕ suggests the majority of tephras were dispersed in the north-east quadrant from the vent ($\phi > 0$), with wide variability in estimated volumes. Tephras dispersed further north (as $\phi \to \pi/2$) or south (as $\phi \to -\pi/2$) require longer ellipses. As a result there are no points that have small α or β values but large absolute wind directions ϕ.

Although the likelihood of observing a double-lobed dispersal pattern for unnamed tephras is small, we are interested in the effect of allowing two lobes on the parameter values. There appears to be slightly more variability in the rate of thickness decay (c) and the finite-thickness parameter (d), for multi-lobed dispersals. Interestingly, allowing for multiple dispersal lobes gives a higher rate of thickness decay. This appears to be a compensatory effect. A single lobe can only reach locations in two directions by dispersing it further. The apparent cluster of black points to the left of the volume versus shape (α or β) plots, in Figure 5.11, suggests a more circular dispersal ellipse for single lobed events (smaller α and β values). Assuming one lobe also tends to force the wind direction further north or south (larger absolute ϕ values). Two lobes allow for more 'targeted' dispersal directions, hence permitting a faster decay (larger c). These findings prompt investigation into the locations at which tephras were observed.

Changes in the posterior mean parameter values, for the likely case of just one dispersal lobe, are shown in Figure 5.12. In this figure, parameter values are separated according to the locations at which corresponding events were recognized. Events observed in both northern and southern sites have greater volumes, consistent with larger eruptions having wider dispersal. Events observed only in the northern records appear to have comparatively smaller volumes, which is a likely result of the location of sampling sites. The small cluster of northern events > 0.05 km^3 correspond to those observed only in the Auckland record.

The wind parameter is largely forced in the direction of the observation sites ($\phi_1 > 0$ for events observed only in northern locations, and $\phi_1 < 0$ for those only observed south of the volcano). There is more variability in the parameter values for tephras observed in both directions. Although wind direction and volume appear to be associated with the location at which tephra is observed, there does not appear to be any great differences in the remaining event-dependent parameters.

5.6 Sensitivity analyses

In analyzing the tephra dispersal a number of assumptions were made, particularly to do with the prior distributions for the model parameters. It is wise to do a sensitivity analysis to see how the conclusions are influenced by these assumptions. As the focus is to estimate the eruptive volumes of Mt Taranaki events, investigation into the sensitivity of volume estimates to changes in the chosen prior distributions is necessary.

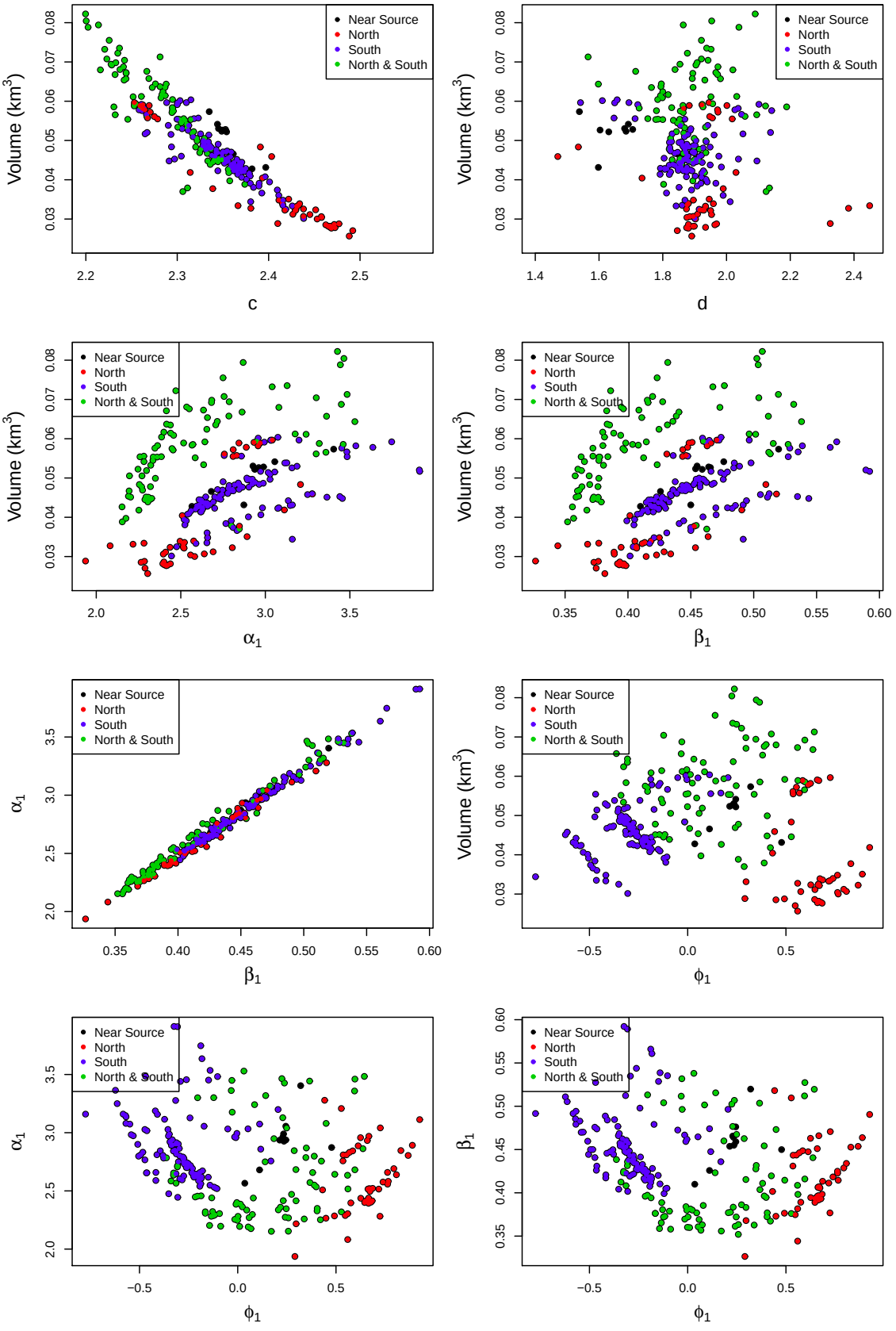
5.6.1 Prior distribution for thickness precision

One of the obvious sensitivity analyses to consider is the use of a prior for the thickness precision $\tau = 1/\sigma^2$ in (5.10). The original model specified σ as the residual standard error (1.78) from Rhoades et al. (2002), explicitly stating $\tau = 1/1.78^2 = 0.3156$. Here a more flexible approach is considered. A prior distribution is imposed on the thickness precision, which is semi-informed by the residual standard error from Rhoades et al. (2002). A gamma prior distribution is used

$$\tau \sim \text{Gamma}(2, 2\sigma^2), \tag{5.33}$$

where $\sigma = 1.78$ from Rhoades et al. (2002).

Figure 5.12: Posterior mean parameter values of unnamed events ($i = 23, \ldots, 270$). Green circles denote tephras observed in at least one southern location (Lake Rotokare, Eltham Swamp, Lake Rangatauanui) and at least one northern location (Lake Umutekai and Auckland). Red circles denote events observed only in northern location(s). Blue circles denote events observed in southern location(s). Black circles denote events only observed in the Near-Source record.



The posterior mean and standard deviation obtained for τ are 0.421 and 0.027, respectively. This suggests a greater precision (smaller variability) in thickness measurements than that obtained by Rhoades et al. (2002) for the Taupo deposits. This is as expected, Rhoades et al. (2002) modeled the *average* relationship between tephra thickness and erupted volume. Therefore, the residual error of their model is much larger than would be expected for individual events (see page 120). Quite pleasingly, the use of a prior appears to have very little effect on the posterior distributions for both event-dependent parameters and site-specific effects (Figure 5.13). Comparison between the DICs (Table 5.4) suggests the extra complexity introduced by using a prior for the thickness precision results in far worse fit (DIC = 3597.7, compared with the original DIC of 3408.15).

Table 5.4: Comparison in DICs for the various sensitivity analyses.

Model	DIC
Baseline	**3408.15**
Hyperprior on thickness precision	3597.70
Circular contours	3582.52
Reference priors on c and d	3446.0
Setting $\alpha = 4\beta$	3439.85
Inflating the volume	3413.48
Prior on probability of two lobes	3407.80

5.6.2 Prior probability of two lobes

In Section 5.4.5, a Bernoulli random variable was introduced to allow for the possibility of observing only one dispersal lobe. The probability of observing two lobes (q_i) in (5.17) was semi-arbitrarily set at 0.1 for the small unnamed events ($i = 15, \ldots, 270$). As an alternative, a uniform prior distribution is imposed on the unnamed events:

$$q_i \sim U(0, 1). \tag{5.34}$$

This avoids presuming that the likelihood of a double lobed dispersal pattern is small.

The posterior mean probability of two dispersal lobes, combined for all unnamed events, is 0.504 with standard deviation 0.289. Interestingly, this makes very little difference with respect to goodness-of-fit (DIC = 3407.8 in comparison to the original DIC of 3408.15, Table 5.4). Figure 5.14 shows that when two lobes occur under (5.34), they usually have very similar directions. So the dispersal lobes overlap, producing a narrow fan rather than two

Figure 5.13: Comparison between prior and posterior distributions for the various sensitivity analyses. Site-specific effects (A-F), event dependent model parameters (G-L). Solid lines for posterior distributions, dotted lines for corresponding prior distributions. Baseline model in black, volume inflated model (Section 5.6.4) in red, reference priors on c and d (Section 5.6.5) in green, circular contours (Section 5.6.6) in blue, the model which imposes a prior on the thickness precision (Section 5.6.1) in cyan, and the model setting $\alpha = 4\beta$ (Section 5.6.3) in grey.

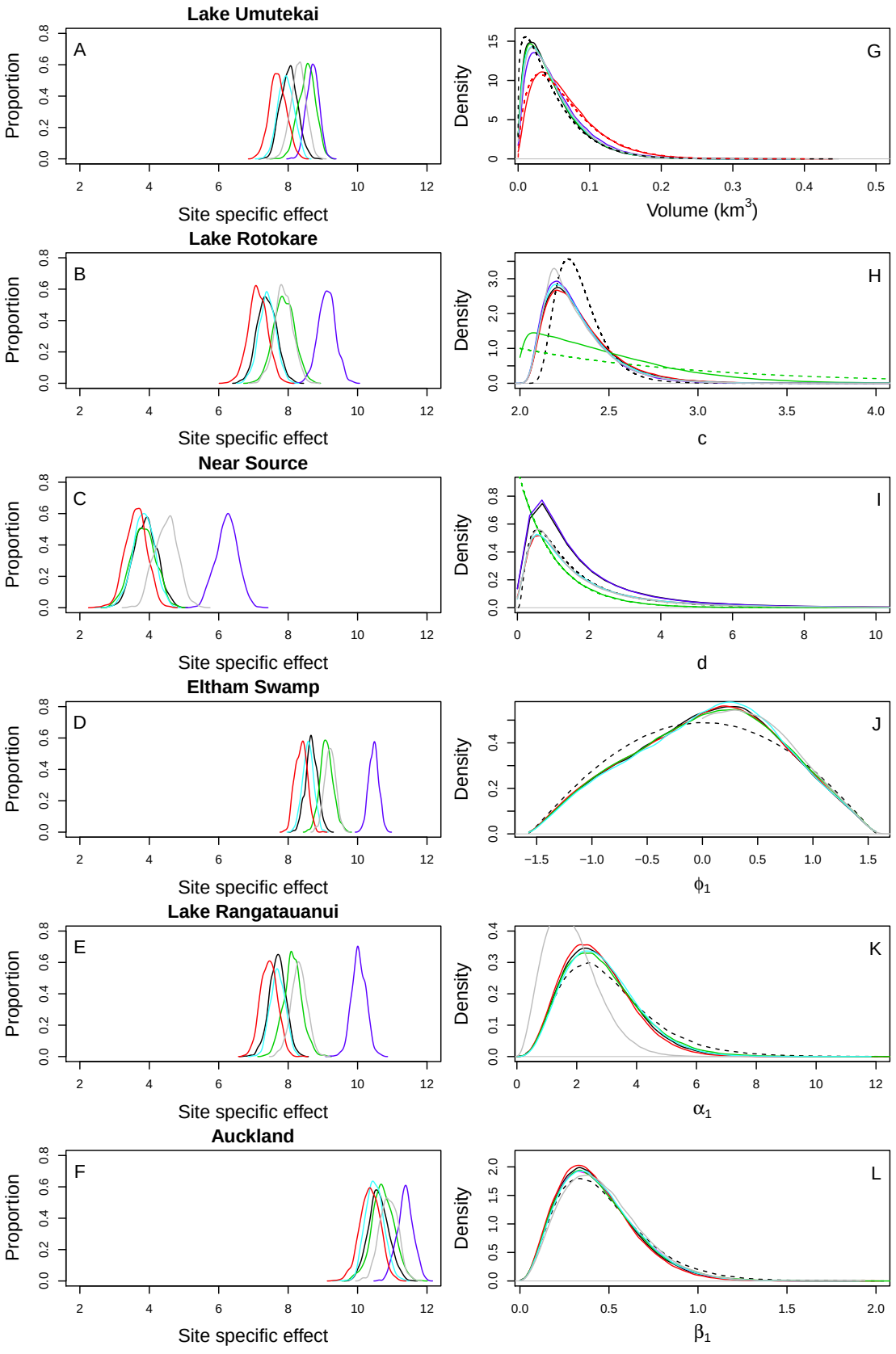
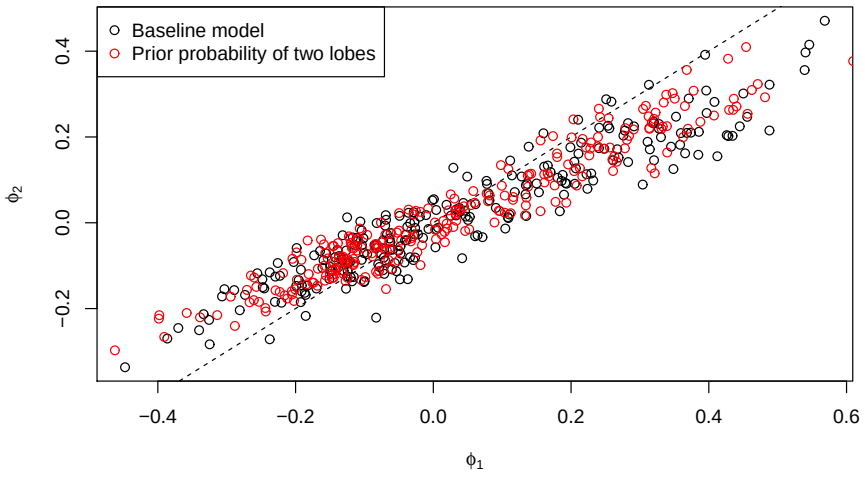


Figure 5.14: Comparison between posterior mean wind directions ϕ_{1i} and ϕ_{2i} (in radians) for the case of two dispersal lobes for the unnamed events. The dashed line represents the case where $\phi_{1i} = \phi_{2i}$.



distinguishable ellipses. Given the limited data available, this is a degree of freedom we can do without, as is indicated by the DIC. Also, to be consistent with this, the major tephras would need four or more lobes. As the estimated volumes are unaffected by imposing a prior distribution on the probability of two lobes, the simpler model will be used.

5.6.3 Setting $\alpha = 4\beta$

The strong linear relationship observed between the two shape parameters α and β in Figure 5.11, suggests the number of model parameters could be reduced by forcing α_{1i} and α_{2i} to be a function of β_{1i} and β_{2i}, respectively. We will consider the effect of setting $\alpha_{ki} = 4\beta_{ki}$ for $k = 1, 2$. Alternatively, one could set $\alpha_{ki} = m\beta_{1i}$ where a prior distribution for m is specified such that $m \geq 4$.

Although the posterior distribution for β is unaffected by this change, the resulting distribution for α has a smaller mode and less variability than the baseline model (Figure 5.13K). Larger α values give more circular dispersal patterns. Forcing $\alpha = 4\beta$ results in smaller α values, hence the dispersal ellipse will be elongated. As a result, the site-specific effects for thickening are over inflated to compensate (Figure 5.13(A-F)).

Despite the large reduction in the number of parameters obtained by setting $\alpha_{ki} = 4\beta_{ki}$ for all $(i = 1, \ldots, 270)$ and $(k = 1, 2)$, the resulting fit is much worse (DIC = 3439.85 in comparison to the baseline model DIC of 3408.15, Table 5.4).

5.6.4 Inflating the prior distribution for volume

Expert opinion was used to elicit the prior distribution for the volume of unnamed events in (5.16). This was based on the assumption of a maximum eruption volume of 1 km^3, a mode at 0.01 km^3, and an expected 90% of events producing less than 0.1 km^3.

Comparison between the posterior and prior volume distributions are given in Figure 5.13G. The posterior distribution for the original model (in black) tends towards a mode larger than 0.01 km^3. As a sensitivity analysis an alternative prior distribution is considered for the volume of the unnamed events. This is based on a mode at 0.03 km^3, with 80% of the events expected to be less than 0.1 km^3:

$$V_i \sim \text{Beta}(1.73, 24.37). \tag{5.35}$$

Although the choice of prior distribution influences the posterior volume estimates, there appears to be a trade-off with the site-specific effects. Figure 5.13(A-F) suggests inflating the prior distribution for volume decreases the estimated site-specific effects. Given the model formulation in (5.5), if the eruption volume V_i increases, there must be a compensatory effect in order to achieve the same observed thickness at each location. Imposing the alternative prior distribution on the volume, does not appear to influence the posterior distributions for any of the other event dependent parameters, as seen in Figure 5.13(H-L). This suggests the compensatory effect comes from changes to the site-specific parameters a_j alone.

The original baseline model imposes a reference prior on the site-specific parameters (see Section 5.4.3). It would be preferable to place a more informative prior distribution on these site-specific effects. However, as of now very little is known about the relative over- or under-thickening properties at each unexposed location and even less about these characteristics thousands of years ago. The site-specific parameters act additively on the logarithm of the observed tephra thickness (see (5.5)). So in addition to understanding the *relative* site-specific effects, the *absolute* size for each location needs to be quantified. If more was known about the site-specific effects a more informative prior could be imposed on a_j, which may provide a better constraint for V_i, ultimately improving estimates of the 'true' eruption volumes. However, at this point in time, we must think carefully about the prior distribution for volume to ensure that realistic volume estimates are obtained.

The smallest named event (Tareki F) has an estimated volume of approximately 0.3 km^3 (Table 5.3). This suggests almost all unnamed events should produce less than 0.3 km^3. Otherwise they would be widespread enough to be easily recognized, and subsequently classified as named events. A heavily right skewed distribution, with a very small probability of volume exceeding 0.3 km^3, is warranted for the unnamed events. Goodness-of-fit comparison suggests the original baseline model is preferred. Inflating the prior distribution for the volume of unnamed events produces a worse DIC of 3413.48 in comparison to the baseline model DIC of 3408.15 (Table 5.4).

Although the volume estimates for the unnamed events appear unfavorably sensitive to the choice of prior distribution for V, the important question is 'how much' the prior effects the posterior. By inflating the volumes the *relative* size of the site-specific effects remains unaffected, i.e. the ordering of site-specific effects from Near-Source to Auckland is unchanged (Figure 5.13(A-F)). Quite pleasingly, the relative sizes of each unnamed event and the variability around the estimated volumes are also maintained (Figure 5.15). The distribution of estimated volumes is right skewed. There is a strong positive relationship between posterior means and standard deviations (Figure 5.15A). Figure 5.15C suggests that there is slightly more variability around the posterior mean volumes under the baseline model. Table 5.4 however, suggests the model with the volume inflated prior gives a worse fit. One could conclude that shifting the prior distribution over-constrains the volumes. Unfortunately, the question remains whether the baseline model does in fact capture the 'true' eruptive volume. It should be noted however, that the differences in posterior mean volumes obtained here are much less than typical volume uncertainty (cf. Bonadonna and Costa, 2012).

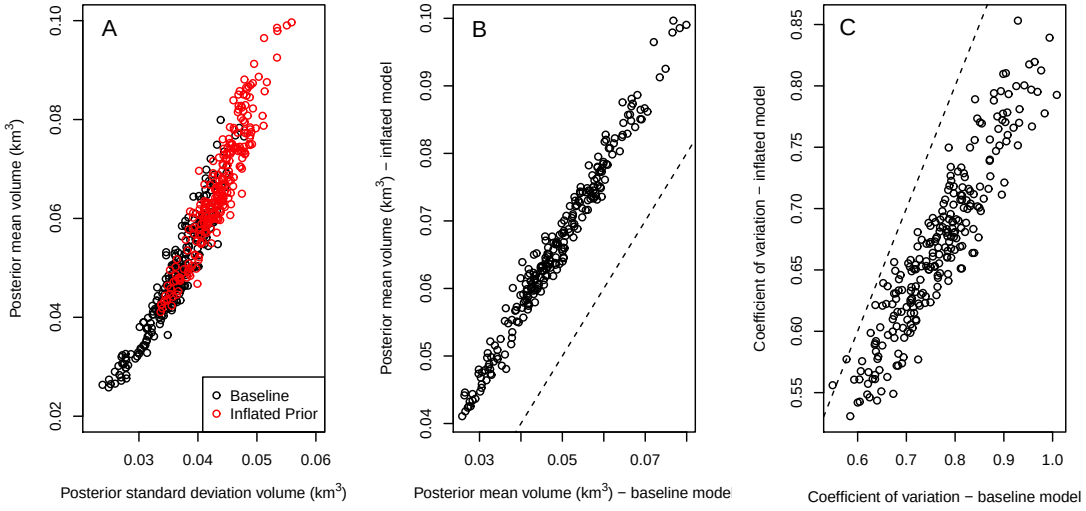
5.6.5 Reference prior distributions for c and d

The event dependent rate of thickness decay c_i, and the finite thickness parameters d_i, were specified with prior distributions based on the estimates given by Rhoades et al. (2002) (see Section 5.4.7). As an alternative, less informative prior distributions are considered, while still maintaining the non-negatively constraint on d, and the constraint that c must be at least 2. The following prior distributions are imposed

$$c_i - 2 \sim \text{Gamma}(1, 1) \tag{5.36}$$

$$d_i \sim \text{Gamma}(1, 1). \tag{5.37}$$

Figure 5.15: Comparison between posterior volume estimates of unnamed events for the baseline and volume inflated models. (A) Posterior mean volume versus posterior volume standard deviation; (B) Posterior mean volumes, on a log scale; (C) Coefficient of variation, the ratio between the posterior standard deviation and mean volume.



In a practical sense, the resulting estimates for d are of little importance. They are only included to constrain the thicknesses very close to the vent, and have very little effect on the dispersal shape. Although there is a noticeable difference in the resulting posterior distribution for c (Figure 5.13H), the posterior volume estimates and other event dependent and site-specific parameters remain relatively unaffected. Comparison between the prior and posterior distribution for c, based on the specification in (5.36), suggests a thinner tailed distribution tending towards a mode more in line with the initial results. Similar to the other sensitivity analyses, the resulting DIC of 3446.0 does not suggest any improvement in fit (Table 5.4).

5.6.6 Circular contours

In all of the sensitivity analyses considered so far, there has been very little impact on the posterior distributions for the wind direction (ϕ) and elliptical shape parameters (α and β). Possibly the largest assumption made by the model specification is that tephra is dispersed under the presence of sizeable winds. An elliptical dispersal pattern is implied, which is elongated according to the shape parameters α and β, and oriented in the direction ϕ in which the wind is blowing. The alternative assumption of 'no wind' would imply tephra deposited in a circular pattern around the vent, where $f_1(.)$ and $f_2(.)$ in (5.11) are set equal to zero. It is surprising that the posterior volumes obtained under the assumption of 'no

Figure 5.16: Distribution of tephra thicknesses for named events. The triangle denotes the position of the vent, axis units are in kilometers from the vent (0,0), isopachs are in centimeters (20 cm, 5 cm, 2 cm). Isopachs are constructed using posterior mean parameter values. Solid lines represent the preferred model, dotted lines represent circular contours resulting from the sensitivity analysis under the assumption of no wind. Points indicate nearby exposed and unexposed locations (Lake Umutekai (Um), Eltham Swamp (El), Lake Rotokare (Ro)) which observed the given tephra fall.

