

Microseismicity associated with actively exploited
geothermal systems: earthquake detection and
probabilistic location at Rotokawa and statistical
seismic network design at Kawerau

Zara Rawlinson

A thesis submitted to Victoria University of Wellington
in partial fulfilment of the requirements for the degree of
Master of Science in Geophysics.

School of Geology, Environment and Earth Sciences
Victoria University of Wellington
2011

Abstract

Geothermal power has progressively been recognised as an important energy resource due to the depletion of old power sources, and as a more environmentally aware population pushes for an increase in renewable energy sources. Monitoring microseismicity occurring in active geothermal systems is one means of both characterising the system's fault architecture and characterising fluid/rock interaction in response to production. This study focuses on better understanding seismicity in two active geothermal fields, through the development and implementation of two different algorithms: an automated microearthquake detection algorithm using a matched filter technique (improving earthquake detection), and an optimal seismic network design algorithm (improving earthquake location). Both algorithms have been implemented in codes that are easily adaptable to other data sets.

The first of these algorithms has been applied to five months of continuous seismic waveform data spanning a fluid injection operation in the Rotokawa geothermal field. The cross-correlation of 14 high-quality master events with the continuous seismic data yields 2461 newly detected earthquakes spanning the magnitude range $-0.4 \leq M \leq 2.6$ with a mean magnitude of $M = 0.47$. The earthquakes detected with each master event exhibit high waveform similarity over approximately three orders of magnitude, and appear to follow a Gutenberg-Richter power law with a catalogue completeness down to $M \sim 0$.

Hypocentres for these detected events computed using the probabilistic earthquake location algorithm *NonLinLoc* reveal the dominant locus of seismicity to lie between 1.0–2.5 km depth, a location consistent with that of the Rotokawa Andesite which forms the Rotokawa reservoir. Focal mechanism solutions for the master events are predominantly normal, with

half displaying a large strike-slip component, and the stress parameters obtained for this suite of focal mechanisms imply a northeast–southwest oriented maximum horizontal stress: both of these results are consistent with the extensional regime of the TVZ. Seismicity occurring within a 300 m horizontal radius of the injection well’s feed-zones, and extending to 5 km depth, initially exhibits a correlation with injection flow rates with a ~ 2 day lag, and seismicity rates decrease ~ 10 weeks after injection. We surmise that seismicity within the injection region and close to the injection well is likely to be injection-induced, with one portion of the injectate returning to the production region, while the other either migrates southeastward out of the field or remains within the injection region; the origin of seismicity within the production region in relationship to production and injection processes is unclear.

The second of these algorithms involves the derivation of a design criterion, which we apply to inform the expansion of the existing seismic monitoring programme at Kawerau geothermal field; we also apply an early version to the short-term/rapid-response network design following the M7.1 September 2010 Darfield earthquake. Unlike previous seismic network design algorithms, the new algorithm incorporates methods for the realistic representation of 3D velocity structures and attenuation models for both P and S travel times, a surface noise model, and the ability to apply complex weighting functions to the earthquake set. The results demonstrate the utility of this algorithm in even simplistic cases, and show how each new parameter incorporated into the design model affects the optimal network design obtained, identifying the need for accurate input data to provide optimal results.

For Ma & Pa

for always striving to live the dream

Acknowledgements

This work has hinged on the guidance of my supervisors, whose efforts can not be overstated. I gratefully thank John Townend for his support and encouragement throughout both undergraduate and postgraduate; for providing me with research-assistant work during undergrad, without which the computer programming involved in this study would have been a completely new and terrifying prospect; and throughout this study for providing his aggressively thorough editor's pen, for always being patient and available, and maintaining an infectious enthusiasm for geophysics — even when swamped with work. I thank Richard Arnold for being the backbone of the experimental design aspect of this project, for spouting seemingly effortless statistical genius and then slowly explaining it to a non-statistician, for being the source of extremely helpful *R* tips and codes, and for always exuding affability. I thank Stephen Bannister for sharing his seismological and geothermal experience and knowledge throughout this project, and for not only being the first to be excited about me wanting to carry out geothermal power related research, but for then also going out and finding funding for me.

This study has been made possible (and a lot more pleasant) through the generous funding of Mighty River Power (MRP). I gratefully thank Steven Sewell for being my point of contact and the source of much needed data and information. I would also like to thank Joe Gammon, Linda Price, Ben Pezaro and Jeremy O'Brien for the help they have provided me, as well as Tom Powell for his valuable encouragement when this project was first proposed.

From GNS Science I would like to thank Steven Sherburn and Sandra Bourguignon for allowing me access to their Rotokawa data. A special thanks to Andrew Curtis, for providing

an unpublished version of the *Coles and Curtis* [2011] paper, which formed the basis of our experimental design analysis. Of key import is Charlotte Rowe, whom I gratefully thank for providing me with her preliminary codes aimed at waveform cross-correlation.

This project has relied on a number of freely available software packages and I am extremely grateful to all the authors who have made these available. A number of people have aided me in software related issues and I give thanks to them all: Yannik Behr for *Python* help; Carolin Boese for *NonLinLoc* knowledge; Jess Johnson for assistance with *BCSEIS* and *hypoDD*; Mark Henderson for assistance with ray-tracing in *MATLAB*; Euan Smith for codes outlining Jeffreys weighting; Andrew Rae for GIS help; Andrew Mellanby for maintaining the geophysics computer servers; Kevin Buckley for setting me up and assisting me on the MSOR computer grid; and in particular, Adrian Benson, for not only the untold number of times he has assisted me with various small issues related to Linux, Makefiles, Fortran, etc, but also for always being willing to take the time to help, for being company in the office through late nights and weekends, and for letting me play my music loud: I would have thrown my computer out the window long ago without him.

Thank you to all the staff and tutors at SGEES I have had dealings with in my time here, and a big thanks to my fellow graduate students for being around for coffees, beers, and mutual stress-out sessions (you know who you are).

Thank you must go to: Ramona White, for being a friend first, a flatmate second, a fellow-postgrad student third...and for combining all of these into you; Ramona White, Jaya Dias-Wanigasekera and Elaine Brent for making the world a better place with their laughter; Jessica Bruce for being the emotional counterpoint to my stoicism, and for always caring about the little/random things in life; Jarek and Skye for giving the love and wrestles only a dog can provide; and my other unnamed friends for their love and support.

I could never have got this far without Will Aitken from Wellington Osteopaths, who always provides a glimmer of hope through the pain.

And finally, an immeasurable thank you to Andrew Kolodziej for being the constant source of my delight.

Contents

Abstract	i
Acknowledgements	v
Contents	vii
List of Figures	xi
List of Tables	xvii
1 Introduction	1
1.1 Motivation and objectives	1
1.2 Geological, tectonic and seismological setting	5
1.2.1 Taupo Volcanic Zone	5
1.2.2 Rotokawa	8
1.2.3 Kawerau	9
1.3 Seismological investigations undertaken	11
1.3.1 Injection-induced microearthquake detection at Rotokawa	11
1.3.2 Optimal network design at Kawerau	23
1.4 Thesis structure	30
2 Earthquake Detection	33
2.1 Introduction	33
2.2 Waveform cross-correlation	36

2.3	Method	38
2.3.1	Master templates	38
2.3.2	Cross-correlation	43
2.3.3	Network correlation coefficients	50
2.3.4	Detection thresholds	52
2.4	Synthetic testing	56
2.5	Results	62
3	Earthquake Location and Spatiotemporal Analysis	71
3.1	Preliminary location	72
3.1.1	Phase-pick refinement	72
3.1.2	Phase-pick weighting	75
3.1.3	<i>Velest</i>	76
3.2	Double-difference relocation	80
3.3	Probabilistic earthquake location: <i>NonLinLoc</i>	83
3.4	Spatiotemporal analysis	95
3.4.1	Focal mechanisms	95
3.4.2	Relation to injection	103
3.5	Summary	108
4	Statistical Experimental Design	111
4.1	Derivation of the design criterion	112
4.2	Implementation and preliminary testing	118
4.3	Application to Kawerau	124
4.3.1	Uniform velocity case	125
4.3.2	Seismological model	126
4.3.3	Noise sources	145
4.4	When is enough enough?	154
4.5	Final designs	157

4.6	Summary	163
5	Discussion and conclusions	165
A	Signal Processing	171
A.1	Multi-taper method	171
A.2	Coherence filtering	173
B	Code	175
B.1	Convert miniseed to SAC	178
B.2	Create master events	178
B.3	Cross-correlation loop	179
C	Location	189
C.1	Double-difference relocation	189
C.1.1	Cross-correlation delay time estimates	189
C.1.2	<i>hypoDD</i>	192
C.2	<i>NonLinLoc</i> cluster locations	196
C.3	Focal mechanisms	204
C.4	Time slices	208
D	Network design	211
D.1	Marginal and conditional means and variances	211
D.2	Code implementation	212
D.3	Darfield earthquake	226
	Bibliography	231

List of Figures

1.1	Cartoon of a geothermal power plant.	2
1.2	Geological setting of the Taupo Volcanic Zone	6
1.3	Cross-sections through the Rotokawa geothermal field	10
1.4	Conceptual model of the Kawerau geothermal field	12
1.5	Data from the 1962 Rocky Mountain Arsenal well waste fluid injections. . .	14
1.6	Map of Rotokawa	18
1.7	An example of earthquake detection using our wavelet matching software .	20
1.8	Double-difference locations of earthquakes that occurred during the 1997 seismic crisis in the Long Valley caldera	22
1.9	Uncertainties in an earthquake hypocentre for a triangular quadripartite net- work	25
1.10	Basis of <i>Coles and Curtis</i> [2011] optimal network design criterion	28
1.11	GIS data for Kawerau	29
2.1	Four examples of delay time estimation using cross-correlation.	39
2.2	Diagram explaining master event and master template terminology	40
2.3	Master template window size example	41
2.4	The constituents of a single master <i>template</i>	42
2.5	An example of the constituents of a single master <i>event</i>	44
2.6	Representation of our library of 14 master events	44
2.7	Cross-correlation method	45

2.8	Example of Network Correlation Coefficient	51
2.9	Network Correlation Coefficient with a line at $\Omega = 11.76\text{MAD} + \delta$	53
2.10	Example showing the deviation of cross-correlation values from a pure normal distribution	54
2.11	A simple example of taking a padded waveform template and randomising the phases	57
2.12	A simple example of taking a padded waveform template with a sine wave imposed	57
2.13	Master events used in synthetic testing	58
2.14	Average percentage of false detections and missed earthquakes vs. detection factor from a one station network.	60
2.15	Average percentage of false detections and missed earthquakes vs. detection factor from a six-station network	60
2.16	Network Correlation Coefficient with line at the synthetically determined thresholds	62
2.17	Detection results from cross-correlation for one day	64
2.18	Four-minute-long record with 23 detected events	64
2.19	Vertical component for six detected events including the master event	65
2.20	East component for six detected events including the master event	65
2.21	Normalised frequency spectrum for 30 events	66
2.22	Cumulative number of earthquakes vs. magnitude for all detected events.	68
2.23	Log of the cumulative number of earthquakes vs. magnitude.	69
3.1	Phase pick refinement and location flow	72
3.2	Earthquake locations after the fourth <i>Velost</i> run	78
3.3	<i>Velost</i> RMS residuals	79
3.4	<i>hypoDD</i> locations.	82
3.5	V_p 3D velocity model	84
3.6	V_p 3D velocity model: zoom-in on the geothermal reservoir	85

3.7	Vs 3D velocity model: zoom-in on the geothermal reservoir	86
3.8	Example of a 3D PDF for a poorly constrained earthquake.	88
3.9	Example of a 3D PDF for a well constrained earthquake.	88
3.10	Maximum likelihood hypocentres determined using <i>NonLinLoc</i> , scaled and coloured by location uncertainty.	90
3.11	Every earthquake's PDF estimate combined into a PDF and coloured by density	91
3.12	Histograms and empirical cumulative density functions of hypocentre east, north and depth uncertainties.	92
3.13	Histogram and empirical cumulative density function of hypocentre uncer- tainty $\sqrt{\lambda}$	93
3.14	Histogram and empirical cumulative density function of asphericity of the PDF volumes.	94
3.15	Histogram and empirical cumulative density function of hypocentre RMS time residuals.	94
3.16	Histograms displaying the distribution of earthquake locations in the East, North and depth planes.	96
3.17	Earthquakes coloured based on each slave event's respective master event. .	97
3.18	Focal mechanism solutions for the 14 master events.	98
3.19	Geological map of Rotokawa with overlaid earthquakes.	101
3.20	Details of the stress inversion results.	102
3.21	Injection flow rate history and daily number of detected earthquakes.	104
3.22	Earthquake locations and injection well tracks.	105
3.23	Two regions of earthquakes surrounding RK12 and RK20.	106
3.24	Temporal pattern of seismicity for within two different regions.	107
3.25	Production and injection regions at Rotokawa with overlaid earthquakes . .	109
4.1	Cartoon of our network design problem	112
4.2	Schematic illustration of the epicentral PDFs of two earthquakes constrained by P-wave travel time measurements at three stations.	113

4.3	Depiction of the basis of <i>Coles and Curtis</i> [2011] optimal network design criterion	114
4.4	Simple synthetic network design testing	123
4.5	Three different earthquake regions used in Kawerau analysis.	124
4.6	Optimal network design: uniform velocity and attenuation models	127
4.7	Comparison of optimal networks produced by using different Q values	128
4.8	Two different design networks produced by varying the size of the search space.	129
4.9	Kawerau velocity schematic	132
4.10	Kawerau velocity models	133
4.11	Schematic of the upper 2.5 km of the Kawerau velocity model	134
4.12	Q factor versus temperature from <i>Jaya et al.</i> [2010].	136
4.13	Empirical estimate of Q factor versus temperature.	136
4.14	Kawerau temperature cross-section	137
4.15	Percentage difference between travel times provided by the analytic solution and ray tracing	138
4.16	Optimal seismic network: analytical solution and ray tracing through uniform velocity and attenuation models.	140
4.17	Comparison of optimal networks produced by using analytical solution and ray tracing.	141
4.18	Optimal network design: 3D P-wave velocity model and uniform attenuation.	142
4.19	Optimal network design: 3D P-wave velocity and attenuation models with a varying attenuation scaling.	144
4.20	Optimal network design: 3D P- and S-wave velocity and attenuation models.	146
4.21	Noise levels at Kawerau (dBA)	152
4.22	Optimal network design: including surface noise model.	155
4.23	Optimal seismic stations: production region and local region.	158
4.24	Optimal seismic stations: full region and comparison of the three regions.	159

4.25	Optimal seismic stations: downweighting earthquakes outside the produc- tion region.	161
4.26	Optimal seismic stations: downweighting earthquakes outside the produc- tion region and comparison of results of the three different weighting schemes.	162
A.1	The five lowest-order multi-taper functions.	172
A.2	A Hann window	173
A.3	Illustration of adaptive coherency pre-filtering.	174
B.1	Original code flow	175
B.2	The code flow we have developed	176
B.3	Our code flow with input and output files	177
B.4	Data directory structure	178
B.5	Master directory structure	179
B.6	Corr_scan result directory structure.	184
B.7	Shifty result directory structure	185
B.8	Mad result directory structure.	186
B.9	Clone detection directory structure.	186
B.10	Clone result file directory structure.	187
C.1	P and S window sizes used in <i>BCSEIS</i>	191
C.2	Interevent distance reweighting functions for <i>hypoDD</i>	195
C.3	<i>hypoDD</i> locations.	196
C.4	901 cluster	197
C.5	902 cluster	197
C.6	903 cluster	198
C.7	904 cluster	198
C.8	905 cluster	199
C.9	906 cluster	199
C.10	907 cluster	200

C.11 908 cluster	200
C.12 909 cluster	201
C.13 910 cluster	201
C.14 168 cluster	202
C.15 170 cluster	202
C.16 238 cluster	203
C.17 290 cluster	203
C.18 Focal mechanisms for 901 and 902.	204
C.19 Focal mechanisms for 903, 904, 905 and 906.	205
C.20 Focal mechanisms for 907, 908, 909 and 910.	206
C.21 Focal mechanisms for 168, 170, 238 and 290.	207
C.22 Pre-RK20 injection, 29 September to 4 October 2008.	208
C.23 First 3.5 weeks of RK20 injection, 5 October to 29 October 2008. Flow rate is 650 t/h.	209
C.24 Very seismically active period, 30 October to 10 November 2008. Flow rate has increased by 8% to 700t/h.	209
C.25 Seismically inactive period (compared to Fig. C.24), 11 November 2008 to 26 January 2009.	210
C.26 Last month of data, 27 January to 28 February 2009. RK20 reduces flow rate and RK11/12 begin pumping again.	210
D.1 Network design code flow.	213
D.2 Example of insufficient sampling during ray tracing.	225
D.3 First application to the Darfield earthquake.	227
D.4 Second application to the Darfield earthquake.	228
D.5 Final stations deployed after the Darfield earthquake.	229

List of Tables

2.1	Master template nomenclature	43
2.2	Cross-correlation nomenclature	46
2.3	Network Correlation Coefficient nomenclature	52
2.4	Detection threshold nomenclature	55
2.5	Results from synthetic testing for one and six station networks	61
2.6	Summary of master events and their detections.	69
3.1	1D Vp and Vs velocity models used in <i>Velost</i> for Rotokawa	77
3.2	Parameters used in each <i>Velost</i> run.	79
3.3	Focal mechanism solution strike, dip, rake and uncertainty, all in degrees, for the 14 master events.	99
4.1	Nomenclature for network design	115
4.2	Nomenclature for network design implementation	119
4.3	Region 1 velocity model	130
4.4	Region 2 velocity model	131
4.5	Region 3 velocity model	131
4.6	Summary of terms used in noise variation with distance equations of the form $L_{10} = \text{Constant} - \text{Gradient} \times \log d$	151
4.7	Optimal seismic stations for locating earthquakes in the three different anal- ysis regions.	157

4.8	Optimal seismic stations to locate all earthquakes with different weighting schemes applied to each region.	163
C.1	1D Vp velocity model used in <i>hypoDD</i> for Rotokawa	192
C.2	Weighting scheme used in <i>hypoDD</i>	193
C.3	Weighting scheme used in <i>hypoDD</i>	195

Chapter 1

Introduction

1.1 Motivation and objectives

This study addresses methods of seismologically characterising actively exploited geothermal fields, motivated by two questions: 1) how can we best detect low-magnitude earthquakes and represent the uncertainties in their calculated hypocentres; and 2) how can we best design seismic observational networks to clearly resolve earthquake hypocentres?

Actively exploited geothermal fields provide seismologically unique challenges and opportunities, as well as being a globally important application. Geothermal systems are regions where high temperatures exist at shallow depths (< 5 km) [Blackett *et al.* 2004; Burnham *et al.* 1993]. Each geothermal system is unique, with attributes governed by numerous geological and physical factors such as tectonic setting, rock type, porosity, heat flow and permeability [Burnham *et al.* 1993]. Our geothermal study fields of Rotokawa and Kawerau, in the Taupo Volcanic Zone (TVZ), New Zealand [Fig. 1.2], are both liquid-dominated (water-dominated convective systems), high-temperature fields ($> 150^{\circ}\text{C}$ at < 2 km depth [Blackett *et al.* 2004]) and are both exploited for energy production. This energy production involves the extraction of hot fluid and steam, and the injection of cold post-processing fluid [Fig. 1.1]. These actions perturb the pressure, temperature, and chemical conditions, and the prevailing stress regime within the reservoir. The overall stress changes are often sufficient

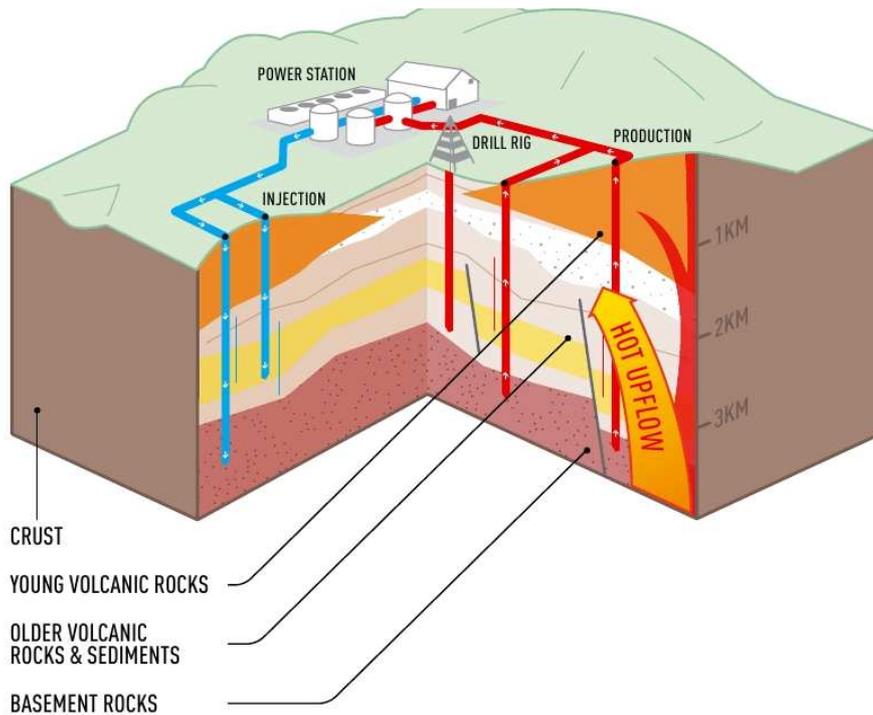


Figure 1.1: Cartoon of a geothermal power plant. Image from *Mighty River Power* [2010] *Geothermal Generation Pamphlet*.

to cause fault ruptures and microseismicity [Majer *et al.* 2007, and references therein].

In this study, we focus on injection-induced microseismicity. Such microseismicity consists largely of low-magnitude earthquakes, whose detection is complicated by low-signal-to-noise ratios (SNRs) due to high attenuation (hot, wet reservoir) and high ambient noise (industrial operations). Past studies demonstrate that hydraulic stimulation of a reservoir can induce microseismicity through pore-pressure increases [e.g. El Hariri *et al.* 2010; Ghassemi *et al.* 2007; Majer *et al.* 2007; Nemoto *et al.* 2008; Shapiro *et al.* 2010; Zoback and Harjes 1997], temperature decreases [Ghassemi and Sureshkumar 2007; Ghassemi *et al.* 2007; Majer *et al.* 2007; Rutqvist and Oldenburg 2007], and volumetric changes and geochemical alterations of fracture surfaces [Ghassemi and Sureshkumar 2007; Ghassemi *et al.* 2007; Majer *et al.* 2007; Nemoto *et al.* 2008]. Location of this microseismicity and characterisation of any spatiotemporal migration pattern may elucidate the three-dimensional distributions of zones of enhanced permeability and fracturing, allowing fluid flow to be tracked, and in some cases the prevailing stress-regime to be determined [Arnold and Townend 2007; Dyer

et al. 2008a; *Kumano et al.* 2006; *Lippitsch et al.* 2005; *Nemoto et al.* 2008; *Phillips* 2000; *Prejean et al.* 2002]. Because of the unique physical properties of each geothermal field due to tectonic setting, rock type, porosity, heat flow and permeability etc., as well as the effects of any production history in terms of both quantity and location of extraction and injection, a particular field's response to stimulation may not be the same as another's.

In the past two decades, microseismic monitoring has been increasingly recognised as a promising tool in reservoir imaging, driven by increased interest in unconventional gas, heavy oils and Enhanced Geothermal Systems (EGS) [*Bjornsson et al.* 2003; *Duncan* 2007; *Dyer et al.* 2008a; *Jupe et al.* 2000; *Le Calvez et al.* 2007], as well as by advances in data storage and computing power. In this study, we consider the monitoring and use of this microseismicity as a means of characterising a geothermal reservoir's geological properties and its response to injection [*Cuenot et al.* 2006; *Majer et al.* 2007; *Niitsuma* 1999; *Simiyu* 1999]. This characterisation is important in maintaining the efficiency of a geothermal power plant, as well as minimising its environmental impact [*Armannsson and Kristmannsdottir* 1992; *Bodvarsson* 1972; *Burnham et al.* 1993; *Sarmiento* 1986; *Stopa and Wojnarowski* 2006; *Tsang* 1980]. Reservoir simulation has been adopted worldwide as a prudent practice in analysing geothermal reservoirs, in which the development and adaptation of such a model requires sub-surface structural, geological, hydrological, stress field, diffusivity and conductivity information, as well as the ability to track production and injection flows accurately [e.g. *Acuna et al.* 2008; *Hayashi* 1999; *Holt* 2007; *Kohl and Megel* 2007; *Mighty River Power* 2007; *O'Sullivan* 2001; *O'Sullivan et al.* 2009, and references therein]. Improved knowledge of these reservoir characteristics and responses is therefore advantageous for reservoir management.

Geothermal power has become increasingly recognised as important locally and globally. This comes with the depletion of old power sources, and as a more environmentally aware population pushes for an increase in renewable energy sources. In the 1950s and 1960s, New Zealand was a leader in liquid-dominated geothermal energy research and development [*Ellis* 1999; *Harvey et al.* 2010; *Rowland and Sibson* 2004]. Research dwindled after the discovery

of the Maui gas field in 1969, but as this field is progressively depleted, the importance of geothermal power development in New Zealand is once again being emphasised. This also comes in response to recent government policies, notably: *the National Energy Efficiency and Conservation Strategy*; ¹; *Sustainable Development Programme of Action for Energy* ²; and *the Resource Management (Energy and Climate Change) Amendment Bill* ³. *The New Zealand Energy Strategy* ⁴ aims to increase New Zealand's renewable energy contribution from 73% of electricity generation in 2009 [Ministry Of Economic Development 2010] to 90% by 2025. Geothermal power currently contributes 10% of renewable generation [Harvey et al. 2010]. Geothermal energy offers base-load energy that is not sensitive to climatic variations, is close to major North Island load centres, has a competitive development cost [Harvey et al. 2010] and leaves a minimal environmental footprint [Fridleifsson et al. 2008].

The funding for this study has been provided by Mighty River Power (MRP), which is undertaking a major geothermal expansion and production programme. MRP's recent expansion has included the commission of a 100 MW power station at Kawerau in August 2008, and a second 140 MW power station at Rotokawa, Nga Awa Purua, in April 2010 [Mighty River Power 2010], in the latter case increasing the Rotokawa field's output by more than a factor of five to 174 MW. MRP is also expanding production from existing facilities, including those at Rotokawa, as deeper drilling accesses higher-temperature geothermal systems and more extensive consents are granted for exploratory drilling [Mighty River Power 2008b]. As many of New Zealand's geothermal assets lie within Maori-owned lands, the development of these resources must fulfil obligations of the Treaty of Waitangi, by forming partnerships with local Iwi Trusts. Rotokawa and Nga Awa Purua are carried out in partnership with Tauhara North No.2 Trust [Mcloughlin et al. 2010], whilst Kawerau resources are accessed with permission from the Ngati Tuwharetoa (Bay of Plenty) Settlement Trust and Putauaki Trust [Mighty River Power 2008a].

¹<http://www.eeca.govt.nz/node/2639> 2007

²<http://www.beehive.govt.nz/release/sustainable-development-programme-action> 2003

³<http://www.parliament.nz/en-NZ/PB/Legislation/Bills/BillsDigests/a/7/0/a7070ebdb99442b29249e038b451c533.htm> 2003

⁴http://www.med.govt.nz/templates/ContentTopicSummary_19431.aspx 2007 & 2010

Given the above, we assert that: geothermal power is an important asset; current understanding is not sufficient to maximise the efficiency of geothermal resources; geothermal power production requires the resource to be continuously monitored to track any evolution and maximise efficiency; microseismicity is a tool that can be used for monitoring and to aid our understanding of geothermal resources; the more microseismicity we can detect and accurately locate the better this tool can be utilised. We therefore focus in this study on better understanding seismicity in two active seismological fields, with respect to two specific objectives:

1. Automated microearthquake detection using a matched filter technique and robust hypocentre estimation for earthquakes associated with fluid injection at Rotokawa, to elucidate how spatiotemporal patterns of induced events relate to injection processes;
2. A pilot study of optimal seismic network design that will guide the expansion of the existing seismic monitoring programme at Kawerau geothermal field.

1.2 Geological, tectonic and seismological setting

1.2.1 Taupo Volcanic Zone

The Rotokawa and Kawerau geothermal fields are both located in the Taupo Volcanic Zone (TVZ) [Fig. 1.2]. The TVZ is the locus of Quaternary volcanism in New Zealand, and forms a region that is 300 km long and up to 60 km wide defined by vent positions and caldera structural boundaries that have been active, or are inferred to have been active during the last 2.0 Ma [Houghton *et al.* 1995; Wilson *et al.* 1995]. The 2.0 Ma date is based on the onset of andesitic calc-alkalic volcanism in the zone [Houghton *et al.* 1995; Wilson *et al.* 1995]. Since ca. 1.6 Ma, rhyolitic volcanism has been volumetrically dominant [Wilson *et al.* 1995]. Volcanism divides the TVZ into three segments: the northeast and southwest segments are composed of andesitic to dacitic composite volcanoes without calderas, whilst the central section is rhyolite-dominated [Wilson *et al.* 1995]. The thickness of the volcanic deposits in

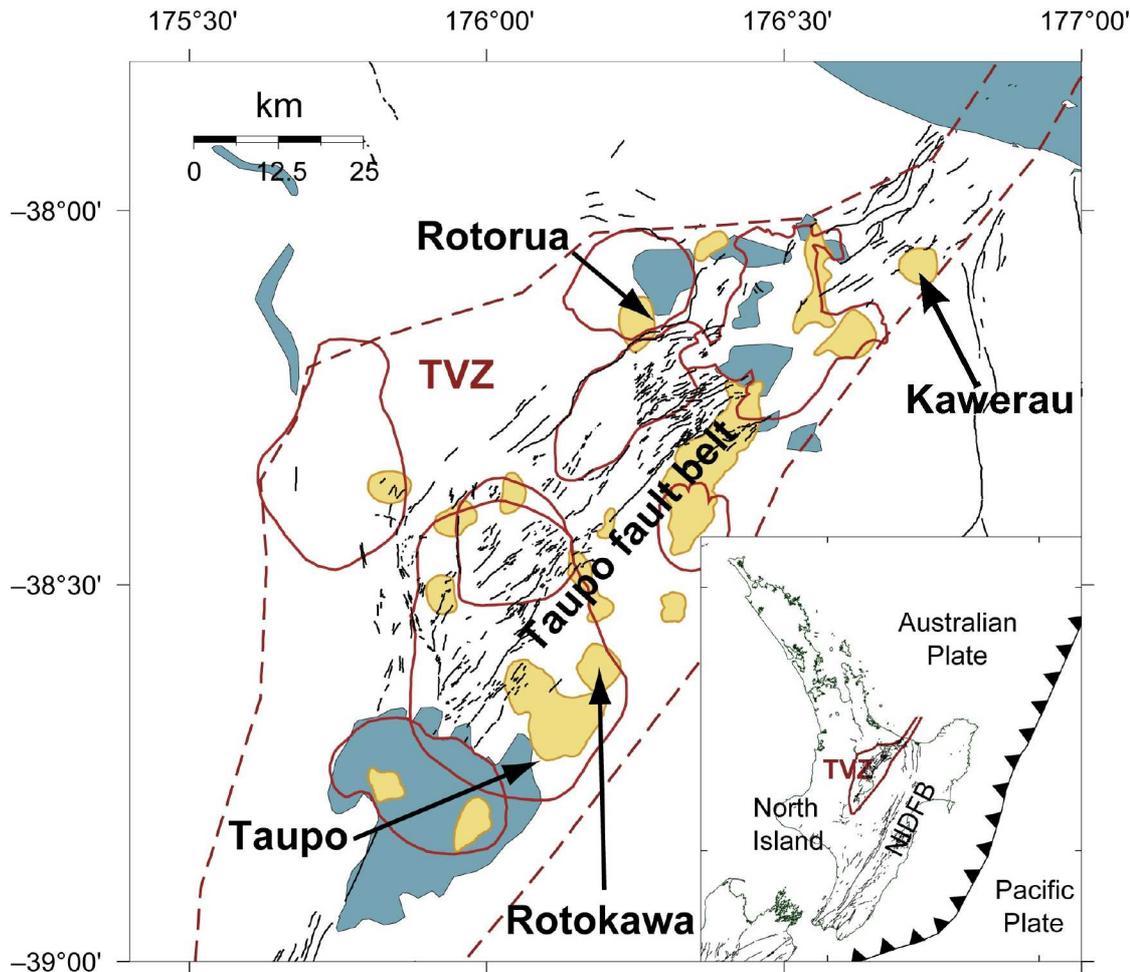


Figure 1.2: The central Taupo Volcanic Zone (TVZ, dashed brown lines). Geothermal systems are coloured yellow. Black lines are faults. Solid brown lines border calderas. The inset shows the location of the TVZ in the North Island and the Hikurangi subduction zone boundary (ticked line) between the Australian and Pacific plates. Black lines mark faults. NIDFB is the North Island Dextral Fault Belt. Figure modified after *Clarke et al.* [2009].

the TVZ is controversial, with inferences from seismic studies, gravity, petrogenetic models, and heat flow models in some disagreement [*Behr et al.* 2011; *Bibby et al.* 1995; *Davey et al.* 1995; *Hochstein et al.* 1993; *Hochstein* 1995; *Rowland et al.* 2010; *Stern* 1987; 1985; *Wilson et al.* 1995, and references therein]. The contact between volcanics and greywacke basement has only been encountered in a few geothermal drillholes, at Rotokawa, Kawerau and Ohaaki, whereas in most places drillholes to ca. 2.5 km depth terminate in volcanics [*Wilson et al.* 1995].

The eastern margin of the TVZ is marked by steep gravity gradients associated with

downfaulting of the greywacke basement to the northwest [*Hochstein and Hunt* 1970], whilst the western margin is not clearly defined by gravity: extending to the west is a region of low residual gravity [*Stern* 1985]. The TVZ is associated with the Hikurangi subduction zone, where the Pacific plate subducts beneath the Australian plate in a northwesterly direction at ca. 43 mm/yr [*Beavan et al.* 2002]. Accompanying this subduction is northwest–southeast extension (presently $\sim 8\text{--}10$ mm/year [*Bibby et al.* 2002; *Rowland and Sibson* 2004]), a thinned crust [*Stratford and Stern* 2006], very high heat flow (700 mW/m^2 [*Stern* 1985]), and extensive geothermal activity. Twenty-three active and two extinct geothermal fields have been identified within the Quaternary volcanics of the central rhyolite-dominated section using electrical resistivity methods [*Bibby et al.* 1995; 1999; 2002]. The locations of these geothermal fields are inferred to have been stable over the last 200,000 years, with an average distance of 10–15 km between fields, average extent of 20 km^2 , and no direct spatial correlation with volcanic features [*Bibby et al.* 1995; *Rowland and Sibson* 2004].

Seismicity in the TVZ occurs in two primary zones: a northwest-dipping zone of deep earthquakes (Benioff Zone) marking the top of the subducted Pacific plate, and a shallow zone (< 20 km) within the Taupo Fault Belt (TFB). The TFB is a band, subparallel to the strike of the TVZ, of dominantly normal faulting [*Bibby et al.* 1999] in a northeast direction, with a small component of dextral strike-slip [*Bryan and Sherburn* 1999; *Clarke et al.* 2009; *Hurst et al.* 2002]. Shallow seismicity within the TFB is characterised by bursts of localised activity followed by long periods of relative quiescence [*Bryan and Sherburn* 1999; *Hurst et al.* 2002; *Sherburn et al.* 1999]. Of this activity, 80% occurs at depths of < 6 km, which defines the base of the seismogenic zone beneath the central TVZ, and coincides with the depth at which heat transfer changes from being dominantly convection-driven to being dominantly conduction-driven [*Bibby et al.* 1995; *Bryan and Sherburn* 1999]. This interpretation is supported by flow parameter estimations that put the base of the geothermal convection system at between 5 and 8 km [*Bibby et al.* 1995].

1.2.2 Rotokawa

The Rotokawa Geothermal field is located 12 km northeast of the town of Taupo in the south-eastern part of the TVZ [Fig. 1.2] and spans an area $\sim 25 \text{ km}^2$ [Hunt and Bowyer 2007]. Rotokawa is a high-temperature, liquid-dominated geothermal field that has been carbon dated at younger than 20,000 years [Krupp and Seward 1987]. Surface manifestations occur primarily on the south-eastern side of the Waikato river in the form of a northeast-alignment of hydrothermal eruption vents, the largest of these being a shallow acid lake (Lake Rotokawa). The surface geology is dominated by rhyolite domes, pumice alluvium, Wairakei breccia and hydrothermal eruption breccias [Rae 2007]. Between 1965 and 1986, seven deep (0.9–2.7 km) exploration wells were drilled by New Zealand government agencies. Two of these encountered temperatures in excess of 300°C [Hunt and Bowyer 2007; Krupp and Seward 1987]. The drillholes penetrated a sequence of silicic volcanic rocks and andesitic lavas, with the deepest wells intersecting Mesozoic basement greywacke of the Torlesse terrane [Browne *et al.* 1992]. Rotokawa is situated above a northeast–southwest-trending structural graben that has block-faulted both the greywacke basement and the overlying Rotokawa Andesite, a 0.9–2.1 km thick sequence that constitutes the reservoir [Fig. 1.3; Rae 2007]. This contact identifies the Rotokawa Andesite as some of the earliest volcanic rocks on the eastern margin of the central TVZ [Arehart 2002].

The extensional stress regime of the TVZ has created a set of northeast–southwest-striking normal faults. Superimposed on these are a series of radial fractures concentrated about the topographic high of the Oruahineawe rhyolite dome north of the Waikato river [Krupp and Seward 1987]. The Rotokawa system contains 10 identified explosion craters of up to 1.5 km in diameter, and at least 12 hydrothermal explosions represented by breccia deposits [Krupp and Seward 1987]. The locations of these explosion craters directly above where the Waiora Formation intersects a set of faults identify this area as a major upflow zone of a two-phase geothermal fluid (aqueous liquid and gas) [Krupp and Seward 1987]. The styles of hydrothermal alteration indicate that permeability is largely fracture-controlled within the Rotokawa Andesite and basement greywacke [Rae 2007].

Rotokawa Geothermal field was purchased in 2000 by Mighty River Power (MRP), the current owner and operator [Hunt and Bowyer 2007]. The Rotokawa geothermal resource stems from an ascending two-phase fluid that is separated into steam and brine at the surface. Energy for electricity is extracted and all post-production geothermal fluid (condensed steam and brine) is reinjected into the ground through dedicated wells [Mighty River Power 2007]. Microseismicity associated with this reinjection is of particular interest to MRP for informing decisions about power production and the sustainable use of the reservoir [Mighty River Power 2007]. For the first eight years of production at Rotokawa, injection took place into a shallow aquifer. In 2005, based largely on gravity data, this injection was significantly reduced due to concerns about over-pressuring and liquid saturation, and deep injection wells were commissioned on the western edge of the field in RK16 and RK18 [Hunt and Bowyer 2007; Mighty River Power 2007]. Tracer tests carried out in 2006 demonstrated that RK16 and RK18 were not sustainable injection wells in the long term as fast returns to production wells caused cooling [Grant 2007]. A new injection well, RK20 (initiated in 2008), is part of subsequent plans to shift injection to the southeast, across the southwest–northeast permeability axis, in an attempt to delay these returns [Grant 2007].

1.2.3 Kawerau

The Kawerau Geothermal field is located in the Eastern Bay of Plenty in the north-eastern part of the TVZ [Fig. 1.2] and spans an area 19–35 km² [Bignall and Harvey 2005], although its surface manifestations span only about half this area [Allis 1997]. Like Rotokawa, Kawerau is a high-temperature, liquid-dominated geothermal field, but has been dated at older than 280,000 years [Bignall and Harvey 2005]. The surface geology is dominated by rhyolite, dacite and andesite lavas, ignimbrite, tuff and other pyroclastic units [Bignall and Harvey 2005]. Shallow drilling and scientific surveys were carried out in 1951 and 1952 by the Department of Scientific and Industrial Research (DSIR) and Ministry of Works (MOW). These surveys identified temperatures of up to 310°C at > 1 km, some of the hottest temperatures measured in geothermal fields of the TVZ [Bignall and Harvey 2005]. Drill-

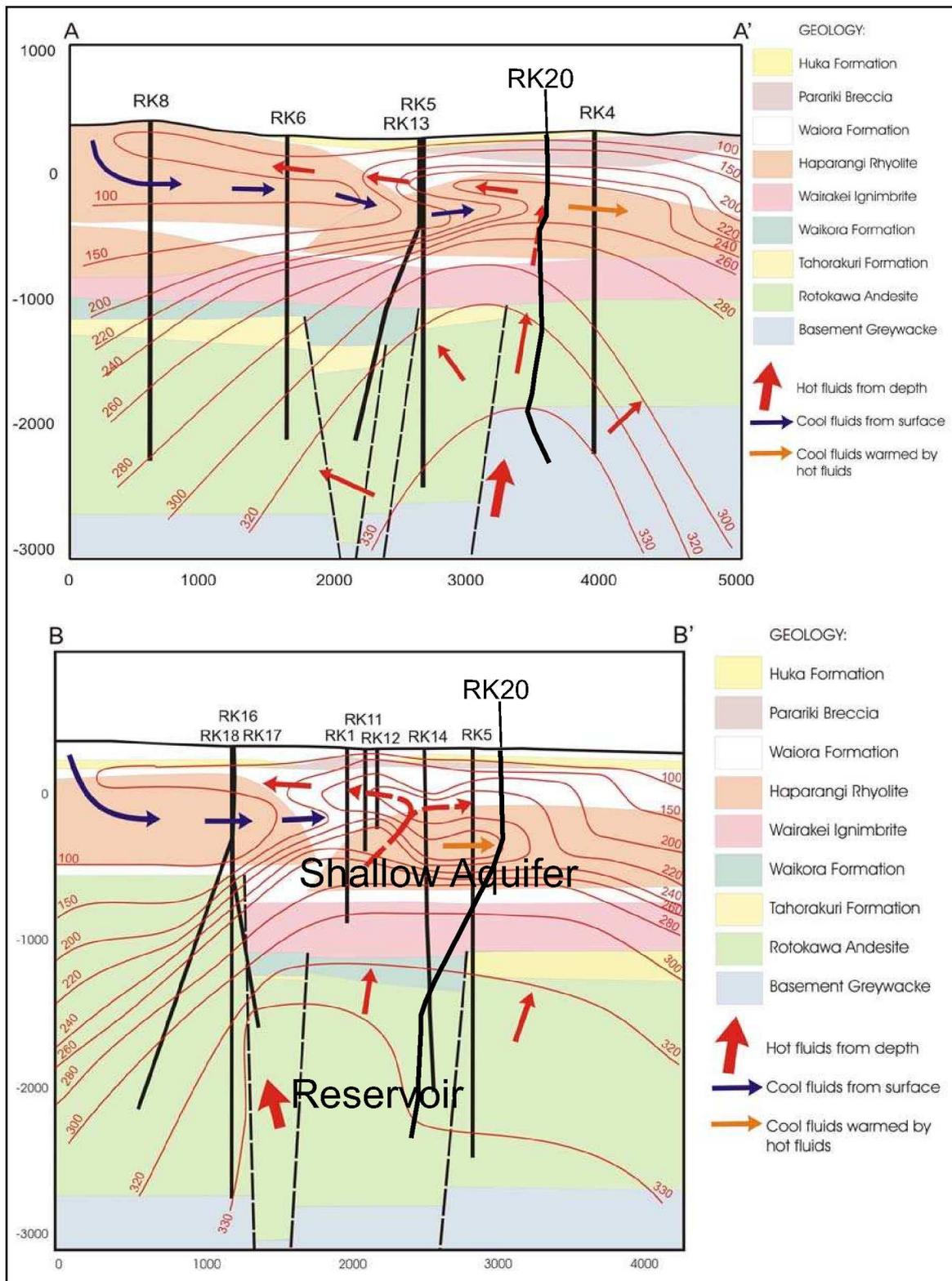


Figure 1.3: NW–SE (A–A') and SW–NE (B–B') cross-sections through the Rotokawa geothermal field illustrating the known stratigraphy, inferred fault locations (denoted by black dashed lines), isotherms and hydrology. Lines showing the locations of these cross-sections are in Fig. 1.6. Figure modified from *Rae* [2007].

holes penetrate a sequence up to 2 km in thickness and consists of layered volcanic lavas, welded ignimbrites, pyroclastic rocks and lacustrine sediments, and terminate in fractured Mesozoic greywacke basement [Wood *et al.* 2001]. Differences in the permeability of these volcanics and the sedimentary layers result in a sequence of aquifers and aquicludes, which separate the resource into discrete productive units at different depths [Bignall and Harvey 2005]. The location of Kawerau on the younger eastern side of the TVZ [Wilson *et al.* 1995] places it in the Whakatane Graben; where the northeast-striking rift of the TVZ intersects north–south-trending strike-slip faults of the North Island Dextral Fault Belt (NIDFB) [Fig. 1.2; Nairn and Beanland 1989]. These steeply-dipping northeast-trending normal faults, downthrown to the west, and northwest-trending cross-faults dominate the reservoir’s structure, and largely control the permeability of the area [Fig. 1.4; Bignall and Harvey 2005]. Earthquakes within the field align along northeast-trending lineations [Clarke *et al.* 2009], consistent with the TVZ’s overall extensional regime.

Kawerau has been supplying steam for process heat and electrical power for the Tasman Pulp and Paper Company since 1957 [Bignall and Harvey 2005]. The new 100 MW power station at Kawerau was commissioned by MRP in 2008 [Mighty River Power 2008a]. Kawerau has an existing seismic network of six permanent telemetered stations. There are plans to extend this seismic network by up to 10 seismometers in the next few years [Steven Sewell, MRP, pers. comm., 2010].

1.3 Seismological investigations undertaken

1.3.1 Injection-induced microearthquake detection at Rotokawa

1.3.1.1 Past research

Research into artificially-induced earthquakes began in the late 1960s [Healy *et al.* 1970; 1968; Hoover and Dietrich 1969; Kisslinger 1976; Raleigh *et al.* 1976; Rothe 1969] when it was recognised that a number of anthropogenic activities could produce earthquakes. Open-

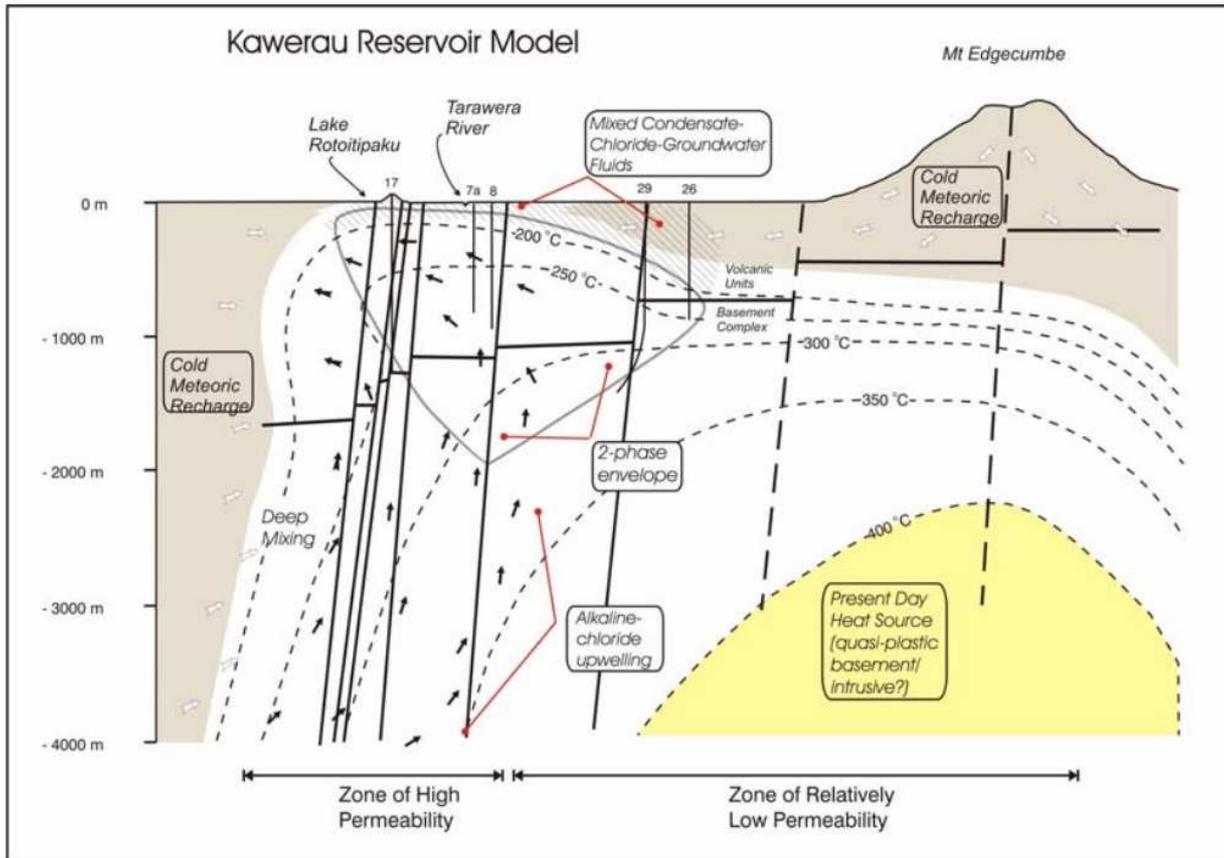


Figure 1.4: Conceptual model of the Kawerau geothermal field, showing the inferred hydrology, thermal profile and magmatic heat source. Image from *Bignall and Harvey* [2005].

ing underground cavities for mining has been known for some time to cause localised low-magnitude seismicity [*Kisslinger* 1976]. The impoundment of artificial lakes such as the Hoover Dam (U.S.); Lake Kremasta (Greece); and Koyna (India); caused pronounced seismicity, including damaging earthquakes of magnitudes > 6 [*Alexander* 1993; *Healy et al.* 1970]. Underground nuclear explosions at the Nevada Test Site of the U.S. Atomic Energy Commission in the late 1960s resulted in a series of earthquakes, as well observable displacements on nearby faults [*Alexander* 1993; *Healy et al.* 1970; *Kisslinger* 1976]. The disposal of contaminated waste fluid into a ~ 3.6 km deep well (the Rocky Mountain Arsenal well) near Denver, Colorado, by the U.S. Army in 1962 caused the first recognised injection-induced seismicity [*Alexander* 1993; *Healy et al.* 1970; 1968; *Hoover and Dietrich* 1969; *Kisslinger* 1976; *Raleigh et al.* 1976]. It was subsequently recognised that the injection of fluid into rocks at depth and the withdrawal of fluids from subsurface formations during

commercial oil and geothermal activities were also causing seismicity [Denlinger and Bufe 1982; Kisslinger 1976; Raleigh et al. 1976; Segall 1989].

Microseismicity due to fluid injection initially incited excitement regarding the possibility of converting potentially destructive earthquakes into numerous non-destructive minor earthquakes by using fluid injection techniques [Healy et al. 1968; Raleigh et al. 1976], however, to date this method is still considered too risky, if indeed possible, to be a feasible method [Alexander 1993]. It was also noted in early studies that the temporal pattern of seismicity coincided with injection pressures [Fig. 1.5; Healy et al. 1970; 1968; Hoover and Dietrich 1969; Raleigh et al. 1976]; spatial patterns appeared to delineate geological structures [Healy et al. 1970]; and that permeability and porosity were influential characteristics, as well as the regional stress field [Kisslinger 1976]. It is these microseismic properties that find useful application in the geothermal industry today.

The mechanism of induced seismicity was initially explained solely by the theory of effective stress and the application of the Coulomb-Mohr theory of shear failure [Healy et al. 1970; 1968; Raleigh et al. 1976; Rothe 1969]. The amount of shear stress required for a fault to rupture can be expressed as follows:

$$\tau = \mu(\sigma_n - P_f) \quad (1.1)$$

where τ is the amount of shear stress required for failure to occur, σ_n is the tectonic normal stress (positive for compression), P_f is the pore fluid pressure and μ is the coefficient of friction [Hubbert and Rubey 1959]. As pore fluid pressure increases, the amount of shear stress required for fault failure decreases. In 1976, Kisslinger suggested that direct physical or chemical alteration resulting in the weakening of rock which fluid has had direct contact with may also play an essential role in fracturing. The dominant cause of induced-seismicity is still under debate today [Majer et al. 2007, and references therein], and may in fact be particular to the region in question and its prevailing stress-field and characteristics. As well as the pore-pressure increases and geochemical alterations of fracture surfaces already men-

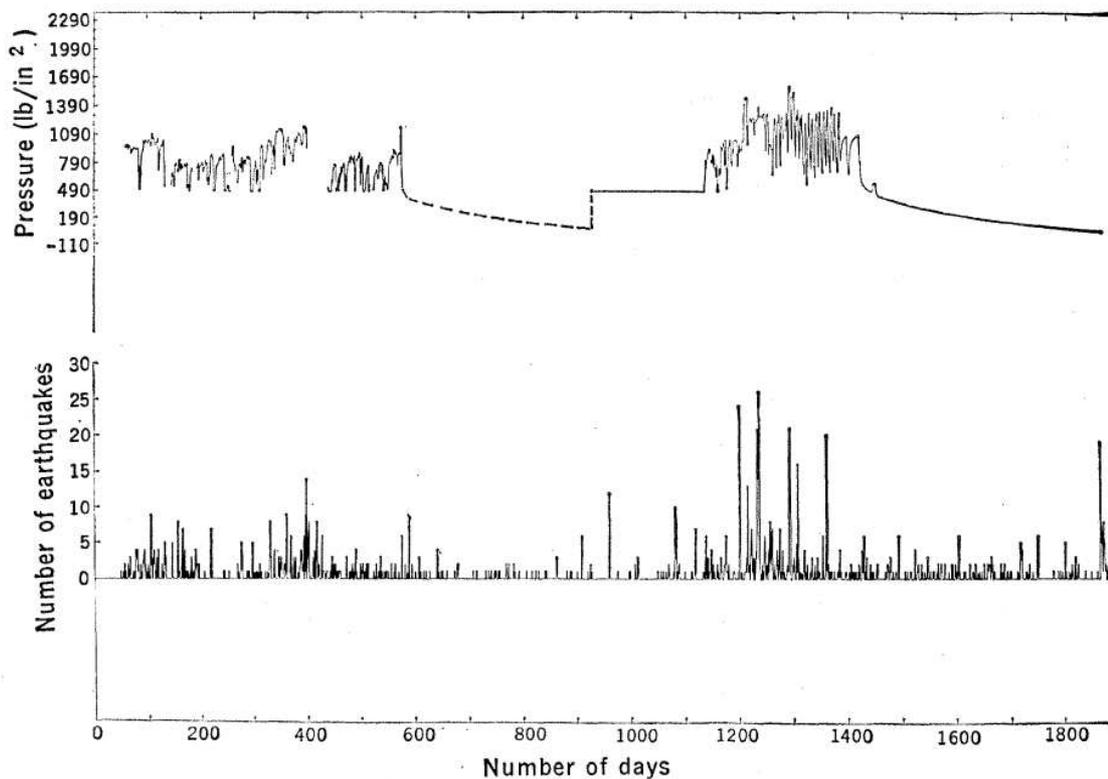


Figure 1.5: Data from the 1962 Rocky Mountain Arsenal well waste fluid injections. Daily pressure is plotted against the number of earthquakes. The dashed portions in the pressure curve are interpolated where only data on the volume of fluid injected were available. Image from *Healy et al.* [1968].

tioned, temperature decreases and volumetric changes have also been suggested as possible causes [*El Hariri et al.* 2010; *Ghassemi and Sureshkumar* 2007; *Ghassemi et al.* 2007; *Majer et al.* 2007; *Nemoto et al.* 2008; *Rutqvist and Oldenburg* 2007; *Shapiro et al.* 2010].

Geothermal injection has recently come under public scrutiny in light of the magnitude 3.4 earthquake caused by injection during the creation of an Enhanced Geothermal System (EGS) in Basel, Switzerland, in 2006 [*Giardini* 2009]. Basel is an industrial centre that is home to more than 700,000 people, and has a history of earthquakes [*Giardini* 2009]. Though not a large earthquake, the fact that the 2006 earthquake occurred in the centre of a large industrial city resulted in international media attention, damage claims of more than \$9 million, halting of the project, a liability court case, public back-lash, and ongoing questions about future earthquake liability issues [*Giardini* 2009; *Ripperger et al.* 2009]. EGS systems

offer the possibility to generate large amounts of clean energy in areas all over the world. It involves the creation of an artificial geothermal reservoir by pumping fluids under high pressure into fractures in non-porous rock at temperatures greater than 100°C. This pumping causes the rock to fracture, generating microearthquakes and thereby increasing the permeability and expanding the fracture reservoir. Once a reservoir of sufficient volume has been created, fluids are circulated through and energy from the heat of the Earth can be extracted [Dyer *et al.* 2008b; Giardini 2009; Majer *et al.* 2007; Ripperger *et al.* 2009]. This method results in a greater risk of larger induced-earthquakes than in a natural geothermal system, but its energy production potential is also greater, since greater spatial extent affords a greater potential source of energy. The possibility of incidents such as the 2006 Basel earthquake damaging the fulfilment of that potential highlights the need to better understand injection-induced seismicity [Majer *et al.* 2008]. Majer *et al.* [2007] provided a thorough review of case histories, technical and public issues, and the current scientific understanding of EGS induced-seismicity.

In natural geothermal systems, the reinjection of spent geothermal fluids is used to safely dispose of waste fluids, as well as to control surface subsidence and maintain pressures in the field to enhance production [Segall 1997]. Siting these injection well locations and determining the response of the field to injection is thus a key problem in geothermal reservoir management, as reservoir pressure must be maintained whilst ensuring this cold injectate does not return to the production region before it has had time to rewarm [e.g. Ungemach *et al.* 2008]. Microseismicity associated with this injection is now used to monitor geothermal reservoirs using the microseismic properties mentioned previously: the temporal pattern of seismicity often coincides with injection pressures, spatial patterns appear to delineate geological structures, and permeability and porosity are influential characteristics, as well as the regional stress field. Recent studies demonstrate that spatiotemporal clusters of seismicity can delineate fractures and highlight locations of enhanced permeability, thus tracking fluid flow, and can also enable the prevailing stress-field to be determined [e.g. Arnold and Townend 2007; Dyer *et al.* 2008a; Kohl and Megel 2007; Kumano *et al.* 2006; Lippitsch *et al.*

2005; Nemoto *et al.* 2008; Phillips 2000; Prejean *et al.* 2002].

Phillips [2000] relocated clusters of microseismicity in the geothermal reservoir at Soultz-sous-Forêts, France, revealing two planar structures consistent with fracture orientations found in core and logging studies. Phillips [2000] also imaged permeable zones likely to be fluid-flow paths and calculated focal mechanisms compatible with the measured stress field. Lippitsch *et al.* [2005] investigated microseismicity associated with the high-temperature Torfajökul geothermal system, in Iceland. Relative relocation was performed using *hypoDD*, the double-difference relocation technique of Waldhauser and Ellsworth [2000], followed by probabilistic hypocentre estimation using *NonLinLoc* [Lomax *et al.* 2000]. The high-resolution locations obtained exhibited a tighter clustering of hypocentres when compared with original locations. Relocated microseismic clusters of Kumano *et al.* [2006] (also calculated using *hypoDD*) clearly delineated the reservoir structure of the Deep Heat Mining project at Basel, Switzerland. Illuminated fractures were generally consistent with local tectonic stress analyses. The spatiotemporal pattern of seismicity corresponded clearly to the history and injection volumes of the hydraulic stimulation. Another study conducted as part of the EGS project in Basel, Switzerland by Dyer *et al.* [2008a] revealed two distinct orientations seen within the induced microseismicity. These orientations are similar to the maximum horizontal stress directions, S_{Hmax} , estimated from borehole breakouts and from drilling-induced tensile fractures, as well as with the majority of natural fracture orientations in the granite reservoir. A series of laboratory injection-induced slip experiments using pre-fractured granitic rock were conducted by Nemoto *et al.* [2008]. The hydro-mechanical coupling behaviour of the fractures during fluid injection suggested that dynamic (temporary) permeability changes may occur in geothermal reservoirs subjected to hydraulic stimulations. Such changes cause a temporal increase in fluid flow in slipping (seismically active) fractures.

The first microseismicity surveys at Rotokawa were carried out by GNS Science in 2005 and 2006 using temporary seismic station deployments. These studies addressed the levels of seismicity associated with new deep injection at RK16, RK17 and RK18, and, combined

with an injection tracer survey, were intended to determine the pathways of injection returns. The deployment revealed microseismicity that likely delineates the deeper extent of a fault structure intersected by RK17: this structure is expected to be a lucrative drilling target for production wells [Bannister *et al.* 2008]. The tracer tests also demonstrated that RK16 and RK18 are not sustainable injection wells in the long term because of fast returns to production wells causing cooling [Grant 2007] and the new RK20 injection well that was initiated in the south-east of the field in September 2008 [Fig. 1.6] is part of subsequent plans to shift injection to the southeast (as discussed in Section 1.2.2). The success of the GNS Science study paved the way for the deployment in July 2008 of a nine station seismic array, whose five months of data following the initial injection at RK20 we will be using in this study.

1.3.1.2 Techniques used in this study

The more seismicity that can be detected and accurately located, the more accurate the information provided by that seismicity will be. To achieve a high number of earthquake detections, we implement a waveform template cross-correlation technique. Cross-correlation of waveforms was originally used as a test of similarity between already detected earthquakes so that families of events could be identified [Aster and Scott 1993; Baisch *et al.* 2008; Barani *et al.* 2007; DeShon *et al.* 2007; Ferretti *et al.* 2005; Lees 1998; Maurer and Deichmann 1995]. A waveform's characteristics are governed by the earthquake's focal mechanism and the source-receiver path travelled to the station. Similar waveforms, therefore, identify events that are tightly clustered in space relative to the dimensions of near-source heterogeneities and to the dominant seismic wavelengths [Maurer and Deichmann 1995]. Waveform families (or clusters) can be used for a number of applications: phase-picking, relative earthquake location, focal mechanism determination, evaluation of source time functions, rupture characteristics, characterisation of Green's functions and their spatial derivatives, multi-source and multi-receiver processing for Earth structure, among other topics. [Aster and Scott 1993, and references therein]. The cross-correlation (or 'matched filter') methods developed for cluster identification have recently been extended to earthquake

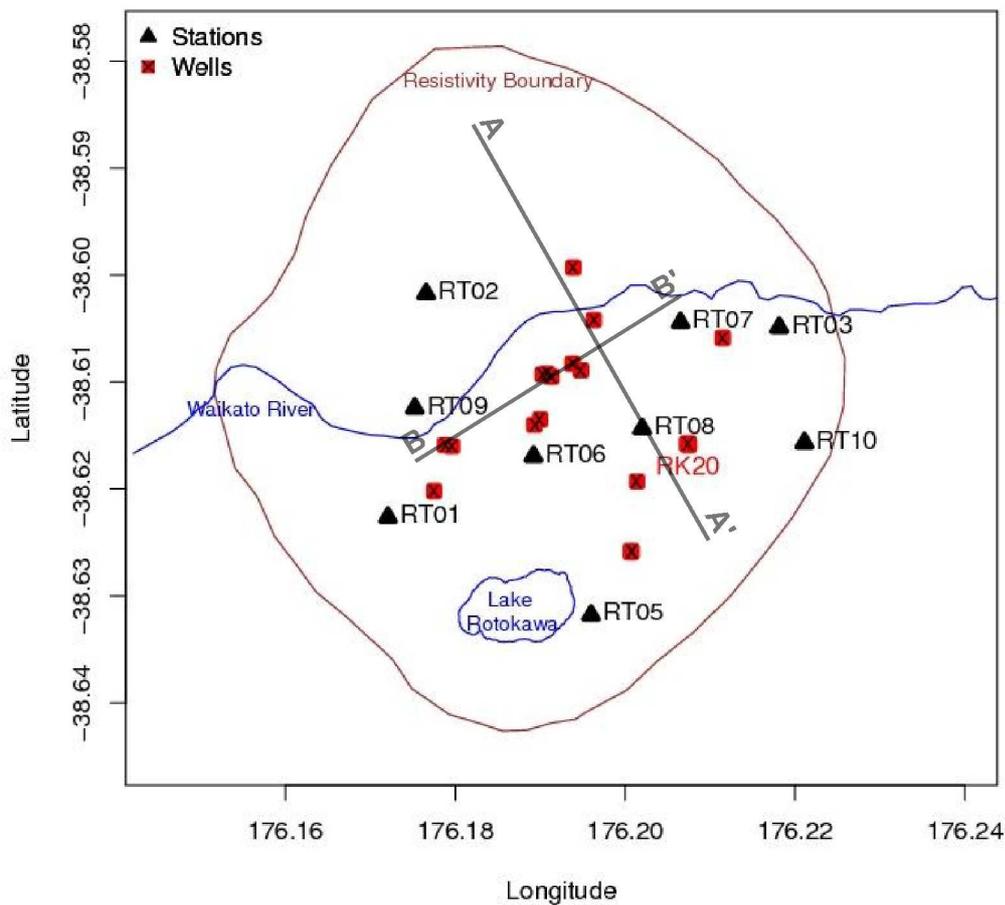


Figure 1.6: Map of Rotokawa showing the 2008–2010 seismic monitoring stations (black triangles), current production and injection wells (red squares) and the new injection site, RK20. Resistivity boundary is the outer reservoir outline of Risk [2000]. Grey lines depict the locations of the A–A’ and B–B’ cross-sections in Fig. 1.3

detection [Brown et al. 2008; Eisner et al. 2008; Gibbons et al. 2007; Gibbons and Ringdal 2006; Peng and Zhao 2009; Schaff and Waldhauser 2008; Schaff 2009; Shelly et al. 2007].

Although it has been recognised for some time that matched filter techniques are the best method for detecting a known signal within noise [e.g. Anstey 1966; Gibbons and Ringdal 2006; van Trees 1968], not much attention was paid to this technique until recently in the seismological community at large as it is rare the waveforms of interest are known *a priori*. Unknown signals are typically detected on the basis of signal power, using a STA/LTA detector, for example, as proposed by Freiburger [1963]. With this approach the energy in

a short-term window (STA) is divided by that in a long-term average window (LTA) and a detection is triggered when this ratio exceeds some SNR threshold. One drawback of this method is that false alarm rates go up dramatically for lower SNR thresholds [*Schaff and Waldhauser* 2008].

Recent studies using matched filter detection methods have been generally carried out in confined geographical regions, where there is a limited set of source and path functions affecting waveforms. The aim in these studies is often to increase the completeness of data sets by detecting a greater portion of the low-magnitude earthquakes missed using conventional methods, and so gain a better understanding of seismicity patterns. Among other recent studies, matched filter detection has been used by *Gibbons and Ringdal* [2006] to look at seismic events occurring at the Barentsburg coal mine in Spitsbergen; by *Eisner et al.* [2008] to investigate microseismicity induced by hydraulic fracturing; by *Brown et al.* [2008] and *Shelly et al.* [2007] to detect low-frequency earthquakes within seismic tremor; by *Peng and Zhao* [2009] to detect missing aftershocks following the 2004 Parkfield earthquake; and by *Gibbons et al.* [2007] to detect missing aftershocks following magnitude 2.0–3.5 events occurring in the Rana region of northern Norway and recorded by the sparse Fennoscandian IMS seismic array NORSAR. The approach has now also been applied at much larger scales: *Schaff* [2009] used a catalogue of Chinese seismicity spanning 20 years to create a large set of waveform templates that were then correlated with the data, resulting in a 70% increase in detected events. As seismic networks densify and computing storage and power increase, applications such as this become more widely feasible.

The Rotokawa geothermal field offers us a confined geographical region ($\sim 25 \text{ km}^2$) in which to perform matched filter detection. In this study, we develop software to perform this task for low-magnitude earthquake detection by cross-correlating master templates, which are waveforms of high-signal-to-noise earthquakes, with seismic data [Fig. 1.7]. To perform accurate hypocentre determination we then use *Velost* [*Kissling et al.* 1994], *hypoDD* [*Waldhauser and Ellsworth* 2000], and *NonLinLoc* [*Husen* 2003; *Lippitsch et al.* 2005; *Lomax et al.* 2009; *Walsh et al.* 2009].

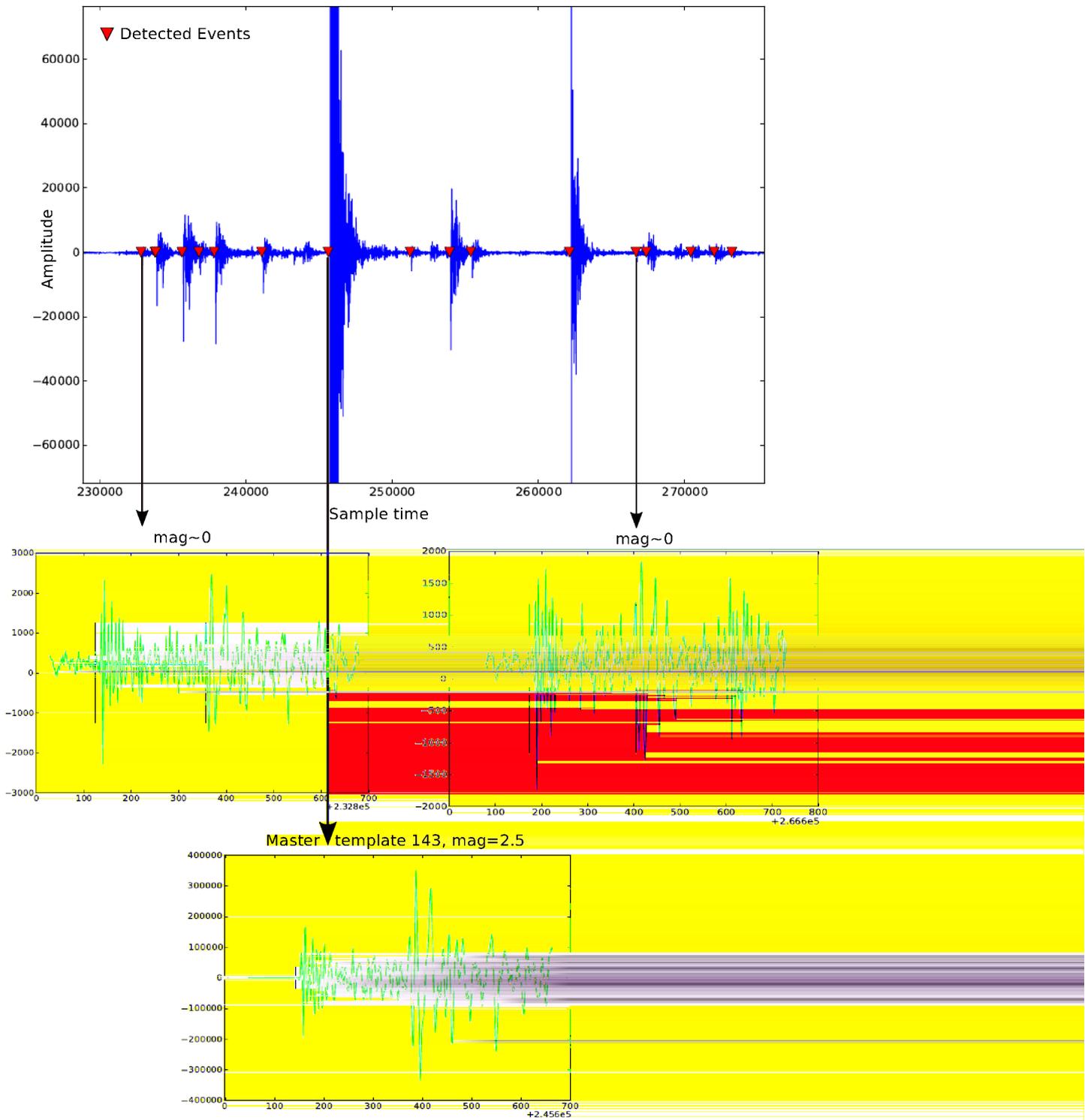


Figure 1.7: A ~ 4 minute long seismic record (top) containing a number of similar earthquakes. Using the magnitude 2.5 master template (bottom) in our wavelet matching software 15 other events of magnitudes as little as 0 can be detected in the record. Note the strong similarity between each of the two detected events illustrated and the template event that is two orders of magnitude larger (all shown on the East component). See Chapter 2 for further details.