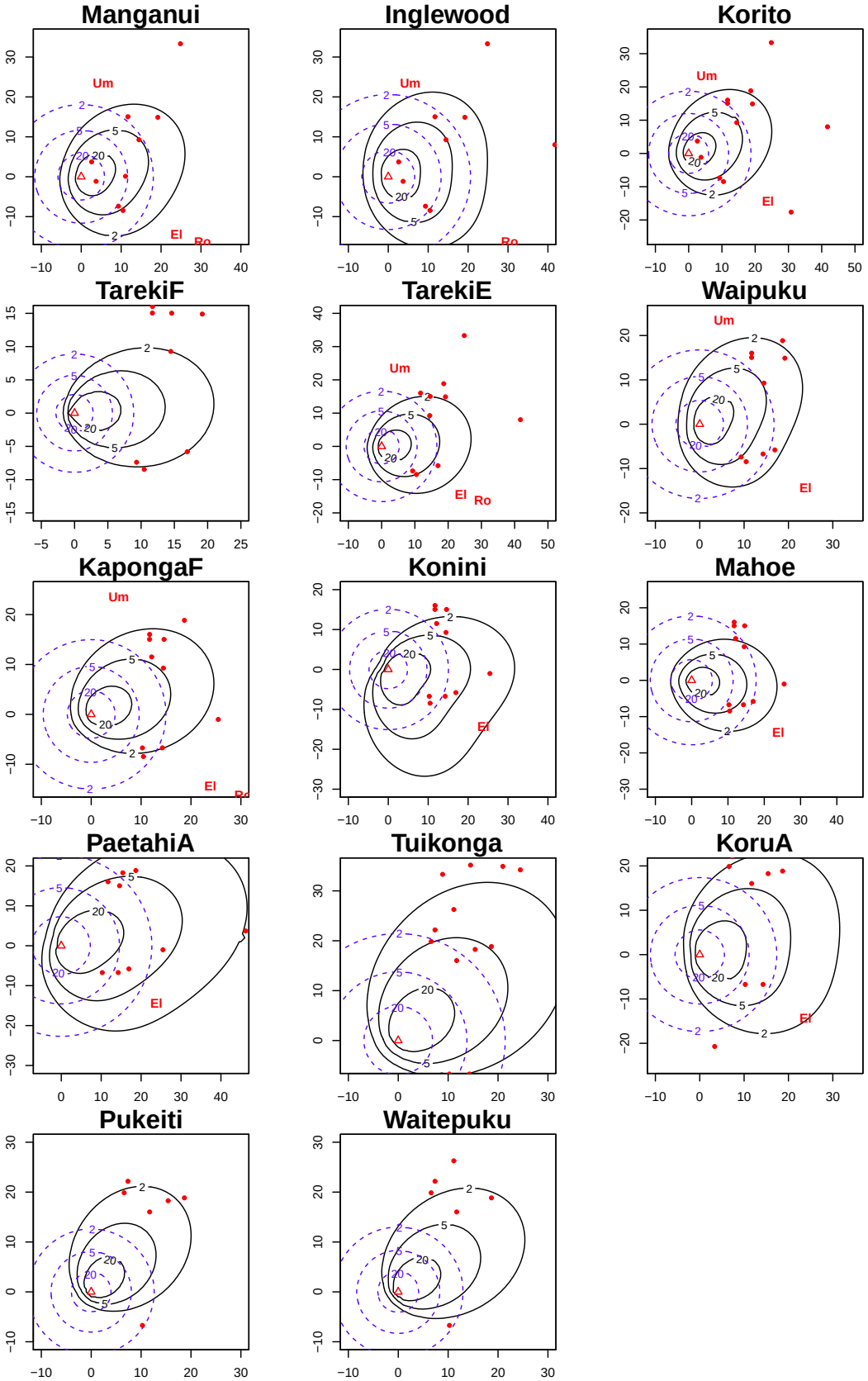
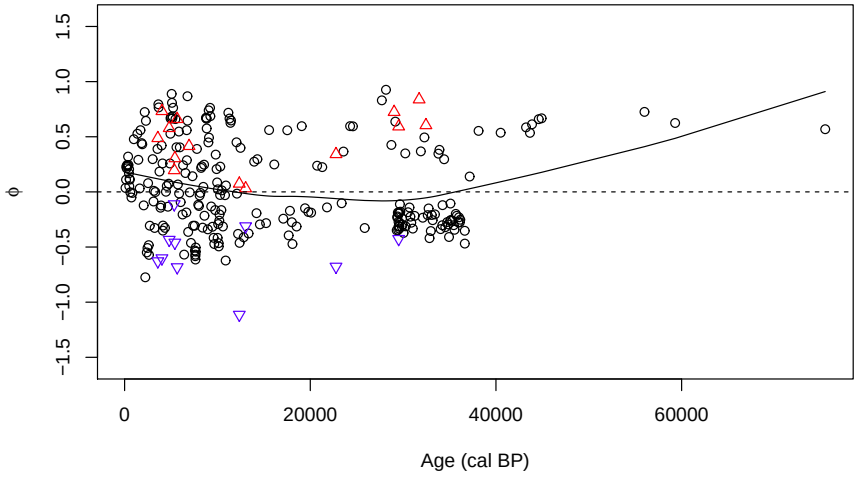


Figure 5.17: Changes in posterior mean wind direction ϕ (in radians) over time for all 270 events. The dashed line at $\phi = 0$ represents due east. Solid line denotes a loess smoothed line. Upwards triangles indicate ϕ_{1i} and downward triangles indicate ϕ_{2i} for the named events ($i = 1, \ldots, 14$). Circles denote ϕ_{1i} for the remaining events ($i = 15, \ldots, 270$).



wind' appear to be similar. However, Figure 5.13(A-F) suggests that the site-specific effects have likely been overstated to compensate.

Examining the dispersal pattern for the large named events in Figure 5.16, it is immediately apparent that a circular dispersal assumption is not applicable for this data. The parameterization of thickness under the 'no wind' assumption only considers the distance of the deposit sites from the vent, not the angular direction. Given the locations at which each tephra is observed (denoted in red in Figure 5.16), allowing tephra dispersal west of the vent is unrealistic. Not surprisingly, the DIC is far worse (3582.52 in comparison to the original model DIC of 3408.15, Table 5.4), despite the substantial reduction in the number of model parameters.

5.7 Discussion

The original parameterization specified a practically symmetric prior distribution for the wind direction, centered around a tephra dispersal east of the vent ($\phi = 0$). Figure 5.13J suggests a posterior distribution with a mode further north (approximately 0.25 radians or 15 degrees anticlockwise from east). This result is similar to that of Rhoades et al. (2002), who likewise found Taupo eruptions to be deposited somewhat north of east.

Given the mean age of each event in Table D.1, changes in the posterior mean wind direction over time can be examined. Figure 5.17 indicates the majority of events deposit tephra

north-east of the vent ($\phi > 0$). This observation is consistent with known isopach diagrams of widespread tephra, which suggest a large dispersal lobe north-east of the vent with a smaller less-prominent lobe south-east of the vent. The record of unnamed events was obtained from the six unexposed locations which differ in age length (see Table 5.1). The cluster of $\phi < 0$ points observed between 25 and 40 cal ka BP, is driven by the fact than events older than 25.6 cal ka BP are only observed in the Eltham Swamp and Auckland records. Similarly, events older than 38.1 cal ka BP are only observed in the Auckland record. This is a limitation of the data. With only one or two records spanning far enough back to observe the oldest tephras, the wind parameter is forced in the direction of the observation site. This is regardless of what the actual wind pattern may have been. For a better understanding in the changes in wind over time, longer records collected from many locations at varying distances and angular directions from the vent would be preferred. However, finding sites suitable for preserving tephra is an ongoing challenge.

5.7.1 Implications for hazard models

The estimated eruptive volumes from the baseline model can be used to establish hazard forecasts for future Mt Taranaki events. A general distribution for modeling the size/volume is required. The presence of both small and very large widespread eruptions at Mt Taranaki, suggests a bimodal distribution for modeling the logarithm of eruption volumes. Turner et al. (2008a) proposed a mixture of Weibull distributions for modeling the bimodal distribution of the *time* between eruptions. Here we consider the fit of a mixture of Weibull distributions, a mixture of normal distributions, and a mixture of lognormal distributions, to model the logarithm of the posterior mean eruption *volumes*. For the mixture of normal distributions:

$$\begin{aligned} f(Z) &= p f_1(Z) + (1-p) f_2(Z) \\ &= \frac{p}{\sigma_1\sqrt{2\pi}} \exp\left(-\frac{(Z-\mu_1)^2}{2\sigma_1^2}\right) + \frac{1-p}{\sigma_2\sqrt{2\pi}} \exp\left(-\frac{(Z-\mu_2)^2}{2\sigma_2^2}\right). \end{aligned} \tag{5.38}$$

where Z is the logarithm of eruptive volume $\log(V)$ and p is the mixing proportion ($0 < p < 1$). Since the Weibull and lognormal distributions only support positive values, the posterior mean volume estimates are scaled by dividing by an arbitrarily small number (0.001 km^3) before taking the logarithm (i.e. the Weibull and lognormal mixture distributions are fit to

$Z = \log(V/0.001)$). For the mixture of Weibull distributions:

$$\begin{aligned} f(Z) &= pf_1(Z) + (1-p)f_2(Z) \\ &= pk_1\delta_1^{k_1} Z^{k_1-1} \exp(-(\delta_1 Z)^{k_1}) + (1-p)k_2\delta_2^{k_2} Z^{k_2-1} \exp(-(\delta_2 Z)^{k_2}). \end{aligned} \tag{5.39}$$

For the mixture of lognormal distributions:

$$\begin{aligned} f(Z) &= pf_1(Z) + (1-p)f_2(Z) \\ &= \frac{p}{Z\sigma_1\sqrt{2\pi}} \exp\left(-\frac{(\ln Z - \mu_1)^2}{2\sigma_1^2}\right) + \frac{1-p}{Z\sigma_2\sqrt{2\pi}} \exp\left(-\frac{(\ln Z - \mu_2)^2}{2\sigma_2^2}\right). \end{aligned} \tag{5.40}$$

The parameters in each model are estimated by numerical maximization of the log-likelihood function

$$\log L(Z_1, \ldots, Z_n) = \sum_{i=1}^{n} \log f(Z_i). \tag{5.41}$$

Model fits are compared using the Akaike Information Criterion (Akaike, 1977)

$$\text{AIC} = -2\log L + 2k, \tag{5.42}$$

where k is the number of parameters in each model. Smaller AICs indicate better models.

Resulting fits are shown in Table 5.5 for the posterior mean volume estimates from the baseline model, and in Table 5.6 for the posterior mean volume estimates from prior volume inflated model. The fitted density functions are shown in Figure 5.18. Although the mixture of Weibull distributions appears to provide a better fit to the baseline model volume estimates (smaller AIC), re-scaling the volumes to ensure positive logarithmic values means that the parameters are not easily interpretable. For this reason the mixture of normal distributions may be preferred. In fact, when fitted to the posterior volume estimates from the volume inflated model the mixture of normal distributions gives a superior fit.

Statistical approaches for forecasting the size of future events has essentially been limited to generalized size-predictable models (Sandri et al., 2005; Marzocchi and Zaccarelli, 2006) (see Section 2.3.6). Turner et al. (2011a) found a size-predictable relationship for the widespread named Taranaki event. They found that, on the scale of large events, the longer the interval between eruptions the smaller the subsequent eruption volume is likely to be. Interestingly, the opposite appears to be true for the smaller events. Figure 5.19 suggests the volume of smaller unnamed events has remained relatively constant through time. However, Figure 5.20

Table 5.5: Baseline model estimates.

Model	Mixture of Weibulls	Mixture of Lognormals	Mixture of Normals
Parameters	$p = 0.919 \pm 0.017$	$p = 0.919 \pm 0.017$	$p = 0.919 \pm 0.017$
	$k_1 = 18.079 \pm 0.831$	$\mu_1 = 1.347 \pm 0.004$	$\mu_1 = -3.055 \pm 0.016$
	$\delta_1 = 0.252 \pm 0.001$	$\sigma_1 = 0.065 \pm 0.003$	$\sigma_1 = -0.245 \pm 0.011$
	$k_2 = 22.266 \pm 3.541$	$\mu_2 = 1.920 \pm 0.013$	$\mu_2 = -0.077 \pm 0.085$
	$\delta_2 = 0.143 \pm 0.001$	$\sigma_2 = 0.061 \pm 0.009$	$\sigma_2 = -0.399 \pm 0.060$
$\log L$	-86.797	-94.650	-90.343
AIC	**183.594**	199.299	190.685

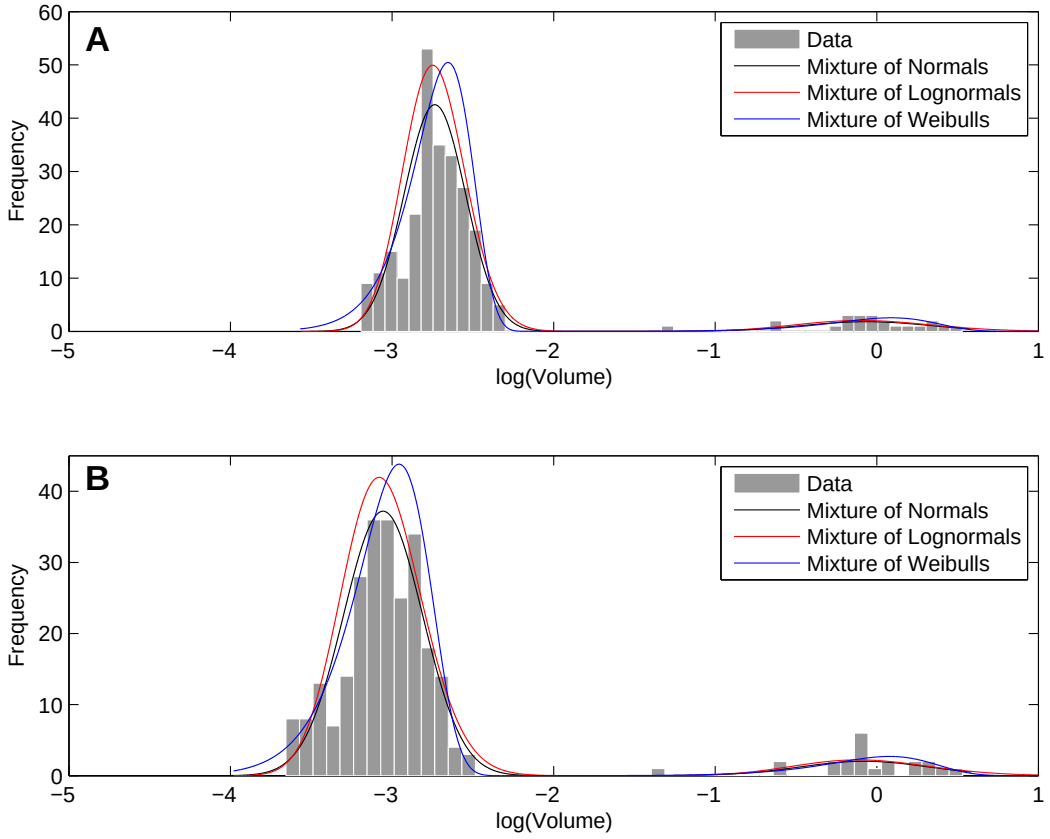
Table 5.6: Prior volume inflated model estimates.

Model	Mixture of Weibulls	Mixture of Lognormals	Mixture of Normals
Parameters	$p = 0.919 \pm 0.017$	$p = 0.919 \pm 0.017$	$p = 0.919 \pm 0.167$
	$k_1 = 24.352 \pm 1.096$	$\mu_1 = 1.428 \pm 0.003$	$\mu_1 = -2.735 \pm 0.012$
	$\delta_1 = 0.235 \pm 0.001$	$\sigma_1 = 0.046 \pm 0.001$	$\sigma_1 = -0.190 \pm 0.009$
	$k_2 = 22.824 \pm 3.651$	$\mu_2 = 1.922 \pm 0.013$	$\mu_2 = -0.063 \pm 0.085$
	$\delta_2 = 0.143 \pm 0.001$	$\sigma_2 = 0.061 \pm 0.009$	$\sigma_2 = -0.396 \pm 0.060$
$\log L$	-28.875	-30.353	-27.169
AIC	67.751	70.707	**65.238**

suggests a tendency for longer intervals between events (repose times) to be terminated by larger eruptive volumes. The slope of a linear regression of the logarithm of eruptive volume ($\log(v_i)$) versus the logarithm of time since the last event ($\log(\tau_i) = \log(t_i - t_{i-1})$), is significantly different to zero (P-value < 0.001). This evidence is useful in a hazard forecasting framework and provides motivation for further exploration into size-predictability of Mt Taranaki eruptions. These findings suggest that Mt Taranaki may in fact, have two mechanisms governing magma recharge.

In Chapter 3, it was found that the geochemistry of the previous eruption could be used to explain the distribution of the *time* between Mt Taranaki eruptions. The obvious extension arising from this work, would be to explore whether variations in geochemistry can also be tied to eruptive volumes. If so, we must consider whether the same combination of geochemical precursors can be used to explain both eruptive volumes and inter-event times.

Figure 5.18: Density functions for a range of distributions fitted to the logarithm of the posterior mean volume estimates. (A) Baseline model, (B) Volume prior inflated model.



5.7.2 Data and model limitations

In carrying out this investigation there were limitations due to the amount of thickness data available for the smaller named events. Eruptive volumes were estimated for a large number of events (270) where, in some cases, a particular event may only have a single thickness measurement from one location. This has an influence on the estimated wind direction during the time of eruption. Dispersal in the direction of the sampling location is implied. Ideally, a large number of unexposed records at various angles and distances from the volcano would be available, to provide more reliable volume estimates. If new records become available, they can easily feed into the model. This may provide better insights into the dispersal direction for Mt Taranaki events.

Understanding the dispersal and size of small events is challenging due to the scarcity of data, and complexities involved in the physical processes that govern tephra dispersal. Bayesian statistical theory provides an ideal framework for estimating eruptive volumes. Figure 5.21 demonstrates the power of this method. Reliable volume estimates are not achievable via a

Figure 5.19: Changes in posterior mean volume (km^3) over time for all unnamed events. Events older than 38000 cal BP are excluded as they were only observed in the Auckland record.

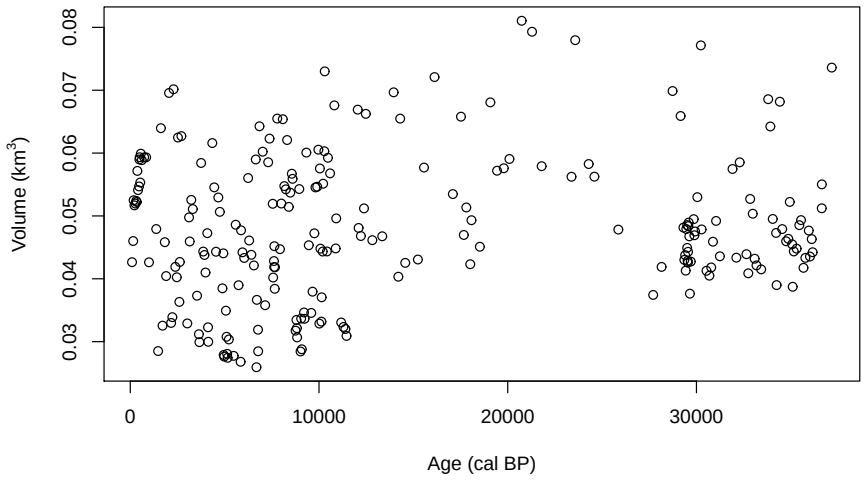
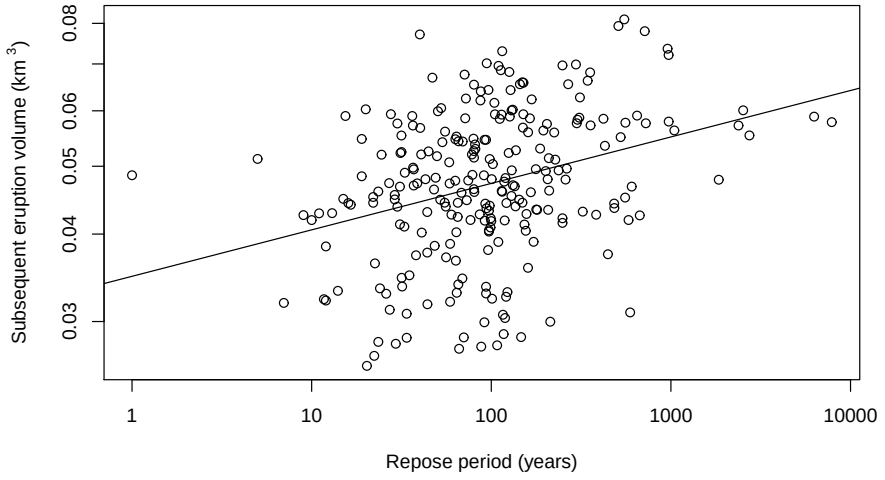


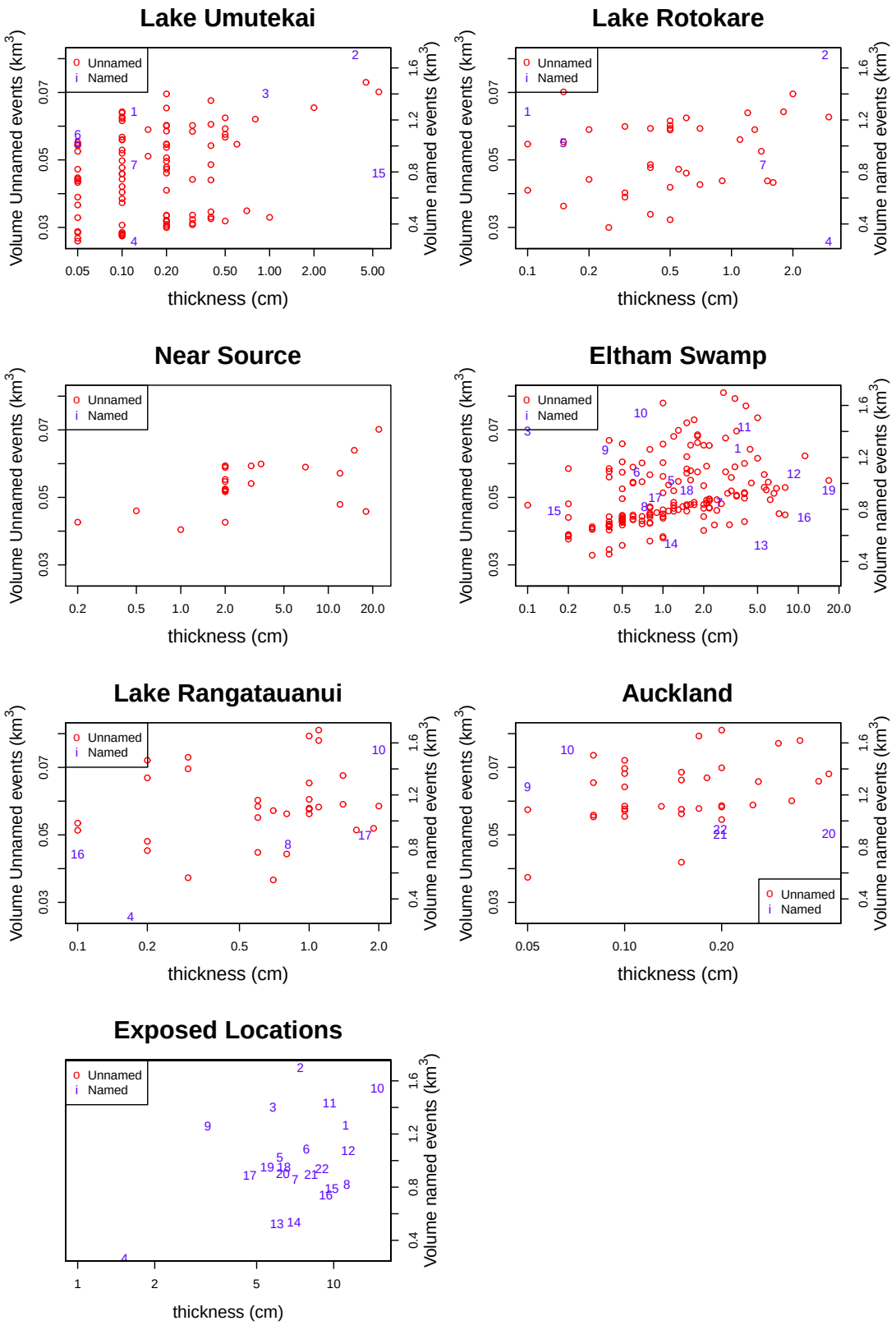
Figure 5.20: Repose times and subsequent eruption volumes (on a log scale) for unnamed events. Best fit regression line added.



simple analysis of the thickness data alone. A parametric physical model approach must be taken to incorporate other factors that influence dispersal, such as wind conditions.

Although the model has some undesirable inferential properties, due to difficulties in separating model parameters, it does provide a tractable framework for estimating volumes. It is unknown whether an improved separable model could be formulated that has the right physical properties. Gonzalez-Mellado and De la Cruz-Reyna (2010) suggest an alternative physical model for tephra dispersal (see Section 2.4.5). However, their model was also found to have problems with separating wind velocity and rate of thickness decay (Kawabata et al., 2013). Additionally, the model of Gonzalez-Mellado and De la Cruz-Reyna (2010) does not include volume as a model parameter.