Velost is used to calculate preliminary hypocentres with a fixed velocity model for Rokawa provided by GNS Science. We then attempt to refine those hypocentres using *hypoDD*, which implements the double-difference earthquake location algorithm of *Waldhauser and Ellsworth* [2000]. This algorithm capitalises on the fact that if the hypocentral separation between two earthquakes is small compared to the event–station distance and the scale length of velocity heterogeneity, then the ray paths between the source region and a common station are similar along almost the entire ray path. The iterative least-squares procedure employed in *hypoDD* uses travel time differences for pairs of earthquakes observed at common stations to remove mutual errors resulting from these near-coincident ray paths traversing similar velocity heterogeneities. This mitigates the need for station corrections or accurately calculated travel times for the portion of the ray path lying outside the focal volume. *Menke and Schaff* [2004] demonstrated that the double-difference algorithm not only resolves absolute earthquake locations, but is capable of improving the absolute locations obtained using traditional methods, especially when high-precision cross-correlation-derived differential travel times are used. *hypoDD* has been extensively used with data from permanent networks and aftershock arrays around the world [e.g. *Clarke et al.* 2009; *Kumano et al.* 2006; *Lippitsch et al.* 2005; *Waldhauser and Ellsworth* 2000; *Wallace et al.* 2004]. The ‘clouds’ of hypocentres commonly observed with routine location methods typically sharpen into networks of discrete fault planes after relocation with *hypoDD* [Fig. 1.8]. However, the technique is dependent on an event being recorded by a number of stations; this restriction can reduce the relocatable catalogue of seismicity, particularly when low-magnitude earthquakes are being analysed. Mindful of this restriction, we also use *NonLinLoc* to obtain earthquake location improvements.

Probabilistic hypocentre estimation is undertaken using *NonLinLoc* software [*Lomax et al.* 2009]. This non-linear, global-search location method gives a relatively complete picture of the location uncertainties; this is particularly important for depth determinations and for event locations outside of a station network. It does this by producing a misfit function, “optimal” hypocentres, an estimate of the posterior probability density function (PDF) for

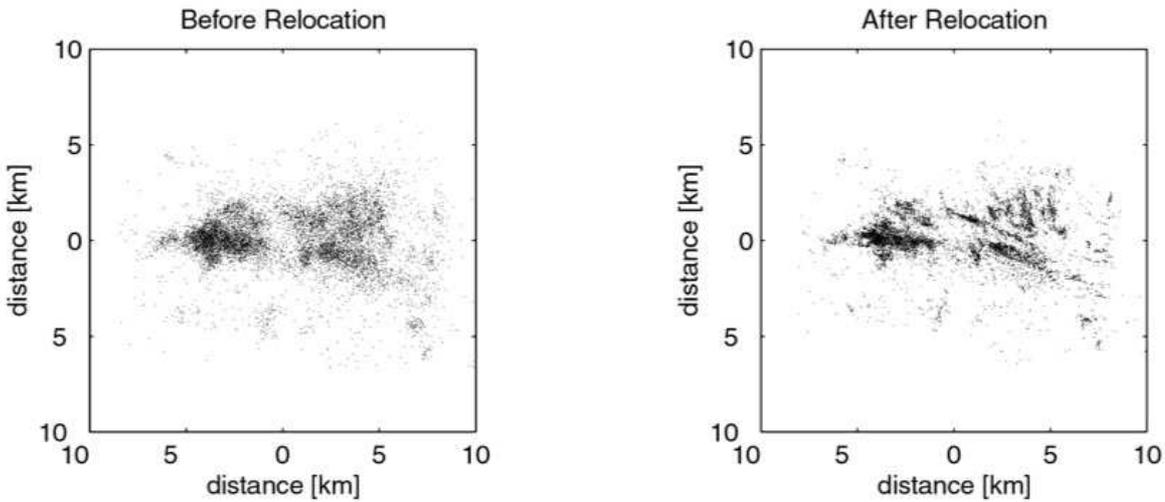


Figure 1.8: Map view of Joint Hypocentre Determination (JHD) locations (left panel) and double-difference locations (right panel) of about 10,000 earthquakes that occurred during the 1997 seismic crisis in the Long Valley caldera. The same P-phase data from the Northern California Seismic Network are used in both cases. The size of the system of double-difference equations in this case is about 1 million equations for the 10,000 events. Note the appearance of linear structures in the relocation results which correspond with faults. Figure from *Waldhauser* [2001].

the hypocentre location, and other results using either a systematic grid-search or a stochastic Metropolis-Gibbs sampling approach. The errors in the observations (phase time picks) and in the forward problem (travel-time calculation) are assumed to be Gaussian. This assumption allows the direct, analytic calculation of a maximum likelihood origin time given the observed arrival times and the calculated travel times between the observing stations and a point in space. Thus the 4D problem of hypocentre location reduces to a 3D search over the three spatial coordinates. The grid-search earthquake location is expressed probabilistically through a 3D, spatial PDF [*Lippitsch et al.* 2005; *Lomax et al.* 2000; 2009; *Walsh et al.* 2009; *Zhang et al.* 2009]. In contrast to the single “optimal” location and associated, local, ellipsoidal (Gaussian) hypocentral statistics produced by linearised location methods, the non-linear PDF can define multiple, maximum-likelihood hypocentres and highly non-ellipsoidal (non-Gaussian) errors volumes.

After these hypocentre determination methods have been performed, we investigate the spatiotemporal patterns of these events. These patterns are compared with injection flow-

rates/pressures and injectant temperatures, as well as existing fracture information.

1.3.1.3 Data

Continuous waveform data for the period 29 September 2008 to 28 February 2009 have been accessed from a GNS Science database in close collaboration with the GNS Science staff responsible for ongoing microearthquake monitoring at Rotokawa. The Foundation for Research, Science and Technology has funded the data collection from a local seismic array consisting of nine surface stations. Each station comprises a Taurus recorder sampling at 200 Hz. The data is being worked on by Steven Sherburn, Sandra Bourguignon and Stephen Bannister (GNS Science) in parallel with this study. Sandra Bourguignon and Stephen Bannister are currently processing the Rotokawa dataset by detecting and locating events with a median magnitude of ~ 1 using a combination of manual and automatic phase picks. The principal objective of their research is 3D delineation of seismic properties inside the reservoir, for comparison to electrical conductivity information. They have provided phase pick information, origin times and locations for 81 events. A 3D velocity model has also been provided for use as the velocity model in our location calculations. In consultation with MRP, flow-rates for RK20 have been collated for comparison with the spatiotemporal characteristics of seismicity.

1.3.2 Optimal network design at Kawerau

‘‘For installation of a permanent network to monitor seismicity in a given area the basic question is: what is the minimum number of stations required, and where should they be installed to achieve a desired precision in locating the earthquakes?’’

– *Uhrhammer* [1980]

In Chapter 4, we address this question, with the goal of designing an optimal extension of the existing seismic network at Kawerau geothermal field.

Geothermal fields are complex geological environments, both at and beneath the surface, and a number of complexities restrict the siting of measuring equipment in such environments. At the surface, we routinely encounter natural and artificial noise sources, such as power plants, roads, water-bodies; ground conditions may be hostile because of geothermal surface features or geothermally altered/unstable ground, etc; and in some situations there may be issues of land ownership and access. Beneath the surface, high temperatures and fluid saturation and distribution may cause high seismic attenuation [Gudmundsson 2004; Gurevich *et al.* 2007; Jones *et al.* 1980; Mavko and Nur 1979; Winkler and Nur 1982]. In spite of such challenges, and for the reasons outlined above, geothermal operators are particularly interested in *microseismicity* induced through exploitation of the field. We must therefore contend with the seismicity of interest having both a low intrinsic energy and propagation paths through a high-attenuating medium.

In a simple environment with no attenuation, and a relatively homogeneous velocity structure, we can use intuition to effectively choose seismometer locations in order to constrain earthquake hypocentres. Such an approach typically favours a layout involving distant stations arranged in a triangular quadripartite network [Fig. 1.9; Uhrhammer 1980]. In more complicated environments, however, such as in a geothermal field, the optimal configuration may not be intuitive and an algorithm for evaluating various locations is desirable.

The major uncertainties in earthquake locations stem from two sources: the geometric configuration of the network with respect to earthquake locations (data acquisition), and the accuracy of phase picks and the crustal model used during modelling (data analysis). Whilst the importance of improving the analysis component is well recognised (typically in inversion methods), the acquisition component (which relates to experimental design) is largely ignored. Curtis and Maurer [2000] pointed out that more than 10,000 papers had been published since 1955 on inversion methods alone, whilst a mere 100 had been published on experimental design.

Since no amount of sophisticated data analysis can overcome a lack of information inherent in the initial data, it seems prudent to ensure that the best possible data are being

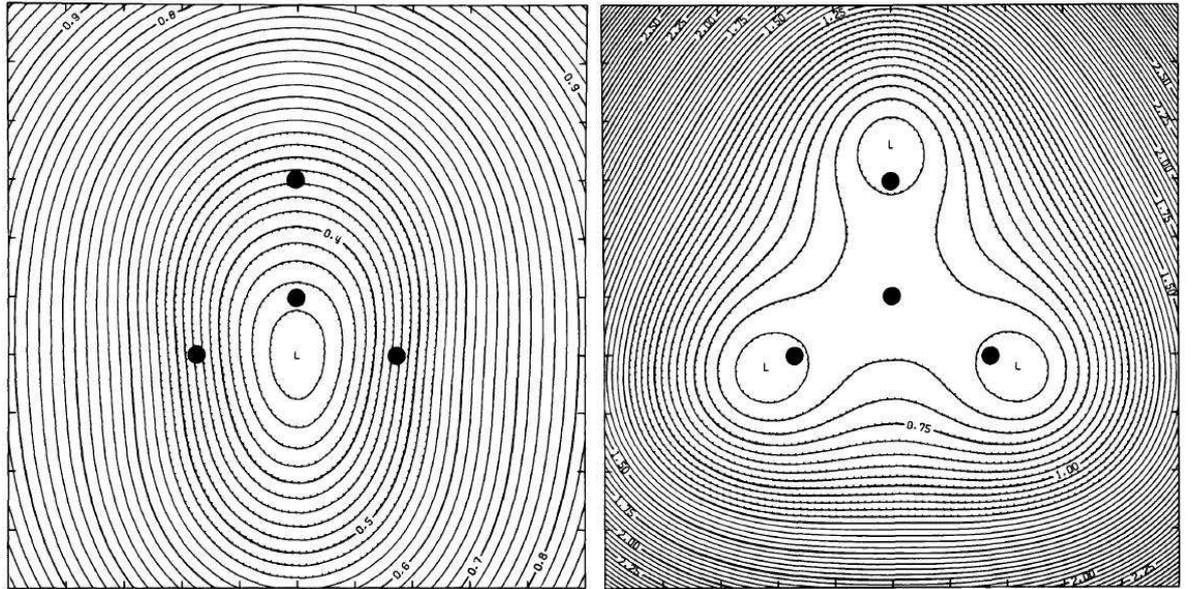


Figure 1.9: Uncertainties in an earthquake hypocentre for a triangular quadripartite network. (left) Contour plot of the uncertainty in the x coordinate of the hypocentre, (right) Contour plot of the uncertainty in the depth estimate of the hypocentre. Note the uncertainty in x and z increases outside the network. The minimum x is at the centre of the widest aperture of the network in the x direction, and the minimum z is about twice the minimum x . Comparison of (left) and (right) shows that uncertainty in the depth estimates increases most rapidly outside the network. Figures from *Uhrhammer* [1980].

acquired [*Curtis and Maurer* 2000]. This becomes especially evident when considering the immense cost of deploying, maintaining, and processing the data from a large-scale seismic network. If cost were not an issue, we could clearly deploy as dense an array of seismometers as needed to acquire the data required: however, funds are typically limited, and hence we face a trade-off between the cost of a network and its resolution capability. The decisive advantage of planning surveys using experimental design techniques is that information and assumptions can be examined, tested and considered before incurring data acquisition expenses [*Maurer and Boerner* 1998].

Consider the basic model-data relationship $\mathbf{d} = \mathbf{F}_\xi(\theta)$, where \mathbf{F}_ξ is a function defined by the experimental design ξ (the measuring equipment locations) and represents the relationship imposed by the acquisition system between the Earth model θ (the earthquake locations) and data \mathbf{d} (seismic travel times). Inversion methods address the task of estimating θ given \mathbf{d} and the ξ implicit in \mathbf{d} . In statistical experimental design, however, the object is to search

the experiment space (the set of all experiments, ξ) for the highest quality mapping between data space \mathbf{d} (the set of all data corresponding to individual observations) and model space θ (the set of all Earth models) [Coles and Morgan 2009]. In other words, we wish to infer θ (earthquake locations), and calculate \mathbf{d} (travel times) in order to identify the best ξ (station geometry).

Statistical experimental design (SED) techniques have been applied before to the problem of determining seismometer locations to locate earthquakes with a minimum uncertainty. *Rabinowitz and Steinberg* [1990] used a linearised problem with a single point source earthquake, extending the earlier algorithm (DETMEX) of *Mitchell* [1974]. This algorithm works by generating an initial random network model from all the possible locations of receivers and then sequentially adding and removing sites and reassessing the quality of the network, until no notable quality improvement occurs with further alterations. As this algorithm depends upon the initial network model, a series of runs using different initial models is required to ensure convergence to an optimal network. *Rabinowitz and Steinberg* [1990] addressed the problem of stations clustering at optimal locations (meaning that redundant data are collected) by assuming station errors were correlated when the angle between two stations from an earthquake epicentre is $< 180^\circ$.

Rabinowitz and Steinberg [1990]’s work was extended by *Steinberg et al.* [1995] to incorporate fault lines and multiple earthquake sources. In this case ‘fault lines’ were represented by three point sources along the line of a fault — in other words, preferential locations of distributed seismicity. In such cases, a ‘compromise’ network is obtained that may not be optimal for each individual source but which provides the most effective coverage for all sources of interest; as a refinement weights can be applied to individual sources based on their perceived importance. Attenuation effects were not considered by either *Rabinowitz and Steinberg* [1990] or *Steinberg et al.* [1995], and it was assumed that each station recorded the P-wave phase arrivals from all events.

More recently, *Curtis* [1999] and *Curtis et al.* [2004] investigated SED techniques using multiple earthquake sources, an adjustment for constant attenuation, and a new algorithm.

The algorithm used by Curtis and colleagues analyses a grid of all possible receivers, each assigned a quality function representing the information contributed by that receiver. The receiver with the smallest quality function is removed and the quality functions of the remaining receivers are reassessed. At the end of this process, each receiver has been assigned a rank based on the order of its removal in the sequence of removals. A network design can now be created using the highest-ranked receivers up to the number required to either obtain sufficient information or to reach a cost threshold. Curtis et al.'s [2004] algorithm is simple, deterministic (always converges to a solution), and is not influenced by the initial input model. However, the algorithm relies on there being a finite number of possible designs and the model-data relationship being approximately linear [Curtis et al. 2004]. The second criterion is an assumption introduced to simplify the problem, but imposes some restrictions on the geometries that can be investigated. It is only strictly pertinent when considering relatively simplistic earth models that assume all parameter distributions are Gaussian [Menke 1984].

Deviations from the Gaussian assumption motivated investigation of non-linear Bayesian survey design by Coles and Curtis [2011], who derived a new survey design criterion by maximising the relative entropy — or Kullback-Leibler divergence [Cover and Thomas 1991] — of the hypocentre probability distributions. The Kullback-Leibler divergence is a measure of the distance between hypocentre probability distributions. With this approach, the goal is to maximise this distance (that is, minimise the overlap of hypocentre distributions) so that each hypocentre is clearly resolved [Fig. 1.10]. Maximisation is carried out by maximising the determinant criterion 'D-criterion' [Pukelsheim 2006] of a nonlinear function that relates the expected data uncertainties and the hypocentre variance. We provide further details in Chapter 4.

We have investigated optimal seismic network design for the extension of Kawerau's seismic network using a modified version of the design criterion developed by Coles and Curtis [2011]. The same criterion has been derived, but in a different manner that allows for an even more straightforward implementation of nonlinear Bayesian survey design. The

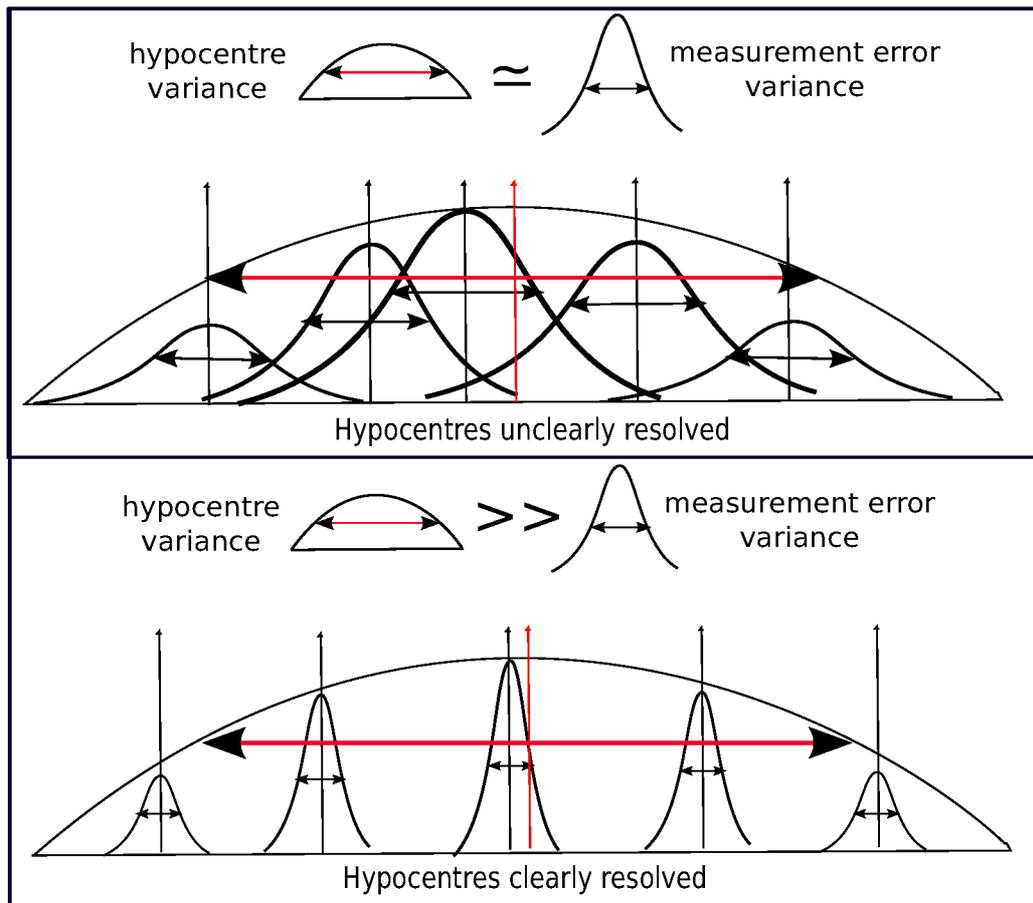


Figure 1.10: Basis of *Coles and Curtis* [2011]’s optimal network design criterion, which we base our criterion on. The aim of the criterion is to maximise hypocentre resolution. We do this by ensuring that the distance between hypocentres (the hypocentre variance) is sufficiently larger than the measurement error, so that the hypocentres are uniquely resolved. The top image shows an example where the hypocentre variance and measurement error variance are similar, so that hypocentres cannot be distinguished from each other. The bottom image shows the ideal case where hypocentres are clearly resolved. For a detailed description and explanation of the mathematics involved see Chapter 4.

criterion works by assessing and returning a quality value for an entire experimental design. For our immediate purpose, an extension of an existing design, this is computationally simplistic. We fix the existing stations and then create an experimental design set encompassing all possible additions of one more station. When the optimal new station location is found, it is fixed in place. This can be iterated through until the number of required new stations is found. The quality value used represents how well the experimental design can resolve earthquake hypocentres. In the single new station designation this encompasses the new station’s importance (whether it duplicates information provided by other receivers or pro-

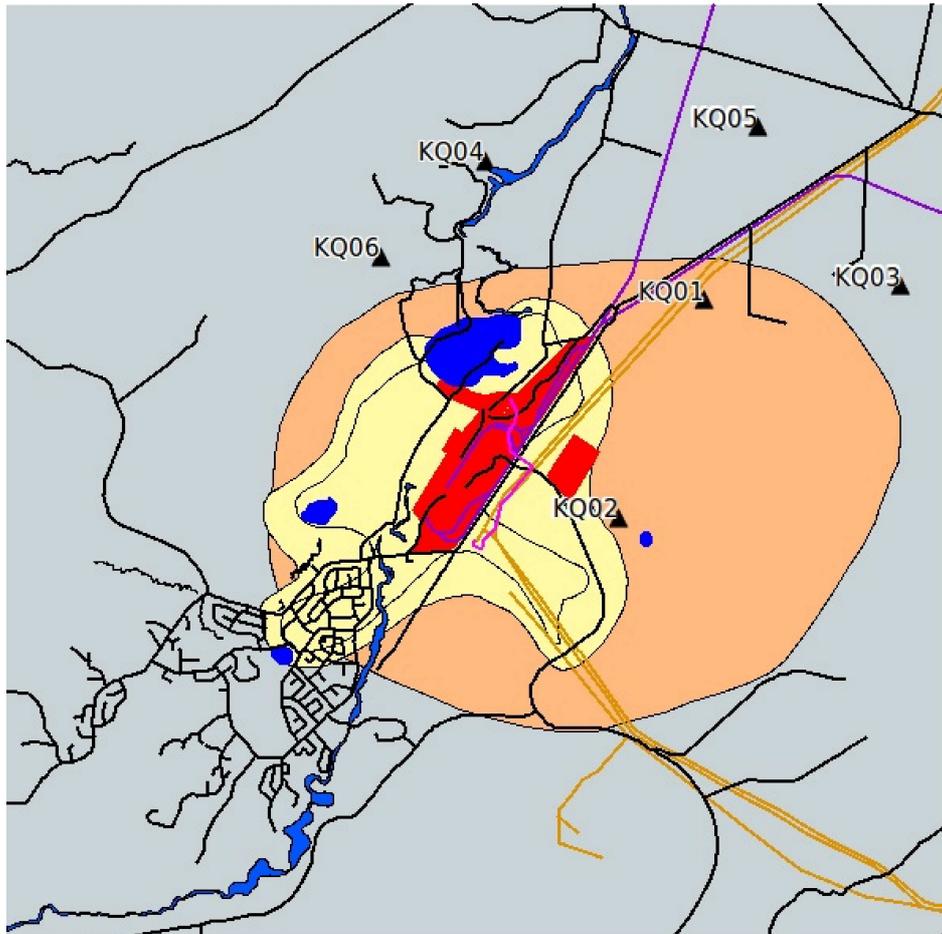


Figure 1.11: GIS data for Kawerau. The red regions are industrial areas, dark orange are power lines, black lines are roads, purple lines are railway lines, blue regions are thermal surface features and the Tarawera River, black triangles show existing seismometer locations, the Salmon coloured region is the reservoir identified by the distribution of resistivity at a depth of 700 m from 3D-MT modelling in 2003 [Bignall and Harvey 2005], and the pale yellow region is the reservoir identified by the 1970 Schlumberger survey with $AB/2=457$ m [Bignall and Harvey 2005].

vides unique information) and quality (signal-to-noise ratio and sensitivity to perturbations in the positions and origin times of events) of the data able to be recorded by the station. We implement the calculation of this value using the following information: locations of the seismicity of interest, a velocity model of the region, attenuation values, surface noise source locations and levels, and possible receiver locations.

1.3.2.1 Data

Mighty River Power has supplied the information needed to apply our network design algorithm to the specific Kawerau situation. This information has been supplied in the form of GIS data detailing: industrial areas, power lines, roads, railway lines, thermal surface features, resistivity boundaries, well locations, and existing seismometer locations [Fig. 1.11]. The locations of seismicity of interest to MRP have also been supplied. A velocity model has been estimated using the GNS Science 3D velocity model of Rotokawa, *Clarke* [2008]'s 1D velocity models for the Kawerau region, and geological information on the Kawerau reservoir supplied by MRP. Attenuation has been estimated from Kawerau reservoir temperatures supplied in *Bignall and Harvey* [2005]. Noise source levels and extent have been estimated based on empirical relations described in a number of papers. Further details are provided in Section 4.3.3.

1.4 Thesis structure

Chapter 2 and Appendix B address the development and testing of software to carry out earthquake detection using a matched filter technique and its application to five months of data from the Rotokawa geothermal field. In Chapter 3 and Appendix C we perform hypocentre location calculations and investigate the spatiotemporal patterns of the events detected at Rotokawa and their implication. Chapter 4 and Appendix D describe the derivation of a design criterion for optimal experimental design, the development of an algorithm to solve this criterion for a seismic network, and its application to Kawerau geothermal field. The key results of this project and their ramifications for further research are summarised in Chapter 5. The methods presented here have been developed for immediate application to geothermal fields, but could similarly be used in other situations: we demonstrate this with a small application to the 2010 Darfield Earthquake [Appendix D.3].

All coordinates in this study are presented in either New Zealand Map Grid (NZGM) Eastings and Northings or New Zealand Geodetic Datum 1949 Latitude and Longitude.

All of the work presented in this thesis has been conducted by me (Zara Rawlinson) unless explicitly stated otherwise: however, it has been written in the first person plural to be consistent with common scientific practice, and to reflect the crucial guidance of my supervisors.

Chapter 2

Earthquake Detection

2.1 Introduction

The detection of earthquakes is typically performed using some form of STA/LTA detector, where the energy in a short-term average (STA) window is divided by a long-term average (LTA) window and a detection is triggered when this ratio exceeds some signal-to-noise ratio (SNR) threshold [Freiberger 1963]. The drawback of such an approach is that false alarm rates (i.e. false detections) increase dramatically for lower SNR thresholds [Schaff and Waldhauser 2008]. Due to background noise and the often sparse density of existing seismic networks, not all earthquakes will necessarily be detected. An earthquake catalogue provided by a seismic network, therefore, has an associated completeness magnitude M_c that defines the lowest magnitude of events that can be reliably and completely recorded by the network. At magnitudes below M_c , events will be missing from the catalogue. This completeness magnitude depends upon the seismic array density and geometry, the geological siting of the instruments, the level of background noise, and the detection and data processing methods used [Schorlemmer and Woessner 2008]. In earthquake catalogues provided by some of the densest regional seismic arrays in the world — for example, the Japan Meteorological Agency Catalogue (JMAC) and the Southern California Seismic Network (SCSN) — the completeness magnitude reaches levels of $M_c = 1.0$ within the densest regions of the arrays

[Nanjo *et al.* 2010; Schorlemmer and Woessner 2008].

In this study, we are making use of the small (nine instrument) local recording array at Rotokawa, which is sited in a region of volcanic deposits [Figs. 1.2 & 1.6]. Volcanic deposits are highly attenuative, and lack the ideal high elastic-strength properties that hard rock supplies for seismometer siting. As discussed in Chapter 1, we are interested in low-magnitude microearthquakes ($M \lesssim 3$) whose detection and identification are complicated by low-SNRs due to their occurrence in a region of high attenuation (a hot, wet geothermal reservoir) and high noise (an industrial zone). As we cannot alter the current instrument siting geometry or the background noise, we aim to decrease the completeness magnitude for this data set through the detection method used. This chapter provides a description of the work undertaken to perform automated earthquake detection using a matched filter technique, with the aim of reliably detecting earthquakes of magnitude $M \sim 0$.

The matched filter technique, in which data are cross-correlated with a predetermined waveform template, is recognised as the best method for detecting a known signal within noise [Anstey 1966; Gibbons and Ringdal 2006; van Trees 1968]. We can apply this technique due to the confined geometry of the Rotokawa geothermal field, as it offers a relatively limited set of sources and propagation paths affecting the seismic waveforms recorded. We create an *a priori* set of waveform templates which we treat as representing the set of earthquake locations and focal mechanisms characterising the area. As this set is chosen by us, albeit with knowledge of the local seismicity, the reduction in completeness magnitude will only be in reference to this particular set of earthquakes, and not the entire catalogue. We chose this set with the aim of representing the catalogue as fully as possible. These ‘master events’ are high-SNR events identified manually during analysis by GNS Science, and are employed in the matched filter process to detect lower-magnitude earthquakes with similar waveform characteristics (slave events) within the field. The general approach is as follows [after Brown *et al.* 2008; Gibbons and Ringdal 2006; Maurer and Deichmann 1995]: Waveform templates of the master events are cross-correlated with continuous seismic data, and signal detectability is enhanced by summing correlation coefficients from single sen-

sors across the entire network to create a network correlation coefficient. High correlations are interpreted as corresponding to approximately co-located events (occurring on the same fracture plane, with similar focal mechanisms, and emitting waves travelling along similar paths).

Waveform cross-correlation is being increasingly used in the seismological community for a number of purposes: high-accuracy phase-picking, high-accuracy relative earthquake locations, high-accuracy focal mechanism determinations, evaluation of source time functions, rupture characteristics, characterisation of Greens function and their spatial derivatives, multi-source and multi-receiver processing for Earth structure etc. (as outlined in Chapter 1). For earthquake detection purposes, waveform cross-correlation has been employed in a number of recent studies [e.g. *Brown et al.* 2008; *Eisner et al.* 2008; *Gibbons et al.* 2007; *Gibbons and Ringdal* 2006; *Peng and Zhao* 2009; *Schaff and Waldhauser* 2008; *Schaff* 2009; *Shelly et al.* 2007]. However, to our knowledge no software implementing waveform cross-correlation for the purpose of earthquake detection has been made freely available for general use.

Preliminary codes aimed at this task were originally written by Charlotte Rowe (ca. 2000) [based on cross-correlation methods for phase pick refinement outlined by: *Aster and Rowe* 2000; *Rowe* 2002]. These were very kindly provided to us by Rowe in the early stages of this project, with the inclusion of many useful subroutines, and served as the basis for subsequent code development. Modifications have been made, however, to the mathematical framework and the programme structure overall.

The remainder of Chapter 2 describes our development of a matched filter detection method that extends on the approach of Rowe (codes ca. 2000) and the application of this method to the Rotokawa seismic network. The mathematical basis of the cross-correlation procedure used is outlined in Sections 2.3 and 2.2; the results of synthetic testing are detailed in Section 2.4; and the results of applying this method to the Rotokawa data set are displayed in Section 2.5. Details of the code's structure and implementation are provided in Appendix B.

2.2 Waveform cross-correlation

Waveform cross-correlation is a procedure used to quantitatively describe the similarity between two functions, and to determine the time delay between them [Phillips *et al.* 2007]. The cross-correlation of functions $a(m)$ and $b(m)$, at lag g is defined as:

$$R_{ba}(g) = \int b(m)a(g+m) dm \quad (2.1)$$

In our case, these functions are seismograms that have been recorded on the same component at a seismograph station. In an autocorrelation, which is the cross-correlation of a signal with itself, there will always be a peak at a lag of zero, unless the signal is the trivial constant signal.

We use a normalised waveform cross-correlation method performed in the frequency domain and outline its derivation below.

The sample cross-correlation operation for traces $a(m)$ and $b(m)$ where $m = 0, 1, \dots, 2N$, at lag g is described in the time domain [Phillips *et al.* 2007] by

$$R_{ba}(g) = \frac{1}{2N} \sum_{m=1}^{2N} b(m)a(g+m), \quad g = 0, 1, \dots, 2N-1 \quad (2.2a)$$

$$= \frac{1}{2N} [a(m)*b(-m)] \quad (2.2b)$$

where $*$ denotes the convolution operation. Taking the Fourier transform yields

$$\mathcal{F}\{R_{ba}(g)\} = \mathcal{F}\left\{\frac{1}{2N} [a(m)*b(-m)]\right\} \quad (2.3a)$$

$$= \frac{1}{2N} [\mathcal{F}\{a(m)\} \times \mathcal{F}\{b(-m)\}] \quad (2.3b)$$

$$= \frac{1}{2N} A(m)B(-m) \quad (2.3c)$$

where $A(m)$ and $B(m)$ are the Fourier transforms of $a(m)$ and $b(m)$, respectively. We can

therefore describe the cross-correlation operation in the frequency domain as,

$$R_{ba}(g) = \mathcal{F}^{-1} \left\{ \frac{1}{2N} A(m) B(-m) \right\} \quad (2.4)$$

As the discrete-time sequence $b(m)$ is real-valued, it can be shown that $B(-m) = B^*(m)$, where $*$ denotes the complex conjugate [Phillips *et al.* 2007]. This means that,

$$R_{ba}(g) = \mathcal{F}^{-1} \left\{ \frac{1}{2N} A(m) B^*(m) \right\} \quad (2.5)$$

The fully normalised cross-correlation coefficient can then be defined as,

$$\text{CC}(g) = \frac{R_{ba}(g)}{\sqrt{R_{aa}(0)R_{bb}(0)}} = \frac{\mathcal{F}^{-1} \left\{ \frac{1}{2N} A(m) B^*(m) \right\}}{\sqrt{\mathcal{F}^{-1} \left\{ \frac{1}{2N} B(1) B^*(1) \right\} \mathcal{F}^{-1} \left\{ \frac{1}{2N} A(1) A^*(1) \right\}}} \quad (2.6)$$

where $R_{aa}(0)$ and $R_{bb}(0)$ are the autocorrelations of $a(m)$ and $b(m)$ at zero lag. If $b(m) \equiv a(m)$ then,

$$\text{CC}(g) = \frac{R_{aa}(g)}{\sqrt{R_{aa}(0)R_{aa}(0)}} = \frac{R_{aa}(g)}{R_{aa}(0)} \quad (2.7)$$

$$\text{and } \text{CC}(0) = 1 \quad (\text{zero lag}) \quad (2.8)$$

Figure 2.1 shows four examples of delay time estimation using cross-correlation. In the first example, the seismograms $a(m)$ and $b(m)$ are identical, but $b(m)$ has been shifted by 163 samples: we therefore obtain a maximum cross-correlation value of 1 at a lag of 163 samples. The second example uses the same two traces, but now $b(m)$ has had random noise added to it in addition to the 163 sample shift. It can be seen that the maximum cross-correlation value drops a small amount (from 1.0 to 0.9), but the procedure still performs well in the presence of noise. In the third example the seismograms are not similar, and so the cross-correlation takes much lower values, and a relatively low maximum. The fourth example uses the same two traces as (c), but this time the second seismogram has random noise added to it, yielding low cross-correlation values as expected.

The cross-correlation we need to perform involves cross-correlating master templates with an average length of ~ 2 seconds, with five months of data. Due to this large length difference between the time series being compared it would be computationally inefficient to zero-pad the master template and perform cross-correlation the standard way. We therefore extract windows of the data the same length as the master template to perform cross-correlation on, and shift through the data until we have examined windows from the entire record [Fig. 2.7]. The information we extract from each window is the start time of the window, the maximum cross-correlation value and its lag within the window (the bottom panels in Fig. 2.1: $CC(\hat{g})$, where \hat{g} is the lag that provides the maximum cross-correlation value).

2.3 Method

The waveform cross-correlation procedure detailed in this section is intended to identify and extract waveforms of similar earthquakes. This involves four key steps: determining master templates; cross-correlating the master templates with data; calculating network correlation coefficients; and calculating detection thresholds. We discuss each of these steps individually below.

2.3.1 Master templates

We start by introducing the terminology used throughout the remainder of this chapter. For every master *event* (high-SNR earthquake) there is a different master *template* (a seismic waveform) for every seismograph component (E, N, Z) at every seismograph station that records this event [Fig. 2.2]. Earthquakes detected by a *master* event are termed *slave* events.

The utility of the cross-correlation method is governed by the characteristics of the master templates used [Baisch *et al.* 2008; Maurer and Deichmann 1995]. Of particular importance are the window length of the master template and the seismic phases included in this window. The maximum correlation coefficient within a window of data is a function of the

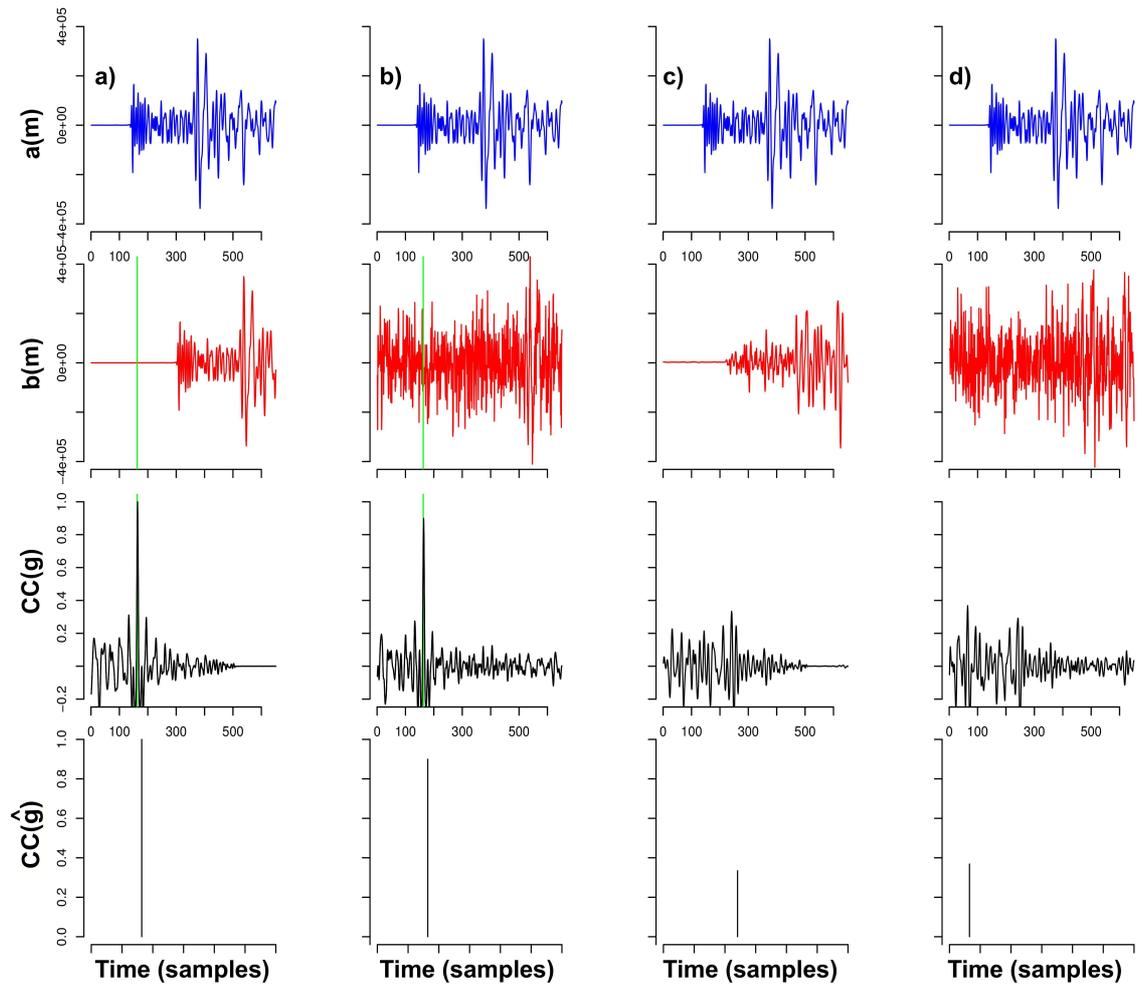


Figure 2.1: Four examples of delay time estimation using cross-correlation. The top panels show an earthquake signal $a(m)$ in sample time. The second panels show a) the same earthquake signal shifted by 163 samples (this shift amount is marked by the green line). b) the shifted signal in a) has random noise added to it. c) a different signal, d) random noise is added to c). The third panels show the cross-correlation obtained via frequency cross-correlation (the method we perform in this study) versus sample time. The bottom panels show the maximum cross-correlation at the delay time, \hat{g} , that it is obtained in each example, this is the output we calculate for each data window in this study. Note that in a) and b) the delay time is accurately calculated, shown by the highest peaks matching the location of the green line, and that the frequency cross-correlation performs well even in the presence of high amplitude noise. In c) and d) the waveforms have poor correlation, and this is shown by the lower CC value obtained.

length of this correlation window: the smaller the window, the more likely it is that an arbitrary portion of the waveform will yield a high correlation and risk a spurious detection; conversely, too long a window will include the low-SNR coda of an event record, which

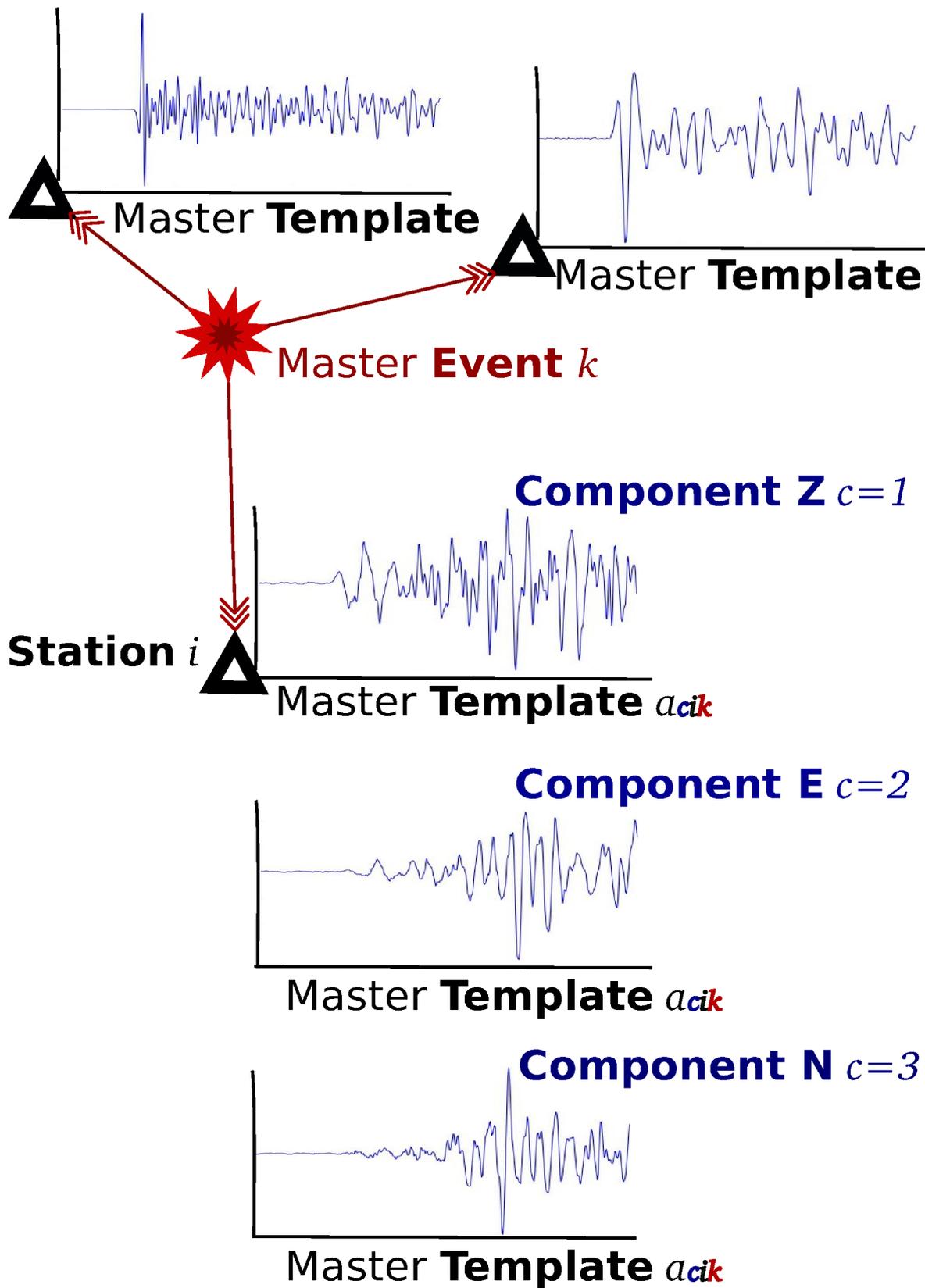


Figure 2.2: A master event is a relatively large-magnitude earthquake (generally $M > 1$ in this study). This earthquake generates different waveforms at different stations and components. These waveforms are cut at each station/component to create a suite of master templates specific to that event.

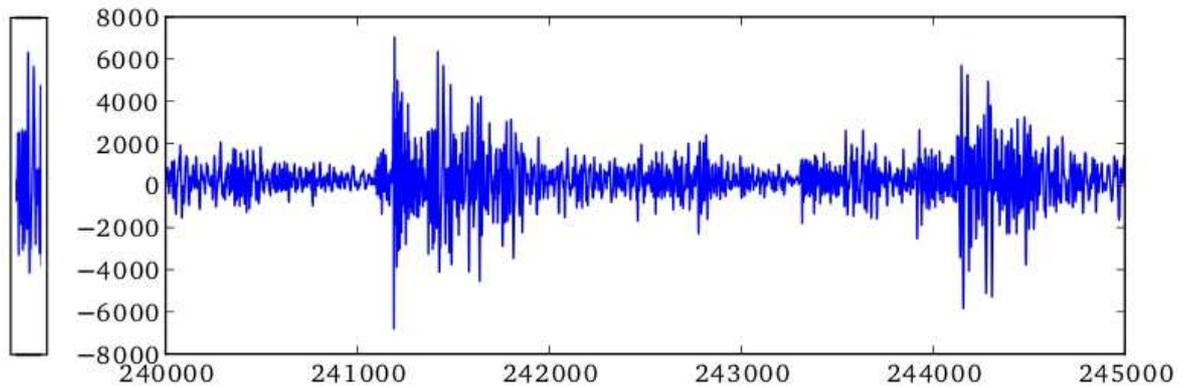


Figure 2.3: Master template window size example. (left) small window, likely to correlate highly with many waveform portions. (right) long window, unlikely to correlate highly with any other waveforms.

may result in a reduction of the computed correlation coefficient even for identical events [Fig. 2.3; *Baisch et al.* 2008]. In general, the P-wave form is structurally rather simple, and inclined to yield high-correlation, whereas the S-wave form is structurally more complex and more inclined towards low-correlation [*Baisch et al.* 2008]. One upshot of this is that a high threshold for P-wave correlation may still lead to false associations, and conversely that a high threshold for S-wave correlations may lead to missed associations. Past cross-correlation studies have used a single phase [*DeShon et al.* 2007; *Shelly and Hardebeck* 2010; *Shelly et al.* 2009], both phases separately [*Aster and Rowe* 2000; *Maurer and Deichmann* 1995], or a window including both phases [*Du et al.* 2004; *Ferretti et al.* 2005; *Gibbons et al.* 2007; *Lees* 1998; *Schaff* 2009; *Stankova et al.* 2008]. Single-phase studies have used just the P-phase for P-wave phase pick refinement (with the P-wave chosen as it provided more numerous and accurate data) [*DeShon et al.* 2007], and just the S-phase for matched filter detection of low-frequency earthquakes (LFEs), for which the S-wave is the strongest arrival [*Shelly and Hardebeck* 2010; *Shelly et al.* 2009]. Studies using both phases separately have focused on the detection of families of events with similar waveforms, with pre-selection using a maximum distance threshold [*Maurer and Deichmann* 1995], and phase pick refinement [*Aster and Rowe* 2000]. Studies using a window including both phases have addressed tasks such as: cluster identification where both phases provide a distance restriction without needing to choose *a priori* a maximum distance separation restriction [*Ferretti et al.* 2005],

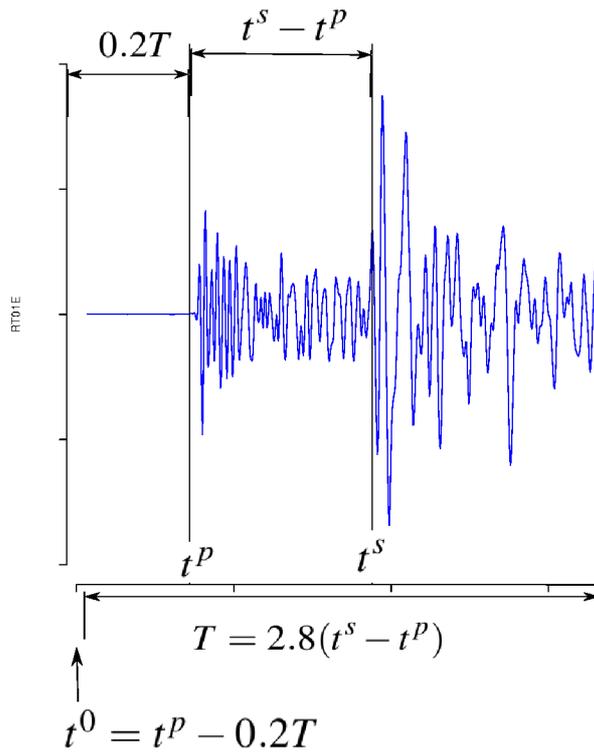


Figure 2.4: The constituents of a single master *template*.

earthquake detection in microseismic swarms [Stankova *et al.* 2008], regional earthquake detection [Gibbons *et al.* 2007; Schaff 2009], and relative relocation [Du *et al.* 2004; Lees 1998]. To capture the desired phases, the window length beneficial for a study depends upon the dominant frequency of the events of interest, and the distance of stations from events.

This study focuses on nearby (typically <10 km), high-frequency (dominant frequency of ~ 10 Hz) microearthquakes, and hence we have chosen to adopt the approach of Baisch *et al.* [2008], who examined the effects of different correlation window lengths at a fluid injection experiment at the Kontinentale Tiefbohrung der Bundesrepublik Deutschland (KTB) deep drilling site. Baisch *et al.* [2008] used a window encompassing both the S and P onsets and only the early portion of the S-coda, that is not affected by backscattering. Including both phases places constraint on the hypocentral distances suitable for cluster/family identification. Including the early S-coda introduces waveform complexity beyond that of the P-wave, but without the disadvantages of the less-coherent production (due to scattering) of the later S-coda.

Table 2.1: Master template nomenclature

Symbol	Explanation
$a_{cik}(t)$	master template for component c , station i , event k , and $t = (0, 1, 2, \dots, T_{ik})$ sec
$t_{ik}^0 = t_{ik}^P - (0.2T_{ik})$	start time for station i , event k
$c = (1, 2, 3)$	components (Vertical, North, East)
$i = (1, 2, \dots, I)$	stations
$k = (1, 2, \dots, K)$	events
t_{ik}^S	time of S-pick for event k and station i
t_{ik}^P	time of P-pick for event k and station i
$T_{ik} = 2.8(t_{ik}^S - t_{ik}^P)$	total window length calculated for station i and event k
η_k	reference time for event k
ϕ_{ik}	shift time for station i from event time η_k

Baisch et al. [2008] determined the best window length to be $T_{ik} = 2.8(t_{ik}^S - t_{ik}^P)$, where t_{ik}^S is the time of the S phase pick and t_{ik}^P is the time of the P phase pick of the arrival of the wave at the i^{th} station from the k^{th} event. The window is calculated so that 20% of the total window length lies ahead of the P-phase onset. These aspects of a master template are identified in Figure 2.4. As illustrated in Figs. 2.2, 2.4 and 2.5 and with notation summarised in Table 2.1, the master templates $a_{cik}(t)$ are created for components $c = (1, 2, 3)$ (Vertical, North, East), stations $i = (1, 2, \dots, I)$, events $k = (1, 2, \dots, K)$ and $t = (1, 2, \dots, T_{ik})$ seconds, with a start time $t_{ik}^0 = t_{ik}^P - 0.2T_{ik}$. For each event k , the reference time η_k is set as the earliest P-arrival, $\eta_k = \min \{t_{ik}^P | i = 1, 2, \dots, I\}$. Then for each station i , the station shift ϕ_{ik} is calculated as the difference between the reference time η_k and that station's P-arrival, $\phi_{ik} = t_{ik}^P - \eta_k$. This shift information is stored for later use during the Network Correlation Coefficient calculation. A list of 81 high-SNR events with accurate phase picks was provided by GNS Science. Highly correlated events within this list were stacked [*Stein and Wysession* 2003] to provide us with a set of 14 master events that constitutes our master template library [Fig. 2.6]. Details of this procedure follow in Section 2.5.

2.3.2 Cross-correlation

As discussed in Section 2.2, we perform cross-correlation between our set of master templates and the five months of data from Rotokawa by extracting windows of the data the

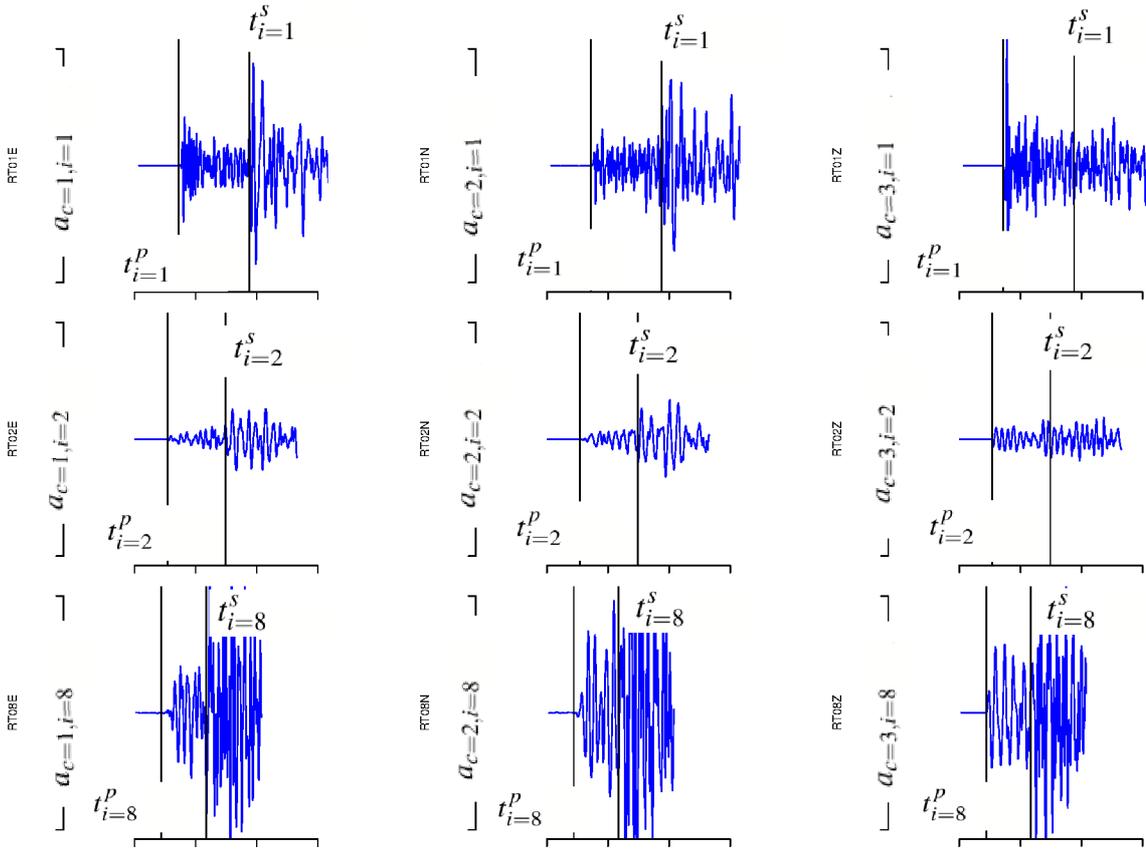


Figure 2.5: An example of the constituents of a single master event, $k = 1$ (the index k has been omitted for clarity). Each row shows the master templates from a different seismograph station i , and each column shows the master templates from a different seismograph component c . The bottom row (station RT08, $i = 8$) has the earliest t^p , therefore this is set as the reference time, $\eta = t_{i=8}^p$, and each station's shift is calculated as $\phi_i = t_i^p - \eta$.

$i=$	$k=1$	2,	3,	...	14
1,	$c=1,$ 2, 3			...	
2,				...	
...
9				...	

Figure 2.6: Representation of our library of 14 master events with their corresponding master templates over the Rotokawa nine station array, each with 3 component seismographs. Note that some events may not be recorded on every station.

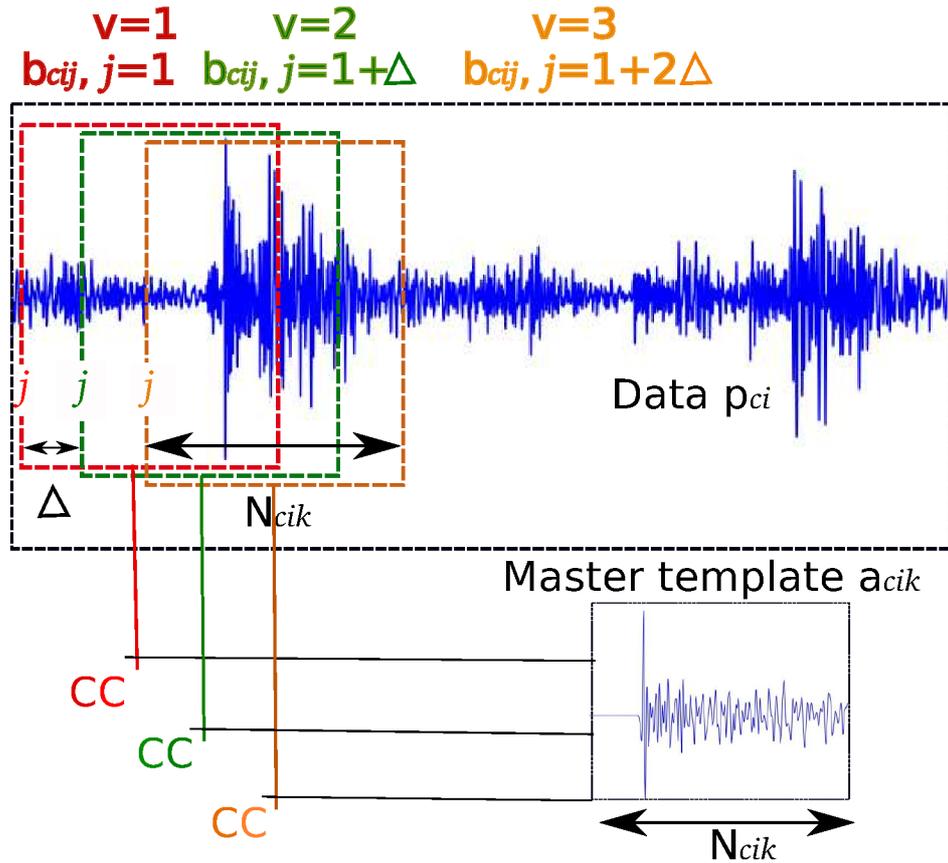


Figure 2.7: Cross-correlation method. A temporary window of the data the same length as the master template is created and the two waveforms cross-correlated. Another temporary window of the data is then created, with the start position j shifted by Δ . Cross-correlation is again performed. This is continued until the end of the data is reached.

same length as the master template to perform cross-correlation on, and shift through the data until we have examined windows spanning the entire length [Fig. 2.7]. The information we extract from each window is the start time of the window, and the maximum cross-correlation value and its lag within the window ($CC(\hat{g})$ in Fig. 2.1).

We implement the following method to perform cross-correlation between a master template $a_{cik}(n)$ and data $p_{ci}(\ell)$, where $n = (1, 2, \dots, N_{cik})$ samples and $\ell = (1, 2, \dots, L_{ci})$ samples, with $N_{cik} \ll L_{ci}$. The cross-correlation is performed sequentially on data windows the same length as the master template. As $N_{cik} \ll L_{ci}$, a temporary N_{cik} -length portion of the data $p_{ci}(\ell)$ is created, $b_{cij}(n)$, where $b_{cij}(1) = p_{ci}(j)$. The start time of the window (in samples) is $j = (v - 1)\Delta + 1$, where v corresponds to the window number $v = (1, 2, \dots, (1 + (L_{ci} - N_{cik} - 1)/\Delta))$. Once the cross-correlation process described below is

Table 2.2: Cross-correlation nomenclature

Symbol	Explanation
$a_{cik}(n), n = (1, 2, \dots, N_{cik})$	master template
$a(m), b(m), m = (1, 2, \dots, 2N)$	traces padded to $2N_{cik}$
$A(\omega), B(\omega)$	FFT of $a(m)$ and $b(m)$
$\text{Re}(A)$	real part of FFT of $a(m)$
$\text{Im}(A)$	imaginary part of FFT of $a(m)$
\widetilde{AB}^*	coherency filtered cross-spectrum of $a(m)$ and $b(m)$
$b_{cij}(n), b(1) = p(j)$	temporary trace window of $p(\ell)$ starting at sample j
$\text{CC}_{cikj}(\hat{g})$	maximum fully normalised cross-correlation coefficient of $a(m)$ and $b(m)$ in data window beginning at j
\hat{g}	index of maximum $R_{ab}(g)$
$j = (v - 1)\Delta + 1$	start time (in samples) of window v
N_{cik}	length of trace $a(m)$ and window correlation length
$2N_{cik}$	length of padded traces
$p_{ci}(\ell), \ell = (1, 2, \dots, L_{ci})$	data
$R_{ab}(g), g = 0, 1, \dots, 2N - 1$	cross-correlation of $a(m)$ and $b(m)$
$s(\omega) = AB^*$	cross-spectrum of $a(m)$ and $b(m)$
$S(\omega)$	amplitude of the cross spectrum in frequency bin widths q
$T[a(n)]$	multi-taper window applied to $a(n)$
$v = (1, 2, \dots, (L_{ci} - N_{cik})/\Delta)$	window number from data $p_{ci}(\ell)$
$w^{(r)}$	eigentaper of order r
X	real part of the cross-spectrum of $a(m)$ and $b(m)$
Y	imaginary part of the cross-spectrum of $a(m)$ and $b(m)$
\widetilde{X}	X with coherency filtering applied
\widetilde{Y}	Y with coherency filtering applied
Δ	window shift step
φ_{cikj}	timing (in samples) of $\text{CC}_{cikj}(\hat{g})$

finished for window $v = 1$, a new temporary N_{cik} -length window of the data is created, $b_{cij}(1) = p_{ci}(j)$ and $v = 2$ [Fig. 2.7 & Table 2.2]. The algorithm is repeated until the end of the data is reached, at which point $v = (1 + L_{ci} - N_{cik} - 1)/\Delta$, and $j = L_{ci} - N_{cik}$ [Stankova et al. 2008]. We have chosen to use $\Delta = 100$ samples (0.5 sec); this ensures that successive windows overlap, avoiding window-edge effects, and provides the opportunity for events very close in time to be detected. We have $L_{ci} = 17280000$ samples (one day), and $225 \leq N_{cik} \leq 768$ samples, giving a value of $44\% \geq \frac{\Delta}{N_{cik}} \geq 13\%$ overlap of successive windows.

The following steps are performed on each window in preparation for cross-correlation, with nomenclature summarised in Table 2.2:

1. Signal processing

- (a) Normalise the two waveforms $a_{cik}(n)$ and $b_{cij}(n)$ by removing the mean and dividing by the maximum value of the centred data.
- (b) Taper each waveform using multi-taper window method. Eigentapers $w^{(r)}$ are calculated up to the eighth order, $r = 0, 1, \dots, 8$, by discrete prolate spheroidal sequences (DPSS) [Bell *et al.* 1993] and applied to the waveforms concerned if their corresponding eigenvalues are > 0.5 . If one order of DPSS does not converge during the calculation of these tapers [Bell *et al.* 1993], then a Hann window [Blackman and Tukey 1959] is used instead. See Appendix A.1 for further details. The tapered window is calculated as,

$$T[a_{cik}(n)] = \sum_{r=0}^8 a_{cik}(n)w^{(r)}(n) \quad (2.9)$$

- (c) Pad each waveform with zeros to length $2N$ (here the subscript N_{cik} is suppressed for clarity) to ensure that the circular convolution result is the same as the linear convolution result (remove any wraparound effects), allowing the Discrete Fourier Transform to be used as a proxy for linear convolution [Phillips *et al.* 2007]. This creates traces $a(m)$ and $b(m)$ with $m = (1, 2, \dots, 2N)$ samples.

2. FFT

- (a) Transform $a(m)$ and $b(m)$ into the frequency domain using the Fast Fourier Transform (FFT),

$$A(\omega) = \mathcal{F}[a(m)](\omega) \quad (2.10a)$$

$$B(\omega) = \mathcal{F}[b(m)](\omega) \quad (2.10b)$$

where $\omega = (1, 2, \dots, 2N)$. $A(\omega)$ and $B(\omega)$ are split into their real components $\text{Re}(A)$, $\text{Re}(B)$ and imaginary components $\text{Im}(A)$, $\text{Im}(B)$.

3. Cross-spectrum

- (a) Calculate the cross-spectrum of $A(\omega)$ and $B(\omega)$ using conjugate multiplication, where $*$ denotes the complex conjugate,

$$AB^* = [\text{Re}(A) + i\text{Im}(A)][\text{Re}(B) - i\text{Im}(B)] \quad (2.11)$$

$$= [\text{Re}(A)\text{Re}(B) + \text{Im}(A)\text{Im}(B)] + i[\text{Im}(A)\text{Re}(B) - \text{Re}(A)\text{Im}(B)] \quad (2.12)$$

$$= \text{Re}(AB^*) + i\text{Im}(AB^*) \quad (2.13)$$

$$= X(\omega) + iY(\omega) \quad (2.14)$$

$$= s(\omega) \quad (2.15)$$

This conjugate multiplication is also used to calculate the magnitudes $|A|$ and $|B|$. The cross-spectrum is employed in the remaining steps.

- (b) Calculate the moving average of the amplitude of the cross spectrum in frequency bin widths q ,

$$S(\omega) = \frac{1}{q} \sum_{\omega'=\omega-q/2}^{\omega+q/2} |s(\omega')| \quad (2.16)$$

where the frequency index $\omega = q/2, 1 + q/2, \dots, N - q/2$. This total signal amplitude in each frequency bin is independent of phase. As in-phase signals add constructively and out-of-phase signals add destructively, this amplitude estimate of the cross-spectrum represents the phase coherency of the two signals. This ‘coherence filtering’ [from codes provided by Charlotte Rowe, pers. comm. 2009; cf. *Aster and Rowe* 2000; *Rowe* 2002] is used to zero-phase pre-filter the template and data spectra, emphasising coherent and/or high-energy frequency bands prior to cross-correlation. It is applied by multiplying the real and imaginary components of the cross spectrum,

$$\tilde{X}(\omega) = S(\omega)X(\omega) \quad (2.17)$$

$$\tilde{Y}(\omega) = S(\omega)Y(\omega) \quad (2.18)$$

This is also applied to each of the spectra A and B . In the absence of correlated noise, this adaptive pre-filtering on the basis of high phase-coherence [Aster and Rowe 2000; Rowe 2002] may yield superior results to *a priori* filters [DeShon et al. 2007] as it down-weights the incoherent frequency bands whilst retaining the useful signal. See Appendix A.2 for an example. Re-compute the cross-spectrum from the filtered real and imaginary components,

$$\widetilde{AB^*} = \widetilde{X}(\omega) + i\widetilde{Y}(\omega) \quad (2.19)$$

4. Cross-correlation

- (a) Perform the Inverse Fast Fourier Transform (IFFT) on the cross-spectrum to calculate the (filtered) cross-correlation of $a(m)$ and $b(m)$ at lag g , $R_{ba}(g)$, where $g = 0, 1, \dots, 2N - 1$ [Eq. 2.2a].

$$R_{ba}(g) = \mathcal{F}^{-1} \left\{ \frac{1}{2N} \widetilde{AB^*} \right\} \quad (2.20)$$

5. Normalisation

- (a) Using the same method (steps 5–7), calculate the filtered autocorrelations $R_{aa}(g)$ and $R_{bb}(g)$,

$$R_{aa}(g) = \mathcal{F}^{-1} \left\{ \frac{1}{2N} \widetilde{AA^*} \right\}, R_{bb}(g) = \mathcal{F}^{-1} \left\{ \frac{1}{2N} \widetilde{BB^*} \right\} \quad (2.21)$$

- (b) Find the maximum, $R_{ba}(\hat{g})$, of the cross-correlation $R_{ba}(g)$, where \hat{g} is the index of this maximum.
- (c) Calculate the fully normalised cross-correlation coefficient, $\text{CC}_{cikj}(\hat{g})$, for the maximum $R_{ba}(\hat{g})$,

$$\text{CC}_{cikj}(\hat{g}) = \frac{R_{ba}(\hat{g})}{\sqrt{R_{aa}(0)R_{bb}(0)}} \quad (2.22)$$

The autocorrelations at zero lag, $R_{aa}(0)$ and $R_{bb}(0)$, provide normalisation. If $a_{cik}(m) = b_{cij}(m)$, then $CC_{cikj}(\hat{g}) = 1$ (see Section 2.2).

6. Output

- (a) Output the maximum cross-correlation coefficient $CC_{cikj}(\hat{g})$ and its corresponding timing (in samples, relative to the start of the data $v = 1$) of this maximum $\varphi_{cikj} = \hat{g}_{cikj} + j$.

These steps are calculated for each component, station, event and for each window j .

Note that the output in each window j is just the maximum CC and its location: not the full CC function. This places a limitation on the algorithm's ability to detect extremely closely spaced events, in which case the largest event will dominate the window. Shift steps, Δ , must be chosen so as to avert this possibility as much as possible. We believe our shift step of $\Delta = 0.5$ s is sufficient to minimise the effects of this limitation. The method has been executed in this manner as it was the method performed by the original codes written by Charlotte Rowe [see Fig. B.1]. This method reduces the amount of data produced, and should be more efficient than using the full CC in situations where earthquakes are unlikely to be extremely close in time. In injection data we have this possibility, and therefore the need to impose small shift steps decreases efficiency. It would be interesting to compare efficiency with a programme using the full CC, which would allow larger shift steps.

2.3.3 Network correlation coefficients

The use of Network Correlation Coefficients (NCCs), produced by stacking Cross-correlation Coefficients (CCs) over all the stations and components in the network, provides a means of increasing the gain of small events due to constructive interference of CCs, and decreasing the gain of spuriously high CCs via destructive interference [Brown *et al.* 2008; Gibbons and Ringdal 2006; Maurer and Deichmann 1995]. Events belonging to the same cluster should exhibit waveform similarities across the whole network, and the corresponding NCC should attain high values only when individual stations interfere constructively.