Figure 5.21: Observed tephra thicknesses versus posterior mean volumes. Thickness are represented on a log scale. The bottom left plot shows mean tephra thicknesses observed across all exposed locations.



Further developments, seeking more informative prior information for the site-specific effects or wind directions, may provide better constraints for the eruption volumes. Using volume estimates from analog volcanoes (those exhibiting similar eruption characteristics), to construct prior distributions for Mt Taranaki eruption volumes could also be a possibility. But that may present challenges due to differing wind conditions and a lack of other detailed available records.

5.8 Conclusions

In this chapter I have presented a Bayesian approach to estimating the eruption volumes for a series of small and large eruptions from Mt Taranaki. Point thickness observations are modeled as a function of the distance and angular direction of each location. Isopach maps are used to construct prior distributions for model parameters. The model is an adaptation of the attenuation model of Rhoades et al. (2002). Rather than aggregating multiple events, eruptions are modeled individually. Eruptive volumes are estimated, in addition to the wind and site-specific effects on the thickness deposits.

Most studies focus on large widespread eruptions using isopach maps constructed from observations at *exposed* locations. Instead in this chapter a unique approach is taken, raw thickness measurements are incorporated from additional *unexposed* lake and swamp records. Unexposed sites can preserve very fine tephra falls, which facilitates investigation into the dispersal pattern and eruption volume of much smaller events. Until now, the size of these smaller events has not been widely understood or quantified.

Resulting estimates of eruption volume provide a comprehensive record suitable for supporting hazard models. Preliminary examination suggests a size-predictable relationship for the smaller events. The findings of this work also motivate exploration into other relationships, for example between geochemical precursory data and eruption volume.

Additional Mt Taranaki event records can be easily accommodated into this analysis, should they become available. The Bayesian framework provides flexibility. Prior distributions for model parameters can be adjusted, as knowledge of the physical processes that govern tephra dispersal improve. Of course, application is not limited to Mt Taranaki event records. Model details could be adjusted to suit records for other polygenetic volcanoes.

Chapter 6

Conclusions and future research directions

This thesis focuses on novel approaches to analysing volcanological data. I have developed and applied statistical methodologies to aid in volcanic hazard estimation. In doing so, many interesting and challenging statistical problems were encountered. This chapter concludes the thesis by way of a review of the work completed, a discussion of the key advances made in statistical volcanology, and suggestions for future research.

6.1 Conclusions

There were three main objectives I set out to achieve in this thesis. As identified in Chapter 1, the objectives were to:

- Formulate a probabilistic temporal hazard model for Mt Taranaki, incorporating ancillary geochemical data, with the hope of improving forecasts of future Mt Taranaki eruptions.
- Create an automated statistical procedure, through application to a series of long sediment core records, to match tephra sequences observed at multiple deposition sites.
- Develop a statistical model, informed by the physical processes that govern tephra dispersal, to estimate the eruptive volume of Mt Taranaki events.

In light of these objectives, I provide a review of the efforts made to address them, and a discussion of the most interesting findings.

Chapter 3 was devoted to achieving the first objective. Mt Taranaki has an activity record punctuated by long periods of quiescence and subsequent re-awakening. Thus, the distribution of inter-onset times is bimodal with the possibility of anomalously long reposes. Prior to this research, it had been hypothesized that cycles in magma geochemistry could be linked to eruption rate (Turner et al., 2008b). I demonstrated, through application to two time-homogeneous event records of Mt Taranaki (referred to as the 'single-source' and 'multi-source' record), how magma geochemistry could be incorporated into a temporal hazard model to improve eruption forecasts. A series of different renewal and proportional hazard models were considered. The preferred model (a proportional hazards model with a Weibull baseline hazard function) is described by the hazard function

$$\lambda(\tau, z) = \alpha\beta^{\alpha}(\beta\tau)^{\alpha-1}\exp(\delta_1 z_1 + \delta_4 z_4 + \delta_5 z_5), \tag{6.1}$$

where τ refers to the repose length, or time between eruptions, and the geochemical data is incorporated through the covariates z_1 =(Mean(TiO_2), z_4 = SD(TiO_2), and z_5 = SD(Al_2O_3). Eruptions high in Mean(TiO_2) or SD(Al_2O_3) were found to decrease the subsequent hazard function, thus lengthening the next repose period. On the other hand high values of SD(TiO_2) point to a higher subsequent hazard, and hence a shorter repose. This model and combination of covariates provided a suitable fit for both event records considered. The model outperformed previous renewal models, particularly in the forecasting of long reposes.

Prior to the model development in Chapter 3, the most suitable model for Mt Taranaki eruptions was a bimodal (mixture of Weibull distributions) renewal model (Turner et al., 2008a). In that model, longer reposes are assumed to occur at random. Therefore, the estimated hazard function is only updated to reflect the likelihood of a long repose after the present repose-length exceeds the shorter mode. The proportional hazards model proposed is able to modulate the renewal model such that the mixture distribution for the bimodal inter-event times is no longer needed. Therefore in addition to improving goodness-of-fit, the proportional hazards approach greatly improves the ability to forecast long reposes. As the geochemical data is available at the beginning of the repose to be forecast, one no longer needs to wait to see if they pass through the first mode to know they are in a long repose. Rather than having to wait many years, the predictive information is available *now*.

In Chapter 4 I proposed a novel approach to correlate tephra records from several locations. Prior to this, there was no standard means of obtaining and evaluating the most reasonable arrangement of tephra across multiple deposition sites, particularly for large numbers of tephra. I developed an automated matching algorithm using stochastic local search techniques. The age and associated age error of each event forms the basis of the matching procedure. The algorithm ensures that the stratigraphy of events is maintained when finding the statistically most likely amalgamated event record.

Initially the matching algorithm was intended for application to Mt Taranaki event records. However the decision was made to develop the automated procedure on higher profile data from the Auckland Volcanic Field, to stimulate a wider interest in the methods of the thesis. Although the algorithm was implemented on Mt Taranaki event records (in Chapter 5), some interesting findings also emerged from the Auckland Volcanic Field case study.

When compared with a previously published arrangement from the Auckland area (Molloy et al., 2009) the record compiled using the matching algorithm is statistically far more likely. The new arrangement alters matches among the AVF-sourced tephras across the cores. It also reduces the hazard to the Auckland region from the distal center of Mt Taranaki by approximately 23%. These findings have implications for hazard assessment, suggesting that the risk from distal volcanic centers may not be as great as that suggested by Molloy et al. (2009).

Using the amalgamated Auckland record, I formulated a likelihood model to quantify the variability in observance probabilities at particular locations. It was found that the preservation of tephras at Orakei Basin is unusual relative to the other sites. The estimated thicknesses of tephras observed in Orakei Basin are much greater in comparison to the other locations, with some exceptions that may be related to erosion (when stream flows into and through the basin are high).

The automated matching procedure is particularly attractive from the standpoint that it is easy to implement, using the code provided, and can be easily applied to a wide range of volcanological datasets. Users can eliminate (or force) certain matches by specifying a list of implausible matches (or matches to force). The procedure is flexible in its ability to incorporate auxiliary information, such as the geochemistry or petrology (origin, structure, and composition) of tephras, to constrain the arrangements.

The final objective of estimating eruptive volumes was, from a statistical perspective, the most challenging. It required estimating eruptive volumes from *very* sparse thickness measurements for tephra deposits located at various distances and directions from the vent. Tephra dispersal depends on (but is not limited to) the height of the eruption plume and the wind speed, strength, and direction during an eruption. Tephra deposited at different locations can be preserved under different environmental conditions. Therefore site-specific effects for over- or under-thickening are also present. Given the general scarcity of data, and the physical phenomena governing tephra attenuation, a high-dimensional complex model was required. Although dispersal from large Mt Taranaki eruptions has been well documented (Alloway et al., 1995), before this thesis very little was known about the eruption volume and tephra dispersal of smaller events. Using isopach maps and individual point thickness observations, I formulated a model (in Chapter 5) in a Bayesian framework, to estimate eruptive volumes, wind, and site-specific effects for a large number of Mt Taranaki eruptions.

The challenge was finding the correct functional form for the semi-elliptical wind terms, such that resulting dispersal shapes are consistent with geologists interpretations of observed thicknesses, yet still mathematically tractable. Although the model has some undesirable properties, due to difficulties in separating model parameters, it does provide an implementable framework for estimating volumes. I have created a means of extracting plausible volume estimates from really sparse data. As the knowledge of the physical processes that govern tephra dispersal improve, prior distributions for model parameters can be updated, and the variance in the estimated volumes hopefully reduced. Additional Mt Taranaki records can also be incorporated as and when they become available. I have demonstrated a paradigm that opens up new vistas for hazard estimation at Mt Taranaki. Knowledge of the size of previous events, particularly the smaller events, enables models for forecasting the size of future events to be investigated.

There is a shortage of long volume records on which to develop models for volume forecasts. Not only have I provided one, for Mt Taranaki, but the methods in Chapter 5 can also be applied elsewhere to dramatically increase the length and number of known records.

6.2 Suggestions for future research

Quantitative forecasting of volcanic hazard has long been a key aim in volcanology. Statistics plays an important part in developing models to better understand the history of volcanic

processes as well as providing insight into the likelihood of future events. In this thesis I have established and implemented statistical models to aid in volcanic hazard estimation, however I recognize that this work is only a step towards the greater goal of better understanding eruptions and eruption histories. The outcomes of this research give rise to new challenges for volcanic hazard estimation. There are a number of possible extensions and further analyses which are not covered in this thesis, but would be worthwhile investigating. The following is by no means an exhaustive list, but it provides some suggestions for future avenues of research.

The field of volcanology is ever advancing and more and more data is becoming available to support hazard models. Work is being currently being done by other students in the Volcanic Risk Solutions group at Massey University, to assemble tephra records from additional locations around Mt Taranaki. In particular, additional sediment cores have recently been extracted, and are currently being examined, from four new sites in northern and eastern Taranaki. With the acquisition of this new data, the correlated record of Mt Taranaki events established in Chapter 5 (Section 5.2) could be updated. Using the algorithm developed in Chapter 4 a more detailed history of events could be compiled.

In Chapter 5 volume estimates were obtained for a number of Mt Taranaki eruptions. Many of the events, particularly smaller and/or older events, only left deposits in one sampling location. The addition of the extra records may improve estimates of eruptive volumes as well as providing better constraints over the wind directions through time. Some of the new sites are located in the north-east quadrant from the vent. This region previously lacked lake or swamp records to record small events. Therefore, the inclusion of the new sites could go a long way towards understanding the multi-lobed dispersal of events. There are also implications for climatology. More data provides a window into changes in wind strength and direction over time. However, there are identifiability problems surrounding the last glacial maximum; frozen lakes are unable to record volcanic eruptions. Large earthquakes or landslides can also change the lake characteristics, through damming or diversion of inflow.

Given the volume estimates obtained in Chapter 5, the obvious extension to this research is to investigate whether the volume estimates can be used to establish hazard forecasts for future Mt Taranaki events. This could involve models such as the Generalized Time Predictable model of Marzocchi and Zaccarelli (2006) (see Section 2.3.3) or the volume-history dependent model of Bebbington (2008). However, the problem is that the records in Chapter 5 suffer from uneven time coverage. The Eltham swamp core extends much further

than the remaining unexposed records. As a result, the eruption occurrences are certainly not homogeneous in time. This was one of the problems encountered with the records of Chapter 3 leading to the analysis of separate (time homogeneous) 'single-source' and 'multi-source' records. If the renewal models in Chapter 3 are to be revisited (using the additional lake and swamp records from Chapter 5) we might consider an approach such as that proposed by Wang and Bebbington (2012) to correct the repose periods for time inhomogeneity. The development of proportional hazards models corrected for inhomogeneous records would be a new development.

In Chapter 3 missing geochemical data was imputed to be that of the previous eruption. This was justified by the fact that the eruptions with missing geochemistry are almost certainly small, and thus are unlikely to result from a major change in the magmatic or volcanic system. An alternative approach, albeit challenging, would be to consider creating a model for the evolution of geochemistry through time based on the findings of Turner et al. (2011a). This could involve time-series techniques or state space modeling, as has been suggested for earthquakes by Takanami et al. (2013). Incorporating changing geochemistry into hazard forecasting would require a 'see-saw' approach. The geochemistry of an eruption would be estimated through a 'geochemistry evolution' model and subsequently fed into the proportional hazards framework in Chapter 3 to create a temporal forecast. As a new eruption is predicted, the geochemistry changes. This necessitates a return to the 'geochemistry evolution' model. The modeling and incorporating of evolving geochemistry into hazard estimation would be a new development in volcanology and a potentially fruitful area of research. We would need to address the question of whether static or evolving geochemistry best describes eruption occurrences.

Similar to Marzocchi and Bebbington (2012) I think the most important future research direction lies with forecasting the size of impending eruptions. In Section 5.7.1 I gave a brief investigation into size-predictability of Mt Taranaki events, but it is worthy of further investigation. The analysis in Chapter 3 found that geochemical data could be used to modulate a renewal process for the *timing* of future events. It would be interesting to explore whether geochemistry can also be used to explain eruption size/volume. If so, the question follows as to whether the same geochemical precursors (namely mean(TiO_2), SD(TiO_2), and SD(Al_2O_3)) can be used to explain both eruption volume and timing, or whether different precursors are required. In a wider sense the methodology in Chapter 3 is also applicable to other sorts of ancillary data. Geochemistry is useful in elucidating eruption timings,

but alternative ancillary data, such as variations in the grain size of tephras, may be more informative for eruption volumes. Chapter 3 provides a paradigm for incorporating extra covariate information.

Volcanic eruptions are inevitable and unfortunately there is no way of controlling them now, or in the foreseeable future. However, the risk to society can be reduced by preparedness. In addition to immediate crisis responses, mitigation procedures involve long-term land-use and development planning. Some indication of when an imminent eruption is likely to occur is useful. However, identification of 'impact zones' is also of particular importance. This is usually done through the construction of hazard maps, which document the areal distribution of volcanic hazards, typically distinguishing low to high risk zones. Although not covered in this thesis, new methods of hazard 'mapping' are an important research direction; the work in Chapter 5 provides a first step towards forecasting the size and/or impact of Mt Taranaki events. Knowledge of the size/volume and dispersal of previous events, particularly the smaller events which still pose a threat to communities and infrastructure, can provide insight into future occurrences and enable risk scenarios to be established.

This thesis has focused largely on the timing and volume of Mt Taranaki eruptions, but the models developed are also applicable to other re-awakening volcanoes. Mt Merapi in Indonesia is known as Taranaki's sister volcano, due to its similar size, shape, and composition. Unlike Mt Taranaki, which is currently in a long period of quiescence, Mt Merapi is in an active phase. It would be interesting to see whether the geochemistry model developed in Chapter 3 is useful for elucidating eruption hazard at Mt Merapi, and in these types of volcanic systems in general, or whether the form of the relationship with magma geochemistry is specific to Taranaki. These ideas may however be hindered by the lack of geochemical data to support the analysis. Although suitable geochemical data is not currently available, it is hoped that it will become accessible in the near future.

As a final remark we recall the words of physicist Neils Bohr,

> "It's hard to make predictions, especially about the future."

Forecasting volcanic eruptions is by no means an exception to this, but it is hoped that the methodologies developed and findings in this thesis will go some way towards understanding volcanic eruptions and their potential consequences.

Bibliography

Akaike, H. (1977). On entropy maximization principle. In Krishnaiah, P., editor, *Applications of Statistics*, pages 27–41. North-Holland, Amsterdam, The Netherlands.

Alexander, D. (2013). Volcanic ash in the atmosphere and risks for civil aviation: a study in European crisis management. *Journal of International Disaster Risk Science*, 4(1):9–19.

Allen, S. and Smith, I. (1994). Eruption styles and volcanic hazard in the Auckland Volcanic Field, New Zealand. *Geoscience Reports of Shizuoka University*, 20:5–14.

Alloway, B. (1989). The late Quaternary cover bed stratigraphy and tephrochronology of north-eastern and central Taranaki, New Zealand. Unpublished PhD thesis, Massey University.

Alloway, B., Neall, V., and Vucetich, C. (1995). Late Quaternary (post 28,000 year B.P.) tephrostratigraphy of northeast and central Taranaki, New Zealand. *Journal of the Royal Society of New Zealand*, 25(4):385–458.

Bacon, C. (1982). Time-predictable bimodal volcanism in the Coso range, California. *Geology*, 10:65–69.

Bain, L. and Engelhardt, M. (1986). Approximate distributional results based on the maximum likelihood estimators for the Weibull distribution. *Journal of Quality Control*, 23:304–311.

Bebbington, M. (2007). Identifying volcanic regimes using hidden Markov models. *Geophysical Journal International*, 171:921–942.

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

Bebbington, M. (2009). *Volcanic eruptions: Stochastic models of occurrence patterns*, volume 9. Encyclopedia of Complexity and System Science, Springer, New York. Editor B. Meyers.

Bebbington, M. (2010). Trends and clustering in the onsets of volcanic eruptions. *Journal of Geophysical Research*, 115:B01203.

Bebbington, M. (2012). Models for temporal volcanic hazard. *Statistics in volcanology*, 1:1–24.

Bebbington, M. (2013). Assessing spatio-temporal eruption forecasts in a monogenetic volcanic field. *Journal of Volcanology and Geothermal Research*, 252:14–28.

Bebbington, M. (2014). Long-term forecasting of volcanic explosivity. *Geophysical Journal International*, 197:1500–1515.

Bebbington, M. and Cronin, S. (2011). Spatio-temporal hazard estimation in the Auckland Volcanic Field, New Zealand, with a new event-order model. *Bulletin of Volcanology*, 73:55–72.

Bebbington, M., Cronin, S., Chapman, I., and Turner, M. (2008). Quantifying volcanic ash fall hazard to electricity infrastructure. *Journal of Volcanology and Geothermal Research*, 177:1055–1062.

Bebbington, M. and Lai, C.-D. (1996a). On nonhomogeneous models for volcanic eruptions. *Journal of Mathematical Geology*, 28:585–600.

Bebbington, M. and Lai, C.-D. (1996b). Statistical analysis of New Zealand volcanic occurrence data. *Journal of Volcanology and Geothermal Research*, 74:101–110.

Bebbington, M. and Marzocchi, W. (2011). Stochastic models for earthquake triggering of volcanic eruptions. *Journal of Geophysical Research*, 116:B05204.

Bonadonna, C. and Costa, A. (2012). Estimating the volume of tephra deposits: A new simple strategy. *Geology*, 48:415–418.

Bonadonna, C., Ernst, G., and Sparks, R. (1998). Thickness variations and volume estimates of tephra fall deposits: the importance of particle Reynolds number. *Journal of Volcanology and Geothermal Research*, 81(3–4):173–187.

Bonadonna, C. and Houghton, B. (2005). Total grain-size distribution and volume of tephra-fall deposits. *Bulletin of Volcanology*, 67:441–456.

Brenna, M. (2012). Geological evolution and magmatic models for spatially and temporally variable modes of distributed volcanism, Jeju Island, Republic of Korea. Unpublished PhD thesis, Massey University.

Brown, F. and McDougall, I. (2011). Geochronology of the Turkana depression of northern Kenya and southern Ethiopia. *Evolutionary Anthropology*, 20(6):217–227.

Buddington, A. and Lindsley, D. (1964). Iron-titanium oxide minerals and synthetic equivalents. *Journal of Petrology*, 5:310–357.

Burden, R., Chen, L., and Phillips, J. (2013). A statistical method for determining the volume of volcanic fall deposits. *Bulletin of Volcanology*, 75(6):1–10.

Bursik, M. and Rogova, G. (2006). Use of neural networks and decision fusion for lithostratigraphic correlation with sparse data, Mono-Inyo Craters, California. *Computers and Geosciences*, 32:1564–1572.

Burt, M., Wadge, G., and Scott, W. (1994). Simple stochastic modelling of the eruption history of a basaltic volcano: Nyamuragira, Zaire. *Bulletin of Volcanology*, 56:87–97.

Černý, V. (1985). Thermodynamical approach to the traveling salesman problem: an efficient simulation algorithm. *Journal of Optimization Theory and Application*, 45(1):41–51.

Charman, D. and Grattan, J. (1999). An assessment of discriminant function analysis in the identification and correlation of distal Icelandic tephras in the British Isles. *Geological Society (London) Special Publications*, 161:147–160.

Claeskens, G. and Hjort, N. (2008). *Model Selection and Model Averaging*. Cambridge University Press, Cambridge.

Collett, D. (2003). *Modelling Survival Data in Medical Research*. Chapman & Hall, 2nd edition.

Connor, C., Bebbington, M., and Marzocchi, W. (2015). Probabilistic volcanic hazard assessment. In Sigurdsson H, et al., editor, *Encyclopedia of volcanoes*, chapter 51. 2nd edition.

Connor, C., Sparks, R., Mason, R., Bonadonna, C., and Young, S. (2003). Exploring the links between physical and probabilistic models: the Soufriere Hills volcano, Monserrat. *Geophysical Research Letters*, 30:1701.

Cox, D. (1972). Regression models and life tables (with discussion). *Journal of the Royal Statistical Society, Series B*, 34:187–220.

Cox, D. and Isham, V. (1980). *Point Processes*. Chapman and Hall, London.

Cronin, S., Neall, V., Palmer, A., and Stewart, R. (1997). Methods of identifying late Quaternary tephras on the ring plains of Ruapehu and Tongariro volcanoes, New Zealand. *New Zealand Journal of Geology and Geophysics*, 40:175–184.

Cronin, S., Wallace, R., and Neall, V. (1996). Sourcing and identifying andesitic tephras using major oxide titanomagnetite and hornblende chemistry, Egmont volcano and Tongariro Volcanic Centre, New Zealand. *Bulletin of Volcanology*, 58(1):33–40.

Daggitt, M., Mather, T., Pyle, D., and Page, S. (2014). AshCalc - a new tool for the comparison of the exponential, power-law and Weibull models of tephra deposition. *Journal of Applied Volcanology*, 3:7:1–8.

Danišík, M., Shane, P., Schmitt, A., Hogg, A., Santos, G., Storm, S., Evans, N., Fifield, K., and Lindsay, J. (2012). Re-anchoring the late Pleistocene tephrochronology of New Zealand based on concordant radiocarbon ages and combined $^{238}U/^{230}Th$ disequilibrium and (U-Th)/He zircon ages. *Earth and Planetary Science Letters*, 349–350:240–250.

De la Cruz-Reyna, S. (1991). Poisson-distributed patterns of explosive eruptive activity. *Bulletin of Volcanology*, 54:57–67.

De la Cruz-Reyna, S. and Carrasco-Nunez, G. (2002). Probabilistic hazard analysis of Citlaltépetl (Pico de Orizaba) Volcano, eastern Mexican Volcanic Belt. *Journal of Volcanology and Geothermal Research*, 113:307–318.

Decker, R. (1986). Forecasting volcanic eruptions. *Annual Review of Earth and Planetary Sciences*, 14:267–291.

Devine, J., Rutherford, M., Norton, G., and Young, S. (2003). Magma storage region processes inferred from geochemistry of Fe-Ti oxides in andesitic magma, Soufriere Hills Volcano, Montserrat, WI. *Journal of Petrology*, 44:1375–1400.

Engwell, S., Sparks, R., and Aspinall, W. (2013). Quantifying uncertainties in the measurement of tephra fall thicknesses. *Journal of Applied Volcanology*, 2(5):1–12.

Faenza, L., Marzocchi, W., and Boschi, E. (2003). A non-parametric hazard model to characterize the spatio-temporal occurrence of large earthquakes; an application to the Italian catalogue. *Geophysical Journal International*, 155:521–531.

Fierstein, J. and Hildreth, W. (1992). The plinan eruption of 1912 at Novarupta, Katmai National Park, Alaska. *Bulletin of Volcanology*, 54:646–684.

Fierstein, J. and Nathenson, M. (1992). Another look at the calculation of fallout tephra volumes. *Bulletin of Volcanology*, 54(2):156–167.

Franks, A. (1984). Soils of Eltham County and the tephrochronology of Central Taranaki. Unpublished PhD thesis, Massey University.

Froggatt, P. and Lowe, D. (1990). A review of late Quaternary silicic and some other tephra formulations from New Zealand: their stratigraphy, nomenclature, distribution, volume, and age. *New Zealand Journal of Geology and Geophysics*, 33:89–109.

Garcia-Aristizabal, A., Marzocchi, W., and Fujita, E. (2012). A Brownian model for recurrent volcanic eruptions: an application to Miyakejima Volcano (Japan). *Bulletin of Volcanology*, 74:545–558.