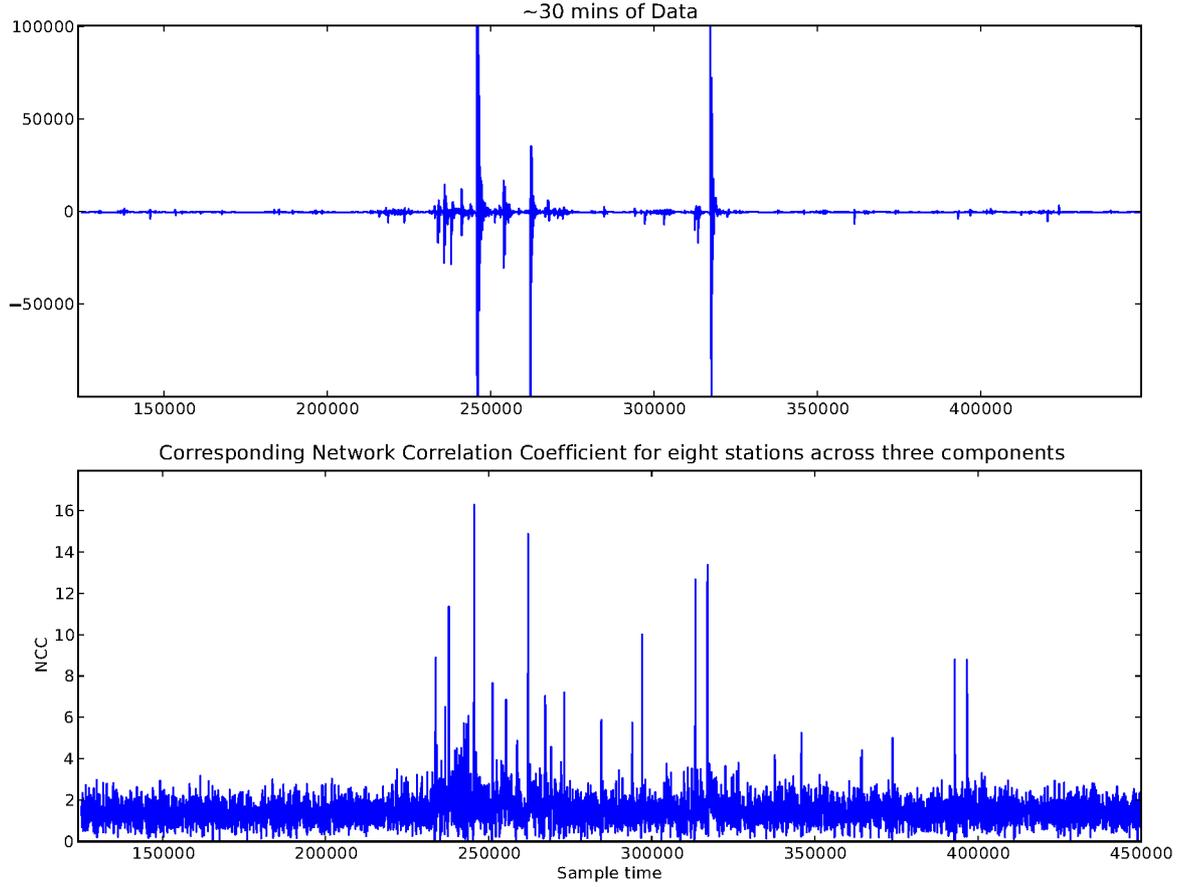


Figure 2.8: (top) ~ 30 minutes of data from one station/component. (bottom) example of the Network Correlation Coefficients produced by cross-correlating the above data with an $M = 2.5$ master template of the event at sample time 245600.

In order to calculate a NCC at each epoch, we must sum the individual channels' CCs after first adjusting for different arrival times caused by different event–station distances. During the master template creation process, the reference time η_k is set for each event k as the earliest P-arrival within the set of master templates associated with that event: that is, $\eta_k = \min \{t_{ik}^p | i = 1, 2, \dots, I\}$ [Fig. 2.5]. For each station i , the station shift ϕ_{ik} is then calculated as the difference between the reference time η_k and that station's P-arrival, $\phi_{ik} = t_{ik}^p - \eta_k$. These ϕ_{ik} values are now used to shift all stations' results to align with the station that has $\phi_{ik} = 0$, and the $CC_{cikj}(\hat{g})$ values are stacked accordingly: For an event k , correlation results across all stations $i = (1, 2, \dots, I)$ and all components c are read in and the station shift, ϕ_{ik} , added to the timing, $\psi_{cikj} = \phi_{ik} + \varphi_{cikj}$, for each correlation maximum calculated in each data window $b_{cij}(n)$. If the difference between the different ψ_{cikj} are less than 50 samples

Table 2.3: Network Correlation Coefficient nomenclature

Symbol	Explanation
$CC_{cikj}(\hat{g})$	maximum fully normalised cross-correlation coefficient of $a(m)$, for event k , and $b(m)$ in data window beginning at j
$CC_{cikj}(\psi)$	maximum $CC_{cikj}(\hat{g})$ within a set of equivalent (within 0.25 seconds) ψ_{cikj} 's
$NCC_{k\Psi}$	Network Correlation Coefficient for event k in bin Ψ
η_k	reference time for event k
ϕ_{ik}	shift time for station i from event time η_k
φ_{cikj}	timing (in samples) of $CC_{cikj}(\hat{g})$
ψ_{cikj}	Station-shift adjusted $CC_{cikj}(\hat{g})$ timing
Ψ	50-sample wide bins spanning the data-length

(0.25 s) then these are treated as equivalent times. As data windows overlap, the maximum $CC_{cikj}(\hat{g})$ within a set of equivalent ψ_{cikj} 's, denoted $CC_{cikj}(\psi)$, is calculated and output at the ψ_{cikj} of this maximum $CC_{cikj}(\hat{g})$.

Next, 50-sample (0.25 s) wide bins spanning the entire data-length are created and labelled with their mid-point, $\Psi = (25, 75, \dots, L_{ci})$. The Network Correlation Coefficient $NCC_{k\Psi}$ is now calculated for event k for each of these bins so that,

$$NCC_{k\Psi} = \sum_{i=0}^I \sum_{c=1}^3 CC_{cikj}(\psi) \mathbf{I}[(\psi_{cikj} - \Psi) < 25] \quad (2.23)$$

where $\mathbf{I}(\cdot)$ is an indicator function, which is 1 if its condition is true, and 0 otherwise. In other words, the maximum cross-correlation coefficient in each 0.25 s is determined for each station and component individually. These are then summed over all $3I$ stations and components in 0.25 s bins. These $NCC_{k\Psi}$ are output at timing $\Psi - \phi_{ik}$ for each station (accommodating each station's shift) [Fig. 2.8 & Table 2.3].

2.3.4 Detection thresholds

Having now computed network-wide cross-correlations, we must identify signals corresponding to real detections. We do this using a detection threshold, Ω , such that if $NCC_{k\Psi} > \Omega$ then this point is interpreted to be the start of an earthquake [Fig. 2.9]. A new trace of this event is created for all components c and stations $i = (1, 2, \dots, I)$ with a start time

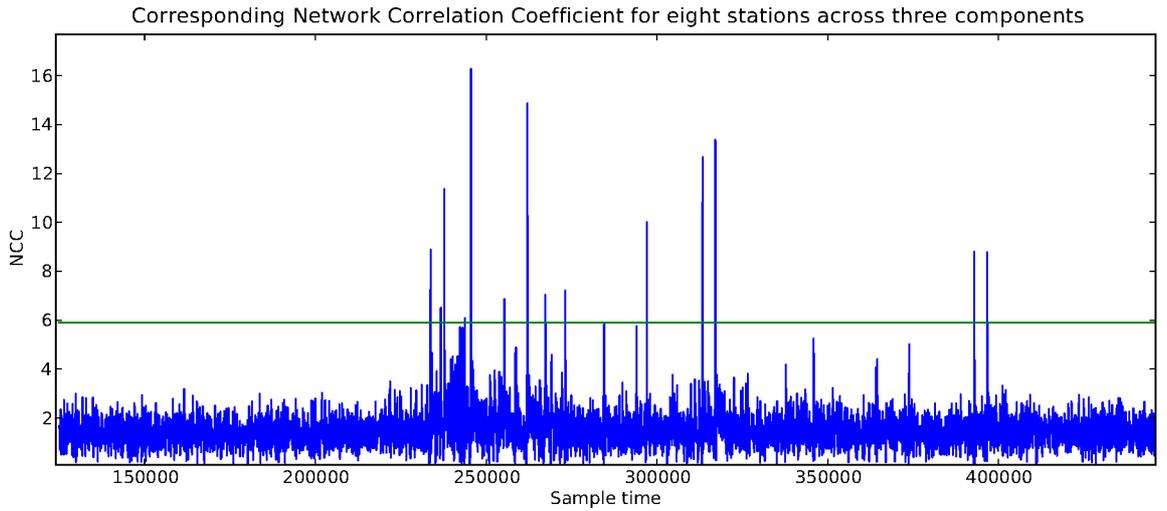


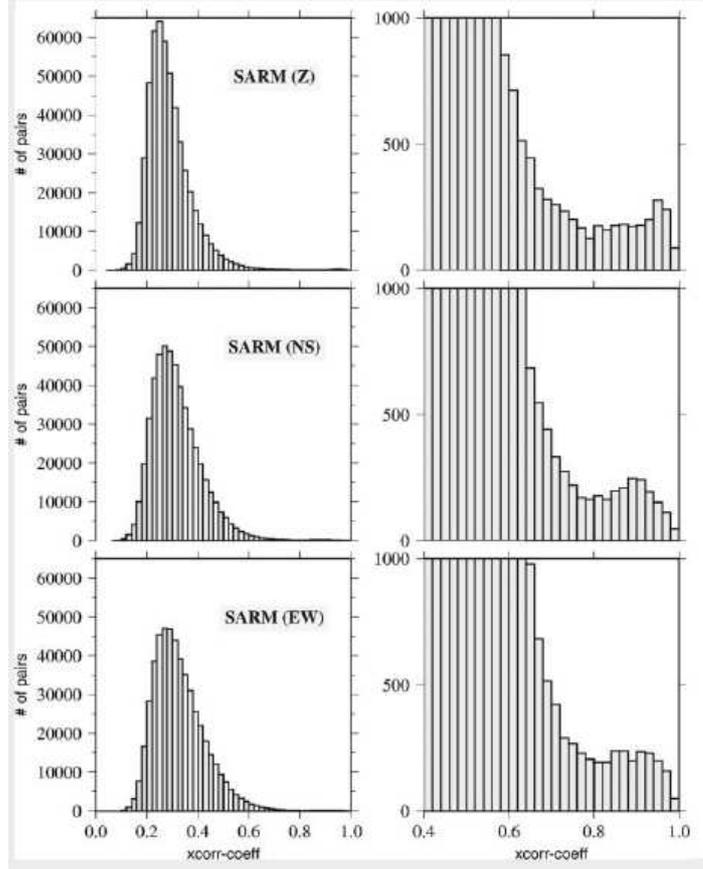
Figure 2.9: The same Network Correlation Coefficient from Fig. 2.8 with a line at $\Omega = 11.76\text{MAD} + \delta$. All NCCs above this line are identified as events. Note that there are some locations with NCCs just below this threshold, this suggests a slightly lower threshold may be more favourable. The determination of the best threshold to use is the goal of the synthetic testing carried out in section 2.4.

$\Psi - \phi_{ik} - 200$ samples and an end time of $\Psi - \phi_{ik} + N_{ik} + 200$ samples (200 samples = 1 second). This creates an event with the length of the master template plus an additional second before and after, to ensure a sufficient length of pre-event noise is available for noise analysis by phase pickers.

The value of Ω is estimated as follows [Table 2.4]. Assuming that the NCC distribution for a day with no detections (pure noise) follows a Gaussian distribution [Ferretti *et al.* 2005; Gibbons *et al.* 2007; Maurer and Deichmann 1995], the events of interest will lie where the NCC distribution deviates from being purely Gaussian [Fig. 2.10; Ferretti *et al.* 2005; Gibbons *et al.* 2007; Maurer and Deichmann 1995]. Maurer and Deichmann [1995] and Ferretti *et al.* [2005] found this assumption held for CC distributions, whilst Gibbons *et al.* [2007] found that NCCs, similarly defined to ours, were almost perfectly normally distributed for data segments without detections, and that earthquake signals resulted in a deviation from this normal distribution.

Our task is to determine a value of Ω above which there is a negligible probability, $P(\epsilon)$, of an NCC existing in this Gaussian NCC distribution, so that any existence will correspond

Figure 2.10: Example (station SARM) showing the number of pairs versus cross-correlation values for the vertical and horizontal components (top to bottom). On the right-hand side, a zoom shows the area of the deviation from a pure normal distribution; the correlation values for this area are targeted as potential thresholds in the following steps of the process. Note that in our NCC calculation these distributions are added together. Figure from *Ferretti et al. [2005]*



to a heavy distribution tail induced by detections. The Median Absolute Deviation (MAD),

$$\text{MAD} = \text{median}|\text{NCC}_k - \text{median}(\text{NCC}_k)| \quad (2.24)$$

is a measure of variation that is (unlike the standard deviation) particularly robust to long distribution tails [*Brown et al. 2008; Shelly et al. 2007*]. We combine the use of MAD, with that of the standard statistical z-score, which indicates how many standard deviations a value is from the mean. A z-score for the value Ω is defined by the mean μ and standard deviation σ ,

$$z = \frac{\Omega - \mu}{\sigma} \quad (2.25)$$

The detection threshold value Ω we are searching for is therefore defined by,

$$\Omega = \mu + z\sigma \quad (2.26)$$

Table 2.4: Detection threshold nomenclature

Symbol	Explanation
MAD	median absolute deviation of NCC_k
$P(\varepsilon)$	negligible probability we wish to calculate the location of
β	detection factor
δ	median of NCC_k
Ω	threshold value for a detection

For a Gaussian distribution, the mean and median are equivalent, and $\sigma = \frac{MAD}{0.6745}$, we therefore have $\mu = \delta$, where $\delta = \text{median}(NCC_k)$, and,

$$\Omega = \delta + \frac{z}{0.6745}MAD \quad (2.27)$$

$$= \delta + \beta MAD \quad (2.28)$$

where $\beta = \frac{z}{0.6745}$. We therefore need to search for the appropriate β in order to define the threshold Ω , and we hence name β the ‘detection factor’. We do this by choosing a desired probability of a false detection $P(\varepsilon)$, and finding the corresponding z-score. Note that $P(\varepsilon)$ is the (number of false detections)/(number of calculations that correspond to non-events). We choose to start with a very strict value of $P(\varepsilon) \simeq 0$, with $P(\varepsilon) = 1 \times 10^{-15}$. This corresponds to a z-score of 7.93, giving $\beta = 11.76$, and hence a detection threshold of $\Omega = 11.76MAD + \delta$, above which very few spurious detections should occur.

We are assuming that the NCC distribution constitutes a Gaussian noise distribution plus earthquake signals. If the signals are weak then we would have to lower the threshold Ω to detect them, but this brings the possibility that the tail of the Gaussian noise distribution now extends above Ω , leading to false detections. We choose a 5% false detection rate to be tolerable in this situation, and perform synthetic testing in the following section [Section 2.4] to refine the preliminary ‘detection factor’ estimate and the corresponding detection threshold. Note that this 5% false detection rate is the (number of false detections)/(number of detections) $\times 100$: the percentage of the detection data which does not correspond to a true event, which is not the same as $P(\varepsilon)$ above.

2.4 Synthetic testing

To determine a suitable detection threshold we must first estimate the detection factor [Eq. 2.28]. We do this by carrying out 2000 synthetic tests: 1000 synthetic tests on a one-station network, and 1000 synthetic tests on a six-station network. The results not only enable us to constrain the detection threshold, but also demonstrate the advantage over individual CCs that NCC calculations provide in distinguishing noise from signal.

Synthetic testing is carried out by embedding a known waveform template in a noise record. To ensure the testing is as relevant as possible to the problem at hand, we require a realistic synthetic noise field that does not contain any instances of the waveform template being targeted. We create this realistic synthetic noise field by maintaining the amplitude of the frequency spectrum of the noise field but randomising the phases, ensuring any short-period signals are lost while retaining long period signals [Figs. 2.11 & 2.12]: randomising the phases results in the coherence of neighbouring frequency components being lost, causing destructive interference, and destroying the signal; long period signals persist because they have so few cycles in the window that destructive interference is not complete. The randomisation ensures that the complex balancing of the frequency component amplitudes and phases that is required to build a complex waveform is disrupted, thus destroying the signal.

We take one third of a day of real data, p_l , $l = 1, 2, \dots, L$ and calculate the frequency spectrum, $f_l = \mathcal{F}(p)$. We then calculate the modulus, $r_l = |f_l|$, and argument, $\phi_l = \tan^{-1} \left(\frac{\text{Im}(f_l)}{\text{Re}(f_l)} \right)$, of the spectrum f_l . The phases are now randomly reordered using a uniform distribution (except for those at zero and the Nyquist frequency $(L/2 + 1)$, to ensure that the mean level of the series is not changed) creating $\tilde{\phi}_l$. The modulus is retained, $\tilde{r}_l = r_l$. The new spectrum is created from the new modulus and argument, $\tilde{f}_l = \tilde{r}_l (\cos \tilde{\phi}_l + i \sin \tilde{\phi}_l)$. The inverse FFT now provides the realistic synthetic noise trace, $\tilde{p}_l = \mathcal{F}^{-1}(\tilde{f}_l)$, in which the frequency spectrum has been kept the same, but any short-period signals have been lost. This will ensure that no earthquake waveforms survive in the data.

For both the one-station and six-station scenarios, 1000 tests have been performed with

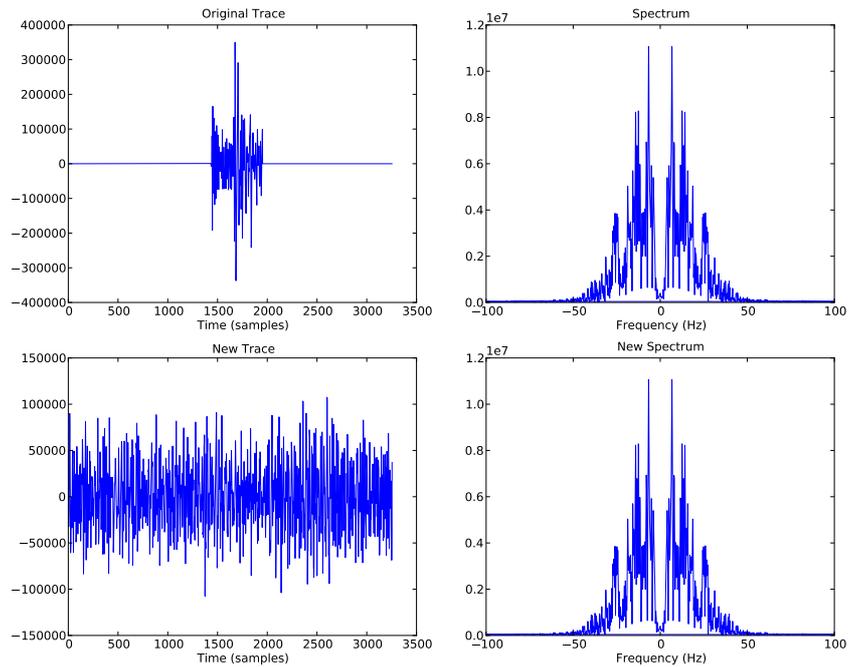


Figure 2.11: A simple example of taking a padded waveform template and randomising the phases. Note the amplitude of the frequency spectrum is retained whilst the template signal is lost in the time domain of the new trace

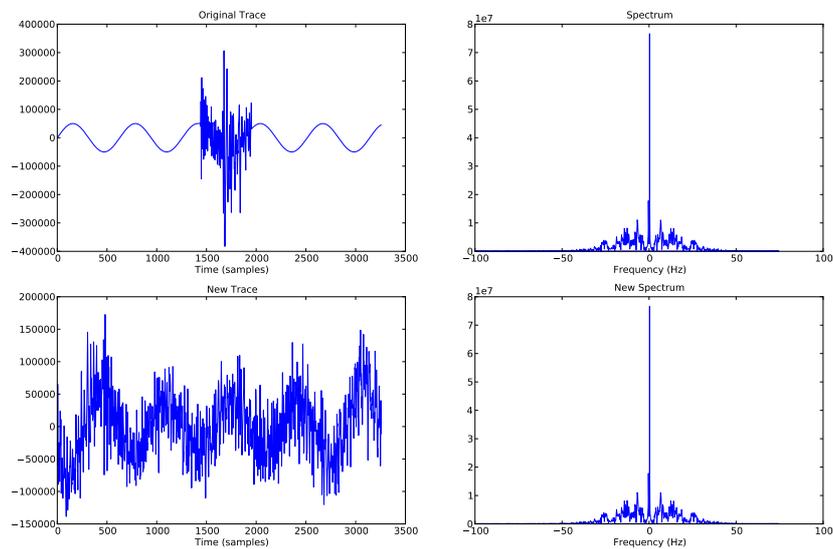


Figure 2.12: A simple example of taking a padded waveform template with a sine wave imposed. Note the amplitude of the frequency spectrum and long period sine wave signal are retained whilst the template signal is lost in the time domain of the new trace. Note that though the phase of the sine wave has shifted, we are randomly embedding waveforms so this is of no consequence.

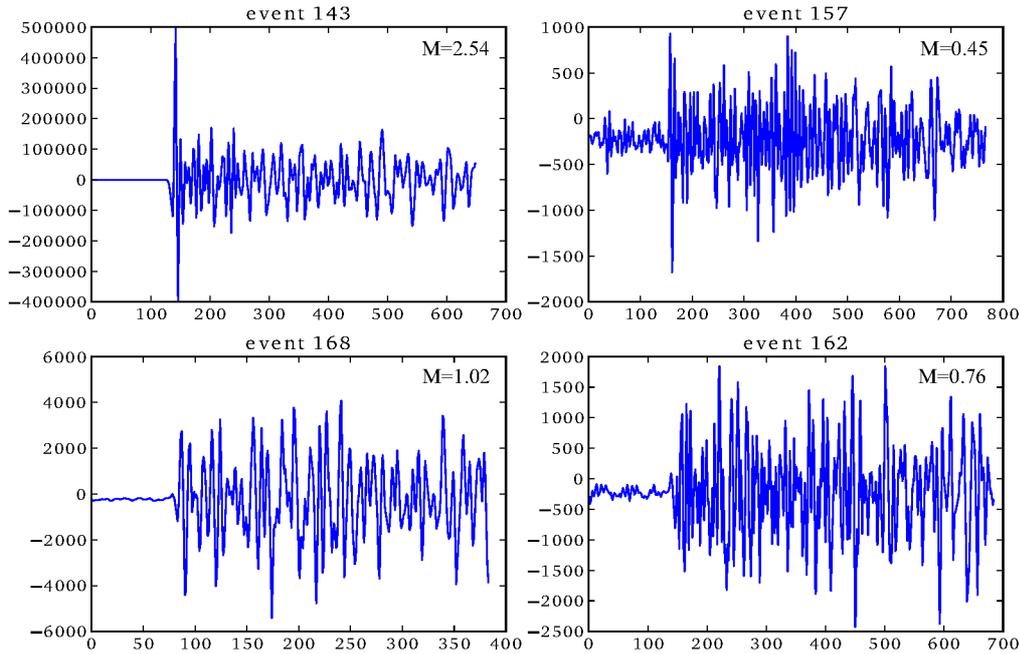


Figure 2.13: Master events used in synthetic testing, amplitude vs. sample time for station RT01 on the Z component.

50 embedded waveforms per test. To minimise any detection bias introduced by the waveform characteristics of a particular event waveform, we have used four different master events [Fig. 2.13] and carried out 250 tests for each event per scenario. The master events used are real events from Rotokawa and were chosen because they were detected at most stations and span a range of magnitudes. To reduce the required computational time, testing has been performed on a third of a day of data, or 5,760,000 samples. Ideally, testing would be performed on full days to include any diurnal variations, and using all days within the data set to include any larger temporal variations, but the time required for such extensive testing was not available. The locations of the embedded waveforms in each test are chosen by a random number generator. The SNR of each embedded waveform is also generated randomly (calculated as the maximum amplitude of the event divided by the root-mean-square amplitude of the data set) and the master template scaled accordingly. Preliminary testing indicated that only earthquakes of $\text{SNR} > 3$ can be consistently detected. SNRs are therefore restricted to be > 3 (any SNR drawn < 3 is increased by a number between 3–10

using a uniform random number generator) and are drawn from a log-normal distribution $Log - N(0, 3)$, so that there are a larger number of low amplitude events compared with high amplitude events. For typical noise levels in this case, an SNR value of 3 equates to a $M \sim 0$ earthquake. This implies that lower magnitude earthquakes may be detected, but that the catalogue of such events may not be complete.

We use the detection factor estimate of 11.76 suggested in Section 2.3.4 as a starting point for testing and examine the effect of varying this number. Due to processing time, we were limited in the number of different detection factors that could be tested. The results of testing a number of different detection factors on a one-station network are shown in Fig. 2.14, and on a six-station network in Fig. 2.15.

We adopt an acceptable false detection rate of $\sim 5\%$ (based on the typical scientifically acceptable value of 95% confidence) and on the basis of the data shown in Figs. 2.14 and 2.15, detection factors of 12.75 and 8.72 were chosen as being optimal for the one-station and six-station networks, respectively. These correspond to $P(\epsilon) = 7.97 \times 10^{-18}$ and $P(\epsilon) = 4.1 \times 10^{-9}$, respectively [see Section 2.3.4]. The difference in these numbers indicates the positive impact an NCC has on distinguishing earthquakes from noise: the lower detection factor provides a lower detection threshold and indicates that the difference between noise and signal is greater [Fig. 2.16]. The details of the varying results using the detection factors are also shown in Table 2.5. Here, a false percentage corresponds to the (false number of detections)/(number of detected earthquakes) and a missed percentage corresponds to (missed number of earthquakes)/(number of earthquakes). This also demonstrates that although there is a great variation between the percentage of false detections for each master event, no master event produces false detections at a rate exceeding 10%, which is a positive outcome.

We have also performed some preliminary tests on the impact of reversed waveform polarisations and using a master event with a decreased SNR, but rigorous examination of these factors is beyond the scope of this study. Waveforms of a fully reversed polarity template will still be detected, but the correlations are 50–85% lower. These correlations still occur due

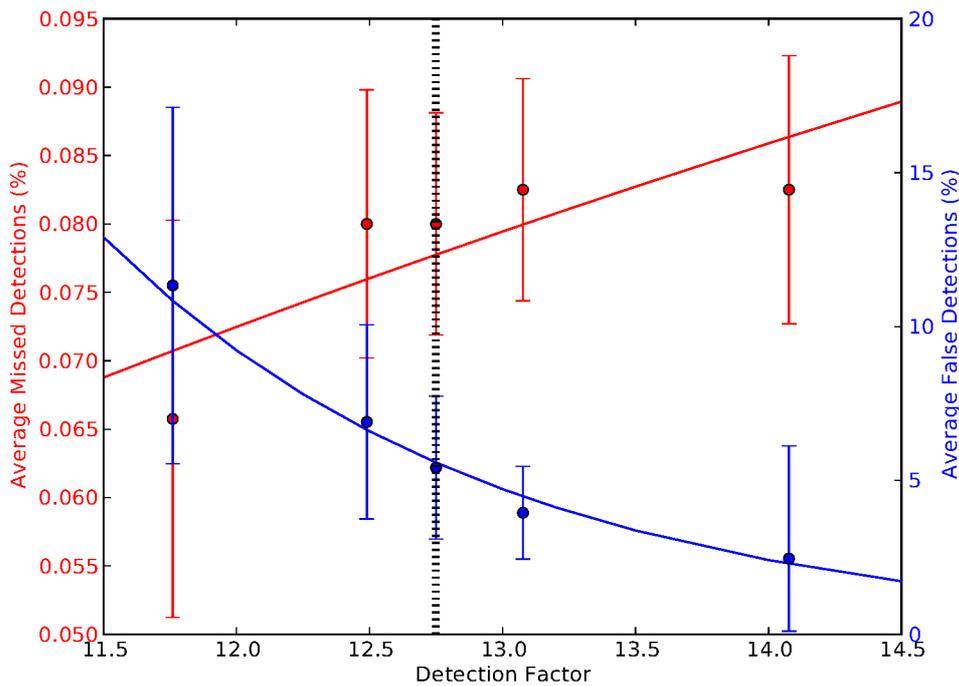


Figure 2.14: Average percentage of false detections and missed earthquakes vs. detection factor. Averaging is performed over 1000 synthetic tests on a one-station network. The black dashed line is the 12.75 detection factor chosen as optimal for this network configuration. Error bars denote the 95% confidence interval.

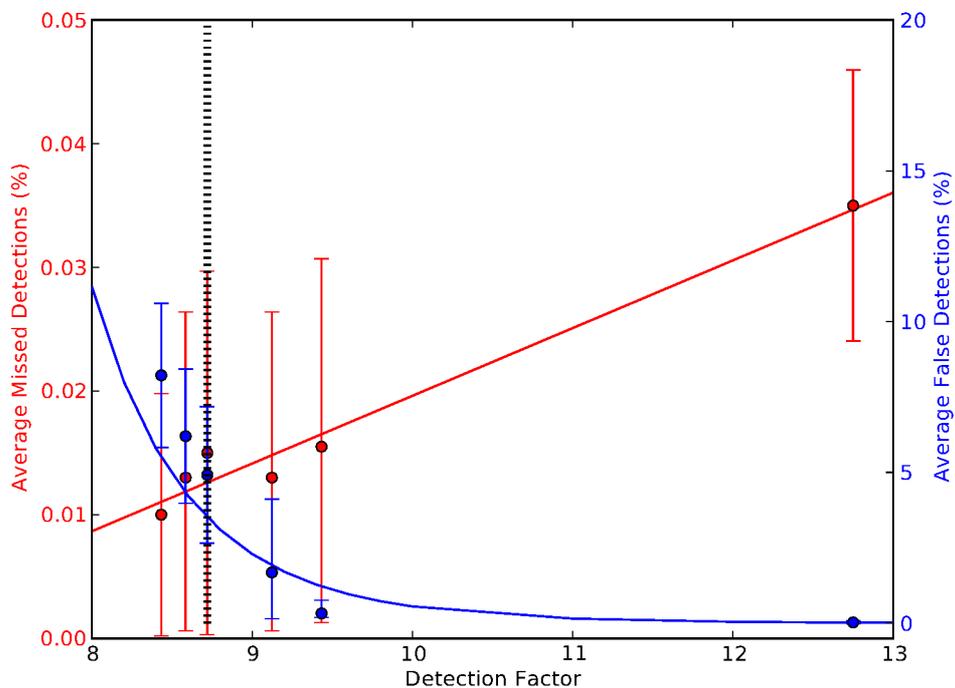


Figure 2.15: Average percentage of false detections and missed earthquakes vs. detection factor. Averaging is performed over 1000 synthetic tests on a six-station network. The black dashed line is the 8.72 detection factor chosen as optimal for this network configuration. Error bars denote the 95% confidence interval.

Table 2.5: Results from synthetic testing for one and six station networks for detection factors 12.75 and 8.72 respectively. These detection factors were chosen as optimal for the two scenarios. Results are from 250 tests per master event, with 50 waveforms embedded in each test. The detection factors satisfy our requirement to average 5% false detections. Note that for every master event false detections are $< 10\%$.

		1 station, $\beta = 12.75$			
Event	Mag	False (%)	Average false detections per test	Missed (%)	Average missed detections per test
143	2.54	0.15	0	0.07	0
168	1.02	9.79	5	0.09	0
162	0.76	3.80	2	0.09	0
157	0.45	7.93	4	0.07	0
mean		5.42 ± 3.7	3 ± 3	0.08 ± 0.01	0 ± 0

		6 stations, $\beta = 8.72$			
Event	Mag	False (%)	Average false detections per test	Missed (%)	Average missed detections per test
143	2.54	4.95	3	0.0	0
168	1.02	7.96	4	0.03	0
162	0.76	1.47	1	0.0	0
157	0.45	5.30	3	0.03	0
mean		4.92 ± 2.3	3 ± 2	0.015 ± 0.015	0 ± 0

to correlation being carried out in the frequency domain, not the time domain. Due to these lower correlations, 22% of earthquakes were missed and there were 6% false detections. A smaller earthquake, SNR=5, was used as a master event and it was found to generate a lot (400%) of false detections due to correlating with noise. Within a network we would expect lower SNRs to correspond to lower-magnitude events. Table 2.5 shows that an increase in false detections is not solely a function of magnitude (and the SNR implied by this magnitude), but seems to be affected more by the characteristics of the waveform templates (e.g. frequency, directivity, impulsivity). This is most likely to be influenced by the similarity between dominant noise-frequencies and template frequencies.

In light of the testing described above, we have chosen to use the detection factor of 8.72 determined for the six-station network to process the complete data set. This corresponds to a detection threshold of $\Omega = 8.72\text{MAD} + \delta$. To ensure validity of this determination, we only use master events that have been detected at ≥ 6 stations. Using this value, the synthetic

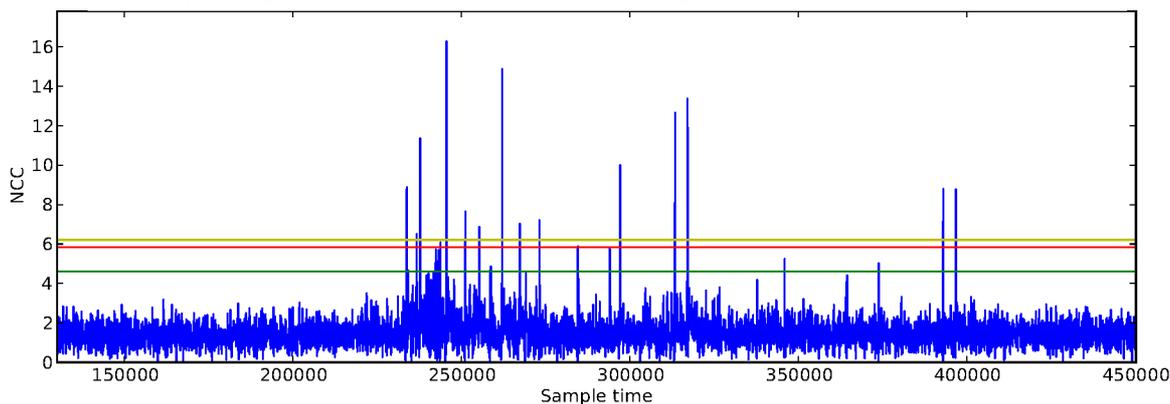


Figure 2.16: Network Correlation Coefficient with a green line at the six-station synthetically determined threshold of $\Omega = 8.72\text{MAD} + \delta$, a yellow line at the one-station synthetically determined threshold of $\Omega = 12.75\text{MAD} + \delta$ and a red line at the Gaussian assumption estimated detection threshold $\Omega = 11.76\text{MAD} + \delta$ (Fig. 2.9).

testing suggests that we expect to detect $99.985 \pm 0.015\%$ of earthquakes with matching waveforms and SNRs larger than 3. We also expect $4.92 \pm 2.3\%$ of the earthquakes to be false detections (Table 2.5). It is important to note that these figures come from synthetic testing of *exactly* the same waveform. In real life we will be dealing with slightly different waveforms and so the results will not conform to these optimal figures (performance will be worse).

This detection factor will be a good starting point for any future studies. However, a more rigorous determination will need to be carried out depending on the differences in noise and earthquake frequencies, and the number of stations in networks being analysed.

2.5 Results

In this chapter, we have detailed the basis of our matched filter detection technique and the premises for our parameter choices. The choice of master templates is one of the key factors in this process. We initially created a suite of master templates containing all high-quality events provided by GNS Science. Preliminary testing (on the first six weeks of data) highlighted two issues:

1. Different master events were detecting the same slave events

2. Significantly more slave events were detected on the same day that their master event occurred than on other days. This could be a result of either different days having significantly different noise spectra, or a large temporal variation in event occurrence.

To address the first of these issues, we cross-correlated all master events with each other and stacked together those with $CC > 0.6$ [Shelly and Hardebeck 2010]. Combined with our restriction that a master event be detected by at least six stations [Section 2.4], this provided us with a refined set of 14 master events [Table 2.6]. This also partially addresses the second issue above, if the cause is a significantly varying noise spectrum, and if stacked master events are from different days.

This suite of master events was then cross-correlated with the five months of data following initial injection, from 29 September 2008 to 28 February 2009. We processed these cross-correlation results using the detection threshold $\Omega = 8.72\text{MAD} + \delta$ determined in Section 2.4. One day of detections resulting from this is shown in Fig. 2.17. This shows 52 slave events that have been detected. We zoom into the densest region of detections in Fig. 2.18, which shows one of the high-SNR events that was stacked to form the master event this day was correlated with, and slave events showing similar waveforms. We zoom in again on six of these events in Figs. 2.19 and 2.20, showing how similar the vertical and horizontal waveforms are over a range of magnitudes. Because cross-correlation is carried out in the frequency domain, we also display the normalised frequency spectrum in Fig. 2.21, for 30 events ranging in magnitude from the 2.5 master to a -0.2 slave event. This shows that throughout this range of magnitudes the frequency spectra are very similar.