Gelman, A. and Rubin, D. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7(4):457–472.

Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:721–741.

Glover, F. (1989). Tabu Search - part I. *ORSA Journal on Computing*, 1(3):190–206.

Glover, F. (1990). Tabu Search - part II. *ORSA Journal on Computing*, 2(1):4–32.

Gonzalez-Mellado, A. and De la Cruz-Reyna, S. (2010). A simple semi-empirical approach to model thickness of ash-deposits for different eruption scenarios. *Natural Hazards and Earth System Sciences*, 10:2241–2257.

Good, I. and Gaskins, R. (1971). Nonparametric roughness penalties for probability densities. *Biometrika*, 58(2):255–277.

Green, R., Bebbington, M., Cronin, S., and Jones, G. (2013). Geochemical precursors for eruption repose length. *Geophysical Journal International*, 193:855–873.

Green, R., Bebbington, M., Cronin, S., and Jones, G. (2014). Automated statistical matching of multiple tephra records exemplified using five long maar sequences younger than 75 ka, Auckland, New Zealand. *Quaternary Research*, 82(2):405–419.

Hajek, B. (1988). Cooling schedules for optimal annealing. *Mathematics of Operations Research*, 13(2):311–329.

Hill, B., Connor, C., Jarzemba, M., La Femina, P., Navarro, M., and Strauch, W. (1998). 1995 eruptions of Cerro Negro volcano, Nicaragua, and risk assessment for future eruptions. *Geological Society of America Bulletin*, 110:1231–1241.

Ho, C.-H. (1991). Nonhomogeneous Poisson model for volcanic eruptions. *Journal of Mathematical Geology*, 23:167–173.

Hogg, A., Hua, Q., Blackwell, P., Niu, M., Buck, C., Guilderson, T., Heaton, T., Palmer, J., Reimer, P., Reimer, R., Turney, C., and Zimmerman, S. (2013). SHCal13 Southern Hemisphere Calibration, 0-50,000 years cal BP. *Radiocarbon*, 55:1–15.

Hoos, H. and Stützle, T. (2005). *Stochastic Local Search: Foundations and Applications*. The Morgan Kaufmann Series in Artificial Intelligence Series. Morgan Kaufmann Publishers, San Francisco.

Hurvich, C. and Tsai, C. (1989). Regression and time series model selection in small samples. *Biometrika*, 76:297–307.

Johnston, D., Houghton, B., Neall, V., Ronan, K., and Paton, D. (2000). Impacts of the 1945 and 1995-1996 Ruapehu eruptions, New Zealand: An example of increasing social vulnerability. *The Geological Society Bulletin*, 112(5):720–726.

Kagan, Y. and Knopoff, L. (1977). Earthquake risk prediction as a stochastic process. *Physics of the Earth and Planetary Interiors*, 14(2):97–110.

Kalbfleisch, J. and Prentice, R. (1980). *The Statistical Analysis of Failure Time Data*. Wiley, New York.

Kawabata, E., Bebbington, M., Cronin, S., and Wang, T. (2013). Modeling thickness variability in tephra deposition. *Bulletin of Volcanology*, 75:738–751.

Kereszturi, G., Németh, K., Cronin, S., Agustin Flores, J., Smith, I., and Lindsay, J. (2013). A model for calculating eruptive volumes for monogenetic volcanoes - implication for the Quaternary Auckland Volcanic Field, New Zealand. *Journal of Volcanology and Geothermal Research*, 266:16–33.

Kermode, L. (1992). Geology of the Auckland urban area. Scale 1:50 000. Institute of Geological and Nuclear Sciences Geological Map 2. Technical report, GNS Science, Lower Hutt, New Zealand.

Kirkpatrick, S., Gelatt, C., and Vecchi, M. (1983). Optimization by simulated annealing. *Science*, 220(4598):671–680.

Klawonn, M., Houghton, B., Swanson, D., Fagents, S., Wessel, P., and Wolfe, C. (2014a). Constraining explosive volcanism: subjective choices during estimates of eruption magnitude. *Bulletin of Volcanology*, 76(2):793–798.

Klawonn, M., Houghton, B., Swanson, D., Fagents, S., Wessel, P., and Wolfe, C. (2014b). From field data to volumes: constraining uncertainties in pyroclastic eruption parameters. *Bulletin of Volcanology*, 76(7):839–854.

Klein, F. (1982). Patterns of historical eruptions at Hawaiian volcanoes. *Journal of Volcanology and Geothermal Research*, 12:1–35.

Latter, J. (1990). Some real and imaginary examples of volcanic hazard and risk assessment. *Pacific Rim Congress 90, Gold Coast, Australia, 6-12 May 1990, Proceedings*, 3:699–711.

Le Corvec, N., Bebbington, M., Lindsay, J., and McGee, L. (2013). Age, distance and geochemical evolution within a monogenetic volcanic field: analyzing patterns in the Auckland Volcanic Field eruption sequence. *Geochemistry, Geophysics, Geosystems*, 14(9):3648–3665.

Le Maitre, R. (1984). A proposal by the IUGS Subcommisson on the Systematics of Igneous Rocks for a chemical classification of volcanic rocks based on the total alkali silica (TAS) diagram. *Australian Journal of Earth Sciences*, 31:243–55.

Legros, F. (2000). Minimum volume of a tephra fallout deposit estimated from a single isopach. *Journal of Volcanology and Geothermal Research*, 96:25–32.

Lindsay, J., Leonard, G., Smid, E., and Hayward, B. (2011). Age of the Auckland Volcanic Field: a review of existing data. *New Zealand Journal of Geology and Geophysics*, 54:379–401.

Lindsey, J. (2004). *Statistical analysis of stochastic processes in time*. Cambridge University Press, Cambridge.

Lowe, D. (2011). Tephrochronology and its application: A review. *Quaternary Geochronology*, 6:107–153.

Lowe, D., Blaauw, M., Hogg, A., and Newnham, R. (2013). Ages of 24 widespread tephras erupted since 30,000 years ago in New Zealand, with re-evaluation of the timing and palaeoclimatic implications of the Lateglacial cool episode recorded at Kaipo bog. *Quaternary Science Reviews*, 74:170–194.

Lowe, D., Shane, P., Alloway, B., and Newnham, R. (2008). Fingerprints and age models for widespread New Zealand tephra marker beds erupted since 30,000 years ago: a framework for NZ-INTIMATE. *Quaternary Science Reviews*, 27:95–126.

Lunn, D., Thomas, A., Best, N., and Spiegelhalter, D. (2000). WinBUGS – a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing*, 10:325–337.

Magill, C. and Blong, R. (2005a). Volcanic risk ranking for Auckland, New Zealand. I: Methodology and hazard investigation. *Bulletin of Volcanology*, 67:331–339.

Magill, C. and Blong, R. (2005b). Volcanic risk ranking for Auckland, New Zealand. II: Hazard consequences and risk calculation. *Bulletin of Volcanology*, 67:340–349.

Martin, D. and Rose, W. (1981). Behavioral patterns of Fuego volcano, Guatemala. *Journal of Geothermal Research*, 10:67–81.

Marzocchi, W. and Bebbington, M. (2012). Probabilistic eruption forecasting at short and long time scales. *Bulletin of Volcanology*, 74:1777–1805.

Marzocchi, W., Sandri, L., and Selva, J. (2008). BET_EF: A probabilistic tool for long- and short-term forecasting. *Bulletin of Volcanology*, 70:623–632.

Marzocchi, W. and Zaccarelli, L. (2006). A quantitative model for the time-size distribution of eruptions. *Journal of Geophysical Research*, 111:B04204.

McCormac, F., Hogg, A., Blackwell, P., Buck, C., Higham, T., and Reimer, P. (2004). SHCal04 Southern Hemisphere calibration, 0-11.0 cal kyr BP. *Radiocarbon*, 46:1087–1092.

Medina Martinez, F. (1983). Analysis of the eruptive history of the Volcan de Colima, Mexico (1560-1980). *Geofisica Internacional*, 22:157–178.

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

Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., and Teller, E. (1953). Equation of state calculations by fast computing machines. *Journal of Chemical Physics*, 21(6):1087–1092.

Miller, L. (1955). Table of percentage points of Kolmogorov Statistics. *Journal of the American Statistical Association*, 51(273):111–121.

Moebis, A. (2010). Understanding the Holocene explosive eruption record of the Tongariro Volcanic Centre, New Zealand. Unpublished PhD thesis, Massey University.

Molloy, C., Shane, P., and Augustinus, P. (2009). Eruption recurrence rates in a basaltic volcanic field based on tephra layers in maar sediments: implications for hazards in the Auckland Volcanic Field. *Geological Society of America Bulletin*, 121(11–12):1666–1677.

Muggeo, V. (2008). Segmented: an R package to fit regression models with broken-line relationships. *R News*, 8(1):20–25.

Mulargia, F., Gasperini, P., and Tinti, S. (1987). Identifying different regimes in eruptive activity: an application to Etna Volcano. *Journal of Volcanology and Geothermal Research*, 34:89–106.

Neall, V. (1972). Tephrochronology and tephrostratigraphy of western Taranaki (N108-109), New Zealand. *New Zealand Journal of Geology and Geophysics*, 15(4):507–557.

Németh, K., Cronin, S., Smith, I., and Agustin Flores, J. (2012). Amplified hazard of small-volume monogenetic eruptions due to environmental controls, Orakei Basin, Auckland Volcanic Field, New Zealand. *Bulletin of Volcanology*, 74:2121–2137.

Newhall, C. and Hoblitt, R. (2002). Constructing event trees for volcanic crises. *Bulletin of Volcanology*, 64:3–20.

Newhall, C. and Self, S. (1982). The Volcanic Explosivity Index (VEI): an estimate of explosive magnitude for historical volcanism. *Journal of Geophysical Research*, 87:1231–1238.

Newnham, R., Lowe, D., and Alloway, B. (1999). Volcanic hazards in Auckland, New Zealand: a preliminary assessment of the threat posed by central North Island silicic volcanism based on the Quaternary tephrostratigraphical record. *Geological Society (London) Special Publications*, 161:27–45.

Nilsson, A., Muscheler, R., Snowball, I., Aldahan, A., Possnert, G., Augustinus, P., Atkin, D., and Stephens, T. (2011). Multi-proxy identification of the Laschamp geomagnetic field excursion in Lake Pupuke, New Zealand. *Earth and Planetary Science Letters*, 311:155–164.

Ogata, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical Association*, 83:9–27.

Passarelli, L., Sandri, L., Bonazzi, A., and Marzocchi, W. (2010a). Bayesian Hierarchical Time Predictable Model for eruption occurrence: an application to Kilauea Volcano. *Geophysical Journal International*, 181:1525–1538.

Passarelli, L., Sanso, B., Sandri, L., and Marzocchi, W. (2010b). Testing forecasts of a new Bayesian time-predictable model of eruption occurrence. *Journal of Volcanology and Geothermal Research*, 198:57–75.

Platz, T., Cronin, S., Cashman, K., Stewart, R., and Smith, I. (2007a). Transitions from effusive to explosive phases in andesite eruptions - a case-study from the AD1655 eruption of Mt. Taranaki, New Zealand. *Journal of Volcanology and Geothermal Research*, 161:15–34.

Platz, T., Cronin, S., Procter, J., Neall, V., and Foley, S. (2012). Non-explosive, dome-forming eruptions at Mt. Taranaki, New Zealand. *Geomorphology*, 136:15–30.

Platz, T., Cronin, S., Smith, I., Turner, M., and Stewart, R. (2007b). Improving the reliability of microprobe-based analyses of andesitic glass for tephra correlation. *Holocene*, 17:573–583.

Plummer, M., Best, N., Cowles, K., and Vines, K. (2006). CODA: Convergence Diagnosis and Output Analysis for MCMC. *R News*, 6(1):7–11.

Pollard, A., Blockley, S., and Lane, C. (2006). Some numerical considerations in the geochemical analysis of distal microtephra. *Applied Geochemistry*, 21:1692–1714.

Press, W., Flannery, B., Teukolsky, S., Vetterling, W., and Chipperfield, J. (1986). *Numerical recipes: the art of scientific computing*. Cambridge University Press, Cambridge.

Pyle, D. (1989). The thickness, volume and grainsize of tephra fall deposits. *Bulletin of Volcanology*, 51(1):1–15.

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

R Core Team (2014). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Reimer, P., Bard, E., Bayliss, A., Beck, J., Blackwell, P., Bronk Ramsey, C., Buck, C., Cheng, H., Edwards, R., Friedrich, M., Grootes, P., Guilderson, T., Haflidason, H., Hajdas, I., Hatté, C., Heaton, T., Hoffman, D., Hogg, A., Hughen, K., Kaiser, K., Kromer, B., Manning, S., Niu, M., Reimer, R., Richards, D., Scott, E., Southon, J., Staff, R., Turney, C., and van der Plicht, J. (2013). IntCal13 and Marine13 radiocarbon age calibration curves, 0-50,000 years cal BP. *Radiocarbon*, 55(4):1869–1887.

Reyment, R. (1969). Statistical analysis of some volcanologic data regarded as series of point events. *Pure and Applied Geophysics*, 74(1):57–77.

Rhoades, D., Dowrick, D., and Wilson, C. (2002). Volcanic hazard in New Zealand: scaling and attenuation relations for tephra fall deposits from Taupo Volcano. *Natural Hazards*, 26:147–174.

Salvi, F., Scandone, R., and Palma, C. (2006). Statistical analysis of the historical activity of Mount Etna, aimed at the evaluation of volcanic hazard. *Journal of Volcanology and Geothermal Research*, 154:159–168.

Sandiford, A., Alloway, B., and Shane, P. (2001). A 28000-6600 cal yr record of local and distal volcanism preserved in a paleolake, Auckland, New Zealand. *New Zealand Journal of Geology and Geophysics*, 42(2):323–336.

Sandiford, A., Horrocks, M., Newnham, R., Ogden, J., and Alloway, B. (2002). Environmental change in the Last Glacial Maximum (ca 25000-16500 years BP) at Mt Richmond, Auckland Isthmus, New Zealand. *Journal of the Royal Society of New Zealand*, 32:155–167.

Sandiford, A., Newnham, R., Alloway, B., and Ogden, J. (2003). A 28000-7600 cal yr BP pollen record of vegetation and climate change from Pukaki Crater, northern New Zealand. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 201:235–247.

Sandri, L., Marzocchi, W., and Gasperini, P. (2005). Some insights on the occurrence of recent volcanic eruptions of Mount Etna volcano (Sicily, Italy). *Geophysical Journal International*, 163:1203–1218.

Santacroce, R. (1983). A general model for the behaviour of the Somma-Vesuvius volcanic complex. *Journal of Volcanology and Geothermal Research*, 17:237–248.

Sarna-Wojcicki, A., Shipley, S., Waitt, J., Dzurisin, D., and Wood, S. (1981). Areal distribution thickness, mass, volume, and grain-size of airfall ash from the size major eruptions of 1980. *The 1980 eruptions of Mount St. Helens, Washington: U.S. Geological Survey Professional Paper*, 1250:577–600.

Shane, P. (2005). Towards a comprehensive distal andesitic tephrostratigraphic framework for New Zealand based on eruptions from Egmont volcano. *Journal of Quaternary Science*, 20(1):45–57.

Shane, P. and Hoverd, J. (2002). Distal record of multi-sourced tephra in Onepoto Basin, Auckland, New Zealand: implications for volcanic chronology, frequency and hazards. *Bulletin of Volcanology*, 64:441–454.

Siebert, L. and Simkin, T. (2002-). Volcanoes of the World: an Illustrated Catalog of Holocene Volcanoes and their Eruptions, Smithsonian Institution, Global Volcanism Program Digital Information Series, GVP-4. (http://www.volcano.si.edu).

Simkin, T. (1994). Distant effects of volcanism - how big and how often? *Science*, 264(5161):913–914.

Small, C. and Naumann, T. (2001). The global distribution of human population and recent volcanism. *Environmental hazards*, 3:93–109.

Spiegelhalter, D., Best, N., Carlin, B., and van der Linde, A. (2002). Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society, Series B*, 64(4):583–639.

Stephens, T., Atkin, D., Augustinus, P., Shane, P., Lorrey, A., Street-Perrot, A., Nilsson, A., and Snowball, I. (2012). A late glacial Antarctic climate teleconnection and variable Holocene seasonality at Lake Pupuke, Auckland, New Zealand. *Journal of Paleolimnology*, 48:785–800.

Stokes, S. and Lowe, D. (1988). Discriminant function analysis of late Quaternary tephras from five volcanoes in New Zealand using glass shard major element chemistry. *Quaternary Research*, 30:270–283.

Stokes, S., Lowe, D., and Froggatt, P. (1992). Discriminant function analysis and correlation of late Quaternary rhyolitic tephra deposits from Taupo and Okataina Volcanoes, New Zealand, using glass shard major element composition. *Quaternary International*, 13–14:103–117.

Stuiver, M., Reimer, P., and Reimer, R. (2013). *CALIB 7.0 Radiocarbon Calibration*. CALIB 14C Calibration Program, Queen's University, Belfast.

Sumner, T. (2014). War zone volcano: An uphill battle to reveal Mount Nyiragongo's fiery past and forecast its future. *Science News*, 186(12):26–30.

Takanami, T., Linde, A., Sacks, S., Kitagawa, G., and Peng, H. (2013). Modeling of the post-seismic slip of the 2003 Tokachi-oki earthquake M8 off Hokkaido: Constraints from volumetric strain. *Earth Planets Space*, 56:731–738.

Tanguy, J.-C., LeGoff, M., Arrighi, S., Principe, C., LaDelfa, S., and Patané, G. (2009). The history of Italian volcanoes revised by archeomagnetism. *EOS, Transactions American Geophysical Union*, 90(40):349–350.

Tanguy, J.-C., Ribière, C., Scarth, A., and Tjetjep, W. (1998). Victims from volcanic eruptions: a revised database. *Bulletin of Volcanology*, 60:137–114.

Thorarinsson, S. (1944). Tefrokronologiska studier pa island. *Geografiska Annaler*, 26:1–217.

Thorarinsson, S. (1954). *The tephra-fall from Hekla on March 29th 1947*. The Eruption of Hekla, 1947-1948. H.F. Leiftur.

Thorarinsson, S. (1981). Tephra studies and tephrochronology: a historical review with special reference to Iceland. In Self, S. and Sparks, R., editors, *Tephra Studies*, pages 1–12. Reidel, Dordrecht.

Tinkler, R. (2013). A high resolution record of late Quaternary climatic and environmental change in Taranaki, New Zealand. Unpublished PhD thesis, Massey University.

Turner, M., Cronin, S., Bebbington, M., and Platz, T. (2008a). Developing a probabilistic eruption forecast for dormant volcanoes; a case study from Mt Taranaki, New Zealand. *Bulletin of Volcanology*, 70:507–515.

Turner, M., Cronin, S., Bebbington, M., and Smith, I. (2011a). Relating magma composition to eruption variability at andesitic volcanoes: A case study from Mount Taranaki, New Zealand. *Bulletin of the Geological Society of America*, 123:2005–2015.

Turner, M., Cronin, S., Bebbington, M., and Smith, I. (2011b). Andesitic tephrochronology: construction of a pyroclastic eruption record for Mt Taranaki, New Zealand. *Quaternary International*, 246:364–373.

Turner, M., Cronin, S., Bebbington, M., and Stewart, R. (2009). Merging eruption datasets: building an integrated Holocene eruptive record for Mt Taranaki, New Zealand. *Bulletin of Volcanology*, 71:903–918.

Turner, M., Cronin, S., Stewart, R., Bebbington, M., and Smith, I. (2008b). Using titanomagnetite textures to elucidate volcanic eruption histories. *Geology*, 36:31–34.

Tyron, C., Logan, M., Mouralis, D., S., K., Slimak, L., and Balkan-Atli, N. (2009). Building a tephrostratigraphic framework for the Paleolithic of Central Anatolia, Turkey. *Journal of Archaeological Science*, 36(3):637–652.

Utsu, T. (1999). Representation and analysis of the earthquake size distribution: a historical review and some new approaches. *Pure and Applied Geophysics*, 155:509–535.

Wadge, G. (1982). Steady state volcanism: evidence from eruption histories of polygenetic volcanoes. *Journal of Geophysical Research*, 87:4035–4049.

Wadge, G. and Burt, L. (2011). Stress field control of eruption dynamics at a rift volcano: Nyamuragira, D.R.Congo. *Journal of Volcanology and Geothermal Research*, 207:1–15.

Wadge, G. and Guest, J. (1981). Steady-state magma discharge and Etna 1971-81. *Nature*, 294:548–550.

Wang, T. and Bebbington, M. (2012). Estimating the likelihood of an eruption from a volcano with missing onsets in its record. *Journal of Volcanology and Geothermal Research*, 243–244:14–23.

Ward, G. and Wilson, S. (1978). Procedures for comparing and combining radiocarbon age determinations: a critique. *Archeometry*, 20(1):19–31.

Wickman, F. (1966). Repose period patterns of volcanoes I. Volcanic eruptions regarded as random phenomena. *Ark. Mineral. Geol.*, 4:291–367.

Zernack, A., Cronin, S., Bebbington, M., Price, R., Smith, I., Stewart, R., and Procter, J. (2012). Forecasting catastrophic stratovolcano collapse: a model based on Mount Taranaki, New Zealand. *Geology*, 40:983–986.

Appendix A

Data used in Chapter 3

The data used in Chapter 3 is summarized in Tables A.1 and A.2 for the single-source and multi-source records, respectively. The asterisk denotes a record that possesses geochemical data. Note that there is no eruption number 27 in either lake record. In both cases this was the rhyolitic 'Stent Ash', sourced from the Taupo Volcanic Center.

Table A.1: Ages and geochemistry for the Mt Taranaki single-source record. NS, U and R denote Near-source, Umutekai and Rotokare derived tephras, respectively.

Age (cal yr BP)		Tephra Number		Mean wt%			SD wt%		
Mean	SD	NS	U	TiO_2	Al_2O_3	MgO	TiO_2	Al_2O_3	MgO
90	1	1							
154	1	2							
199	1	3*		7.50	2.04	1.82	0.45	0.05	0.13
240	34	4							
287	31	5							
324	30	6							
360	41	7							
389	39	8*		6.19	3.72	2.77	0.23	0.33	0.10
421	38	9							
454	28	10							
484	11	11*		7.04	2.81	2.22	0.80	0.28	0.23
504	7	12*		6.37	2.42	1.96	0.50	0.47	0.31
529	18	13							
570	32	14*		7.44	2.70	2.17	1.66	0.55	0.58
615	29	15							
745	38	16*		7.43	2.83	2.49	0.55	0.93	0.46
854	54	17*		7.78	2.96	2.75	0.75	0.51	0.32
990	36	18							
1218	104	19*		6.78	4.72	3.00	0.50	1.75	0.76
1403	56		1*	8.41	2.22	1.97	0.91	0.88	0.30
1479	47		2						
1688	67		3*	9.01	3.95	3.63	0.15	0.23	0.32
1709	68		4						
1714	68		5						
1890	56		6						
2007	62		7*	7.31	3.56	2.77	0.44	0.70	0.76
2168	47		8*	6.74	2.71	2.60	0.37	0.91	0.62
2323	58		9*	7.33	4.34	3.17	0.90	0.79	0.55

Continued on next page

Table A.1 – *Continued from previous page*

Age (cal yr BP)		Tephra Number		Mean wt%			SD wt%		
Mean	SD	NS	U	TiO_2	Al_2O_3	MgO	TiO_2	Al_2O_3	MgO
2534	66		10*	6.72	6.09	4.00	0.33	1.23	0.77
2690	64		11*	6.80	4.67	3.12	0.64	1.44	0.52
3025	57		12						
3120	56		13						
3147	56		14						
3250	54		15*	7.52	1.64	1.62	0.26	0.20	0.19
3302	52		16						
3551	49		17*	7.79	1.64	1.50	0.57	0.29	0.22
3582	52		18						
3628	57		19						
3656	60		20*	7.68	1.89	1.68	0.53	0.28	0.19
3720	65		21						
3898	71		22						
3948	71		23*	7.39	1.98	2.00	0	0	0
3987	71		24*	7.18	2.30	1.90	1.62	0.78	0.56
4034	70		25*	7.41	2.27	1.63	0.31	0.14	0.17
4044	69		26						
4327	57		28*	8.08	1.76	1.84	0.95	0.47	0.28
4448	55		29						
4574	56		30*	7.98	2.36	2.02	0.06	0.12	0.21
4644	58		31						
4741	61		32						
4820	63		33*	7.86	2.58	2.29	0.24	0.25	0.41
4863	64		34						
4892	65		35						
4929	66		36						
4941	66		37						
4969	67		38						
5054	68		39*	8.79	1.95	1.91	0.63	0.46	0.48
5088	68		40						
5121	68		41						
5144	68		42						
5208	68		43*	8.22	2.92	2.42	0.30	0.44	0.32
5332	68		44*	8.22	2.76	2.53	0.15	0.25	0.31
5422	67		45						
5459	67		46*	8.40	2.97	2.53	0.36	0.14	0.52
5561	66		47*	8.19	2.62	2.61	0.12	0.16	0.15
5645	65		48*	9.70	1.75	1.93	1.32	0.82	0.49
5726	64		49*	8.18	2.71	2.56	0.25	0	0.06
5829	62		50						
5853	62		51						
5888	62		52*	10.48	1.75	1.74	1.80	0.35	0.61
5987	60		53						
6237	58		54						
6288	57		55*	8.58	2.67	2.49	0.74	0.12	0.37
6589	56		56						
6646	56		57						
6676	57		58						
6698	57		59						
6761	57		60*	9.54	2.78	2.85	0.45	0.13	0.28
6769	57		61*	10.01	2.56	2.83	0.56	0.46	0.28
6799	57		62						
6922	58		63*	8.24	2.74	2.55	0.16	0.23	0.26
7020	58		64*	8.40	2.49	2.40	0.22	0.40	0.55
7339	61		65*	8.82	2.01	1.90	0.35	0.40	0.55
7398	61		66						
7848	63		67*	9.80	2.44	2.82	0.28	0.12	0.12
7914	63		68*	9.31	2.31	2.13	0.31	0.62	0.62
8002	62		69						
8055	62		70*	6.30	2.26	1.70	0.03	0.16	0.15
8177	61		71						

Continued on next page

Table A.1 – *Continued from previous page*

Age (cal yr BP)		Tephra Number		Mean wt%			SD wt%		
Mean	SD	NS	U	TiO_2	Al_2O_3	MgO	TiO_2	Al_2O_3	MgO
8218	61		72*	7.92	2.60	2.23	0.19	0.27	0.47
8279	61		73*	7.94	2.60	2.41	0.54	0.33	0.64
8486	59		74						
8555	58		75*	8.30	2.76	2.82	0.13	0.09	0.14
8579	58		76*	8.50	2.82	2.83	0.16	0.09	0.07
8764	56		77*	9.80	2.38	2.19	1.20	0.90	0.75
8808	56		78*	7.46	2.88	2.61	0.20	0.09	0.23
8832	56		79*	9.04	2.62	2.50	0.83	0.75	0.60
8844	56		80*	9.90	3.19	3.20	0.27	0.43	0.53
8956	55		81						
9026	54		82						
9060	54		83*	9.65	2.26	1.90	1.79	0.31	0.37
9091	53		84						
9209	53		85*	9.87	2.35	2.54	0.53	0.51	0.57
9240	52		86						
9320	52		87*	10.47	1.67	1.75	0.35	0.35	0.26
9771	53		88						
9828	54		89						
9904	55		90*	10.10	2.91	2.90	0.19	0.18	0.52
9981	56		91*	9.48	2.85	2.69	0.53	0.36	0.69
10043	57		92*	9.84	3.24	3.19	0.45	0.32	0.16
10229	63		93*	9.70	3.14	3.08	0.31	0.11	0.08
10292	65		94						
10342	67		95*	9.35	3.25	3.16	0.16	0.17	0.19
10411	69		96						
10493	73		97*	9.65	3.57	3.37	0.67	0.34	0.14
10566	76		98						
10805	89		99*	9.60	3.28	3.46	0.14	0.13	0.10
10911	96		100						
11178	114		101*	9.10	3.49	3.32	0.68	0.48	0.36
11301	124		102						
11401	132		103*	7.65	3.08	2.70	0.18	0.18	0.31
11460	137		104						

Table A.2: Ages and geochemistry for the Mt Taranaki multi-source record. NS, U and R denote Near-source, Umutekai and Rotokare derived tephras, respectively.

Age (cal yr BP)		Tephra No.			Mean wt%			SD wt%		
Mean	SD	NS	R	U	TiO_2	Al_2O_3	MgO	TiO_2	Al_2O_3	MgO
91	0	1								
154	0	2								
200	0	3*			7.50	2.04	1.82	0.45	0.05	0.13
250	39	4								
294	15	5								
313	12	6								
347	31	7								
383	36	8*			6.19	3.72	2.77	0.23	0.33	0.10
427	41	9								
469	17	10								
488	10	11*			7.04	2.81	2.22	0.80	0.28	0.23
506	7	12*	1		6.37	2.42	1.96	0.50	0.47	0.31
510	6		2							
512	7		3							
526	10	13	4							
538	13		5							
569	31	14*	6		7.44	2.70	2.17	1.66	0.55	0.58
610	32	15								
665	20		7							
747	22	16*	8		7.43	2.83	2.49	0.55	0.93	0.46
856	46	17*			7.78	2.96	2.75	0.75	0.51	0.32
993	37	18								
1235	112	19*			6.78	4.72	3.00	0.50	1.75	0.76
1400	52			1*	8.41	2.22	1.97	0.91	0.88	0.30
1480	42			2						
1581	33	20	9*		7.92	3.53	2.60	1.12	0.68	0.43
1682	53			3*	9.01	3.95	3.63	0.15	0.22	0.32
1700	52			4						
1817	57	21*		5	6.52	3.51	3.14	0.23	0.12	0.28
1903	36	22		6						
2086	39			7*	7.31	3.56	2.77	0.44	0.70	0.76
2110	33		10*		9.45	3.55	3.57	0.16	0.17	0.17
2156	38	23*		8*	6.87	4.34	3.13	0.30	1.61	0.65
2270	33		11*		9.68	3.53	3.25	0.20	0.17	0.06
2336	44		12*	9*	7.62	3.95	2.87	1.55	0.97	0.61
2394	53		13*		7.08	3.47	2.79	0.40	1.01	0.55
2464	64		14*		6.92	3.30	2.70	0.30	0.32	0.32
2521	71		15*	10*	6.81	4.79	3.39	0.32	1.74	0.93
2575	75		16*		6.78	2.56	2.02	0.15	0.97	0.33
2592	75		17*		6.87	2.53	2.16	0.43	0.59	0.44
2672	74		18*	11*	6.77	5.06	3.31	0.68	1.60	0.62
3017	57			12						
3116	53			13						
3145	52		19*	14	8.00	5.63	3.76	0.84	0.98	0.37
3251	49			15*	7.52	1.64	1.61	0.26	0.20	0.18
3304	47			16						
3562	49		20	17*	7.79	1.64	1.50	0.57	0.29	0.22
3594	53			18						
3641	59			19						
3668	62			20*	7.68	1.89	1.68	0.53	0.28	0.19
3729	66			21						
3889	63			22						
3933	63		21*	23*	7.57	1.95	1.70	0.12	0.05	0.13
3968	64		22*	24*	7.30	2.12	1.85	1.15	0.59	0.40
4010	66		23*	25*	7.47	2.16	1.78	0.25	0.17	0.22
4019	67		24*	26	7.95	2.30	2.13	0.19	0.16	0.22
4021	67		25*		8.40	2.02	1.76	0.45	0.32	0.14
4033	67		26*		7.60	2.39	1.97	0.30	0.21	0.09
4290	92		28*	28*	8.53	2.12	2.02	0.85	0.48	0.40
4419	84			29						

Continued on next page

Table A.2 – *Continued from previous page*

Age (cal yr BP)		Tephra No.			Mean wt%			SD wt%		
Mean	SD	NS	R	U	TiO_2	Al_2O_3	MgO	TiO_2	Al_2O_3	MgO
4554	68			30*	7.98	2.36	2.02	0.06	0.12	0.21
4630	61			31						
4734	60			32						
4819	63			33*	7.86	2.58	2.29	0.24	0.25	0.41
4864	66			34						
4895	68			35						
4933	70			36						
4946	71			37						
4975	72			38						
5063	76			39*	8.79	1.95	1.91	0.63	0.46	0.48
5098	77			40						
5132	78			41						
5156	79			42						
5222	80		29*	43*	8.26	2.75	2.43	0.37	0.46	0.39
5348	82		30*	44*	8.35	2.59	2.45	0.27	0.42	0.42
5439	83			45						
5476	83		31*	46*	8.19	2.88	2.45	0.33	0.33	0.46
5579	82			47*	8.19	2.62	2.61	0.12	0.16	0.15
5663	81		32*	48*	9.12	1.90	2.01	1.50	0.94	0.50
5744	80		33*	49*	8.53	3.03	2.77	0.82	0.59	0.46
5847	79			50						
5871	78			51						
5906	77		34	52*	10.48	1.75	1.74	1.80	0.35	0.61
6005	75		35*	53	8.27	2.29	2.19	0.10	0.13	0.22
6200	71		36*		8.20	2.45	1.98	0.71	0.90	0.61
6253	70		37*	54	8.20	2.66	2.45	0.19	0.16	0.26
6304	69		38*	55*	8.11	2.72	2.45	0.85	0.78	0.64
6601	64		39*	56	8.92	1.80	1.76	0.97	0.54	0.49
6657	64			57						
6688	63			58						
6709	63			59						
6772	63			60*	9.54	2.78	2.85	0.45	0.13	0.28
6779	63		40*	61*	8.82	2.63	2.42	1.01	0.32	0.37
6809	63		41*	62	9.61	3.09	3.18	0.41	0.48	0.29
6889	65		42*		9.46	2.66	2.84	0.20	0.51	0.44
6930	62			63	8.24	2.74	2.55	0.16	0.23	0.26
7028	62			64*	8.40	2.49	2.40	0.22	0.40	0.55

Appendix B

Matlab code to accompany Chapter 3

B.1 Matlab code to impute geochemical data

This section provides the Matlab code, 'SortGeochem.m', to combine and sort available geochemical data and impute missing geochemical data (see page 55 for an explanation). The code prints two files 'SingleSourceGeochem.txt' and 'MultiSourceGeochem.txt', which are required to run the Matlab code in Appendix B.2. Files are of the form [UmutekaiID, RotokareID, NearSourceID, Mean(TiO_2), Mean(Al_2O_3), Mean(MgO), SD(TiO_2), SD(Al_2O_3), SD(MgO)]. The data file 'TaranakiGeochem.csv' must be placed in the same directory as the Matlab code. It contains the raw geochemical data for each tephra (where available) and the alignment of tephras across the locations. 'TephraID0' specifies the location (1XXX for Umutekai, 2XXX for Rotokare, and 3XXX for Near-source) and tephra number (given by the the remaining three digits -XXX).

```
SingleSourceOn=0; % 1 for the single source record. 0 for the multisource record.

% Read in Raw Geochem, of the form TiO2 Al2O3 MgO, and corresponding TephraID numbers
RawGeochem = csvread('TaranakiGeochem.csv',1,1,[1,1,824,3]);
TephraID0 = csvread('TaranakiGeochem.csv',1,0,[1,0,824,0]);

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Average the geochemistry vectors across eruption number
AverageGeochem=[unique(TephraID0) fix(unique(TephraID0)/1000) rem(unique(TephraID0),1000)];
```

```
for counter=1:length(AverageGeochem(:,1))
    whichOne = find(TephraID0 ==AverageGeochem(counter,1)); % find which raw geochem columns
     correspond to the ID number
    if length(whichOne)==1 %if we only have a single point,  standard deviation = 0
     AverageGeochem(counter,4:9) = [RawGeochem(whichOne,:) 0 0 0];
    else % compute the mean and standard deviation of the raw geochem
     AverageGeochem(counter,4:9)=[mean(RawGeochem(whichOne,:)) std(RawGeochem(whichOne,:))]
    end
end
% AverageGeochem in form [TephraID1 Source ErupNo MeanTi MeanAl MeanMgO SdevTi SdevAl
    SDevMgO]
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% load in Eruption age data in form [MeanAge StdAge NSTephraID RTephraID UTephraID]
AgeData = csvread('TaranakiGeochem.csv',1,5,[1,5,137,8]);
StratigraphicOrder = AgeData(:,1);
AgeData(end,:)=[]; %remove umu104 since no geochem available for that eruption

SourceTephra{1}=AgeData(:,end); % Umutekai Data
if SingleSourceOn
    SourceTephra{2}=[]; % single-source record doesn't include rotokare tephras
else
    SourceTephra{2}=AgeData(:,end-1); % RotokareData
end
SourceTephra{3}=AgeData(:,end-2); % NearSourceData

if SingleSourceOn
    WhichSources=1:2:3; %only use sources Umutekai(1) and NS(3)
else
    WhichSources=1:3; % if multisource use all 3 sources Umutekai(1), Rotokare(2) and NS(3)
end

for i=WhichSources;
    if SingleSourceOn&&i==3
        NoTephra=19; %only use the first 19 NS tephras
    else
        NoTephra=max(SourceTephra{i}); %use all tephras from given location
    end
    OriginalTemplate= [(1:NoTephra)' zeros(NoTephra,7)]; %initialize matrix full of zeros
    for j = 1:NoTephra % for each tephra in given location
        % find age and age error
        if find(SourceTephra{i}==j)>0
            location(j) = find(SourceTephra{i}==j);
            OriginalTemplate(j,2) = StratigraphicOrder(location(j));
        else
            OriginalTemplate(j,2:end) = NaN;  % taupo stent (U27 and R27) not from taranaki
        end
    end
    % fill matrix with Geochem data
```

```
    Data = AverageGeochem(AverageGeochem(:,2)==i,:);
    for j = 1:NoTephra % for each tephra in given location
        if find(Data(:,3) == j)>0 %find if there is geochem data available
            for k = 1:length(Data)
                %find where in Data matrix the geochem is available
                location(k) = find(Data(:,3) == j);
                %fill the Template with available geochem data
                OriginalTemplate(j,3:8) = Data(location(k),4:9);
                %store TephraID number for the geochem data.
                OriginalTemplate(j,9) = Data(location(k),1);
            end
        end
    end
    NameTephra{i}=OriginalTemplate;
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% stack the 3 sources together
%in format [Source ErupNo Age StdAge MeanTiO2 MeanAl2O3 MeanMgO StdevTiO2 StdevAl2O3
    StdevMgO TephraID1]
Umu=NameTephra{1};
Roto=NameTephra{2};
NS=NameTephra{3};
Combined = sortrows([ones(length(Umu),1) Umu;ones(length(Roto),1)*2  Roto;ones(length(NS),1)
    *3  NS],3);
Combined  = Combined(¬isnan(Combined(:,3)),:); %remove NaNs (i.e TaupoStentAsh U27, R27)
SortAge = Combined(:,3); %extract stratigraphic order
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
Template = []; %initialize template as empty
if SingleSourceOn
    Template=Combined;
else
    Template(1:3,:) = Combined(1:3,:); %fill the first 3rows of the template
    TemLocate = length(AgeData); %initialize counters for template location
    m = length(SortAge); %initialize counters for CombSort matrix
    while m>3
    %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
        %if 3 sources have same mean age and std
        if (SortAge(m,:) == SortAge(m−1,:))&(SortAge(m−1,:)== SortAge(m−2,:))==1
            %and if all geochem components are available find mean geochem across the 3
    locations, fill the template.
            if ((Combined(m,end) ≠ 0)&(Combined(m−1,end) ≠ 0)&(Combined(m−2,end) ≠ 0)) == 1
                temp = [find(TephraID0==Combined(m,end)); find(TephraID0==Combined(m−1,end))
    ; find(TephraID0==Combined(m−2,end))];
                Template(TemLocate,3:12) = [Combined(m,3) mean(RawGeochem(temp,:)) std(
    RawGeochem(temp,:)) Combined(m,end) Combined(m−1,end) Combined(m−2,end)];
            end
            m = m−3;
            TemLocate = TemLocate − 1;
```

```
    %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
        %if 2 sources have the same mean and std
        elseif (SortAge(m,:) == SortAge(m-1,:))==1
            %and if 2means but no components are available then leave empty
            if ((Combined(m,end) == 0)&(Combined(m-1,end) == 0)) == 1
                Template(TemLocate,1:10) = Combined(m,:);
                m = m-2;
                TemLocate = TemLocate - 1;
            end
            % and if 2means and all components are available find mean and variance of
    stacked data and fill.
            if ((Combined(m,end) ~= 0)&(Combined(m-1,end)~= 0)) == 1
                temp = [find(TephraID0==Combined(m,end)); find(TephraID0==Combined(m-1,end))
    ];
                Template(TemLocate,3:11) = [Combined(m,3) mean(RawGeochem(temp,:)) std(
    RawGeochem(temp,:)) Combined(m,end) Combined(m-1,end)];
                m = m-2;
                TemLocate = TemLocate - 1;
            % and if 2means but only one component available fill the template with the
    geochem data available
            elseif ((Combined(m,end) == 0)&(Combined(m-1,end) == 0)) == 0
                if (Combined(m,end) > Combined(m-1,end)) == 1 %first mean has data
                    Template(TemLocate,1:10) = Combined(m,:);
                elseif (Combined(m,end) < Combined(m-1,end)) == 1 %2nd mean has data
                    Template(TemLocate,1:10) = Combined(m-1,:);
                    %else fill template with CombSort(m-1,:)
                end
                m = m-2;
                TemLocate = TemLocate - 1;
            end
    %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
        %if no sources have the same mean and std then fill Geochem data with zeros
        %and let the tephra ID = current ID
        else
            Template(TemLocate,1:10) = Combined(m,:);
            m = m-1;
            TemLocate = TemLocate-1;
        end
    end
end
Template = sortrows(Template,3); %sort template by stratigraphic order
if SingleSourceOn
    Unfilled  = [AgeData(Template(:,3),:) Template(:,4:end)]; % exclude those only observed
    in Rotokare
    FinalData = Unfilled(:,2:10);
else
    Unfilled  = [AgeData Template(:,4:end)];
    FinalData = Unfilled(1:97,2:10);
```

```
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% If geochem missing, assume it is the same as the most recent geochem.
m = length(FinalData)-1; % - 1 since the last observation has no geochem data
while m > 0
    if FinalData(m,4) == 0
        FinalData(m,4:end) = FinalData(m+1,4:end);
    end
    m = m-1;
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% save the geochem data
if SingleSourceOn
    dlmwrite('SingleSourceGeochem.txt',FinalData,'delimiter','\t','newline','pc');
else
    dlmwrite('MultiSourceGeochem.txt',FinalData,'delimiter','\t','newline','pc');
end
```

B.2 Matlab code to fit proportional hazard models

This section provides the Matlab code, 'FitModels.m', to fit the lognormal renewal model and the Weibull, exponential, and mixture of Weibull proportional hazards models in Chapter 3, for all 64 possible combinations of geochemical covariates. Combinations range from the model that has no covariates attached, through to the model that has all six covariates attached [1:Mean(TiO_2), 2:Mean(Al_2O_3), 3:Mean(MgO), 4:SD(TiO_2), 5:SD(Al_2O_3), 6:SD(MgO)]. The code requires that the following functions for maximizing the various loglikelihood functions, be placed in the same working directory: 'lognormalLL.m', 'PHexponentialLL.m', 'PHmixLL.m', and 'PHweibullLL.m'.

The data files 'MultiSourceAges.csv' and 'SingleSourceAges.csv' contain 100 realizations (100 rows) of the actual event ages. The files 'SingleSourceGeochem.txt' and 'MultiSourceGeochem.txt' contain the geochemical data, as created in Appendix B.1.

The Matlab code will create four files: SingleSourceLognormal.txt (or MultiSourceLognormal.txt) of the form [μ, σ, $\log L$, AIC]; SingleSourceWeibull.txt (or MultiSourceWeibull.txt) of the form [Run, α, β, $\log L$, AIC, $\boldsymbol{z}$]; SingleSourceExponential.txt (or MultiSourceExponential.txt) of the form [Run, β, $\log L$, AIC, $\boldsymbol{z}$]; and SingleSourceMixture.txt (or MultiSourceMixture.txt) of the form [Run, p, α_1, β_1, α_2, β_2, $\log L$, AIC, $\boldsymbol{z}$], where 'Run' explains the combination of covariates selected (0 for no covariates, 123456 for all six attached) and $\boldsymbol{z}$

denotes the covariate vector values. The code will also produce Figures 3.4, 3.5, 3.6, and 3.7 from Chapter 3.

```
global tau tau0 covVector covVector0
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% FLAGS
SingleSourceOn=0; %1 for single-source record, 0 for multi-source record.
NoRealisations = 100; % number of realizations of actual event ages
convno = 1000; % specify the number of convergences --- amend if you like
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% load in the geochem data
if SingleSourceOn
    Names = 'SingleSource';
else
    Names = 'MultiSource';
end
Geochem = dlmread([Names,'Geochem.txt']);
GeochemData = fliplr(Geochem')'; % arrange data from largest age to smallest age (reverse
    stratigraphic order)
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% read in ages and compute (interevent time) tau and (current repose) tau0.
tau=[];
tau0=[];
tdata = dlmread([Names,'Ages.csv']); % 100 realisations of the actual event ages
presentAge = -61;
for j=1:NoRealisations % for each realisation
    age=tdata(j,:)'; % extract the ages for all events
    % calculate the interevent times
    intertau(j,:) = age(1:length(age)-1)-age(2:length(age)); % size NoRealisations by
    NoEvents
    tau = horzcat(tau,intertau(j,:));  % one row NoRealisations*NoEvents entries.
    % calculate the unfinished interval
    intertau0(j) = tdata(j,end)-presentAge; %size 1 by NoRealisations
    tau0 = horzcat(tau0,intertau0(j)); %size 1 by NoRealisations
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
meanAge = mean(tdata)';  %compute the meanAge - the mean of all realisations
sd = std(tdata)'; %compute the stdAge - the standard dev of all realisations
Run = unique(nchoosek([0 0 0 0 0 0 1:6],6),'rows');
% find all possible combinations of covariates - each row of 'Run' is a
% combination ranging from [0 0 0 0 0 0] (No covariates) to [1 2 3 4 5 6]
% (all covariates attached)
% create blank matrix of parameter estimates for all possible combinations.
expoPHparameters = zeros(size(Run,1),9);
weibPHparameters = zeros(size(Run,1),10);
mixPHparameters = zeros(size(Run,1),13);

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

```
for runNo=1:size(Run,1)
    covariateNo = Run(runNo,:);
    covariateNo(covariateNo==0)=[]; % 1=MeanTi 2=MeanAl 3=MeanMg 4=SdTi 5=SdAl 6=SdMg
    covariate=[GeochemData(:,covariateNo+3)];

    covV0=[];
    covVector=[];
    covVector0=[];
    for j = 1:NoRealisations
        covVector=repmat(covariate(1:end-1,:),j,1);
        covV0(j,:) = [covariate(end,:)];
        covVector0 = covV0;
    end
    %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
    % fit PH exponential model
    exitflag_expo0 = 0;
    while exitflag_expo0 < convno
        % initial points
        betExpo_0 = -log(rand(1))/mean(tau);
        betExpo_0 = log(betExpo_0);
        ΔExpo_0=repmat(0,1,size(covVector,2)); % 1 by numberofCovs - full of zeros
        [x,fval,exitflag] = fminsearch(@PHexponentialLL,[betExpo_0,ΔExpo_0]);
        if exitflag > 0
            exitflag_expo0 = exitflag_expo0 + 1;
            if exitflag_expo0 == 1
                fval_expoPH = fval + 1;
            end
            if fval < fval_expoPH
                fval_expoPH = fval;
                xx_expoPH = x;
            end
        end
    end
    % best fit parameters for exponentialPH model
    betExpoPH = exp(xx_expoPH(1));
    ΔExpoPH = xx_expoPH(2:(1+size(covVector,2)));
    LL_expoPH = -fval_expoPH/NoRealisations; % adjust for the number of realisations
    AIC_expoPH = -2*LL_expoPH+2*(1+size(covVector,2));
    expoPHparameters(runNo,1:(3+size(covVector,2))) = [betExpoPH LL_expoPH AIC_expoPH Δ
    ExpoPH];

    %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
    % fit PH weibull model
    exitflag_weib0 = 0;
    while exitflag_weib0 < convno
        % initial points
        alp_0 = 0.5 + 2*rand(1);
        bet_0 = -log(rand(1))/mean(tau);
```