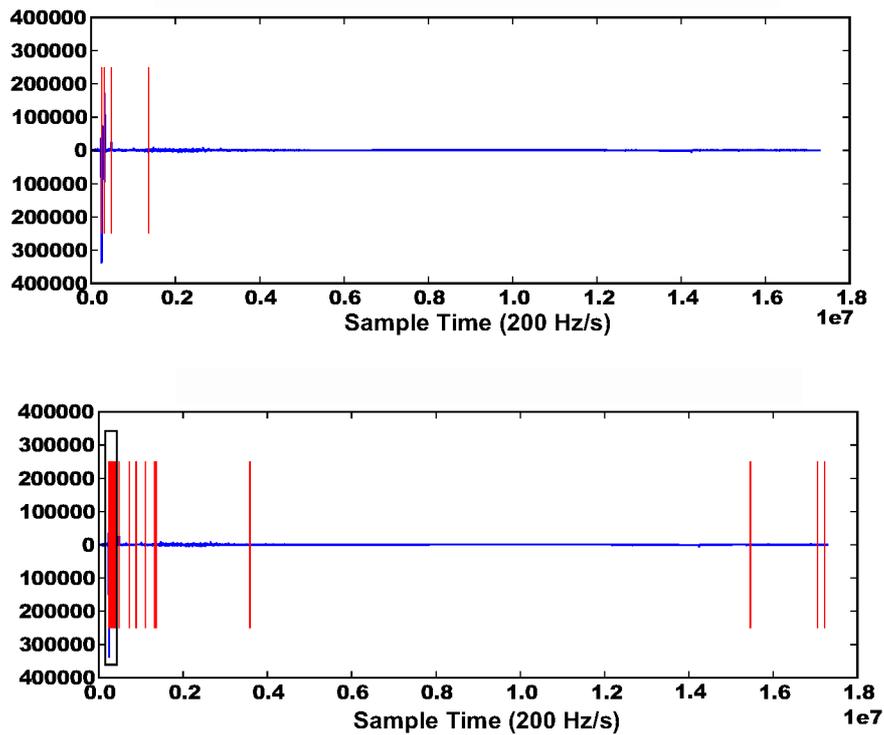


Figure 2.17: (Top) four high-SNR events detected by GNS Science that were stacked to create the master event, the largest of these is the magnitude 2.5 event that is visible. (Bottom) the detection results from cross-correlation for this one day and master event. Fig. 2.18 zooms into the dense region of detections inside the black rectangle.

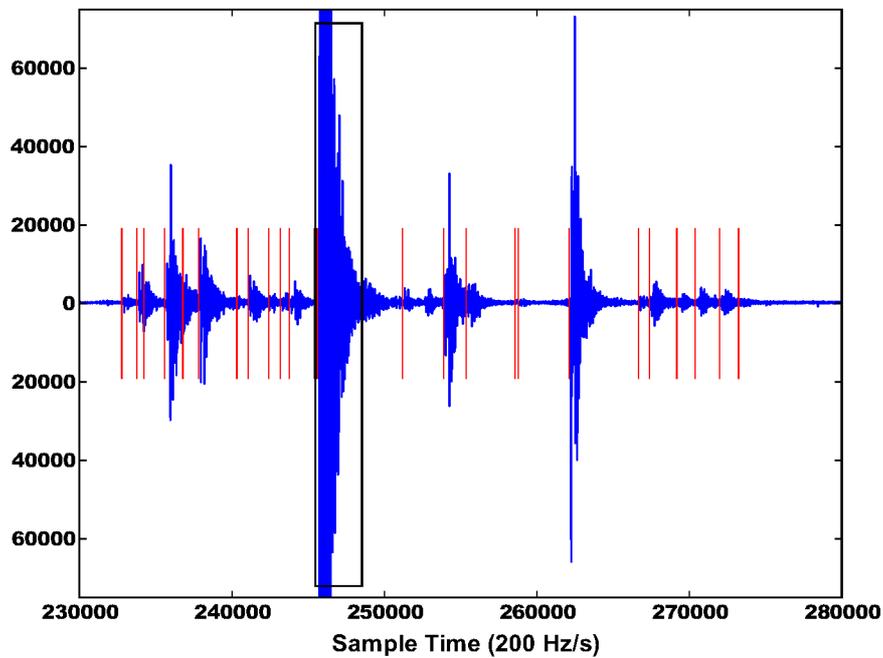


Figure 2.18: Four-minute-long record that contains the magnitude 2.5 high-SNR event (black rectangle), and 23 other events which have been detected. Note that waveforms of detected events that are visible match nicely with the master event.

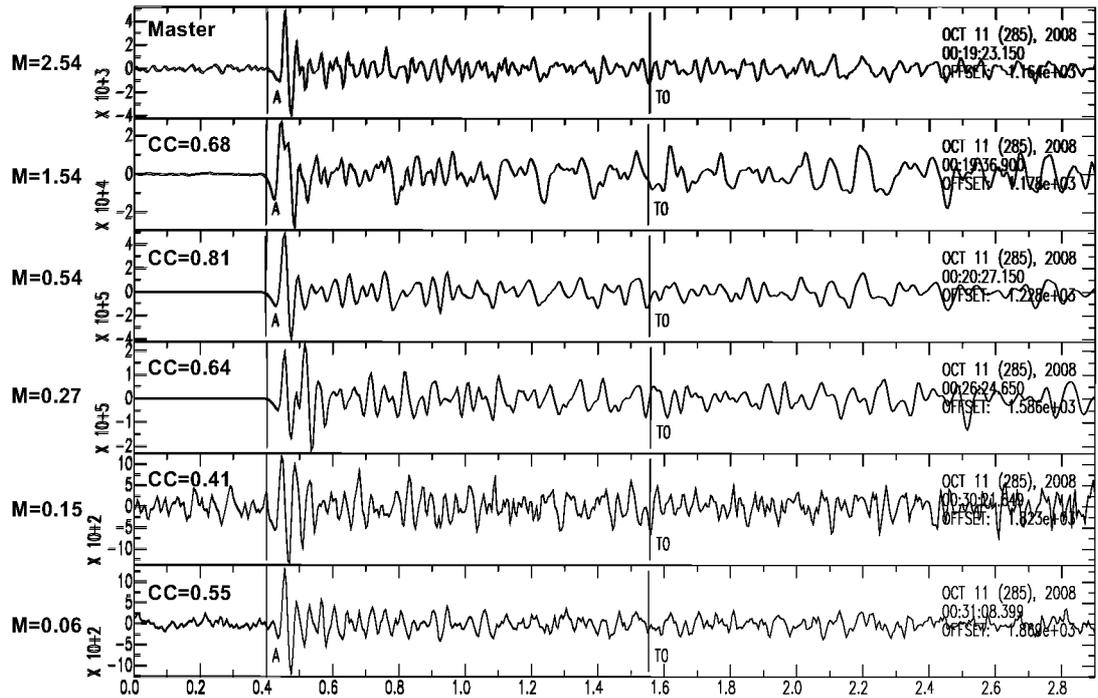


Figure 2.19: Vertical component for six of the events in Fig. 2.18 including the master event. Note that waveforms are highly similar over ~ 2.5 orders of magnitude: notice the shape of the initial P-pulse. CC is just the correlation coefficient for this station/component (RT01/vertical).

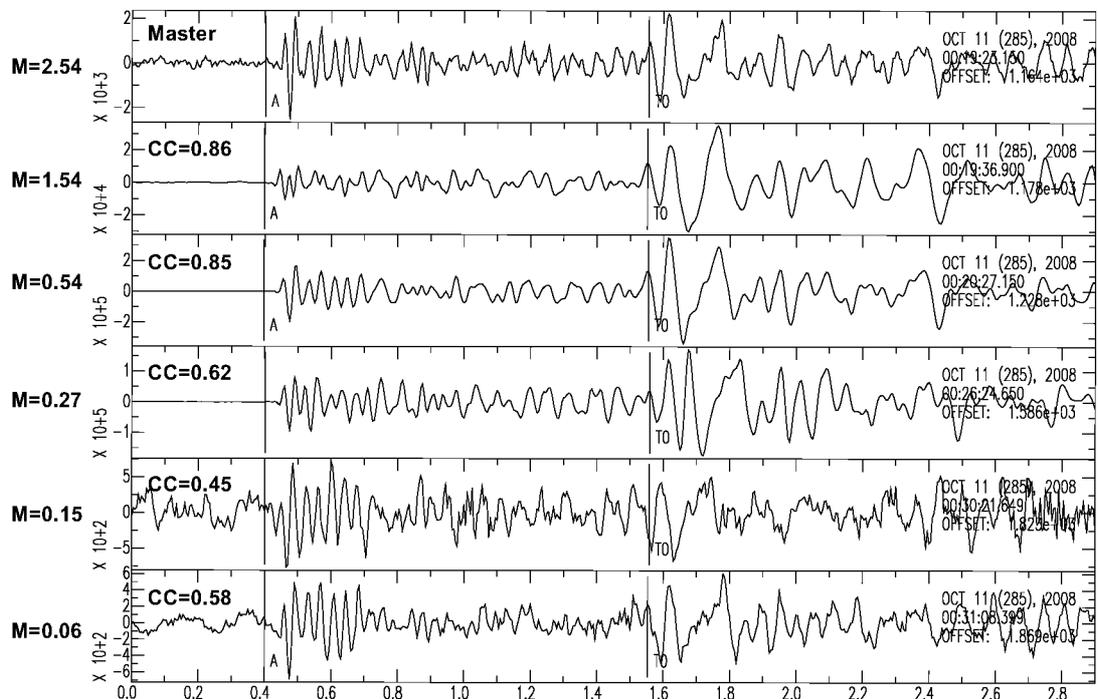


Figure 2.20: East component for the six events in Fig. 2.19 including the master event. Note that waveforms are highly similar over ~ 2.5 orders of magnitude: notice the number of peaks and their shape and frequency component from the P-picks to ~ 0.6 s, and from the S-picks to ~ 2.0 s. CC is just the correlation coefficient for this station/component (RT01/east).

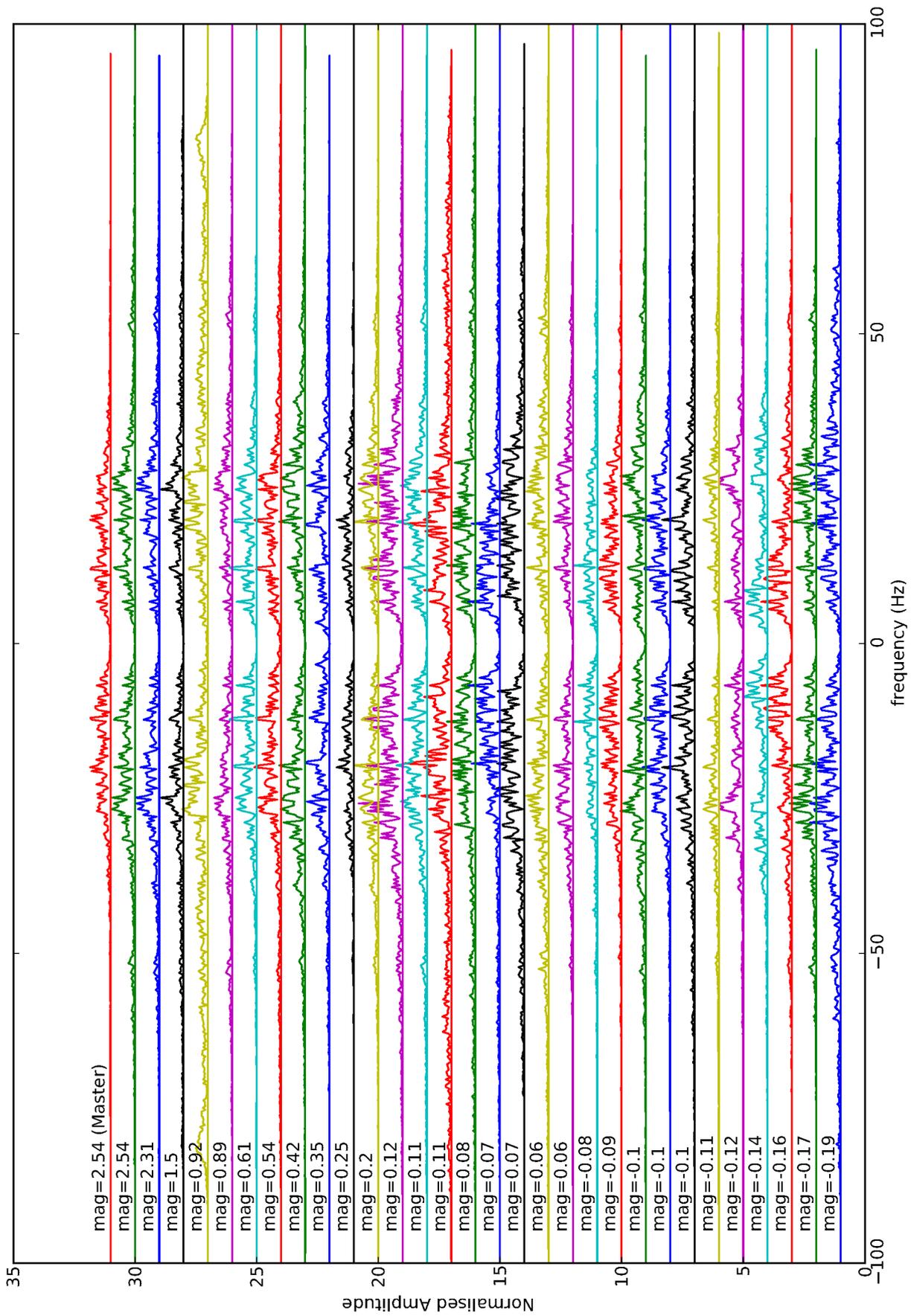


Figure 2.21: Normalised frequency spectrum for 30 events covering three orders of magnitude, from the 2.54 master at the top, to the magnitude -0.19 event at the bottom. Note the similarity of the spectra over this range of magnitudes.

We have also calculated relative magnitudes for the slave events. Due to the calculation of magnitudes being complex for microearthquakes [*Rubinstein and Ellsworth 2010; Shemeta and Anderson 2010*], magnitudes have been estimated using an adaptation of the formula described by *Schaff* [2008],

$$M = \log_{10} \left(\frac{\Lambda}{\Lambda_{\text{master}}} \right) + M_{\text{master}} \quad (2.29)$$

which provides a magnitude estimate based on an earthquake’s maximum amplitude and a reference master event’s maximum amplitude and magnitude. Λ is the maximum amplitude of the earthquake, and the reference master event used is 901. Magnitude estimates are calculated for every component and station and then averaged over all components and stations to provide the final magnitude. The master event 901 was chosen as a reference as it is the third largest event ($M=2.45$) and due to its location it provided the most consistent results across the network (compared with the two larger events that were tested). A single master event is used as a reference so that estimates are consistent over all slave event sets. Master event magnitudes have been provided by GNS Science. The magnitudes of all slave events are shown in Figs. 2.22 & 2.23, in which we have a magnitude range $-0.4 \leq M \leq 2.6$ and a mean slave event $M= 0.47$.

Fig. 2.23 displays the cumulative number of slave events for each master event with a Gutenberg-Richter b value of 1 displayed for comparison. A b value of 1 corresponds to the typical frequency-magnitude distribution in seismically active regions, and represents a factor of 10 increase in seismicity with each decreasing magnitude [*Gutenberg and Richter 1944; Wiemer and Wyss 2000*]. Data which perfectly follows the Gutenberg-Richter law plots to a straight line, with low magnitude data tending to flatten, called “roll-off”, which is generally assumed to be due to incompleteness of the reporting of earthquakes [*Wiemer and Wyss 2000*]. The deviation point from the Gutenberg-Richter line, at which this roll-off begins, is deemed the completeness magnitude M_c of a catalogue [*Schorlemmer and Woessner 2008*]. Each cluster of slave events in Fig. 2.23 exhibits a slightly different slope, but are

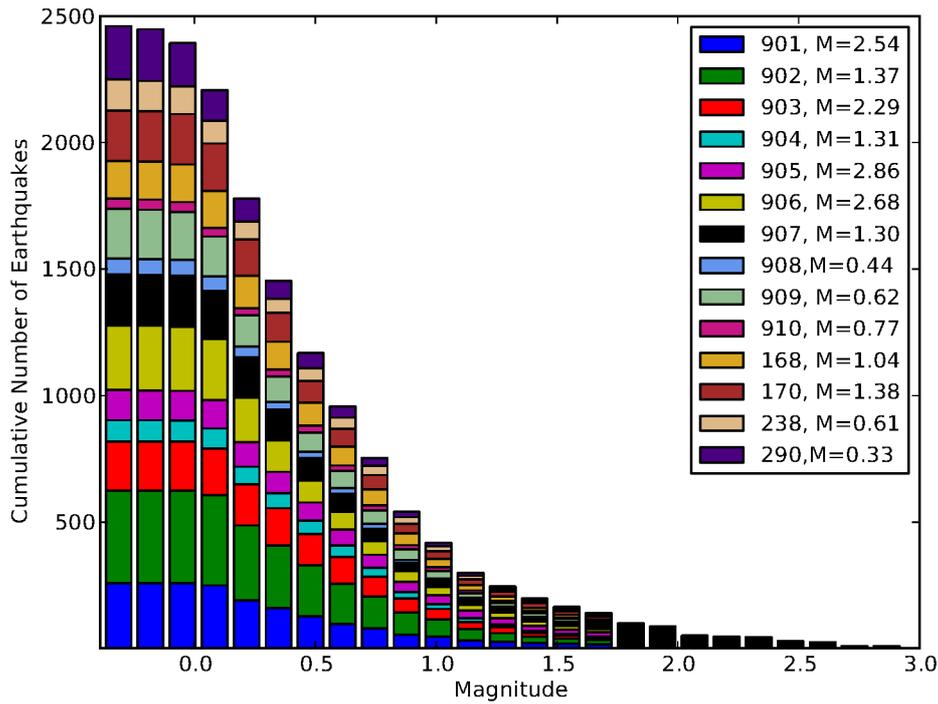


Figure 2.22: Cumulative number of earthquakes vs. magnitude for all detected events (accumulating from high-magnitude to low). Note that magnitudes have been calculated using the amplitude ratio formula described in Eq. 2.29.

overall remarkably similar and are roughly following a Gutenberg-Richter power law. Their slopes are generally gentler than $b = 1$, providing values such as $b \simeq 0.77$ (for master event 901) and $b \simeq 0.95$ (for master event 907). Roll-off begins at $M \sim 0$, suggesting the catalogue has a completeness magnitude $M_c \simeq 0$. We cannot claim this catalogue completeness for the entire Rotakawa region, but only for the 14 earthquake forms used as master events. Table 2.6 provides a summary of the master events used and their detection results.

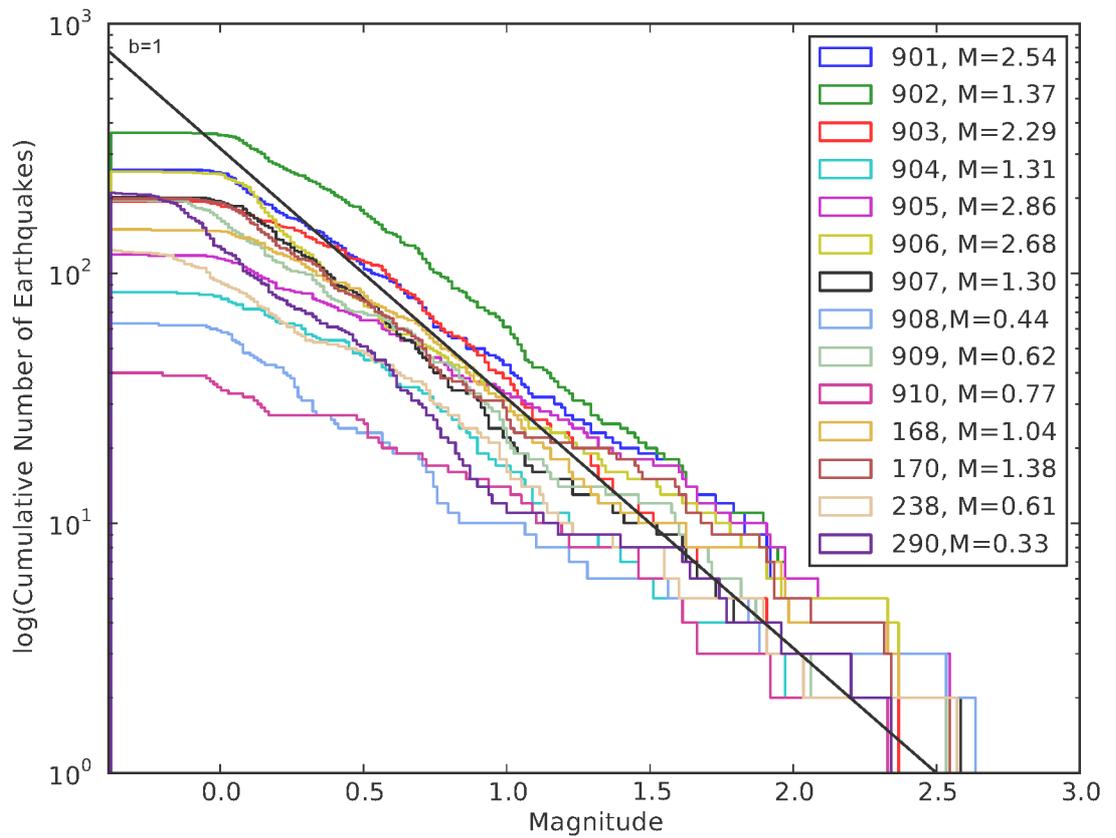


Figure 2.23: Log of the cumulative number of earthquakes vs. magnitude. Note that events appear to be roughly following a power law, and display a completeness to $M \sim 0$. For comparison a line is shown with a Gutenberg-Richter b value of 1. Note that magnitudes have been calculated using the amplitude ratio formula described in Eq. 2.29.

Table 2.6: Summary of master events and their detections.

master event	no. events in stack	max. mag.	no. of slave events
901	3	2.54	259
902	7	1.37	366
903	4	2.29	194
904	2	1.31	84
905	4	2.86	119
906	16	2.68	255
907	23	1.30	202
908	2	0.44	63
909	2	0.62	196
910	3	0.77	40
168	1	1.04	150
170	1	1.38	199
238	1	0.61	124
290	1	0.33	210
total			2461

Chapter 3

Earthquake Location and Spatiotemporal Analysis

This chapter details the steps taken to obtain accurate hypocentre locations [Fig. 3.1] for the Rotokawa events described in the previous chapter, and an analysis of these results. After detection, the detected events have their phase picks refined and initial locations computed [Section 3.1]. Phase pick refinement is carried out using the cross-correlation software *BC-SEIS* [Du *et al.* 2004] and phase pick weights are assigned using Jeffreys' weighting method [Jeffreys 1973]. Initial location is carried out using *Velost* [Kissling *et al.* 1994] and a velocity model supplied by GNS Science. We have experimented with relocating the earthquakes using the *hypoDD* software that implements the double-difference earthquake location algorithm of Waldhauser and Ellsworth [2000] [Section 3.2]. However, the results obtained with *hypoDD* are unsatisfactory and instead we obtain final location using the probabilistic hypocentre estimation method, *NonLinLoc* [Lomax *et al.* 2000]), which yields a relatively complete picture of the location uncertainties [Section 3.3]. We consider the spatiotemporal patterns of the events and their significance in Section 3.4.

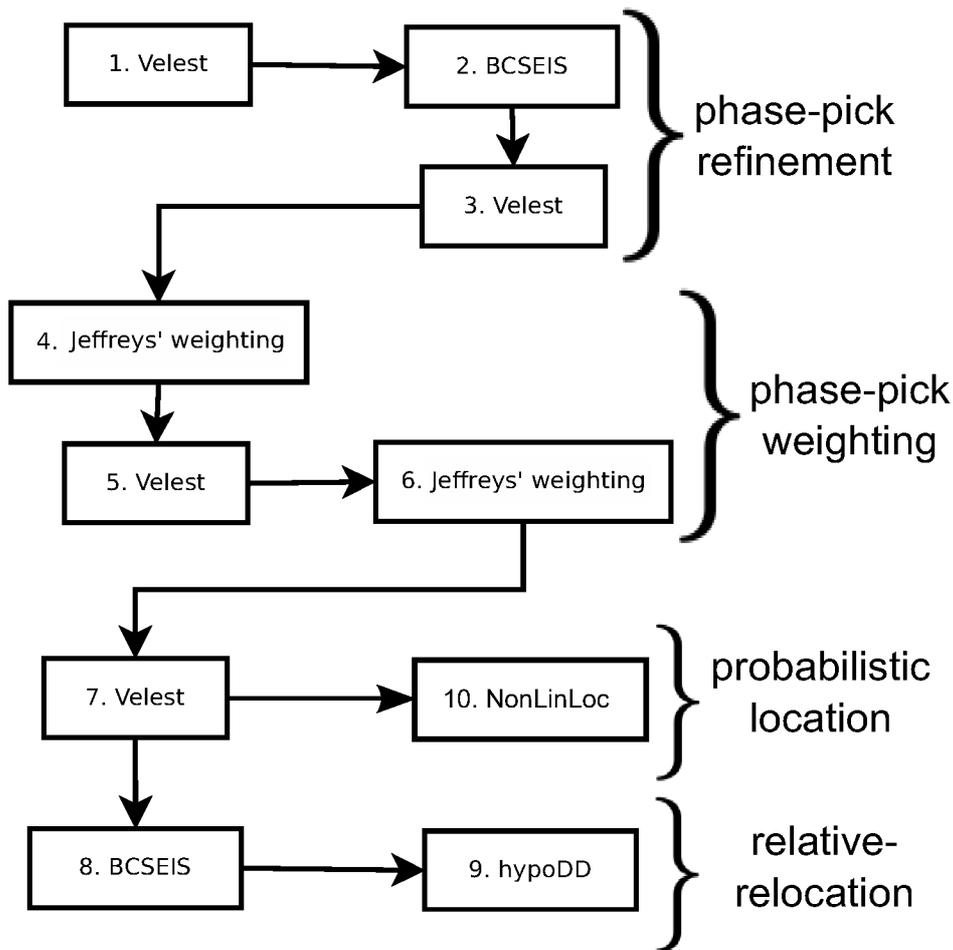


Figure 3.1: Phase pick refinement and location flow described in this chapter.

3.1 Preliminary location

3.1.1 Phase-pick refinement

In the context of this local earthquake study, phase-picks identify the arrival times of the P- and S-phases of each earthquake. Accurate timing of these phase-picks and the estimation of any timing errors are the integral link between earthquake detection and location: no amount of sophisticated location calculations can make up for poor phase-picks, although they are assisted by consistent timing error assessment [Di Stefano *et al.* 2006]. Determination of phase-picks is a fundamental problem in seismology, with manual estimation being time-consuming and human dependent, whilst the reliable automatic detection of phase-picks remains an ongoing research topic [Di Stefano *et al.* 2006; Galiana-Merino *et al.* 2008;

Gentili and Michelini 2006; Leonard 2000; Rowe 2002; Song et al. 2010].

With the Rotokawa data set we are dealing with a large number of low-magnitude, low-SNR earthquakes. This causes manual phase-pick refinement to be both difficult and labour-intensive, and the earthquakes low-SNRs also make automatic phase-picking non-trivial: this is the precise reason accurate automatic detection of phase-picks is an ongoing research topic. Advantageously, however, we know which master event each slave event correlates highly with. We capitalise on this by using the cross-correlation software *BCSEIS* [*Du et al. 2004*] to perform cross-correlation between each slave event and its master event, and then shift the slave event's phase-picks by the determined lag.

During the final step of our detection process, in which the seismic trace of a detected event is created and stored, phase picks are transferred from the master events to the slave events. These transferred phase picks will not be exactly correct due to the 0.25 second bin widths used in the Network Correlation Coefficient (NCC) calculation [Section 2.3.3], and the fact that slave events may not be in exactly the same location as the master events, and hence may have slightly different S–P times. We therefore refine these phase picks using *BCSEIS*. *BCSEIS* [*Du et al. 2004*] is a software package developed to obtain reliable waveform-based differential times for a group of earthquakes. Time delays are calculated using both the cross-correlation technique and the bispectrum method. The bispectrum method works in a similar way to cross-correlation but, as it operates in the third-order spectral domain, is less sensitive to Gaussian noise than cross-correlation [*Du et al. 2004*]. The application of *BCSEIS* requires the specification of a bandpass filter and the sizes of correlation windows spanning the P and S picks. One disadvantage of *BCSEIS* is that the same windows and filtering are applied at all stations. This restricts the size of the windows that we are able to use, as we need to ensure that the P-window does not overlap the S-phase. *Du et al. [2004]* carried out a regional study in New Zealand using a bandpass filter of 1–6 Hz, and a P-window extending from 1 s before to 2.82 s after the P pick. *DeShon et al. [2007]* carried out a local study at Redoubt Volcano, Alaska, using a bandpass filter of 1–15 Hz and a window from 0.3 s before to 1.1 s after the P pick. *Clarke [2008]* carried out a regional study at nearby TVZ

geothermal fields, Kawerau and Rotorua, using a bandpass filter of 1–10 Hz, and a window 0.3 s prior to and 0.96 s following the P pick, and 0.5 s prior to and 1.4 s post the S pick.

We apply a 3-16 Hz bandpass: the lower limit of this bandpass filter has been chosen because there is a lot of background noise at < 3 Hz [Stephen Bannister pers. comm., 20 December 2010]. The upper limit is similar to that of *DeShon et al.* [2007], and we have carried out testing that consistently identifies 16 Hz as a suitable upper limit in our close-proximity reservoir work. We then define time windows around each P and S-wave arrival time pick. The P window begins 0.5 s before the P pick and ends 0.4 s after the P pick. The S window begins 0.8 s before the S pick, and ends 0.8 s after the S pick. The number of samples in a window is required to be 2^N , we set this to a ‘high’ sampling rate of 256 samples per window. These windows were chosen as optimal after testing window lengths within the range 0.25–1 s before and after the P and S picks. Note that our window choice employs more time pre-pick than is standard for use with *BCSEIS*, as we are using *BCSEIS* here for picking, and the ‘true’ pick is equally likely to be pre-current-pick as post-current-pick. As our study is a local one it is not surprising that our parameters significantly differ to those used by *Du et al.* [2004]. Likewise, our bandpass filter and window lengths are similar to those used in the local studies of *DeShon et al.* [2007], and *Clarke* [2008].

We cross-correlate for the P-arrival on the Z component as the P-arrival is most prominent on the vertical component, and cross-correlate for the S-arrival on the E component as the S-arrival is most prominent on the horizontal components. We perform the cross-correlation and pick refinement twice: first we use the cross-correlation lag to refine the pick, secondly, from this new pick location we use the bispectrum method lag to refine the pick. This process was decided upon after testing as it was seen that the cross-correlation lag tended to behave better for coarse-adjustments, whilst the bispectrum method lag tended to behave better for fine-adjustments. As an initial location for each earthquake, we use a preliminary *Velost* solution [see Section 3.1.3].

To check the phase picking results, we take a random sample of slave events from every master event and every station within the event, giving us a sample size of 126, which we

check visually. On the basis of this checking the sample set was found to contain 59% good picks (<0.1 s difference), 24.5% reasonable picks (<0.25 s difference), 16.5% bad picks (>0.25 s). These are not great results, but alternative methods tried — time-domain cross-correlation and ARAIC [Leonard 2000] — gave poorer results. The need for accurate phase picks has been highlighted as the biggest limitation in our method of detecting and locating low-magnitude events and will be the largest source of location uncertainties. To improve on the worst of these phase picks we perform some manual phase-pick adjustment, using the RMS travel time residuals provided by *Velost* as a measure of the pick quality. Those data do not yield a normal distribution, so in order to repick the estimated 16.5% of bad picks, we order the data by RMS residual size to determine the value above which 16.5% of the data lie. This corresponds to $\sim 2.5 \times \text{median}(\text{RMS})$ and we therefore manually repick any event with an RMS exceeding $2.5 \times \text{median}(\text{RMS})$, for each master event set separately.

These provide the final phase picks. We then use these refined phase picks to perform another *Velost* location analysis.

3.1.2 Phase-pick weighting

We use the location residual file created by *Velost* combined with Jeffreys’ weighting method [Jeffreys 1973] to assign uncertainties to the phase picks. Jeffreys’ method employs a weight given by

$$w(r) = \frac{1}{1 + \mu \exp(r^2/2\sigma_0^2)} \quad (3.1)$$

where σ_0 is the standard deviation of the population without extreme values, and μ is related to the rate of occurrence of extreme values, which should be close to the (mean height of the tails)/(height of the central peak).

To calculate these parameters, we first need to determine a suitable weighting with which to remove extreme values; these are estimated for each station independently. To do this, we test for suitable values of P_1 and P_2 , which are the values of the residual distribution below and above which residuals are classed as extreme and removed:

1. Begin with values P_1 and P_2 set as those which 5% and 95% of the distribution lie below, respectively.
2. Calculate the test weighting $\sigma_t = 0.5(P_2 - P_1)$
3. Count the number k of values with $r > 3\sigma_t$, giving us roughly the mean height of the tails
4. Calculate the other test parameter $\mu_t = \frac{k}{n-k}$, where n is the number of data points, and $n - k$ is roughly the height of the central peak
5. Calculate the weight $w(r)$ with the test parameters σ_t and μ_t
6. Calculate the standard deviation σ_0 of the weighted data $rw(r)$
7. Calculate the difference $|\sigma_t - \sigma_0|$

An appropriate weighting should provide $\sigma_t \sim \sigma_0$ [Euan Smith, VUW, pers. comm., 11 January 2011; *Jeffreys* 1973]. We repeat steps 2–7 above, and test all P_1 and P_2 within the range 5–30% and 95–70%. The values are adopted that provide the minimum difference between σ_0 and σ_t , after which we set $\sigma_0 = \sigma_t$ and $\mu = \mu_t$. We now use Equation 3.1 to weight the data using the calculated μ and σ_0 , and provide this weight as the pick uncertainty. This assigns a weight to the picks in the range 0–1, where 1 is the best quality and 0 is the worst quality.

After this phase-pick weight assignment we run *Velest* again. We then extract the new location residual file, and perform a second application of *Jeffreys*' weighting, assigning new pick weights. *Velest* is then run for a final time. We are now in possession of our phase-picks and their corresponding weights and preliminary earthquake locations.

3.1.3 *Velest*

Velest [*Kissling et al.* 1994] is a “joint hypocentre–velocity model” inversion routine, but we use it here simply to obtain preliminary earthquake locations using fixed P- and S-wave

Table 3.1: 1D V_p and V_s velocity models used in *Velost* for Rotokawa. These are the 1D average of the GNS Science 3D-velocity model [Figs. 3.5 and 3.6; GNS Science 3D velocity model, version 2010-dec05a; Stephen Bannister, pers. comm., 18 January 2011]

Depth (km)	V_p (km/s)	V_s (km/s)
-0.40	2.00	1.14
0.10	2.63	1.50
0.60	2.80	1.61
1.00	2.90	1.66
1.30	2.97	1.70
1.60	3.05	1.74
1.80	3.12	1.78
2.00	3.17	1.81
2.20	3.20	1.83
2.60	3.41	1.95
2.90	3.74	2.14
3.40	4.32	2.47
4.00	4.57	2.61
6.00	5.51	3.15
10.00	5.81	3.32
20.00	6.90	3.94
50.00	7.43	4.25

velocity models. The velocity models used are detailed in Table 3.1, and are 1D models created by averaging over a 3D-velocity model for Rotokawa created by GNS Science colleagues [Figs. 3.5 and 3.6; GNS Science 3D velocity model, version 2010-dec05a; Stephen Bannister, pers. comm., 18 January 2011]. We adopt station RT08 as the reference station as it is in the centre of the array and exhibits good SNR. Station elevations are incorporated in the calculations, and locations are output with reference to sea level.

We run *Velost* a total of four times [Fig. 3.1; Table 3.2]. For the first run, we use phase picks transferred from the master event, all assigned a weighting of one, an initial location of the slave event’s corresponding master event, and an initial event-time (synonymous with origin time) of the earliest P-pick. We use the earliest P-pick as the initial event-time because the earthquakes are very shallow and close to the stations. For the second run, we use the refined picks as detailed in Section 3.1.1, all assigned a weighting of one, and the location and event-times determined in the previous *Velost* run. The third run uses the refined picks as detailed in Section 3.1.1 with a weighting determined using Jeffreys’ weighting [Section

3.1.2], and the location and event-time determined in the previous *Velest* run. For the fourth and final run, we use the refined picks detailed in Section 3.1.1 with a weighting determined using a second application of Jeffreys' weighting [Section 3.1.2], and the location and event-times determined in the previous *Velest* run. During each run we run *Velest* iteratively until all RMS residuals have stabilised.

The results for all events are displayed in Fig. 3.2, and a histogram of their RMS time residuals is shown in Fig. 3.3. The events have a mean RMS residual of 0.1 s, with 99.84% of residuals being smaller than 0.35 s.

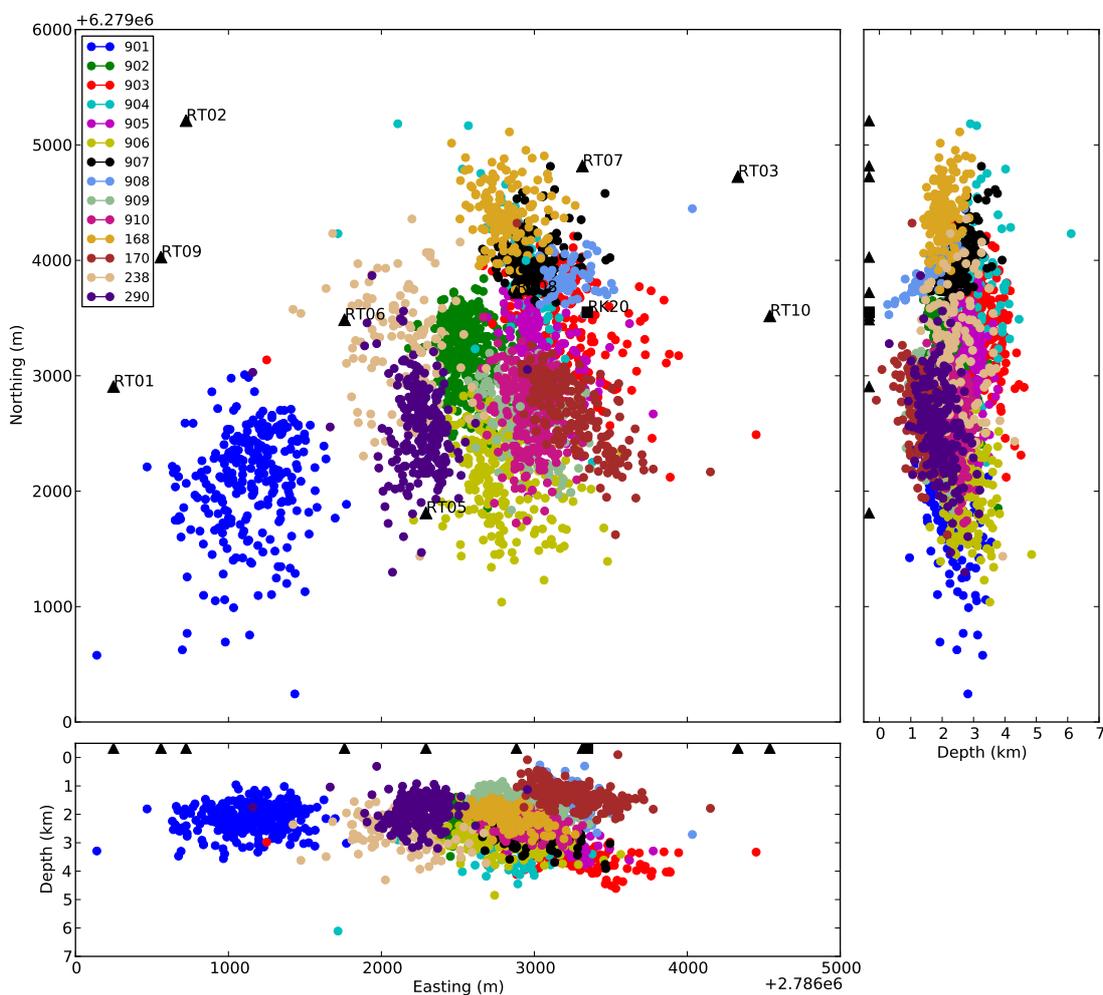


Figure 3.2: Earthquake locations after the fourth *Velest* run. Slave-events are colour coded based on their master event. Black triangles denote seismometers and the black square is the injection well RK20.

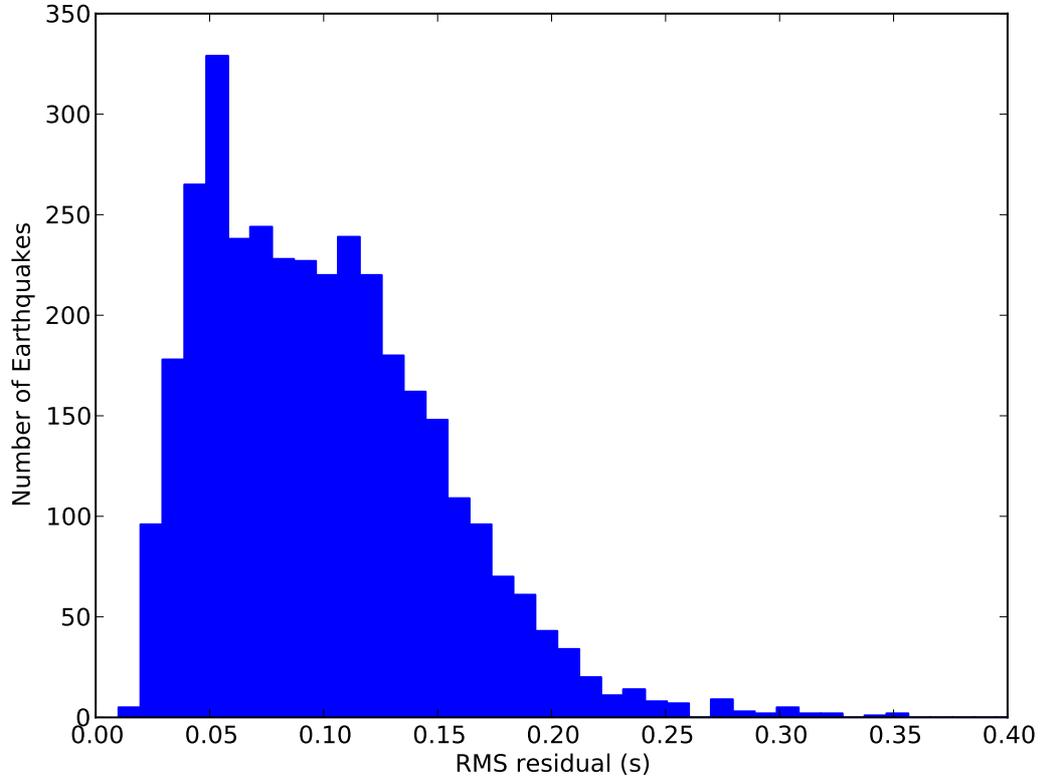


Figure 3.3: RMS residuals for the earthquake locations shown in Fig. 3.2. The events have a mean RMS residual of 0.1 s. Note that four particularly large residuals (of 2461) are not displayed: 0.48, 0.72, 0.98, and 3.86 s.

Table 3.2: Parameters used in each *Velost* run. JW 1 and JW 2 refer to the first and second application of Jeffreys' weighting, respectively.

run	phase-picks	weights	initial locations	initial event-times
run 1	transferred from master	1	master location	earliest p-pick
run 2	refined	1	from run 1	from run 1
run 3	refined	JW 1	from run 2	from run 2
run 4	refined	JW 2	from run 3	from run 3

3.2 Double-difference relocation

Double-difference earthquake location algorithms take advantage of the fact that if the hypocentral separation between two earthquakes is small compared to the distance between each event and a common station and the length scale of velocity heterogeneity, then the ray paths between the source region and the common station are similar along almost the entire ray path. Travel time differences for pairs of earthquakes observed at common stations are used to remove mutual errors resulting from these near coincident ray paths that sample similar velocity heterogeneities. As the detected earthquakes are composed of clusters of events that are highly correlated, by design, we would expect this algorithm to be ideal for their location. We perform calculations of cross-correlation differential travel times using *BC-SEIS* [Du et al. 2004] and perform double-difference relocation using *hypoDD* [Waldhauser 2001]. Details of the parameters used in these processes are discussed in Appendix C.1.

hypoDD implements the double-difference earthquake location algorithm of Waldhauser and Ellsworth [2000], and has been extensively tested with data from permanent networks and aftershock arrays around the world [Clarke et al. 2009; Kumano et al. 2006; Lippitsch et al. 2005; Waldhauser and Ellsworth 2000]. However, the technique is dependent on an event being recorded by a number of stations. In principle, earthquake data should only be used that have at least eight common observations: this threshold is equal to the number of degrees of freedom (two hypocentres, each with four coordinates) in the location of an event pair [Waldhauser 2001]. In this case, however, the majority of our master events, and thus their respective slave events, are only recorded on six stations.

We use both catalogue and cross-correlation-derived travel time differences, and as we have more than 100 earthquakes to solve for, we must use *hypoDD*'s LSQR inversion method, which is the conjugate gradient method applied to the least-squares problem [Paige and Saunders 1982]. Uncertainties reported by the LSQR method are unreliable because the diagonal elements of the covariance matrix are only approximately computed and critically dependent on the proper convergence during the internal iterations [Paige and Saunders 1982; Waldhauser 2001]. Therefore, we assess location uncertainties based on the conditioning

number (CND) of the matrix formed by the system of linear equations describing each location calculation. CND is the ratio of the largest to the smallest eigenvalue and represents how stable the inversion is. A damping factor is used to damp the hypocentral adjustments if the adjustment vector becomes large or unstable. Generally, a damping between 1 and 100 is appropriate, resulting in a condition number of between about 40 and 80 [Waldhauser 2001]. If a very high damping is needed to reduce the condition number, or if the condition number remains high, then the system may not be well conditioned. This might be because of weak links between events, data outliers, or extreme differences in weights, etc. [Waldhauser 2001]. Larger numbers of earthquakes generally produce larger condition numbers, as it only takes one poorly constrained event to drive up the condition number. Because of this reason we have removed the four earthquakes with RMS residuals > 0.35 [Fig. 3.3] from the *hypoDD* analysis.

Using the parameters described in Table C.2, with damping set to 100, the CND is 267. To elicit a CND of 40–80, the lowest damping we can use is 270. This represents extremely high damping, and reflects the fact that the inversion performed on the Rotokawa data set is unstable. We believe this is due to the low number of stations at Rotokawa and the low-magnitudes of the earthquakes.

We display the locations determined using the damping parameter of 270 in Fig. 3.4, but we do not have confidence in these locations because of the high damping required to obtain an acceptable CND, and because of the locations themselves. The locations look suspicious for two reasons. First, the overall distribution of earthquakes is markedly more amorphous than that obtained with *Velost*, despite the use of cross-correlation measurements. Second, we see no clear clustering of the slave events corresponding to any particular master. This is unexpected because the waveforms of the slave events are, by definition, highly similar to those of the master event, implying that they are reasonably closely located. We attribute the poor performance of *hypoDD* in this case to a lack of geometric constraint imposed on each earthquake by our use of only six stations on average. As we show using *NonLinLoc* in the following section, refined hypocentres that do cluster as expected can be obtained.

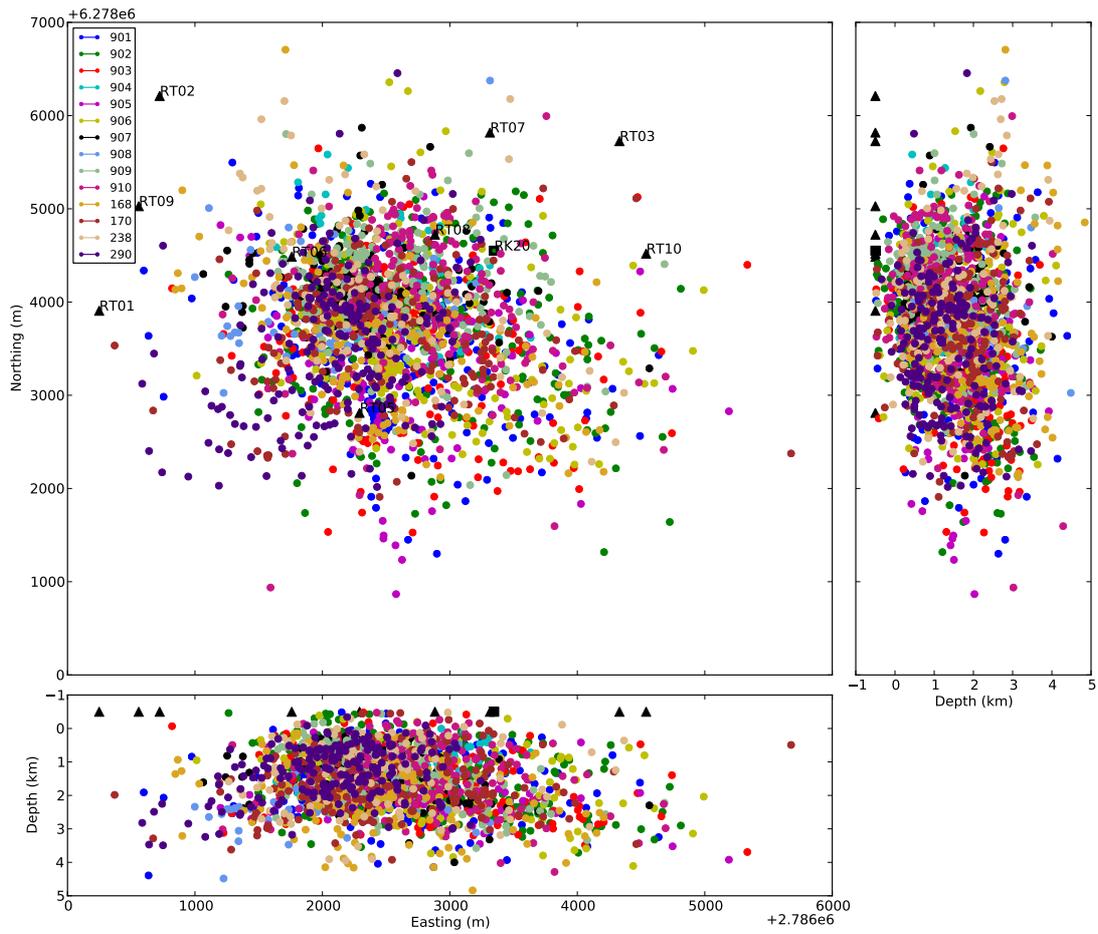


Figure 3.4: *hypoDD* locations with slave-events colour coded based on their master event. Note the lack of slave-event clustering and overall lack of any spatial structure.

3.3 Probabilistic earthquake location: *NonLinLoc*

In view of the unsatisfactory results obtained using *hypoDD*, we also use *NonLinLoc* software [Lomax *et al.* 2000; 2009] to obtain improvement in the earthquake locations. *NonLinLoc* implements the probabilistic non-linear inversion algorithm of *Tarantola and Valette* [1982] to provide an estimate of the posterior probability density function (PDF) for each hypocenter location. The uncertainties in the observations (phase time picks) and in the forward problem (travel-time calculation) are assumed to be Gaussian. This assumption allows the direct, analytic calculation of a maximum likelihood origin time given the observed arrival times and the calculated travel times between the observing stations and a point in space. Thus the 4D problem of hypocenter location reduces to a 3D search over the three spatial coordinates. The grid-search earthquake location is expressed probabilistically through a spatial (3D) PDF [Lomax *et al.* 2000; 2009].

To estimate the PDF, we use the Oct-Tree importance sampling algorithm, as it is faster than the grid-search method by a factor of approximately 100, more robust than Metropolis-simulated annealing (MET) in that it can identify multi-modal solutions, and requires few user-dependant input parameters [see <http://alomax.free.fr/nlloc/>, last accessed 13 June 2011]. Although a complete PDF can only be obtained when an exhaustive grid-search is undertaken, the Oct-Tree importance sampling algorithm suffices to describe the PDF [Lomax *et al.* 2000]. This search method solves the equal differential time (EDT) likelihood function, which is more robust in the presence of outliers than the usual root-mean-square L2 norm [Lomax *et al.* 2009]. In contrast to the single “optimal” location and associated, local, ellipsoidal (Gaussian) hypocentral statistics produced by linearised location methods, the non-linear PDF can define multi-modal hypocentre solutions and highly non-ellipsoidal (non-Gaussian) uncertainty volumes. In this context, we define our earthquake hypocentres as the maximum likelihood earthquake location within the PDF produced: this is the location that has the highest probability within the PDF.

The velocity models we use are the 3D P-wave and S-wave velocity models obtained by GNS Science colleagues [GNS 3D velocity model, version 2010-dec05a; Stephen Bannister,

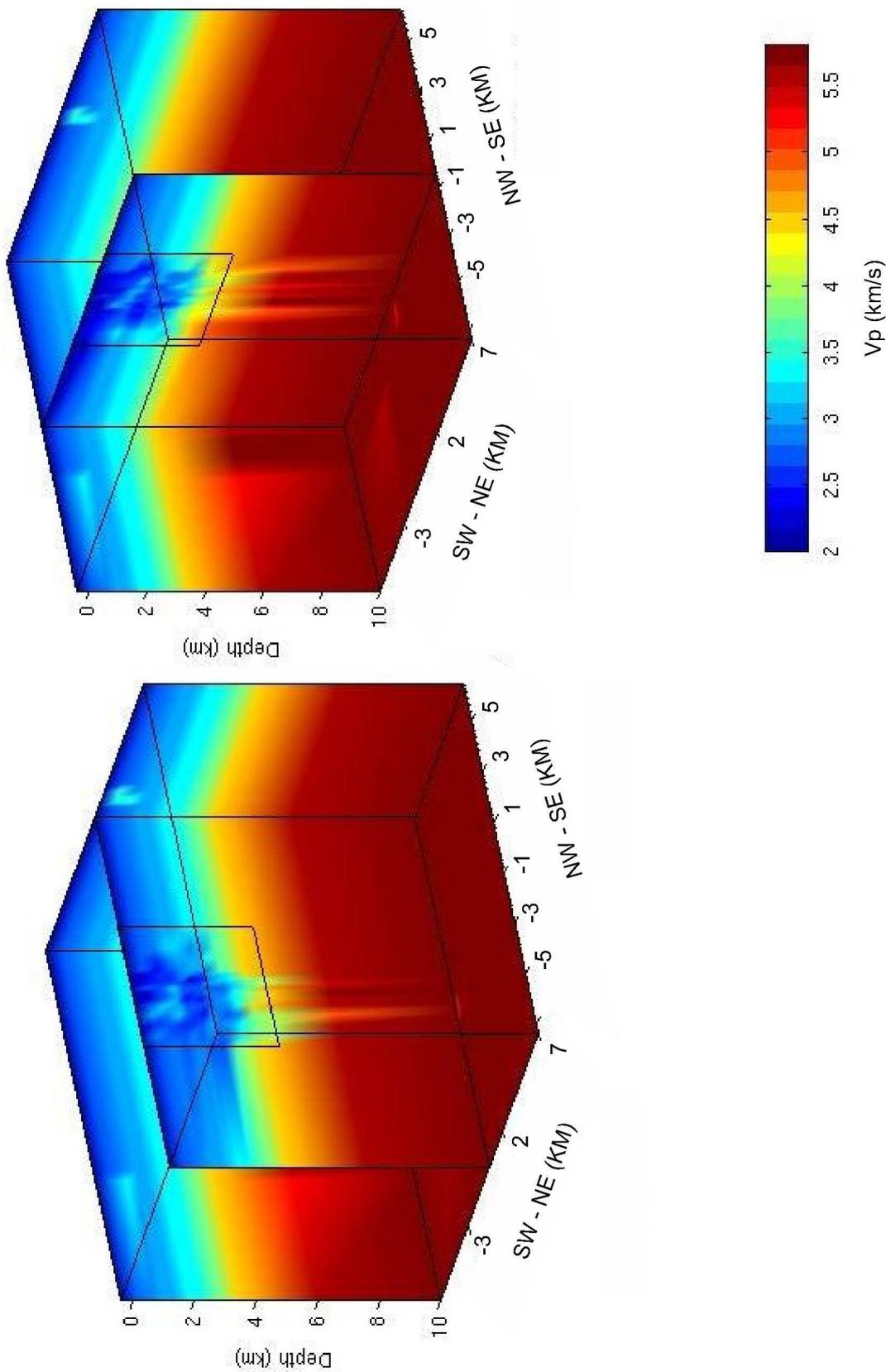


Figure 3.5: Vp 3D velocity model. The coordinate system is centred at an Easting and Northing of 2788000 and 6282500 respectively, and has a 30° rotation, therefore axes are aligned NW-SE and SW-NE: roughly parallel to the geological cross-sections A-A' and B-B' respectively (see Figs. 1.3 and 1.6). Only the central portion of this volume is well resolved due to ray coverage, and the focus being on reservoir velocity determination. Black squares outline the central slices displayed in Fig. 3.6.

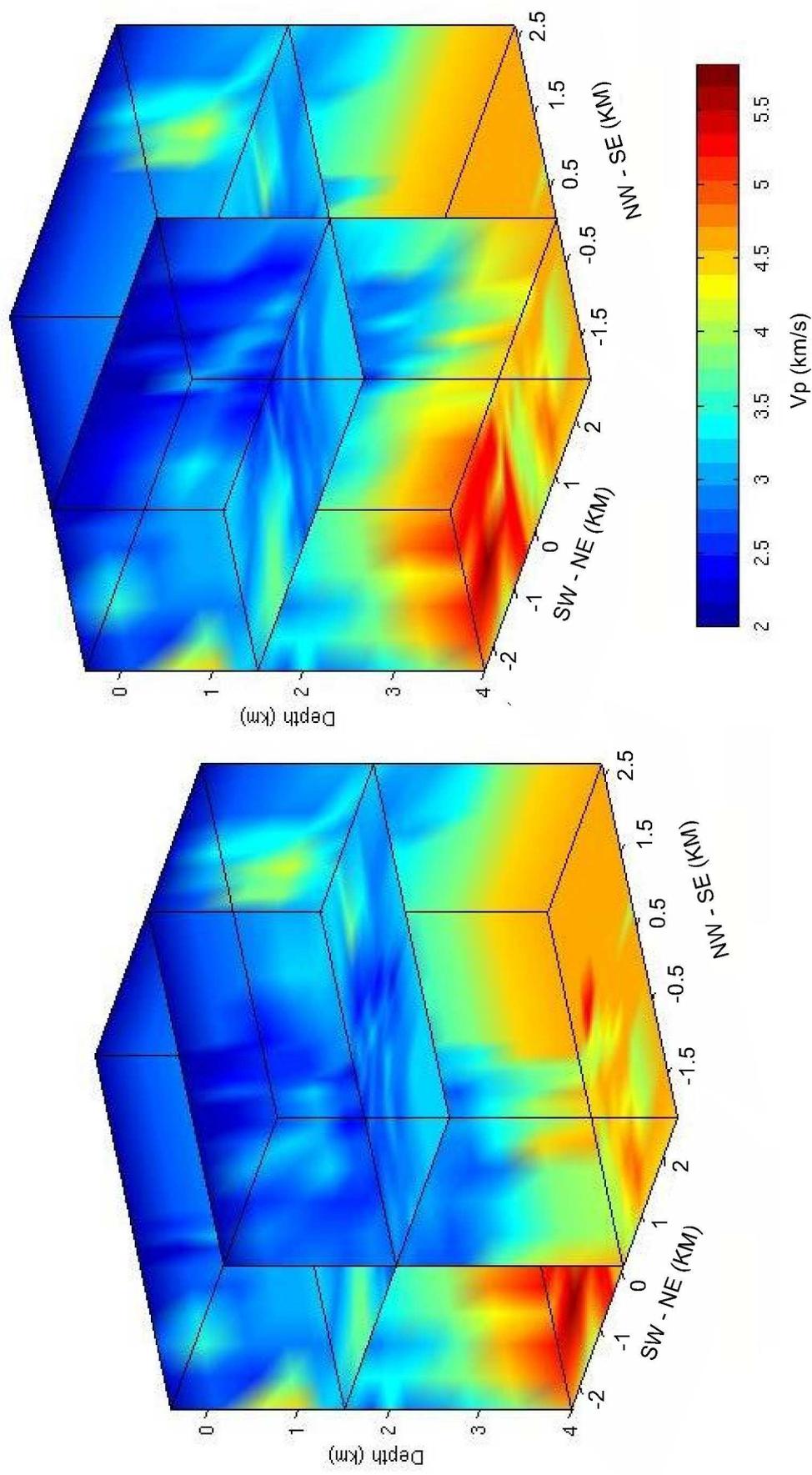


Figure 3.6: Vp 3D velocity model, zoom-in on the geothermal reservoir. The coordinate system is centred at an Easting and Northing of 2788000 and 6282500 respectively, and has a 30° rotation, therefore axes are aligned NW-SE and SW-NE: roughly parallel to the geological cross-sections A-A' and B-B' respectively (see Figs. 1.3 and 1.6).

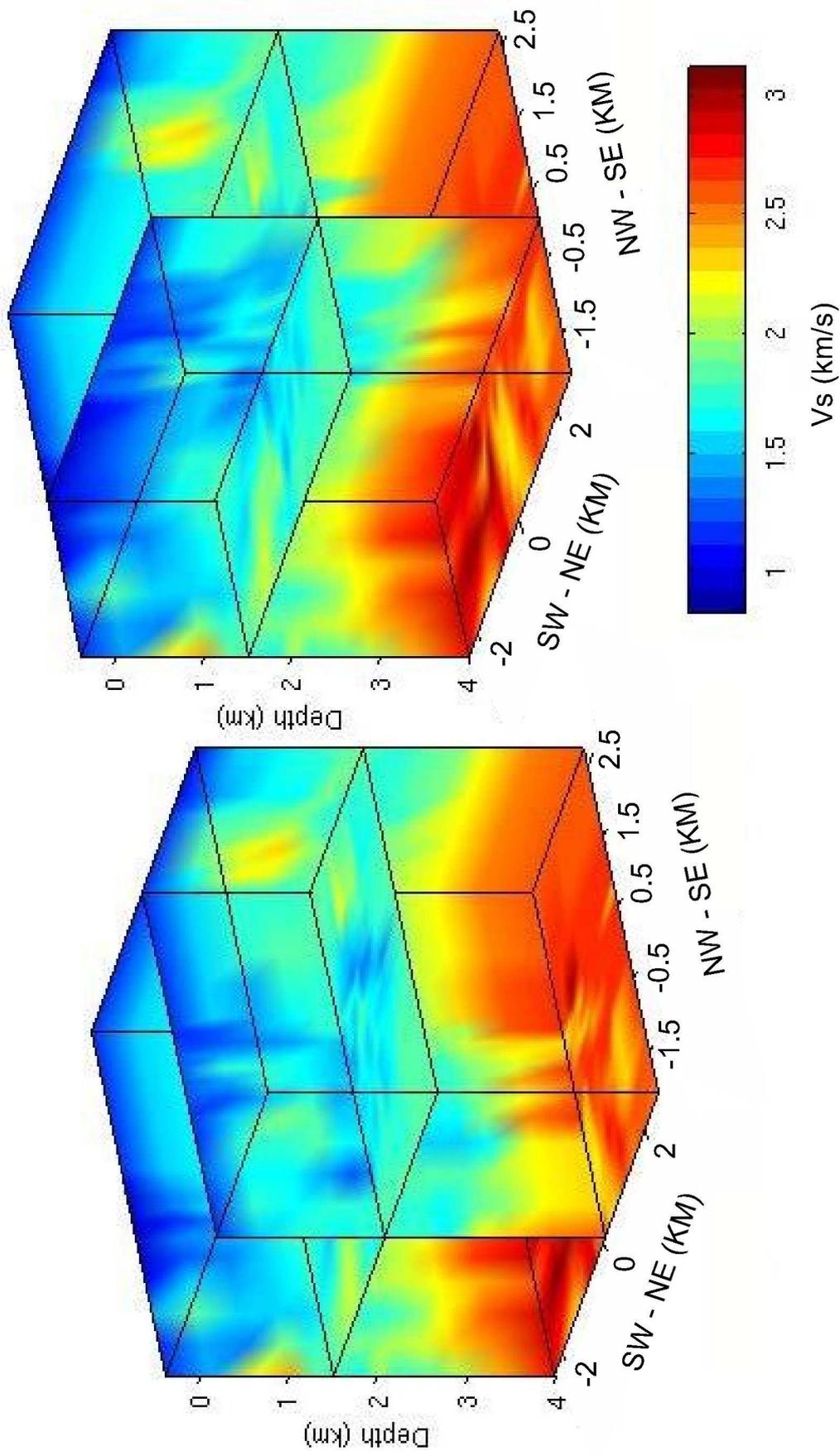


Figure 3.7: V_s 3D velocity model, zoom-in on the geothermal reservoir: the same area is displayed as in Fig. 3.6. The coordinate system is centred at an Easting and Northing of 2788000 and 6282500 respectively, and has a 30° rotation, therefore axes are aligned NW-SE and SW-NE: roughly parallel to the geological cross-sections A-A' and B-B' respectively (see Figs. 1.3 and 1.6). Note the colour scale is different to the V_p velocity models shown previously.

pers. comm., 18 January 2011]. These models are illustrated in Figs. 3.5–3.7, and are resampled onto a 250 m resolution grid. The velocity models provided are part of ongoing work by GNS Science colleagues, and the labelling of these models as version 2010-dec05a reflects that they are constantly being updated and refined as further constraining data is obtained and velocity inversions performed. The models are expected to evolve from the version we have used, and any inaccuracies in this version could degrade our calculated hypocentres.

S- and P-phases retain the same phase pick weights used in *Velest*. The uncertainty in earthquake location is determined by sampling of the PDF. This uncertainty volume is directly influenced by several user-defined parameters within *NonLinLoc*. These parameters are:

1. σ^T , which is the average error in seconds incurred during the travel-time to one station due to model errors;
2. L_{corr} , the correlation length that controls covariance between stations, which may be interpreted as the characteristic length scale of velocity anomalies between two stations;
3. $\Delta\sigma^T$, the fraction of the travel time to use as error, which introduces a station–distance weighting;
4. σ_{min}^T minimum travel time error in seconds; and
5. σ_{max}^T maximum travel time error in seconds.

We assign values of $\sigma^T = 0.1$ s the mean of the RMS residuals from *Velest* locations; $\sigma_{min}^T = 0.001$ s to allow travel time errors near zero; and $\sigma_{max}^T = 0.2$ s, as nearly all RMS residuals from *Velest* locations are less than this value. We set $\Delta\sigma^T = 5\%$, and $L_{corr} = 2.5$ km as this is the mean station distance and so the maximum anomaly length which we would expect to resolve with the station spacing. The minimum number of phases required to locate an earthquake is set to four. As the location PDF is nonlinear and may be irregularly-

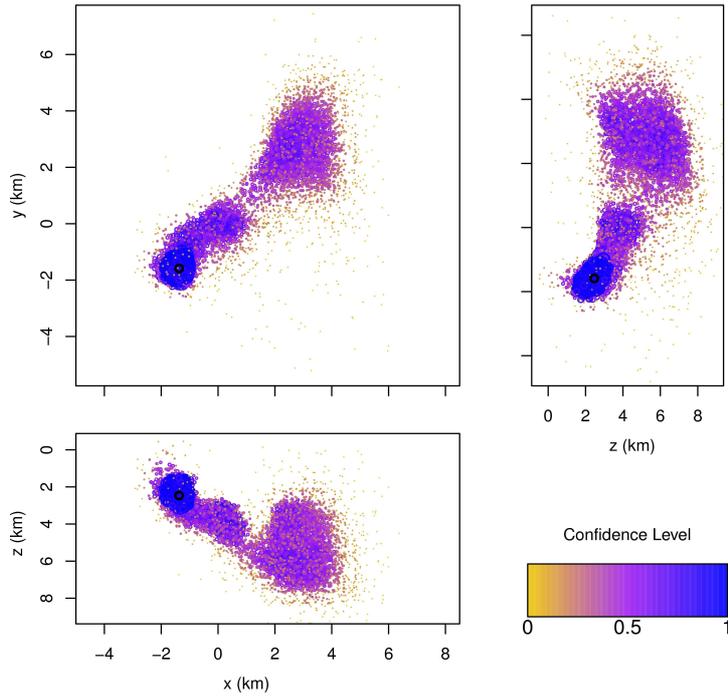


Figure 3.8: Example of a 3D PDF for a poorly constrained earthquake. The black circle denotes the maximum likelihood hypocentre, with an uncertainty $\sqrt{\lambda} = 1.85$ km. Note that the same colour scale has been used as in Fig. 3.9, but because the earthquake is so poorly constrained most of the PDF has a similar confidence level.

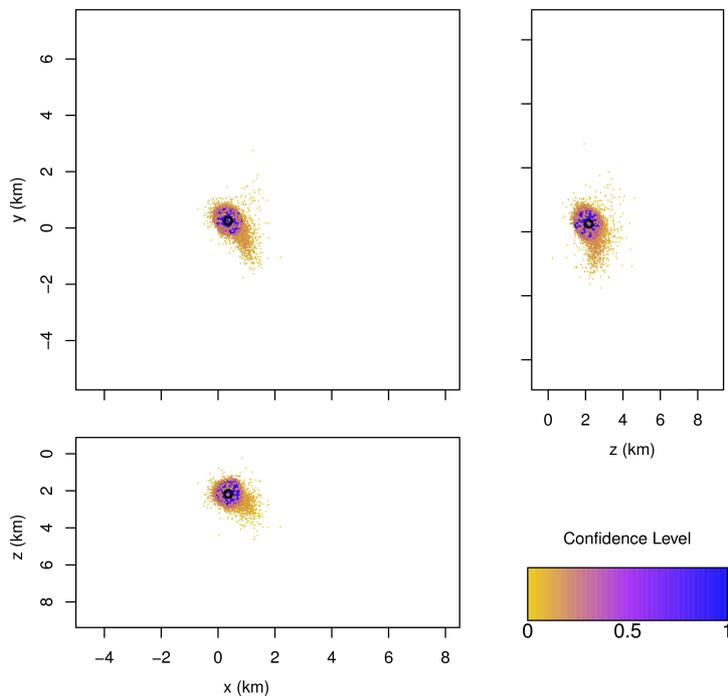


Figure 3.9: Example of a 3D PDF for a well constrained earthquake. The black circle denotes the maximum likelihood hypocentre, with an uncertainty $\sqrt{\lambda} = 0.18$ km. Note that the same colour scale has been used as in Fig. 3.8, but because the earthquake is well-constrained only a small portion of the PDF has a high confidence level.

shaped and/or multi-modal, it is convenient to approximate its volume in order to assign a single-valued uncertainty estimate. We do this by treating the PDF as Gaussian in shape and compute the lengths of the principal axes of the best fitting ellipsoid. The smaller and denser the region covered, the better constrained the earthquake's location is. Each point in the PDF is assigned a confidence level between 0 and 1, and the point with the highest confidence is the maximum-likelihood hypocentre location. Fig. 3.8 displays an example of a poorly constrained earthquake, with an irregular PDF, and Fig. 3.9 displays an example of a well-constrained earthquake with an almost spherical and highly localised PDF.

The mean event location, given the N scatter locations $\{\mathbf{x}_k\}$ that compose the PDF, and their confidence levels $\{\omega_k\}$, is

$$\bar{\mathbf{x}} = \frac{1}{\sum_{k=1}^N \omega_k} \sum_{k=1}^N \omega_k \mathbf{x}_k \quad (3.2)$$

with 3×3 variance covariance matrix

$$\hat{V} = \frac{1}{\sum_{k=1}^N \omega_k} \sum_{k=1}^N \omega_k (\mathbf{x}_k - \bar{\mathbf{x}})(\mathbf{x}_k - \bar{\mathbf{x}})^T \quad (3.3)$$

We solve for the eigenvalues of \hat{V} , λ_i ($i = 1, \dots, 3$), which correspond to the lengths of the principal axes of the best fitting ellipsoid to the PDF. The square root of the mean of these eigenvalues provide us with a measure of the size of the PDF in km

$$\sqrt{\bar{\lambda}} = \sqrt{\frac{1}{3} \sum_{i=1}^3 \lambda_i} \quad (3.4)$$

We refer this parameter $\sqrt{\bar{\lambda}}$ below as the location uncertainty.