```
        alp_0 = log(alp_0);
        bet_0 = log(bet_0);
        ΔWeib_0=repmat(0,1,size(covVector,2));
        [x,fval,exitflag] = fminsearch(@PHweibullLL,[alp_0,bet_0,ΔWeib_0]);
        if exitflag > 0
            exitflag_weib0 = exitflag_weib0 + 1;
            if exitflag_weib0 == 1
                fval_weibPH = fval + 1;
            end
            if fval < fval_weibPH
                fval_weibPH = fval;
                xx_weibPH = x;
            end
        end
    end
    % best fit parameters for weibullPH model
    alpWeibPH = exp(xx_weibPH(1));
    betWeibPH = exp(xx_weibPH(2));
    ΔWeibPH = xx_weibPH(3:(2+size(covVector,2)));
    LL_weibPH = -fval_weibPH/NoRealisations;
    AIC_weibPH = -2*LL_weibPH+2*(2+size(covVector,2));
    weibPHparameters(runNo,1:(4+size(covVector,2))) = [alpWeibPH betWeibPH LL_weibPH
    AIC_weibPH ΔWeibPH];
    %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
    % fit PH mixture model
    exitflag_mix0 = 0;
    while exitflag_mix0 < convno
        % initial points
        p_0 = -log(-log(rand(1)));
        alp1_0 = 0.5 + 2*rand(1);
        bet1_0 = -log(rand(1))/mean(tau);
        alp2_0 = 0.5 + 2*rand(1);
        bet2_0 = -log(rand(1))/(20*mean(tau));
        ΔMix_0=repmat(0,1,size(covVector,2));
        % as both parameters must be positive, transform for fminsearch
        alp1_0 = log(alp1_0);
        bet1_0 = log(bet1_0);
        alp2_0 = log(alp2_0);
        bet2_0 = log(bet2_0);
        [x,fval,exitflag] = fminsearch(@PHmixLL,[p_0,alp1_0,bet1_0,alp2_0,bet2_0,ΔMix_0]);
        if exitflag > 0
            exitflag_mix0 = exitflag_mix0 + 1;
            if exitflag_mix0 == 1
                fval_mixPH = fval + 1;
            end
            if fval < fval_mixPH
                fval_mixPH = fval;
                xx_mixPH = x;
```

```
        end
    end
end
% best fit parameters FOR MIX PH
pMixPH = exp(-exp(-xx_mixPH(1)));
alp1MixPH = exp(xx_mixPH(2));
bet1MixPH = exp(xx_mixPH(3));
alp2MixPH = exp(xx_mixPH(4));
bet2MixPH = exp(xx_mixPH(5));
ΔMixPH = xx_mixPH(6:(5+size(covVector,2)));

% ensure first component of the mixture is the dominant component
% if p is < 0.5 swap alp1 and alp2, and bet1 and bet2
if pMixPH < 0.5
    tempBMixPH = bet2MixPH;
    bet2MixPH = bet1MixPH;
    bet1MixPH = tempBMixPH;
    tempAMixPH = alp2MixPH;
    alp2MixPH = alp1MixPH;
    alp1MixPH = tempAMixPH;
    pMixPH = 1-pMixPH;
end
LL_mixPH = -fval_mixPH/NoRealisations;
AIC_mixPH = -2*LL_mixPH+2*(5+size(covVector,2));
mixPHparameters(runNo,1:(7+size(covVector,2))) = [pMixPH alp1MixPH bet1MixPH alp2MixPH
bet2MixPH LL_mixPH AIC_mixPH ΔMixPH];

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%fit lognormal renewal model, no PH model for the lognormal as the renewal model
provides a poor fit.
if runNo == 1
    exitflag_log0=0;
    while exitflag_log0 < convno
        sig_0 = std(log(tau));
        mu_0 = mean(log(tau));
        [x,fval,exitflag] = fminsearch(@lognormLL,[sig_0,mu_0]);
        if exitflag > 0
            exitflag_log0 = exitflag_log0 + 1;
            if exitflag_log0 == 1
                fval2_log = fval + 1;
            end
            if fval < fval2_log
                fval2_log = fval;
                xx_log = x;
            end
        end
    end
    % best fit parameters FOR LOGNORM
```

```
        sigLog = exp(xx_log(1));
        muLog = exp(xx_log(2));
        LL_log = -fval2_log/NoRealisations;
        AIC_log=-2*LL_log+2*2;
        logparameters=[muLog sigLog LL_log AIC_log];
        dlmwrite([Names,'Lognormal.txt'],[logparameters],'delimiter','\t', 'newline', 'pc',
    'precision',6);
    end

    dlmwrite([Names,'Weibull.txt'],[str2num(num2str(Run,'%d')) weibPHparameters],'delimiter
    ','\t', 'newline', 'pc','precision',6);
    dlmwrite([Names,'Mixture.txt'],[str2num(num2str(Run,'%d')) mixPHparameters],'delimiter',
    '\t', 'newline', 'pc','precision',6);
    dlmwrite([Names,'Exponential.txt'],[str2num(num2str(Run,'%d')) expoPHparameters],
    'delimiter','\t', 'newline', 'pc','precision',6);
end
% extract best fit weibullPH model...
WeibullMatrix = [str2num(num2str(Run,'%d')) weibPHparameters];
BestWeibull = WeibullMatrix(WeibullMatrix(:,5)==min(WeibullMatrix(:,5)),:)
% extract best fit mixturePH model...
MixtureMatrix = [str2num(num2str(Run,'%d')) mixPHparameters];
BestMixture = MixtureMatrix(MixtureMatrix(:,8)==min(MixtureMatrix(:,8)),:)
ExponentialMatrix = [str2num(num2str(Run,'%d')) expoPHparameters];
% extract best fit mixturePH model...
BestExponential=ExponentialMatrix(ExponentialMatrix(:,4)==min(ExponentialMatrix(:,4)),:)

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% plot geochem data vs event ages
j=size(GeochemData,1);
PlottingGeochem(1,1:6)=[GeochemData(1,4:6) log(GeochemData(1,7:9))];
while j>1
    if GeochemData(j,6)≠GeochemData(j-1,6)
        PlottingGeochem(j,1:6) = [GeochemData(j,4:6) log(GeochemData(j,7:9))];
    else
        PlottingGeochem(j,1:6) = NaN;
    end
    j=j-1;
end

figure(1)
symbolcell ={'*','o', 'o'};
symbolcolor ={'k','k',[0.7,0.7,0.7]};
for i=1:3
    subplot(1,2,1)
    plot(meanAge(:),PlottingGeochem(:,i),symbolcell{i},'Color',symbolcolor{i},
    'MarkerFaceColor',symbolcolor{i},'MarkerSize',6)
    xlabel('Mean eruption age (cal. yrs BP)')
    ylabel('Mean %wt')
```

```
    legend('TiO_2','Al_2O_3','MgO')
    hold on
    subplot(1,2,2)
    plot(meanAge,PlottingGeochem(:,3+i),symbolcell{i},'Color',symbolcolor{i},
    'MarkerFaceColor',symbolcolor{i},'MarkerSize',6)
    xlabel('Mean eruption age (cal. yrs BP)')
    ylabel('log(SD %wt)')
    hold on
end
    legend('TiO_2','Al_2O_3','MgO')
hold off

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% plot ages vs interevent times
figure(2)
plot(tdata(:,1:end-1),intertau(:,:),'.b','Color',[0.7,0.7,0.7]) % for all realizations
hold on
plot(meanAge(1:end-1),mean(intertau)','ok','MarkerFaceColor','k') % for mean values
ylabel('Repose length \tau (years)')
xlabel('Mean eruption age (cal. yrs BP)')

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% plot the renewal density functions
maxplottau = 1.1*max(tau);
nobins = 30;
binwidth = maxplottau/nobins;
den2hist = binwidth*length(tau);
fineness = 1000;
plottau = (1:fineness)*maxplottau/fineness;

% extract renewal model parameters
alpWeib = weibPHparameters(1,1);betWeib = weibPHparameters(1,2);
betExpo = expoPHparameters(1,1);
pMix = mixPHparameters(1,1); alp1Mix = mixPHparameters(1,2); bet1Mix = mixPHparameters(1,3);
    alp2Mix = mixPHparameters(1,4);bet2Mix = mixPHparameters(1,5);

fmix=pMix*alp1Mix*(bet1Mix^alp1Mix)*(plottau.^(alp1Mix-1)).*exp(-(bet1Mix*plottau).^alp1Mix)
    +(1-pMix)*alp2Mix*(bet2Mix^alp2Mix)*(plottau.^(alp2Mix-1)).*exp(-(bet2Mix*plottau).^
    alp2Mix);
fweibull=alpWeib*(betWeib^alpWeib)*(plottau.^(alpWeib-1)).*exp(-(betWeib*plottau).^alpWeib);
flognorm=(plottau*sigLog*sqrt(2*pi)).^(-1).*exp(-((log(plottau)-muLog).^2)./(2*sigLog.^2));
fexpo=(betExpo)*exp(-(betExpo*plottau));

figure(3)
hist(tau,nobins);
h = findobj(gca,'Type','patch');
set(h,'FaceColor',[0.8,0.8,0.8],'EdgeColor','w')
hold on;
```

```
plot(plottau,fmix*den2hist,'-k',plottau,fweibull*den2hist,'--k',plottau,flognorm*den2hist,':
    k',plottau,fexpo*den2hist,'-.k','LineWidth',1.5)
legend('Data','Mixture','Weibull','Lognormal','Exponential')
xlabel('Repose length \tau (years)');ylabel('Frequency');
hold off;
```

```
function f = lognormLL(x)
% calculates -loglikelihood of a Lognormal distribution
global tau tau0
sig = exp(x(1));
mu = exp(x(2));
n = length(tau); % number of observed values
F0 = (1/2)*erfc(-(log(tau0)-mu)/(sig*sqrt(2)));

logL = -sum(log(tau)) - (n/2)*log(2*pi)-n*log(sig)-(1/(2*sig^2)).*sum((log(tau)-mu).^2)+sum(
    log(1-F0));

f = sum(-logL); % minimum of -logL = maximum of logL
```

```
function f = PHexponentialLL(x)
% calculates -loglikelihood of a proportional hazards model with a Weibull baseline hazard
global tau tau0 covVector covVector0
bet = exp(x(1));
num = length(tau); % number of observed values
Δ=x(2:(1+size(covVector,2)));
newVector = repmat(0,size(covVector,1),1);
newVector0 = repmat(0,size(covVector0,1),1);

for j = 1:size(covVector,2)
    temp = covVector(:,j).*Δ(j);
    newVector = newVector+temp;
    temp0 = covVector0(:,j).*Δ(j);
    newVector0 = newVector0+temp0;
end

f = bet.*exp(newVector').*exp(-((bet*tau)).*exp(newVector'));
logL = sum(log(f))-sum(-((bet*tau0)).*exp(newVector0'));

f = sum(-logL);
```

```
function f = PHweibullLL(x)
% calculates -loglikelihood of a proportional hazards model with a Weibull baseline hazard
global tau tau0 covVector covVector0
alp = exp(x(1));
bet = exp(x(2));
num = length(tau); % number of observed values
Δ=x(3:(2+size(covVector,2)));

newVector = repmat(0,size(covVector,1),1);
newVector0 = repmat(0,size(covVector0,1),1);
```

```
for j = 1:size(covVector,2)
    temp = covVector(:,j).*Δ(j);
    newVector = newVector+temp;
    temp0 = covVector0(:,j).*Δ(j);
    newVector0 = newVector0+temp0;
end

f = alp.*(bet.^alp).*exp(newVector').*(tau.^(alp-1)).*exp(-((bet*tau).^alp).*exp(newVector')
    );
logL = sum(log(f))-sum(-((bet*tau0).^alp).*exp(newVector0'));

f = sum(-logL);
```

```
function f = PHmixLL(x)
% calculates -loglikelihood of a proportional hazards model with a Mixture baseline hazard
global tau tau0 covVector covVector0
p = exp(-exp(-x(1)));
alp1 = exp(x(2));
bet1 = exp(x(3));
alp2 = exp(x(4));
bet2 = exp(x(5));
num = length(tau); % number of observed values
Δ=x(6:(5+size(covVector,2)));

newVector = repmat(0,size(covVector,1),1);
newVector0 = repmat(0,size(covVector0,1),1);
for j = 1:size(covVector,2)
    temp = covVector(:,j).*Δ(j);
    newVector = newVector+temp;
    temp0 = covVector0(:,j).*Δ(j);
    newVector0 = newVector0+temp0;
end

f1 = alp1.*(bet1.^alp1).*exp(newVector').*(tau.^(alp1-1)).*exp(-((bet1*tau).^alp1).*exp(
    newVector'));
f2 = alp2.*(bet2.^alp2).*exp(newVector').*(tau.^(alp2-1)).*exp(-((bet2*tau).^alp2).*exp(
    newVector'));
fnew = p*f1 + (1-p)*f2;

F0 = 1- (1-p)*exp(-((bet2*tau0).^alp2).*exp(newVector0'))-p*exp(-((bet1*tau0).^alp1).*exp(
    newVector0'));
logL = sum(log(fnew))+sum(log(1-F0));

f = sum(-logL); % minimum of -logL = maximum of logL
```

Appendix C

Matlab code, functions, and data to accompany Chapter 4

C.1 Matlab code to estimate the age of each tephra

This section provides the Matlab code, 'InterpolateCode.m', to estimate the age of each tephra using the method discussed in Section 4.2.2. The data file containing the ages of the marker tephras from which to simulate, is given in 'MarkerAges.csv'.

```
% read in the data in the form [Source, ErupNo, Depth(m), Thick(mm), MeanAge, SdevAge]
Data = csvread('MarkerAges.csv',1,2,[1,2,217,7]);
InterpolatedData = []; % initialize as empty
for Source = 1:5  % for each core: Pupuke, Onepoto, Orakei, Hopua, Pukaki
    SelectData = Data(Data(:,1)==Source,:); %select data for the particular core
    Depth = SelectData(:,3); % depth in m
    Thickness = SelectData(:,4); % thickness in mm
    % age and age error of marker tephras in cal ka BP
    % (as given in Lowe et al (2013), Molloy et al (2009) or Danisik et al (2012))
    Age=SelectData(:,5);
    Sigma=SelectData(:,6);
    No = 100;  % number of Monte Carlo simulations to perform
    SimulateFrom = SelectData(Sigma≠0,:); % extract only the marker horizons
    NewData = [SelectData zeros(size(SelectData,1),No)]; % create matrix with blank columns
    to fill with simulated ages
    NoZeros = NewData(Sigma≠0,1:6); %extract rows for the marker horizons.

    mu1 = SimulateFrom(:,5)*1000;   % mean age in cal BP
    sigma1 = SimulateFrom(:,6)*1000; % standard deviation in cal BP
    agemat=[];
```

```
    for j=7:(No+6) % for each Monte Carlo simulation
        age = zeros(size(mu1)); % initialize sequence of ages as empty
        i = 1;
        while i≤ length(mu1) % for each marker tephra
            age(i) = mu1(i) + sigma1(i).*randn(1,1); % obtain a realization of the actual
    age
            if i>1 % if not for the first tephra
                if age(i) > age(i-1) % check the stratigraphy
                    i = i+1; % if ok, move on to the next tephra
                else
                    i = i-1; % otherwise reset the counter
                end
            else
                i = i+1; % cannot check stratigraphy if there is only one tephra
            end
        end
        agemat(:,j-6)=age; %add simulated sequence to matrix
    end
    for j = 7:(No+6) % for each Monte Carlo simulation
        % using the depths (after subtracting the thicknesses) interpolate
        % (or extrapolate) the the non-marker tephra ages
        NewData(:,j) = (interp1(NoZeros(:,3)-NoZeros(:,4)/1000,agemat(:,j-6),Depth-Thickness
    /1000,'linear','extrap'))/1000;
        if Source==1 % forward extrapolation for the top of the Pupuke core
            % top of core is 57m deep
            D2 = (Depth(1)-Thickness(1)/1000)-57;
            D3 = (Depth(2)-Thickness(2)/1000)-57;
            NewData(1,j) = (D2/D3)*NewData(2,j); % calculate the new age for the non marker
    event
        end
    end

    % find mean and sdev of all Monte Carlo simulated ages.
    Means = mean(NewData(:,8:end),2);
    Sdevs = zeros(size(Means));
    for i = 1:size(NewData,1) % for all events from the particular core
        if NewData(i,6)==0 % for all non marker tephras
            NonLinearityConst = 0.5;
            Sdevs(i) = sqrt(std(NewData(i,8:end)).^2+NonLinearityConst^2); % add a
    nonlinearity constant
        else
            Sdevs(i) = std(NewData(i,8:end)); % do not add the penalty for marker tephras.
        end
    end
    InterpolatedData = [InterpolatedData;NewData(:,1:4) Means Sdevs];
    %[Source ErupNo Depth(m) Thickness(mm) SimulatedAge SimulatedStd]
end
```

C.2 Matlab code for the tephra matching algorithm

This section provides the Matlab code ('MatlabCode.m') to implement the automated statistical matching algorithm presented in Chapter 4. Associated functions required to run the Matlab code are given in 'MatlabFunctions.m' and presented in Appendix C.3.

The ages and associated errors of the tephras are given in 'AgeData.csv' where 'CoreNo' $1, \ldots, 5$ corresponds to the locations Lake Pupuke, Onepoto Orakei Basin, Hopua Crater, and Pukaki Crater respectively. Tephras within each core are numbered ('ErupNo') in stratigraphic order from 1 to n where 1 denotes the most recent event found at the smallest depth. The estimated age ('SimulatedAge') and associated error ('SimulatedStdev') for each tephra are given in cal yr BP.

'PriorMatches.csv' contains the list of tephras forced to be matched by our algorithm. For the case of the Auckland records, the matches correspond to the marker tephras. A zero place holder is used if a particular core does not contain the particular marker tephra.

'IncompatibleMatches.csv' contains the list of all incompatible pairwise matches. For the case of the Auckland records, incompatible matches refer to tephras sourced from different volcanic centers.

The Matlab function file 'MatlabFunctions.m' and all data files must be placed in same file directory as 'MatlabCode.m'.

```
%% MatlabCode.m
% This file implements the stochastic local search procedure of Green et al
% (2014) "Automated statistical matching of multiple tephra records"

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%% READ IN THE DATA FILES FROM CURRENT DIRECTORY %%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% define global variables
global nCores Data AllMatches PriorMatches InconsistentGeochemMatch AcceptedWorse
    MatchMatrixOld

Data = csvread('AgeData.csv',1,0);
% age data of the form [CoreNo ErupNo MeanAge Sdev]
% In this example Core = 1,...,5 correspond to Pupuke, Onepoto, Orakei, Hopua and Pukaki
    cores respectively.
nCores = max(Data(:,1));  % find the number of sediment cores in the record
```

```
% List of the apriori matches which must be forced. In the form [Core1 Erup1,...., CoreN
    ErupN] where zeros are used if no marker tephra was found in the corresponding core.
PriorMatches = csvread('PriorMatches.csv',1,0); % Marker horizons

% read in the list of Inconsistent matches which is of the form [SourceNo1 ErupNo1 SourceNo2
     ErupNo2]
InconsistentGeochemMatch = csvread('IncompatibleMatches.csv',1,0);

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%% INITIALIZE ANNEALING SCHEDULE %%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
AnnealingConst = 0.05;  %initialize the annealing constant.
maxTabuListSize = 200; %initialize the number of runs each selected 2 way match to consider
    will stay in the tabuList for.
maxStoppingCriterion = 3000; %initialize the number of times to cycle before updating the
    annealing constant.
AnnealingMultiplier = 2; %initialize the multiplier for the annealing schedule

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%% CODE BELOW SHOULD NOT BE AMMENDED %%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% calculate age distance matrix
DistanceMatrix = zeros(length(Data),length(Data));
for i = 1:length(Data)
  for j = 1:length(Data)
    if Data(i,1)==Data(j,1)
      DistanceMatrix(i,j)=NaN; % remove all ages distances that are calculated between
    members of the same source.
    else
      DistanceMatrix(i,j) = abs(Data(i,3)-Data(j,3))/(sqrt(Data(i,4)^2+Data(j,4)^2));
    end
  end
end
DistanceMatrix(DistanceMatrix>3) = NaN; % remove all age distances > 3

% store all 2 way matches with age distances < 3 in increasing order into a matrix
    AllMatches.
AllMatches=[];
while min(DistanceMatrix(:))<3
  % find the smallest value of the distance matrix i.e. the closest match
  [minvalue,idxofmax] = min(DistanceMatrix(:));
  % where in the distance matrix is this closest match? row r and col c
  [r,c] = ind2sub(size(DistanceMatrix),idxofmax);
  % this match corresponds to ErupNo(r) and ErupNo(c) - set these 2
  % events to NaN so that the next min value can be found.
  DistanceMatrix(r,c)=NaN;
  DistanceMatrix(c,r)=NaN;
  % create matrix that stores eruption numbers and the minimum value
```

```
  AllMatches = [AllMatches;Data(r,1:2),Data(c,1:2),minvalue];
end

% remove bad geochem from list of all matches.
for i = 1:size(InconsistentGeochemMatch,1)
  Locator = find(ismember(AllMatches(:,1:4),InconsistentGeochemMatch(i,:),'rows'));
  AllMatches(Locator,:) = [];
end

% Remove any 2 way matches that contain members of the prior matches (Marker horizons) from
    the list of AllMatches
ReshapePrior = reshape(nonzeros(PriorMatches'),2,[])';
for k = 1:size(ReshapePrior,1)
  [r,c]=find(ismember(AllMatches(:,1:2:3),ReshapePrior(k,1)).*ismember(AllMatches(:,2:2:4),
    ReshapePrior(k,2)));
  AllMatches(r,:)=[];
end
% AllMatches denotes the list of all plausible two way tephra matches.

%% Create the initial MatchMatrix of the form [ChiStat, Age, Sdev, Source1,ErupNo1, Source2,
     ErupNo2, ....]
MatchMatrixOld = zeros(length(Data),2*nCores+3); % initialize as zeros
Pointer = 1;
for i = 1:nCores % for each data location
  ToReplace = Data(Data(:,1)==i,:);  % extract the age data - of the form [SourceNo ErupNo
    MeanAge Stdev]
  NewPointer = length(ToReplace)-1+Pointer;
  % replace the correct rows in the MatchMatrixOld with age data, source and erupNo's for
    the given location
  MatchMatrixOld(Pointer:NewPointer,[2 3 i*2+2 i*2+3]) = ToReplace(:,[3 4 1 2]);
  Pointer = NewPointer+1;
end
MatchMatrixOld = sortrows(MatchMatrixOld,[2 3]); % sort MatchMatrix according to mean then
    stdev
AICold = sum(MatchMatrixOld(:,1))+2*length(MatchMatrixOld);   %compute the initial AIC
    value.

%% Force the prior matches / marker horizons to be joined
for k = 1:size(PriorMatches,1) % for each known prior match (or match to be forced)
  ReshapePrior = reshape(nonzeros(PriorMatches(k,:))',2,[])'; % reshape the selected prior
    match as a matrix with 2 columns (SourceNo and ErupNo)
  locatePriorPoint = cell(1,size(ReshapePrior,1)); % initialize an empty cell of length =
    the number of events in the selected prior match
  for i = 1:size(ReshapePrior,1) % for each event in that selected prior match.
    locatePriorPoint{i} = find(ismember(MatchMatrixOld(:,ReshapePrior(i,1)*2+2:ReshapePrior(
    i,1)*2+3),ReshapePrior(i,:),'rows')); % locate which row of the current MatchMatrix the
     event lies in
  end
```

```
  TempMatchMatrix = MatchMatrixOld([locatePriorPoint{:}],:); %extract those rows of the
    current match matrix into a temporary match matrix

  Means = TempMatchMatrix(:,2); % vector of the meanAges
  Sdevs = TempMatchMatrix(:,3); % vector of the Stdev's
  x = Means./Sdevs.^2;
  Ap = sum(x(¬isnan(x)))./(sum(1./Sdevs(Sdevs>0).^2)); % calculate the new combined age
  SdevAp = sqrt(1./(sum(1./Sdevs(Sdevs>0).^2))); % calculate the new stdev of this combined
    age
  ChiStat = sum(((Means(Means>0) - Ap).^2)./Sdevs(Sdevs>0).^2); % calculate the chistat. =
    sum( (Ai-Ap)^2 / Si^2)

  % replace the old rows of the current MatchMatrix with one single row that has the
    selected prior match events joined
  MatchMatrixOld(min([locatePriorPoint{:}]),:) = [ChiStat Ap SdevAp PriorMatches(k,:)];
  MatchMatrixOld(setdiff([locatePriorPoint{:}],min([locatePriorPoint{:}])),:)=[];
end

MatchMatrixOld = sortrows(MatchMatrixOld,[2 3]); % sort MatchMatrix according to mean then
    stdev
% sort the rows in the MatchMatrix so that each core has ascending erupNo's
for i = 5:2:size(MatchMatrixOld,2)
  Indx = find(MatchMatrixOld(:,i)≠0);
  MatchMatrixOld(Indx,:) =  sortrows(MatchMatrixOld(Indx,:),i);
end
AICold = sum(MatchMatrixOld(:,1))+2*length(MatchMatrixOld); % recalculate AIC

%% initialize the best match matrix and best AIC equal to the current match and AIC
AICbest = AICold;
MatchMatrixBest = MatchMatrixOld;

%% Find new possible match....
tabuList = zeros(maxTabuListSize,1); % intialize the tabuList as empty
AcceptedWorse = 1; %initialize stopping condition for annealing schedule. Should always be
    set to one!
NewRun = 1; % the number of times the annealing constant has been updated is initialized as
    1.

StoreAICold = [];
StoreAICbest = [];
while AcceptedWorse == 1;
  AcceptedWorse = 0; % this will update to 1 if we accept a worse match during this run
    through with our selected annealing const.
  StoppingCriterion = 1; % initialize the stopping criterion - the number of runs performed
    since the AnnealingConst was last updated
  while StoppingCriterion ≤ maxStoppingCriterion % if the stopping criterion (number of runs
    ) is not more than our prespecified maximum number of runs
```

```
    List = setdiff(1:size(AllMatches,1),tabuList); % find which rows of the AllMatches
    matrix are not in the tabuList
    MatchNo = List(ceil(numel(List)* rand(1))); % select a random MatchNo from those not in
    the tabuList

    if size(nonzeros(tabuList),1)<maxTabuListSize  % if the tabuList is not full
      tabuList(size(nonzeros(tabuList),1)+1) = MatchNo; % then add the selected matchNo to
    the next empty row of the tabuList
    else
      tabuList = [tabuList(2:end);MatchNo]; % otherwise add the selected matchNo to the last
     row of the tabuList and let the matchNo at the top of the tabuList expire
    end
    [AICold,AICbest,MatchMatrixBest] = MatlabFunctions(MatchNo,AICold,AICbest,
    MatchMatrixBest,AnnealingConst); % calculate the rearrangement and what the best AIC is
     in here.
    StoppingCriterion = StoppingCriterion+1; % add 1 to the number of runs performed since
    the AnnealingConst was last updated
    % store the current AIC and the best AIC so far
    StoreAICold = [StoreAICold;AICold];
    StoreAICbest= [StoreAICbest;AICbest];
  end % while stopping criterion < max stopping criterion

  AnnealingConst=AnnealingConst*AnnealingMultiplier; % update the annealing constant.
  NewRun = NewRun+1; % add 1 to the number of times the annealing constant has been updated
end % while accepted worse == 1

%% plot the changes in AIC over time and save the plots and best match matrix
figure(1)
semilogx(1:length(StoreAICold),StoreAICold,'-b')
hold on
semilogx(1:length(StoreAICbest),StoreAICbest,'-r')
xlabel('Iteration number')
ylabel('AIC')
legend('Current arrangement','Best arrangement')
% save the plot of the changes in AIC over time.
print(figure(1),'-djpeg','AICPlot.jpg');
dlmwrite('BestMatchMatrix.txt',MatchMatrixBest,'delimiter','\t', 'newline', 'pc') % save the
     best match matrix
```

C.3 Functions required to run the Matlab code

This section provides the Matlab function file, 'MatlabFunctions.m', required to run the Matlab code presented above.

```
%% MatlabFunctions.m
% This file contains the functions required to run the MatlabCode.m file which implements
    the stochastic local search procedure of Green et al  (2014) "Automated statistical
    matching of multiple tephra records"
function [AICold,AICbest,MatchMatrixBest] = MatlabFunctions(MatchNo,AICold, AICbest,
    MatchMatrixBest,AnnealingConst)

global  nCores Data AllMatches PriorMatches InconsistentGeochemMatch AcceptedWorse minPoint
    maxPoint NewTempMatchMatrix Prior PriorLocation MatchMatrixOld

%% locate which row in the MatchMatrix the match is...
% for the first event in the selected 2 way match
locatePoint1=[];
n1 =2;
while isempty(locatePoint1)
  n1 = n1+2;
  locatePoint1 = find(ismember(MatchMatrixOld(:,n1:n1+1),AllMatches(MatchNo,1:2),'rows'),1);
end

% for the second event in the selected 2 way match
locatePoint2=[];
n2 =2;
while isempty(locatePoint2)
  n2 = n2+2;
  locatePoint2 = find(ismember(MatchMatrixOld(:,n2:n2+1),AllMatches(MatchNo,3:4),'rows'),1);
end
minPoint = min([locatePoint1 locatePoint2]); % locate smallest rowNo
maxPoint = max([locatePoint1 locatePoint2]); % locate largest rowNo

% locate Data is a list of all observations between locatePoint1 and 2 in the form of a 2
    column matrix [SourceNo ErupNo] this specifies the chunk of observations which we will
    look to rearrange
locateData = reshape(nonzeros(reshape(MatchMatrixOld(minPoint:maxPoint,4:end)',2,[])')
    ',[],2);

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
TempLocateData{1} = locateData;
x=1; % initialize the number of segments to divide the selected chunk into
n = size(locateData,1);
while n > 12  % only segments with less than 12 tephras are manageable
  x = x+1; % if > 12 then need to add one to the number of segments
  GroupSize = floor(minPoint:(maxPoint-minPoint)/x:maxPoint);
```

```
  TempLocateData= cell(1,numel(GroupSize)-1); %initialize temporary locate data
  TempLocateData{1} = reshape(nonzeros(reshape(MatchMatrixOld(minPoint:GroupSize(2),4:end)
    ',2,[])')',[],2);
  n = size(TempLocateData{1},1);
  for i = 2:numel(GroupSize)-1
    TempLocateData{i} = reshape(nonzeros(reshape(MatchMatrixOld(GroupSize(i)+1:GroupSize(i
    +1),4:end)',2,[])')',[],2);
    n = max([n size(TempLocateData{i},1)]);
  end
end

for nChunks = 1:numel(TempLocateData) % for each manageable sized segement
  locateData = TempLocateData{nChunks};
  % locate the MIN row number where the segment starts in the MatchMatrix
  minPoint=[];
  nMin =2;
  while isempty(minPoint)
    nMin= nMin+2;
    minPoint = find(ismember(MatchMatrixOld(:,nMin:nMin+1),locateData(1,:),'rows'),1);
  end
  % locate the MAX row number where the segment ends in the MatchMatrix
  maxPoint=[];
  nMax =2;
  while isempty(maxPoint)
    nMax= nMax+2;
    maxPoint = find(ismember(MatchMatrixOld(:,nMax:nMax+1),locateData(end,:),'rows'),1);
  end
  %% marker horizons / prior matches
  Prior = find(ismember(locateData,reshape(nonzeros(PriorMatches'),2,[])','rows')); %locate
    points in the locateData list that are part of the prior matches
  if ~isempty(Prior) % if there terms in the locateData list that are members of a 'prior
    match'
    locateData(Prior,:)=[]; % remove those points from locateData
    PriorLocation=intersect(find(ismember(MatchMatrixOld(:,4:end),PriorMatches,'rows')),
    minPoint:maxPoint); %locate where in the MatchMatrixOld the PriorMatches lie
  end
  if ~isempty(locateData)&&(size(locateData,1)>1)  % if there are still atleast 2 events to
    consider rearranging, create the temporary match matrix of points to rearrange of the
    form [MeanAge, SdevAge,Source,ErupNo]
    NewTempMatchMatrix=zeros(length(locateData),nCores*2+2);
    for i=1:length(locateData)
      LabelEvent = Data(ismember(Data(:,1:2),locateData(i,:),'rows'),[3 4 1 2]);
      NewTempMatchMatrix(i,[1 2 LabelEvent(3)*2+1 LabelEvent(3)*2+2]) = LabelEvent;
    end
% store if any of the events to consider rearranging have inconsistent geochem matches with
    other events. store which row in the InconsistentGeochemMatch these lie
    rowBadGeoMatch = cell(1,size(NewTempMatchMatrix,1)); %initialize
    for n = 1:size(NewTempMatchMatrix,1) % for each row of the temporary match matrix
```

```
      rowBadGeoMatch{n} = [find(ismember(InconsistentGeochemMatch(:,1:2),nonzeros(
    NewTempMatchMatrix(n,3:end))','rows'));... % check the first event in the
    InconsistentGeochemMatch
      find(ismember(InconsistentGeochemMatch(:,3:4),nonzeros(NewTempMatchMatrix(n,3:end))',
    'rows'))]; % check the second event in the InconsistentGeochemMatch
    end
%% Compile a list of 2way, 3way, 4way etc. rearrangements of events in the temporary match
    matrix
    EndPoint = numel(unique(nonzeros(NewTempMatchMatrix(:,3:2:end))));  % events to consider
     rearranging are from how many unique locations?
    Starting= cell(1,EndPoint); %initialize
    for i=1:EndPoint
%run through "funcUnique.m" to remove inconsistent geochem matches and matches consisting
    from members from the same source
      Starting{i} = funcUnique(nchoosek(1:size(locateData,1),i), rowBadGeoMatch);
    end
    emptyCells = cellfun(@isempty,Starting);  % remove any empty cells from Starting cell
    array
    Starting(emptyCells) = [];
    EndPoint = size(Starting{end},2); %max possible match will be equal to the number of
    unique locations in the temp match matrix — or less

% 'Starting' will be a cell array of all possible combinations size 1 by EndPoint where
    EndPoint is the length of the largest allowable match. i.e if the largest allowable
    match is a 4way match and there are 4 separate events to match we would have:
% Starting{1}   Starting{2}  Starting{3}  Starting{4}
% 1          1 2       1 2 3       1 2 3 4
% 2          1 3       1 2 4
% 3          1 4       1 3 4
% 4          2 3       2 3 4
%            2 4
%            3 4
%  where for example 2 3 shows that the we could consider matching the 2nd and 3rd events in
     the TempMatchMatrix
%% Create a list 'Mat' of all possible rearrangements to consider for this chunk i.e. 2 way,
     3 way, 4way, ..., 2*2 way matches, 2*3 way matches, 2—2—3 match etc...
    v = repmat([0 2:EndPoint],1,floor(size(locateData,1)/2));
    if v==0
      Mat = zeros(0,floor(size(locateData,1)/2));
    else
      Mat = nchoosek(v,floor(size(locateData,1)/2));
    end
    Mat = unique(Mat(sum(Mat,2)≤size(locateData,1),:),'rows'); %remove ones for which the
    sum is > than the total number of ob's to match
    Mat = Mat(¬ismember(Mat,zeros(1,size(Mat,2)),'rows'),:); % remove any zero rows.
    for k = 2:size(Mat,2) % for each column in Mat
       Mat = Mat(Mat(:,k)≥Mat(:,k—1),:); %rows must increase. i.e. 2—2—3 allowed but 2—3—2
    not.
```

```
    end
    Mat = [[zeros(1,size(Mat,2)-1) 1];Mat]; % include a single zero row to allow for all
    events to be kept separate / not joined.
% mat is a vector showing all the possibilities.
% i.e. for size of locate data == 5 with length(unique(locateData))==5
% Mat would show:
%    0   2    i.e. one 2 way match and all other events kept separate
%    0   3    i.e. one 3 way match and all other events kept separate
%    0   4    i.e. one 4 way match ''
%    0   5    i.e. one 5 way match ''
%    2   2    i.e. two 2 way matches ''
%    2   3    i.e. one 2 way match and a 3 way match. ''

    BestCombination=[];  %initialize best combination as empty
    BestAICCombination = Inf; % initialize best AIC as something really large.
% first consider the null match where all terms are kept separate - run through the
    rearranegement function "funcMatch.m"
    [BestAICCombination,BestCombination] = funcMatch(BestAICCombination,BestCombination,
    nonzeros(Mat(1,:)));

%% for all other rows of Mat we need to loop over all possible combinations e.g in the above
     example where length of locate data == 5
% for row 5 of 'Mat' [2 2] we could have the arrangements:
% 1 2   -   3 4  i.e. 2x2way matches where the 1st and 2nd ob's are joined, so are the 3rd
    and 4th ob's in TempMatchMatrix
% 1 2   -   3 5
% 1 2   -   4 5
% 1 3   -   2 4
% 1 3   -   2 5
% 1 3   -   4 5
% 1 4   -   2 3
% 1 4   -   2 4
% 1 4   -   2 5
% 1 5   -   2 3
% 1 5   -   2 4
% 1 5   -   3 4
% 2 3   -   4 5
% 2 4   -   3 5
% 2 5   -   3 4
% we run a function called "functionRemoveDuplicates.m" so that 1 2 - 3 4 is considered the
    same as 3 4 - 1 2 to avoid considering the same arrangement more than once.
    for i=2:size(Mat,1) % for all other rows of Mat (the matrix showing the possible
    combinations)
      NonZeroMat = nonzeros(Mat(i,:)); % remove the zeros
% find StartingNum - a cell array that shows the number of events to match
% i.e for row 5 of 'Mat' [2 2] we would have StartingNum = [2] [2] - indicating 2 x 2 way
    matches and starting would be of the form StartWith = [Starting{2}] [Starting{2}] where
     each cell of StartWith shows all possible arrangements of 2 way matches. eg
```

```
%1 2 joins the 1st and 2nd events in the TempMatchMatrix
%1 3 joins the 1st and 3rd events in the TempMatchMatrix
%1 4 joins the 1st and 4th events in the TempMatchMatrix
%...
%3 4 joins the 3rd and 4th events in the TempMatchMatrix
      StartingNum=cell(1,numel(NonZeroMat));%initalize.
      StartWith=cell(1,numel(NonZeroMat));%initalize.
      for n = 1:numel(NonZeroMat)
        StartingNum{n} = NonZeroMat(n);
        StartWith{n} = Starting{StartingNum{n}};
      end
      %% remove the duplicates.
      N = [StartingNum{:}]; % Transform StartingNum into a vector not a cell array
      Len = size(locateData,1);  % let Len be the size of the chunk to rearrange
      Uni = unique(N);  % Uni gives the unique terms in StartingNum
      nUni = numel(Uni); % nUni gives the number of unique terms in StartingNum

% for each unique match size run through the function functionRemoveDuplicates to create a
    cell array 'Output'. Keeping with the above example if N = [2 2] and StartWith = [
    Starting{2}] [Starting{2}] where Starting{2} = [1 2; 1 3; 1 4; 2 3; 2 4; 3 4] then we
    will get Output = [1 2 3 4; 1 3 2 4; 1 4 2 3] where the first row 1 2 3 4 would
    indicate that we have 2x2 way matches — the first of which joins the 1st and 2nd terms
    in TempMatchMatrix and the second of which joins the 3rd and 4th.
      Output = cell(1,nUni);
      for k = 1:nUni
        Index = find(N==Uni(k));
        if numel(Index)==1
          Output{k}= StartWith{Index};
        else
          Output{k} = functionRemoveDuplicates(StartWith{Index});
        end
      end
      Output = Output(¬cellfun('isempty',Output));
      nOutput = numel(Output);

      % now create "MachesCombined" — a list of all possible rearrangements
      if nOutput==nUni   % if the number of cells in Output is still the same as the number
    of unique terms in StartingNum
        if nOutput==1  % and if there is only 1 cell in Output
        % then MatchesCombined = Output
          MatchesCombined = Output{:};
        else
          for k = 1:nOutput-1 % for each cell in "Output" — except the last one
            if size(Output{k},2)+size(Output{k+1},2)== Len
              MatchesCombined = [];
              for j = 1:size(Output{k},1)
                SetDiff = setdiff(1:Len,Output{k}(j,:));
                if any(ismember(Output{k+1},SetDiff,'rows'))
```

```
                MatchesCombined =[MatchesCombined;Output{k}(j,:), SetDiff];
              end
            end % for j
          else % if sum ≠Len
            MatchesCombined = [];
            for j = 1:size(Output{k},1)
              Temp = Output{k+1};
              Temp = Temp(¬any(ismember(Temp,Output{k}(j,:))')',:);
              MatchesCombined = [MatchesCombined;repmat(Output{k}(j,:),size(Temp,1),1)
Temp];
            end %loop over j
          end % if ==Len
          Output{k+1} = MatchesCombined;
        end % loop over k
      end % if numel(Output)==1

      % run the rearrangement to calculate the BestAIC
      MatchToConsider=cell(1,size(StartingNum,2));%initalize.
      for k=1:size(MatchesCombined,1)
        MatchToConsider{1} = MatchesCombined(k,1:StartingNum{1});
        for j = 2:size(StartingNum,2)
          MatchToConsider{j} = MatchesCombined(k,(sum([StartingNum{1:(j-1)}])+1):sum([
StartingNum{1:j}]));
        end
        if ¬isempty(MatchToConsider)
          [BestAICCombination,BestCombination] = funcMatch(BestAICCombination,
BestCombination,MatchToConsider{:});
        end
      end
    end % if numelOutput==nUni
end % for i=1:size(Mat)

% after looping over all possible combinations for that chunk we
% are left with the BestAIC if the rearrangement is accepted and
% BestCombination shows what the rearrangement would look like.
%% Update the new match matrix
if ¬isempty(BestCombination)
  % replace the selected chunk in the old match matrix with what the new 'best'
rearrangement for that chunk is
  MatchMatrixNew = [MatchMatrixOld(1:minPoint-1,:) ; BestCombination ; MatchMatrixOld(
maxPoint+1:end,:)];
  AICnew = sum(MatchMatrixNew(:,1))+2*size(MatchMatrixNew,1); % compute the new AIC
  UpdateProbability = exp(-AnnealingConst*(AICnew-AICold)); % calculate the probability
of accepting this new match
  RandomProbability = rand(1); % draw a random number.
  if AICnew < AICbest % if the new AIC is better than the best AIC then update the
stored best combination and AIC
    AICbest = AICnew;
```

```
        MatchMatrixBest = MatchMatrixNew;
      end

      if AICnew > AICold % if the new AIC is worse than the old AIC
        if UpdateProbability > RandomProbability % then update the stored old combination
    and AIC with some probability = Update Probability
          AcceptedWorse = 1;
          % set the annealing schedule stopping condition  = 1 so that we continue to update
     the annealing constant until we no longer accept a worse match
          AICold = AICnew;
          MatchMatrixOld = MatchMatrixNew;
        end
      else % if the new AIC is better than the old AIC then update the stored old
    combination and AIC
        AICold = AICnew;
        MatchMatrixOld = MatchMatrixNew;
      end % if AICnew < AICold
    end % cannot rearrange without violating constraints.
  end % if there are still atleast 2 events to consider rearranging
end % for nChunks

end % function end
```

```
function [OutMatrix] = funcUnique(InMatrix,rowBadGeoMatch)
% in matrix is the match to consider reducing by removing matches between pairs from the
    same source and pairs with inconsistent geochem.
global NewTempMatchMatrix

AllOk=zeros(size(InMatrix,1),1); %intialize as zeros
for num = 1:size(InMatrix,1)
  % if there are no matches between pairs from the same source then CheckUnique=1
  CheckUnique = (length(unique(nonzeros(NewTempMatchMatrix(InMatrix(num,:),3:2:end))))==
    numel(InMatrix(num,:)));
   % CheckGeo will equal 1 if geochem is consistent
  CombinedGeo = cell2mat(rowBadGeoMatch(InMatrix(num,:))');
  if length(unique(CombinedGeo)) ≠ length(CombinedGeo)
    CheckGeo = 0;
  else
    CheckGeo = 1;
  end
  AllOk(num) = CheckUnique*CheckGeo; %will equal 1 if geochem ok and there are no pairs from
     the same source.
end % looping over possibilities.
OutMatrix = InMatrix(AllOk≠0,:);  % remove any rows from the list that have incompatible
    geochem or would be matching 2 members from the same source
end
```

```
function [BestAICCombinationOut BestCombinationOut] = funcMatch(BestAICCombinationIn,
    BestCombinationIn,varargin)
```

```
% the varargin are the matches to consider i.e. FirstMatch, SecondMatch,ThirdMatch etc.
    check stratigraphy of events is not violated, if ok calculate chi stat
global  MatchMatrixOld NewTempMatchMatrix Prior PriorLocation minPoint maxPoint nCores

%%%%%% initialize the output terms equal to the inputs
BestAICCombinationOut = BestAICCombinationIn;
BestCombinationOut = BestCombinationIn;
% for each Match to consider in the selected chunk - create the match and calculate the
    ChiStat. NewCombination{i} will be of the form [ChiStat Age Stdev Source1 Erup1 Source2
     Erup2, .....]
NewCombination = cell(1,nargin-2); %preallocation
for i=1:nargin-2 %Calculate the ChiStats and set up the NewCombination.
  NewCombination{i} = FunctionChiStat(NewTempMatchMatrix([varargin{i}],:));
end

% check if there are any events that are to be left by themselves (not joint to another
    event)
MissingPoint = setdiff(1:size(NewTempMatchMatrix,1),[varargin{:}]);
NewCombinationMissingPoint = [zeros(numel(MissingPoint),1) NewTempMatchMatrix([MissingPoint
    ],:)];

% create the Combination matrix - each row of this matrix shows events matched (or left by
    themselves if no matches were to be made) of the form [ChiStat Age Stdev Source1 Erup1
    Source2 Erup2, .....]
Combination = [cell2mat(NewCombination');NewCombinationMissingPoint];

% if there were terms in the selected chunk that are from the list of PriorMatches then this
     PriorMatch / marker horizon needs to be added back in our rearrangement of the chunk
if ~isempty(Prior)
  Combination = [MatchMatrixOld(PriorLocation,:); Combination];
end

Combination = sortrows(Combination,[2 3]); % sort MatchMatrix according to mean then stdev.
    sort the rows in the MatchMatrix so that each core has ascending erupNo's
for i = 5:2:size(Combination,2)
  Indx = find(Combination(:,i)~=0);
  Combination(Indx,:) =  sortrows(Combination(Indx,:),i);
end

if ~isequal(Combination(:,4:end),MatchMatrixOld(minPoint:maxPoint,4:end)) % if the new
    rearrangement is not the status quo / the same as the current arrangement
  TrialBestAIC = [MatchMatrixOld(1:minPoint-1,:) ; Combination ; MatchMatrixOld(maxPoint+1:
    end,:)]; % then replace the new rearrangement with the current arrangement
  AICCombination = sum(TrialBestAIC(:,1))+ 2*size(TrialBestAIC,1); % and calculate the AIC
  % check that the  stratigraphy is not violated i.e.
  BadStratig  = checkStratig(Combination); % will = 1 if bad / stratigraphy is violated, 0
    if good.
  if (BadStratig==0)  % if the stratigraphy is not violated
```

```
    if (AICCombination ≤ BestAICCombinationIn)  % and if the AIC using the new arrangement
    is better than AIC of the arrangement that was fed in then make and update to the
    BestCombination.
      BestCombinationOut = Combination;
      BestAICCombinationOut = AICCombination;
    end
  end
end
end % function end
```

```
%% calculate chi statistic function.
function CombinationOut = FunctionChiStat(ConsiderMatches)
% input vector is of form (Source1, Erup1, Age1, Sdev1, ..., etc) this function calculates
    the combined mean age Ap, and new sdev of this combined age Ap and the chistat = sum((
    Ai-Ap)^2/Ei^2); the output vector is of the form [ChiSq Ap Sdev Source1, Erup1, Source2
     Erup2,...]
global nCores

CombinationOut = zeros(1,nCores*2+3);
if size(ConsiderMatches,1)==1  % if tephra does not match to another tephra
   CombinationOut=[0 ConsiderMatches]; % then chi stat is = 0.
else
  Means = ConsiderMatches(:,1); %otherwise calculate the mean ages and stdevs of all tephras
     in a match
  Sdevs = ConsiderMatches(:,2);
  x = Means./Sdevs.^2;
  NumAp = sum(x(¬isnan(x)));  % allows special case where Sdevs are NaN
  DenomAp=sum(1./Sdevs(Sdevs>0).^2);
  Ap = NumAp./DenomAp;  % calculate the pooled mean
  SdevAp = sqrt(1./DenomAp);  % and pooled standard deviation
  % calculate chistat. =sum( (Ai-Ap)^2 / Ei^2)
  ChiStat = sum(((Means(Means>0) - Ap).^2)./Sdevs(Sdevs>0).^2);
  % create the Combination matrix for the selected segement, this is of the form [ChiStat Ap
     SdevAp Source1 ErupNo1, .... SourceN ErupNoN] for each row of the segment
  CombinationOut(1:3) = [ChiStat Ap SdevAp];
  Vals = nonzeros(ConsiderMatches(:,3:2:end));
  for i = 1:numel(Vals)
    CombinationOut(Vals(i)*2+2:Vals(i)*2+3) = nonzeros(ConsiderMatches(:,Vals(i)*2+1:Vals(i)
    *2+2))';
  end
end
end % function end
```

```
%% check stratigraphy function
function BadStratig  = checkStratig(Combination)

BadStratig = 0; % initialize the stratigraphy as good.
for Order = 5:2:size(Combination,2)  % for each core in the record
  if ¬issorted(nonzeros(Combination(:,Order))) % if the eruption numbers are not in order
```

```
    BadStratig=1; % then the stratigraphy is violated
    return  % If you find 1 source with its stratigraphy violated then break out of the
    function — no need to check the rest
  end
end
end   % function end
```

```
function Output = functionRemoveDuplicates(varargin)
%ensures for example that both 12—34 and 34—12 matches are not considered i.e 12—34—56 is
    considered to be the same as 12—56—34 and 34—56—12, 34—12—56, 56—34—12, and 56—12—34.
    This function is created to speed up run time. the input is a cell array StartWith{}
    for example StartWith = [Starting{2}] [Starting{2}] where Starting{2} = [1 2; 1 3; 1 4;
     2 3; 2 4; 3 4] then this function would create MatchesCombined = [1 2 3 4; 1 3 2 4; 1
    4 2 3] where the first row 1 2 3 4 would indicate that we have 2x2 way matches — the
    first of which joins the 1st and 2nd terms in TempMatchMatrix and the second of which
    joins the 3rd and 4th.
Output=[];
for j = 1:nargin—1
  if j>1
    OutputOld = Output;
  else
    OutputOld = varargin{1};
  end
  Output = [];
  for i = 1:size(OutputOld,1)
    NoPoints = max(max(varargin{end}));
    SetDiff = setdiff(1:NoPoints,OutputOld(i,:));
    if ¬isempty(SetDiff)
      if SetDiff==0
        NewSet = zeros(0,size(varargin{j},2));
      else
        if size(varargin{j},2)>size(SetDiff,2)
          NewSet=zeros(0,size(varargin{j},2));
        else
        NewSet = nchoosek(SetDiff,size(varargin{j},2));
        end
      end
      % this is to stop matches not allowed due to matching with self and matching with
    inconsistent geo
      NewSet = NewSet(ismember(NewSet,varargin{j},'rows'),:);
      NewSet = NewSet(OutputOld(i,j*2—1)<NewSet(:,1),:);
      Output = [Output;repmat(OutputOld(i,:),size(NewSet,1),1) NewSet];
    end
  end
end % cycle over nargin
end % end the function
```

Appendix D

Data used in Chapter 5

Table D.1: Resulting arrangement of tephra thicknesses and estimated ages across the six unexposed cores: Lake Umutekai (Um), Lake Rotokare (Ro), Near Source (NS), Eltham Swamp (El), Lake Rangatauanui (Ra) and Auckland (Au). There is no tephra number 27 in the Umutekai and Rotokare records and no tephra number 4 in the Rangatauanui record as these are the rhyolitic 'Stent Ash' sourced from the Taupo Volcanic Centre.

ID i	Tephra	Combined Age	Tephra number						Thickness (mm)					
	Name	(cal yr BP)	Um	Ro	NS	El	Ra	Au	Um	Ro	NS	El	Ra	Au
23		94±0			1						2			
24		154±0			2						5			
25		177±37			3						20			
26		222±48			4						20			
27		272±42			5						20			
28		313±21			6						20			
29		344±27			7						20			
30		376±34			8						120			
31		412±36			9						30			
32		466±23		1	10					1	20			
33		485±12		2	11					2	70			
34		500±10		3	12					4	30			
35		528±13		4	13					1.5	20			
36		559±21		5	14					3	35			
37		610±25		6	15					5	20			
38		737±31		7	16					7	20			
39		841±40		8	17					5	20			
40		992±39			18						20			
41		1376±49	1		19				2		120			
42		1476±43	2						1					
43		1623±31	3	9	20				1	12	150			
44		1710±64	4						4					
45		1831±39	5		21				1		180			
46		1899±36	6		22				1		10			
47		2055±24	7	10			1		2	20			3	
48		2165±44	8						10					
49		2229±53		11						4				
50		2294±35	9	12	23				55	1.5	220			
51		2389±49		13						5				
52		2465±48		14						3				
53		2529±40	10	15					5	6				
54		2601±50		16						1.5				
55		2624±50		17						7				
56		2710±41	11	18					1	30				
57		3021±54	12						0.5					
58		3115±50	13			1			1			40		

Continued on next page

Table D.1 – *Continued from previous page*

ID i	Tephra	Combined Age	Tephra number						Thickness (mm)					
	Name	(cal yr BP)	Um	Ro	NS	El	Ra	Au	Um	Ro	NS	El	Ra	Au
59		3152±50	14			2			1			15		
60		3232±39	15	19		3			0.5	14		5		
61		3312±47	16			4			1.5			40		
62		3539±34	17				2		1				3	
1	Manganui.c	3577±36	18	20		5			2	3		70		
63		3635±58	19						3					
64		3663±61	20						2					
65		3754±33	21			6	3		3			15	6	
66		3866±53	22			7			0.5			20		
67		3931±45	23	21					0.5	15				
68		3986±44	24	22					2	1				
2	Inglewood.b	4019±44	25	23					30	33				
69		4082±39	26	24					2	5.5				
70		4121±48		25						5				
71		4133±47		26						2.5				
72		4346±46	28	28		8			1	5		50		
73		4450±55	29			9			2			60		
74		4543±47	30			10			1			10		
75		4664±42	31			11			1			80		
76		4746±40	32			12			2			35		
3	Korito.b	4804±42	33			13			10			12		
15	Mangatoki.a	4840±43	34			14			40			15		
77		4884±43	35			15			1			2		
78		4932±43	36			16			4			2		
79		4949±68	37						1					
80		4978±70	38						1					
81		5066±75	39						7					
82		5101±76	40						3					
83		5135±78	41						1					
84		5159±79	42						1					
85		5225±81	43						2					
4	Tariki.f	5344±42	44	29			5		2	34			3	
5	Tariki.e	5412±41	45	30		17			1	4		40		
86		5478±85	46						1					
87		5586±65	47	31					4	4				
6	Waipuku	5664±46	48			18			1			30		
88		5739±63	49	32					0.5	3				
89		5848±82	50						0.5					
90		5870±43	51	33		19			1	4		1		
91		5945±62	52	34					3	2				
92		6037±62	53	35					0.5	16				
93		6245±45	54	36		20			1	11		4		
94		6300±60	55	37					2	6				
95		6414±61		38		21				9		10		
96		6551±50	56			22			1			4		
97		6650±44	57	39		23			1.5	13		6		
98		6686±66	58						0.5					
99		6706±60	59				6		0.5				7	
100		6770±65	60						5					
101		6777±65	61						1					
102		6847±38	62	40		24			1	18		44		
7	Kaponga.f	6944±36	63	41		25			2	20		60		
103		7016±42	64	42		26			3	5		7		
104		7148±66				27						5		
105		7308±42	65			28	7		2			2	20	
106		7380±43	66			29			1			112		
107		7548±46				30	8					32	19	
108		7572±50				31						20		
109		7613±50				32						40		
110		7624±50				33						31		
111		7634±50				34						72		
112		7656±50				35						10		
113		7668±50				36						24		

Continued on next page

Table D.1 – *Continued from previous page*

ID i	Tephra	Combined Age	Tephra number						Thickness (mm)					
	Name	(cal yr BP)	Um	Ro	NS	El	Ra	Au	Um	Ro	NS	El	Ra	Au
114		7787±41	67			37			20			16		
115		7931±44	68			38			0.5			5		
116		8004±45	69			39			1			12		
117		8082±34	70			40	9		2			22	10	
118		8162±47	71			41			2			13		
119		8242±48	72			42			4			6		
120		8311±48	73			43			8			15		
121		8398±46				44	10					40	16	
122		8478±47	74			45			2			11		
123		8559±47	75			46			5			8		
124		8600±47	76			47			1			32		
125		8760±60	77						2					
126		8804±59	78						2					
127		8828±59	79						2					
128		8840±59	80						1					
129		8956±44	81			48			0.5			45		
130		9022±57	82						0.5					
131		9056±56	83						3					
132		9087±56	84						0.5					
133		9205±55	85						4					
134		9236±55	86						2					
135		9330±43	87			49			2			40		
136		9460±11				50	11					8	2	
137		9589±73				51						4		
138		9658±74				52						10		
139		9754±44	88			53			0.5			8		
140		9813±44	89			54		1	0.5			6		2
141		9904±46	90			55			6			7		
142		9967±15	91			56	12		4			5	10	
143		10020±82				57						3		
144		10046±48	92			58			5			18		
145		10076±17				59	13					6	6	
146		10128±85				60						4		
147		10142±85				61						8		
148		10198±18				62	14					5	8	
149		10220±17	93			63	15		0.5			16	6	
150		10284±18	94			64	16		2			10	6	
151		10304±19	95			65	17		45			17	3	
152		10419±57	96			66			0.5			6		
153		10474±59	97			67			5			21		
154		10588±61	98			68			1			56		
16	Kaponga.d	10738±69				69	18					122	2	
155		10813±57	99			70	19		4			29	14	
156		10884±101				71						80		
157		10913±71	100			72			2			5		
158		11176±121	101						4					
159		11299±131	102						3					
160		11400±140	103						2					
161		11459±146	104						2					
162		12052±24				73	20	2				4	2	1.8
163		12099±26				74	21					2	2	
164		12219±116				75						22		
165		12384±122				76						66		
8	Konini.b	12354±31				77	22					32	10	
166		12482±120				78		3				18		1.5
167		12826±145				79						25		
9	Mahoe.a	13036±132				80		4				23		1.8
168		13351±174				81						20		
169		13956±190				82		5				35		1
170		14205±214				83						4		
171		14302±186				84		6				20		0.8
172		14570±225				85						8		
173		15241±229				86						7		

Continued on next page

Table D.1 – *Continued from previous page*

ID i	Tephra	Combined Age	Tephra number						Thickness (mm)					
	Name	(cal yr BP)	Um	Ro	NS	El	Ra	Au	Um	Ro	NS	El	Ra	Au
174		15564±209				87		7				4		1
175		16122±152				88	23	8				15	2	1
176		17093±143				89	24					20	1	
177		17522±158				90		9				10		2.6
178		17673±208				91						22		
179		17805±148				92	25					10	1	
180		18014±210				93						8		
181		18079±211				94						62		
17	Kaihouri.h	18316±158				95	26					35	18	
182		18523±216				96						10		
183		19079±171				97		10				12		4.3
184		19434±180				98	27					15	7	
185		19791±186				99	28					29	10	
186		20090±192				100	29					34	14	
187		20738±176				101	30	11				28	11	2
188		21288±169				102	31	12				34	10	1.7
189		21798±199				103	32					16	10	
10	Paetahi.a	22770±174				104	33	13				31	20	2
190		23380±201				105	34					10	10	
191		23575±161				106	35	14				10	11	3.5
192		24290±209					36	15					11	2
193		24592±256					37	16					8	1.5
18	Poto.a	25637±252				107						45		
194		25861±255				108						17		
195		27707±395						17						0.5
196		28153±339						18						1.5
197		28734±272				109		19				13		2
11	Tuikonga.d	29030±293				110						74		
198		29160±209				111		20				5		4
199		29309±217				112						27		
200		29358±202				113						5		
201		29402±191				114						5		
202		29432±186				115						3		
203		29463±183				116						20		
204		29506±183				117						17		
205		29507±184				118						8		
12	Koru.a	29522±185				119						112		
206		29532±186				120						4		
207		29545±188				121						5		
208		29554±190				122						4		
209		29570±194				123						21		
210		29589±198				124						22		
211		29622±207				125						12		
212		29653±217				126						2		
213		29697±231				127						5		
214		29854±284				128						22		
215		29891±296				129						8		
216		29928±308				130						15		
217		30046±341				131						56		
218		30233±300				132		21				41		3
219		30273±393				133						15		
220		30532±436				134						4		
221		30685±455				135						3		
222		30782±465				136						5		
223		30874±473				137						12		
224		31040±485				138						21		
225		31241±495				139						5		
13	Pukeiti	31725±505				140						85		
226		31910±398				141		22				5		0.5
227		32114±502				142						6		
228		32292±394				143		23				4		1
14	Waitepuku.a	32456±495				144						40		
229		32648±491				145						5		

Continued on next page

Table D.1 – *Continued from previous page*

ID i	Tephra	Combined Age	Tephra number						Thickness (mm)					
	Name	(cal yr BP)	Um	Ro	NS	El	Ra	Au	Um	Ro	NS	El	Ra	Au
230		32746±488				146						3		
231		32845±485				147						69		
232		32982±481				148						35		
233		33084±479				149						5		
234		33181±476				150						7		
235		33430±470				151						4		
19	Mangapotoa.a	33679±466				152						150		
236		33801±329				153		24				18		1.5
237		33914±376				154		25				8		1
238		34042±463				155						22		
239		34219±463				156						14		
240		34246±464				157						2		
241		34418±298				158		26				18		1
242		34544±468				159						16		
243		34750±474				160						11		
244		34865±478				161						13		
245		34945±482				162						58		
246		35070±488				163						9		
247		35099±490				164						2		
248		35158±493				165						7		
249		35307±502				166						8		
250		35450±512				167						12		
251		35541±519				168						25		
252		35671±530				169						4		
253		35771±539				170						5		
254		35951±556				171						12		
255		36014±562				172						6		
256		36108±572				173						10		
257		36156±577				174						8		
258		36639±623				175						30		
259		36644±623				176						168		
260		37169±414				177		27				50		0.8
261		38127±632						28						1
262		40505±751						29						0.8
263		43241±728						30						1.3
264		43662±925						31						0.8
265		43886±935						32						1.5
266		44611±1055						33						2
267		44920±1124						34						2.5
20	Waitui	51189±1793						35						4
268		55999±3734						36						3.3
21	Araheke	58524±3609						37						3
269		59301±3793						38						1.7
22	Te Arei	67161±6766						39						3
270		75467±12016						40						1

Appendix E

WinBUGS code to accompany Chapter 5

```
model{

for (j in 1:nSites){
for (i in 1:nEvents){
thick[i,j] ~ dlnorm(mu[i,j],0.3156)I(0,cens[i,j])

mu[i,j] <- log(0.5*g[i,j]*exp(f1[i,j])+(1-0.5)*g[i,j]*exp(f2[i,j]))
g[i,j] <- exp(a[j])*pow(V[i],(c[i]+1)/3)/pow(r[j]+d[i]*pow(V[i],1/3),c[i])

f1[i,j] <- alpha1[i]*cos(theta[j]-phi1[i])+beta1[i]*cos(2*(theta[j]-phi1[i]))
f2[i,j] <- alpha2[i]*cos(theta[j]-phi2[i])+beta2[i]*cos(2*(theta[j]-phi2[i]))
}
}

 ##############################################################################################
##### DATA #####
# cens[i,j] = censoring variable,
# cens[i,j] = 0.05 for 'missing' tephras in the unexposed records, cens[i,j] = 100 otherwise
# r[j] = distance (in km) of location j from the vent
# theta[j] = angular direction (radians) of location j from the vent
# thick[i,j] = thickness (in cm) of event i observed at location j
# q[i]  = probability of observing two lobes
 ##############################################################################################

#####
# prior for the presence of 2 lobes
for (i in 1:nEvents){
X[i] ~ dbern(q[i]) # if x = 0 one lobe, if x = 1 two lobes.
}

#####
# prior for the site-specific effect
# for the unexposed locations
for (j in 1:6){
a[j] ~ dnorm(0,0.0001)
}
# for the exposed locations
for (j in 7:nSites){
a[j] ~ dnorm(5.81150,306.78)
}

#####
# prior for the volume V[i] (in cubic km)
# for the named events
V[1] ~ dlnorm(0.27275,46.391)
V[2] ~ dlnorm(0.54401,81.503)
V[3] ~ dlnorm(0.36218,94.199)
V[4] ~ dlnorm(-1.328,53.477)
```

```
V[5] ~ dlnorm(0.026265,46.547)
V[6] ~ dlnorm(0.10067,81.708)
V[7] ~ dlnorm(-0.17255,62.272)
V[8] ~ dlnorm(-0.20823,54.539)
V[9] ~ dlnorm(0.28387,36.457)
V[10] ~ dlnorm(0.40912,26.467)
V[11] ~ dlnorm(0.34426,31.393)
V[12] ~ dlnorm(0.063743,65.386)
V[13] ~ dlnorm(-0.70093,22.645)
V[14] ~ dlnorm(-0.65574,20.944)

# for the named events that do not have isopach diagrams
for (i in 15:22){
Vbeta[i] ~ dbeta(1.6538,2.3274)
V[i] <- Vbeta[i]*2   # correction to ensure between 0 and 2 cubic km.
}

# for the unnamed tephras
for (i in 23:nEvents){
V[i] ~ dbeta(1.249767834,25.727016156)
}

#####
# prior for c[i]
# for the named events
for (i in 1:14){
clog[i] ~ dlnorm(-1.6549, 37.2806)
c[i] <- clog[i]+2 # constraint c > 2
}

# for the unnamed events
for (i in 15:nEvents){
clog[i] ~ dlnorm(-1.160,2.604)
c[i] <- clog[i]+2 # constraint c > 2
}

#####
# prior for d[i] -included to ensure the thickness is finite at the vent.
# for the named events
for (i in 1:14){
d[i] ~ dlnorm(-0.3334,5.8511)
}

# for the unnamed events
for (i in 15:nEvents){
d[i] ~ dlnorm(0.1798,1.1485)
}

#####
# prior for the wind parameters alpha1[i] alpha2[i] beta1[i] beta2[i]
for (i in 1:nEvents){
V1[i] ~ dgamma(0.9902,0.8280)   # variable for difference alpha1 - 4*beta1
V2[i] ~ dgamma(0.9902,0.8280)   # variable for difference alpha2 - 4*beta2
W1[i] ~ dgamma(3.4759,7.3736) # variable for beta1
W2[i] ~ dgamma(3.4759,7.3736) # variable for beta2

alpha1[i] <- 4*W1[i]+V1[i] # constraint alpha > 4beta
beta1[i] <-  W1[i]

# if there is only 1 lobe (X = 0) then alpha2 = alpha1 and beta2 = beta1.
alpha2[i] <- (1-X[i])*alpha1[i]+ X[i]*(4*W2[i]+V2[i])
beta2[i] <- (1-X[i])* beta1[i] + X[i]* W2[i]
}

#####
# prior for the wind directional parameters phi1[i] and  phi2[i]
# for the named tephras
angle1[1] ~ dbeta(124.71,67.257)
angle1[2] ~ dbeta(161.82,59.233)
angle1[3] ~ dbeta(138.05,66.439)
angle1[4] ~ dbeta(77.457,61.968)
angle1[5] ~ dbeta(216.17,148.77)
angle1[6] ~ dbeta(55.932,23.366)
angle1[7] ~ dbeta(203.39,114.79)
angle1[8] ~ dbeta(118.53,118.53)
angle1[9] ~ dbeta(42.2,42.2)
angle1[10] ~ dbeta(60.078,41.581)
angle1[11] ~ dbeta(73.253,25.618)
angle1[12] ~ dbeta(6.6709,1.6247)
angle1[13] ~ dbeta(24.088,4.9764)
```

```
angle1[14] ~ dbeta(49.224,20.966)

angle2[1] ~ dbeta(13.499,29.849)
angle2[2] ~ dbeta(12.604,26.239)
angle2[3] ~ dbeta(22.171,39.213)
angle2[4] ~ dbeta(9.4297,11.305)
angle2[5] ~ dbeta(12.388,21.287)
angle2[6] ~ dbeta(13.503,31.891)
angle2[7] ~ dbeta(1,1) # one lobe for Kaponga F
angle2[8] ~ dbeta(5.6573,35.927)
angle2[9] ~ dbeta(16.099,39.632)
angle2[10] ~ dbeta(30.627,79.436)
angle2[11] ~ dbeta(1,1) # one lobe for Tuikonga
angle2[12] ~ dbeta(6.0242,13.413)
angle2[13] ~ dbeta(1,1) # one lobe for Pukeiti
angle2[14] ~ dbeta(1,1)   # one lobe for Waitepuku

# for the unnamed tephras
for (i in 15:nEvents){
angle1[i] ~ dbeta(2.0909,2.0776) # beta distribution on interval 0 - 1
angle2[i] ~ dbeta(2.0909,2.0776)
}

# convert beta distributions to -pi/2 - pi/2 interval.
for (i in 1:nEvents){
phi1[i] <- (angle1[i]-0.5)*(3.14159)

# create temporary variable for the absolute difference phi1 - phi2
Y[i] <-abs( (angle1[i]-0.5)*(3.14159) -(angle2[i]-0.5)*(3.14159))

# for only one lobe (X = 0) phi2 = phi1, for two lobes (X = 1) phi2 is some angle Y from phi1.
# If phi1 > 0 then phi2 is south of phi1, If phi1 < 0 then phi2 is north of phi1.
phi2[i] <- phi1[i]+(-step(phi1[i])+(1-step(phi1[i])))*Y[i]*X[i]
}

}
```

Appendix F

Statement of contribution to doctoral thesis containing publications

MASSEY UNIVERSITY
GRADUATE RESEARCH SCHOOL

STATEMENT OF CONTRIBUTION TO DOCTORAL THESIS CONTAINING PUBLICATIONS

(To appear at the end of each thesis chapter/section/appendix submitted as an article/paper or collected as an appendix at the end of the thesis)

We, the candidate and the candidate's Principal Supervisor, certify that all co-authors have consented to their work being included in the thesis and they have accepted the candidate's contribution as indicated below in the *Statement of Originality*.

Name of Candidate: Rebecca Green

Name/Title of Principal Supervisor: Mark Bebbington

Name of Published Research Output and full reference:

Green, R., Bebbington, M., Cronin, S., and Jones, G. (2013). Geochemical precursors for eruption repose length. Geophysical Journal International, 193:855-873

In which Chapter is the Published Work: 3

Please indicate either:

- The percentage of the Published Work that was contributed by the candidate:

  and / or

- Describe the contribution that the candidate has made to the Published Work:

All three co-authors were involved in the editing of the manuscript. Additional contributions by specific co-authors include: Mark Bebbington, for advice with model development, for providing the multi-source and single-source sample catalogs in Appendix B.2 and for the construction of Figure 3.10. Shane Cronin, for providing Figure 3.1 (the map showing the location of Lake Umutekai and Lake Rotokare) and for detailed description of the more geochemical aspects of this work. In particular, for the interpretation of titanomagnetite and its sensitivity to details of magma storage times and conditions (see Section 3.2) and discussion of the way in which geochemical predictors operate on the repose times (see Section 3.5, Case 1 and Case 2 magmas). Geoff Jones, for general advice on proportional hazard models. The remainder of the work was carried out by Rebecca, including performing the analyses and writing the paper.

Rebecca Green
Digitally signed by Rebecca Green
DN: cn=Rebecca Green, o=IFS, ou=Massey University, email=bex.green@hotmail.com, c=NZ
Date: 2014.12.08 09:16:32 +13'00'

Candidate's Signature — 8/12/2014 Date

Mark Bebbington
Digitally signed by Mark Bebbington
DN: cn=Mark Bebbington, c=NZ, o=Massey University, ou=IFS-Statistics, email=m.bebbington@massey.ac.nz
Date: 2014.12.08 09:47:40 +13'00'

Principal Supervisor's signature — 8/12/2014 Date

MASSEY UNIVERSITY
GRADUATE RESEARCH SCHOOL

STATEMENT OF CONTRIBUTION TO DOCTORAL THESIS CONTAINING PUBLICATIONS

(To appear at the end of each thesis chapter/section/appendix submitted as an article/paper or collected as an appendix at the end of the thesis)

We, the candidate and the candidate's Principal Supervisor, certify that all co-authors have consented to their work being included in the thesis and they have accepted the candidate's contribution as indicated below in the *Statement of Originality*.

Name of Candidate: Rebecca Green

Name/Title of Principal Supervisor: Mark Bebbington

Name of Published Research Output and full reference:

Green, R., Bebbington, M., Cronin, S., and Jones, G. (2014). Automated statistical matching of multiple tephra tephra records exemplified using give long maar sequences younger than 75 ka, Auckland, New Zealand. Quaternary Research, 82(2):405-419

In which Chapter is the Published Work: 4

Please indicate either:

- The percentage of the Published Work that was contributed by the candidate:

  and / or

- Describe the contribution that the candidate has made to the Published Work:

All three co-authors were involved in the editing of the manuscript. Additional contributions by specific co-authors include:
- Mark Bebbington, for advice with model development and discussion of the implications for hazard from the AVF (Section 4.5.5).
- Shane Cronin, for interpretation of the preservation of Orakei Basin tephras (see Section 4.5.2) .
- Geoff Jones, for suggestions regarding the model to quantify the variability in observance probabilities (see Section 4.5.3) .
The remainder of the work was carried out by Rebecca, including assembling and cleaning the data, developing the heuristic, performing the analyses and writing the paper.

Rebecca Green — Digitally signed by Rebecca Green DN: cn=Rebecca Green, o=IFS, ou=Massey University, email=bex.green@hotmail.com, c=NZ Date: 2014.12.08 09:18:10 +13'00'

Candidate's Signature — 8/12/2014 — Date

Mark Bebbington — Digitally signed by Mark Bebbington DN: cn=Mark Bebbington, c=NZ, o=Massey University, ou=IFS-Statistics, email=m.bebbington@massey.ac.nz Date: 2014.12.08 09:48:06 +13'00'

Principal Supervisor's signature — 8/12/2014 — Date