Fig. 3.10 displays the maximum likelihood hypocentres determined using *NonLinLoc*, scaled and coloured by location uncertainty $\sqrt{\bar{\lambda}}$. Our catalogue has been reduced to 2101 earthquakes due to the removal of ‘air-quakes’, which are earthquakes which locate above the surface of the Earth. Fig. 3.11 displays the corresponding region with every earthquake's

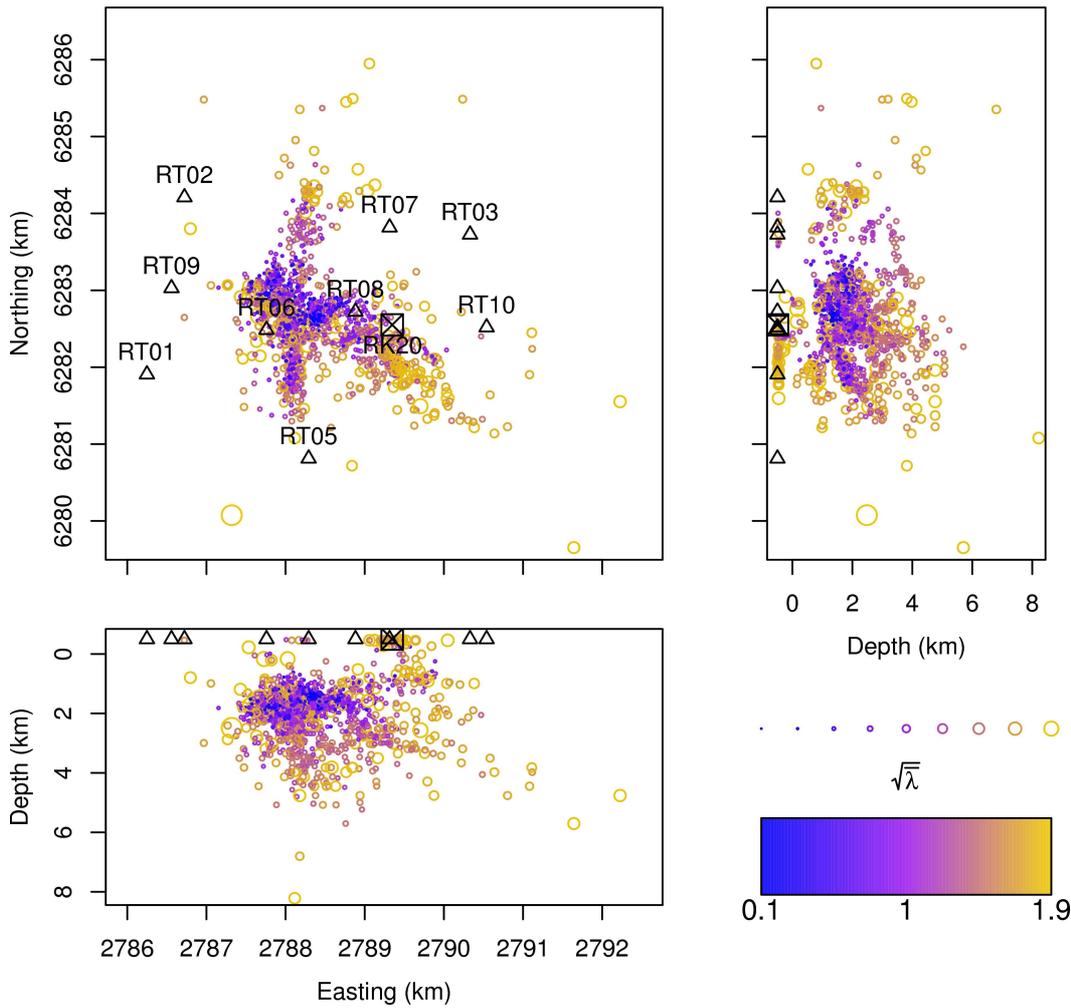


Figure 3.10: Maximum likelihood hypocentres determined using *NonLinLoc*, scaled and coloured by location uncertainty $\sqrt{\lambda}$. Black triangles are seismometers, and the black square with a cross is the injection well RK20.

PDF estimate combined into a single PDF describing the distribution of all 2101 earthquakes. Here, dark blue corresponds to very high density and white to zero density. These figures identify a region of well-located earthquakes at 1.0–2.5 km depths, and in the centre of the seismometer network where two lines of earthquakes strike north–south and northwest–southeast. The worst located earthquakes occur to the north and southeast, due to gaps in the network between RT02 and RT07, and RT05 and RT10. Fig. 3.12 displays histograms of the east, north and depth uncertainties. Due to the station configuration, the hypocentres’ east coordinates have a lower uncertainty than the north coordinates, with a mean east uncertainty of 0.26 km and a mean north uncertainty of 0.34 km. As expected, the earthquakes’ depths

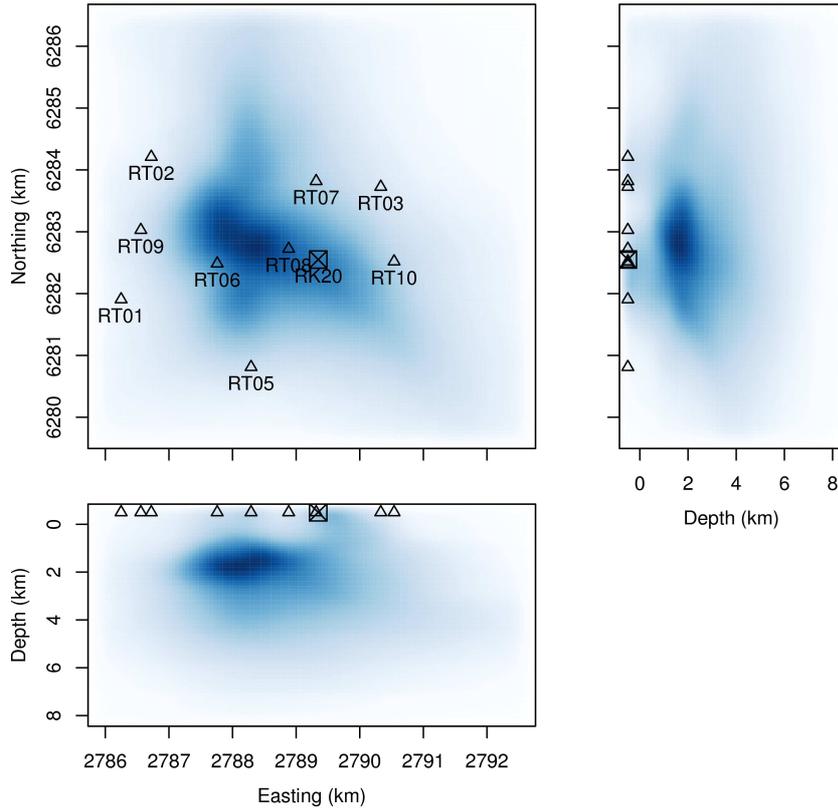


Figure 3.11: Every earthquake’s PDF estimate combined into a single PDF, where dark blue corresponds to a very high density and white to zero density. Black triangles are seismometers and the black square with a cross is the injection well RK20.

are less well determined than either horizontal component, with a mean uncertainty of 0.42 km. Fig. 3.13 displays the hypocentre uncertainty $\sqrt{\lambda}$ which has a mean of 0.36 km. Fig. 3.14 displays the $\lambda_{max}/\lambda_{min}$ which is a rough proxy for the asphericity of the PDF volumes, and has a mean of 2.97. Fig. 3.15 displays the RMS time residuals with a mean of 0.11 s: this is comparable to the mean RMS residual of 0.1 s obtained using *Velost*.

We discuss the spatiotemporal patterns of the *NonLinLoc* earthquake hypocentres and earthquake origin times in the following section.

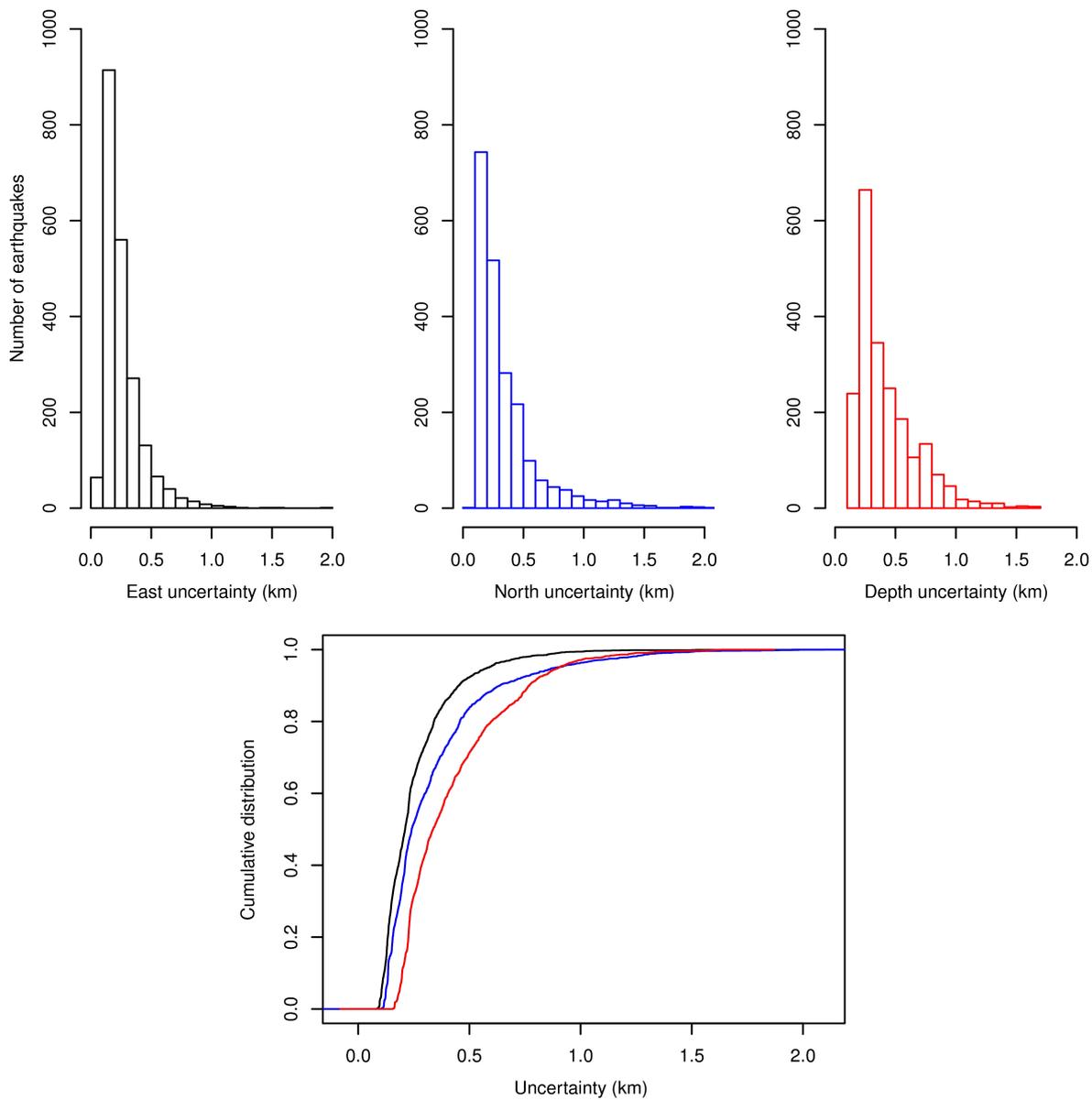


Figure 3.12: Histograms and empirical cumulative density functions of hypocentre east, north and depth uncertainties.

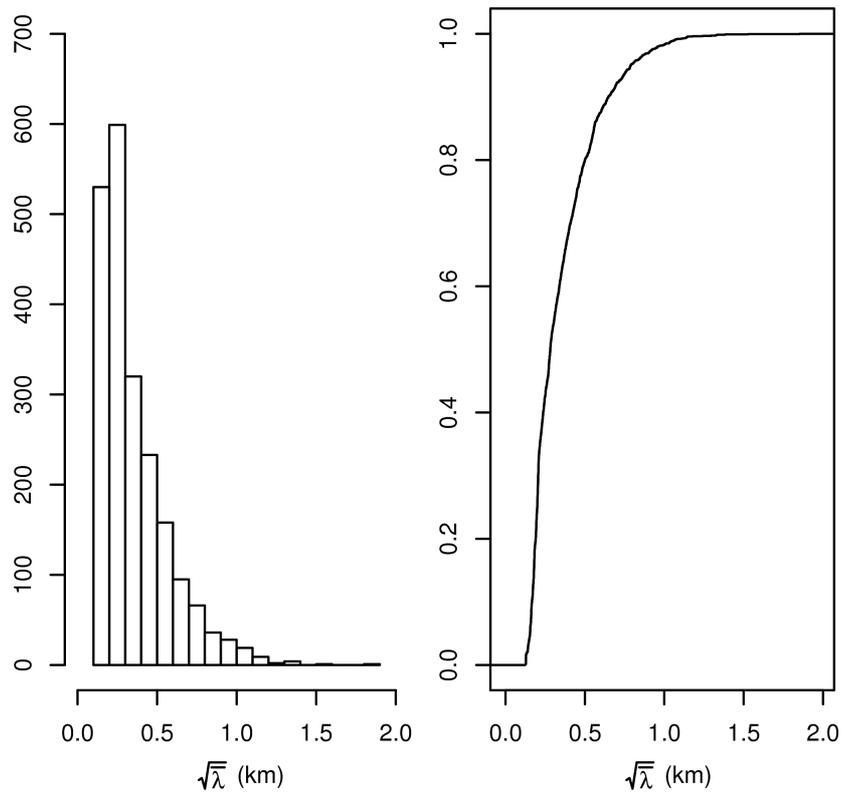


Figure 3.13: Histogram and empirical cumulative density function of hypocentre uncertainty $\sqrt{\lambda}$.

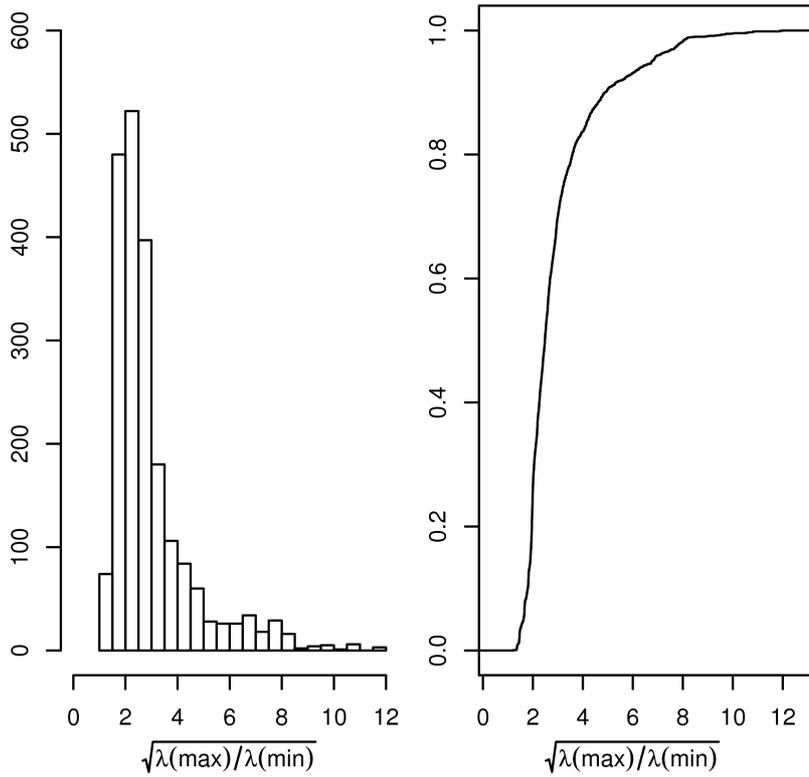


Figure 3.14: Histogram and empirical cumulative density function of asphericity of the PDF volumes, described by the $\lambda_{max}/\lambda_{min}$.

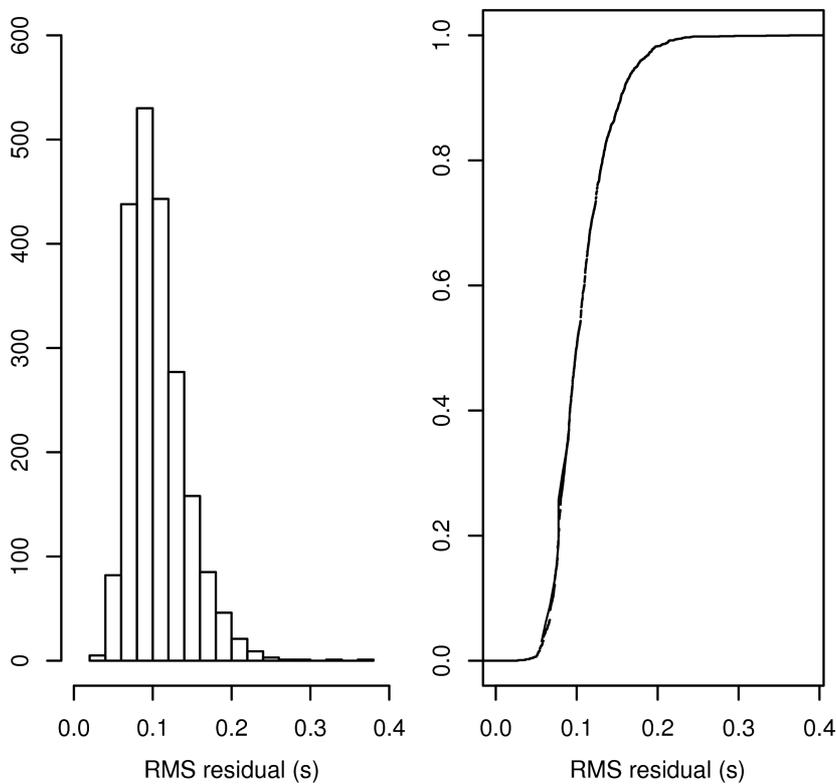


Figure 3.15: Histogram and empirical cumulative density function of hypocentre RMS time residuals.

3.4 Spatiotemporal analysis

In the remainder of this chapter we investigate some of the spatial and temporal characteristics of the detected events.

Fig. 3.16 displays histograms of how the Rotokawa earthquakes are distributed within the East, North and depth planes. Two dominant lineations of seismicity are evident, which intersect in the vicinity of 2788000 E, 6283000 N, and have a median depth of 1.7 km. The dominant locus of seismicity, that lies at 1.0–2.5 km depth, is consistent with the location of the Rotokawa Andesite which forms the Rotokawa reservoir [Fig. 1.3]. Fig. 3.17 displays the earthquakes coloured according to their respective master events. Slave events generally cluster together, with scatter usually produced by the most uncertain earthquakes. As there is substantial overlap between some clusters, we provide location maps displaying each cluster of slave events separately in Appendix C.2. Note we show only events at depths less than 5 km, as there are only ~ 4 poorly constrained hypocentres located deeper than this [Fig. 3.17]. This observation is consistent with observations in the TVZ that 80% of seismicity occurs at depths of < 6 km, which defines the base of the seismogenic zone beneath the central TVZ [Bibby *et al.* 1995; Bryan and Sherburn 1999].

3.4.1 Focal mechanisms

We have determined focal mechanism solutions from P-wave first motion polarities for the 14 master events used in the earthquake detection process. Focal mechanisms have been calculated using the Bayesian method of Walsh *et al.* [2009], which utilises results provided by *NonLinLoc*. The PDF estimate of a hypocentre provided by *NonLinLoc* results in a PDF of different P-wave take-off and azimuth angles, which form a cloud of points on the focal sphere. The take-off angles of shallow earthquakes, such as those analysed here, are particularly sensitive to depth uncertainties and velocity model errors [Hardebeck 2002].

Two other parameters are used to represent observational uncertainties: the prior probability of a reversely wired seismometer, and the first motion amplitude divided by its standard

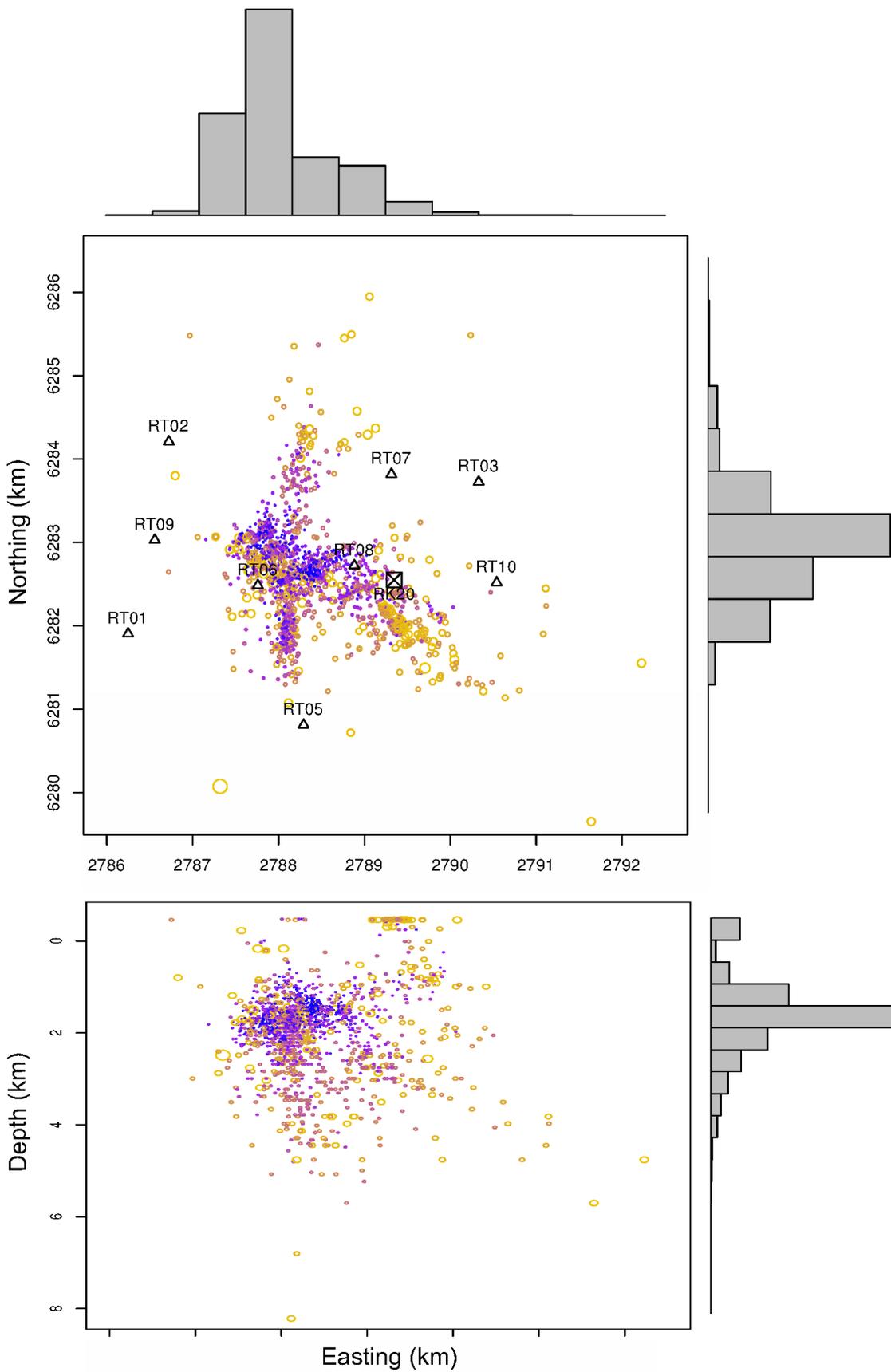


Figure 3.16: Histograms displaying the distribution of earthquake locations in the East, North and depth planes. Earthquakes are coloured and scaled based on uncertainty, as per Fig. 3.10.

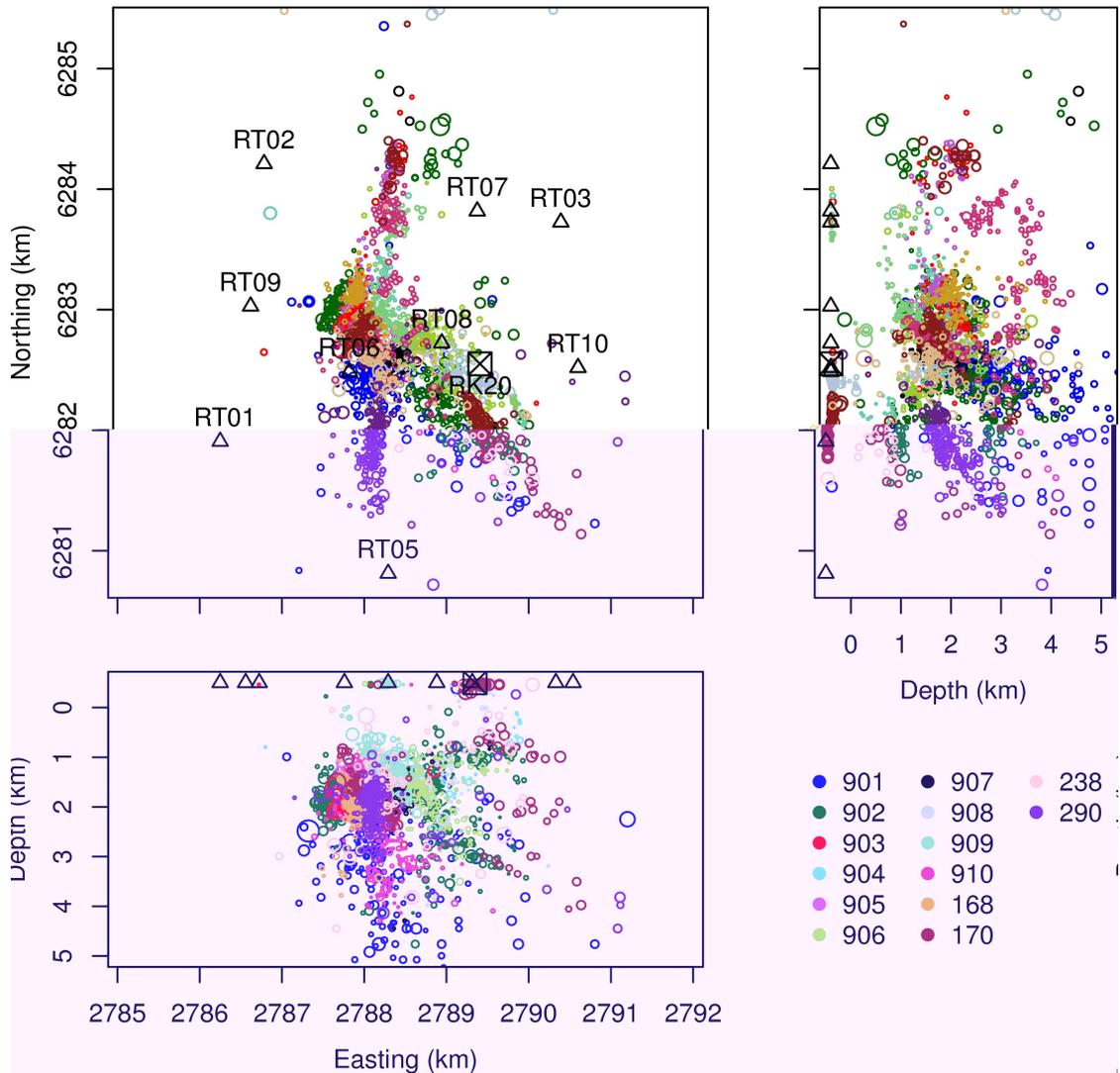


Figure 3.17: Earthquakes coloured based on each slave event’s respective master event. Earthquakes are scaled based on uncertainty as per Fig. 3.10. Note that this is a zoom-in of the region, refer to Fig. 3.16 for the few missing earthquakes.

deviation — the noise amplitude. The wiring of the Rotokawa instruments is set to ‘seismic exploration industry’ standard (positive down), which is opposite to the convention normally used in earthquake seismology. We have corrected for this difference, and accordingly set the prior probability of a reversely wired seismometer to a low value of 0.001: given the small network at Rotokawa that was deployed by a single team over a short time period, we do not expect inconsistencies between instruments. We set the noise amplitude parameter to $1/6$ after *Walsh et al.* [2009], which represents the probability of an incorrect polarity reading due to instrumental or human effects.

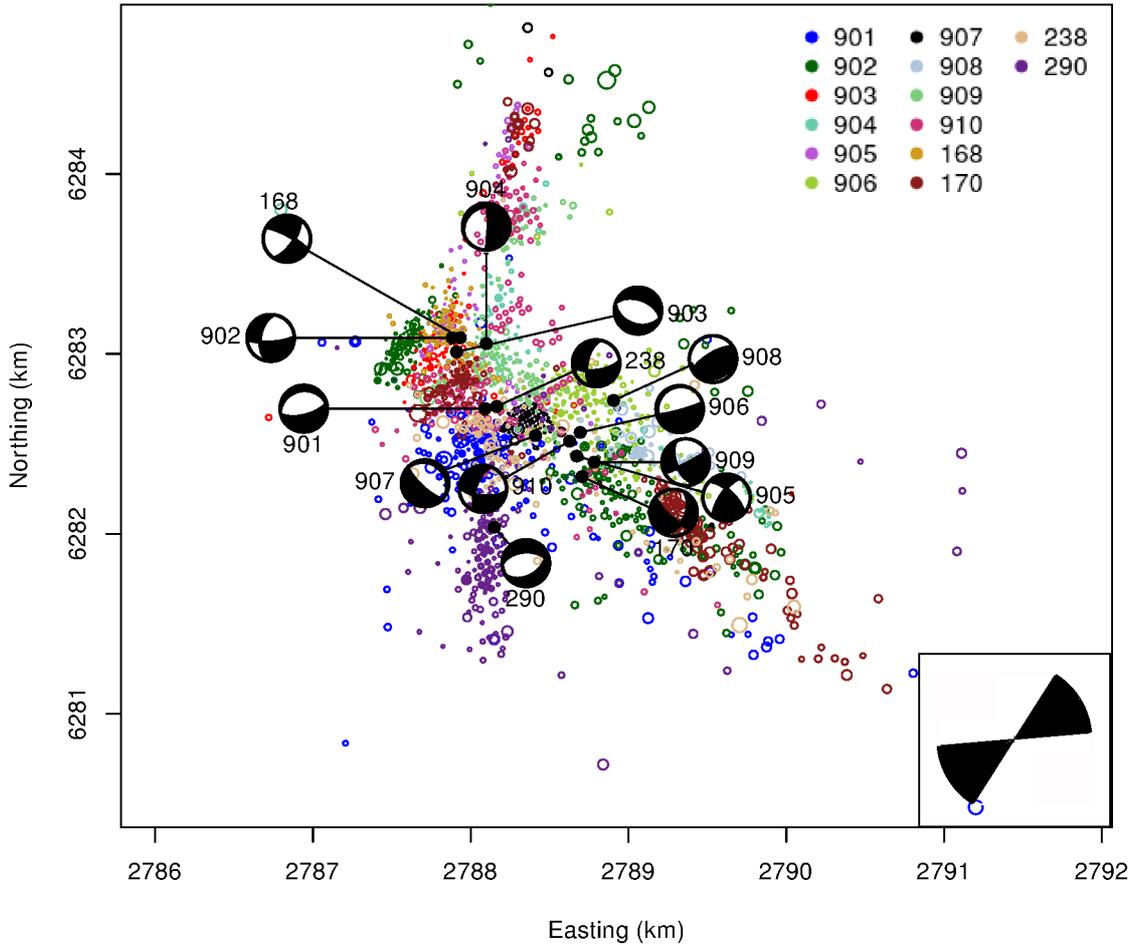


Figure 3.18: Focal mechanism solutions for the 14 master events. See Table 3.3 and Appendix C.3 for details of each focal mechanism. Solid circles indicate the epicentre of the adjoining focal mechanism. In the bottom right hand corner is shown the axis of maximum horizontal compression (S_{Hmax} determined from this suite of focal mechanisms) with the wedge symbol spanning the 80% confidence interval of the estimate. See the following text and Fig. 3.20 for details of this stress calculation. Earthquakes are coloured based on their respective master event and scaled based on uncertainty as per Fig. 3.10.

The 14 focal mechanism solutions are displayed in Fig. 3.18. More detailed plots showing P and T axis contours and the distribution of compressional and dilatational points on each focal sphere are included in Appendix C.3, and a summary of the focal mechanism strike/dip/rake and uncertainty parameters is listed in Table 3.3. The uncertainty parameter has been calculated using a Matrix Fisher Distribution with scalar concentration parameter κ , whose validity as an approximation of the posterior PDF of the focal mechanism parameters has been investigated in some detail by *Walsh* [2008] and *Walsh et al.* [2009]. It is important to note that *Walsh et al.* [2009] found this approximation to be less satisfactory for badly

Table 3.3: Focal mechanism solution strike, dip, rake and uncertainty, all in degrees, for the 14 master events. Depth of the event is also shown. See Fig. 3.18 and Appendix C.3 for visual displays.

event	depth (km)	strike	dip	rake	uncertainty
901	2.30	76	67	267	25
902	2.03	87	75	323	22
903	2.21	290	51	271	26
904	1.97	1	83	255	27
905	2.25	214	70	155	26
906	2.11	74	84	276	27
907	2.19	308	15	265	23
908	2.21	239	71	89	27
909	1.99	63	85	327	27
910	2.09	96	66	332	21
168	1.91	34	67	193	30
170	1.97	132	73	225	29
238	2.07	84	65	325	25
290	2.19	268	50	290	27

constrained focal mechanisms, suggesting that in our case — where we have few stations and thus few polarity observations with which to constrain the solutions — uncertainties may be underestimated. *Arnold and Townend* [2007] established the following approximate relationship between κ and the standard deviation σ_{Θ} of the focal mechanism parameters (in degrees):

$$\sigma_{\Theta} = \exp(3.9155 - 0.5659 \log \kappa) \quad (3.5)$$

This σ_{Θ} is the parameter listed in Table 3.3 as the uncertainty for each focal mechanism.

Given the extension regime prevalent within the TVZ as a whole, normal faulting is the expected dominant faulting mechanism. This is reflected by the majority of the focal mechanisms being either purely normal, or having a large normal component. Half of the solutions have large strike-slip components and one solution presents as a thrust fault. Similarly, *Hurst et al.* [2002] presented focal mechanism solutions from the TVZ which were mainly normal, mostly had strike-slip components and a few had thrust mechanisms.

One focal mechanism that is particularly notable is the almost purely normal mechanism of event 290, with a strike/dip/rake $058^{\circ}/44^{\circ}/248^{\circ}$. The slave events of 290 are coloured purple on the location map [Fig. 3.17] and appear as a north–south lineation in the horizontal

plane, and in the north–depth plane as a linear feature with a dip of $\sim 35^\circ$ towards the south (within the focal mechanism’s 27° uncertainty). The shallowest of these earthquakes appear to coincide with an east–west striking inferred fault circled in black in Fig. 3.19, which shows a geological map of Rotokawa [*Krupp and Seward* 1987].

There are three sets of predominant fault orientations evident within the geological map of *Krupp and Seward* [1987]:

1. Northeast–southwest-striking normal faults consistent with the extensional stress regime of the Taupo Volcanic Zone. The orientation of these faults is expected to provide a northeast–southwest-striking permeability axis [*Grant* 2007].
2. Superimposed on these are radial fractures focused on the topographic high of the Oruahineawe rhyolite dome north of the Waikato river.
3. A series of inferred faults with a west-northwest–east-southeast strike.

It is interesting to note that we observe a northwest–southeast striking lineation of earthquakes that cross the northeast–southwest-striking normal faults. Some of the different clusters appear to be bounded by these faults, and some earthquakes may occur on these faults: ie. the fault to the immediate left of RK20, and the fault to the north of RT06.

We have also used the calculated focal mechanism solutions to perform a stress inversion using the algorithm of *Arnold and Townend* [2007] and the *Lund and Townend* [2007] S_{Hmax} transformation. *Arnold and Townend* [2007] developed a Bayesian method of estimating tectonic stress by assuming that each earthquake in a geographically confined region reflects a single, uniform state of stress, which produces slip on a planar fault in the direction of the maximum resolved shear stress (Wallace-Bott hypothesis). The stress inversion algorithm is intended to find the principal stress axes that are most consistent with the constraints imposed by a suite of focal mechanism solutions. The method of *Lund and Townend* [2007] is then used to estimate the direction of maximum horizontal compressive stress S_{Hmax} in the case that none of the three principal stresses is strictly vertical.

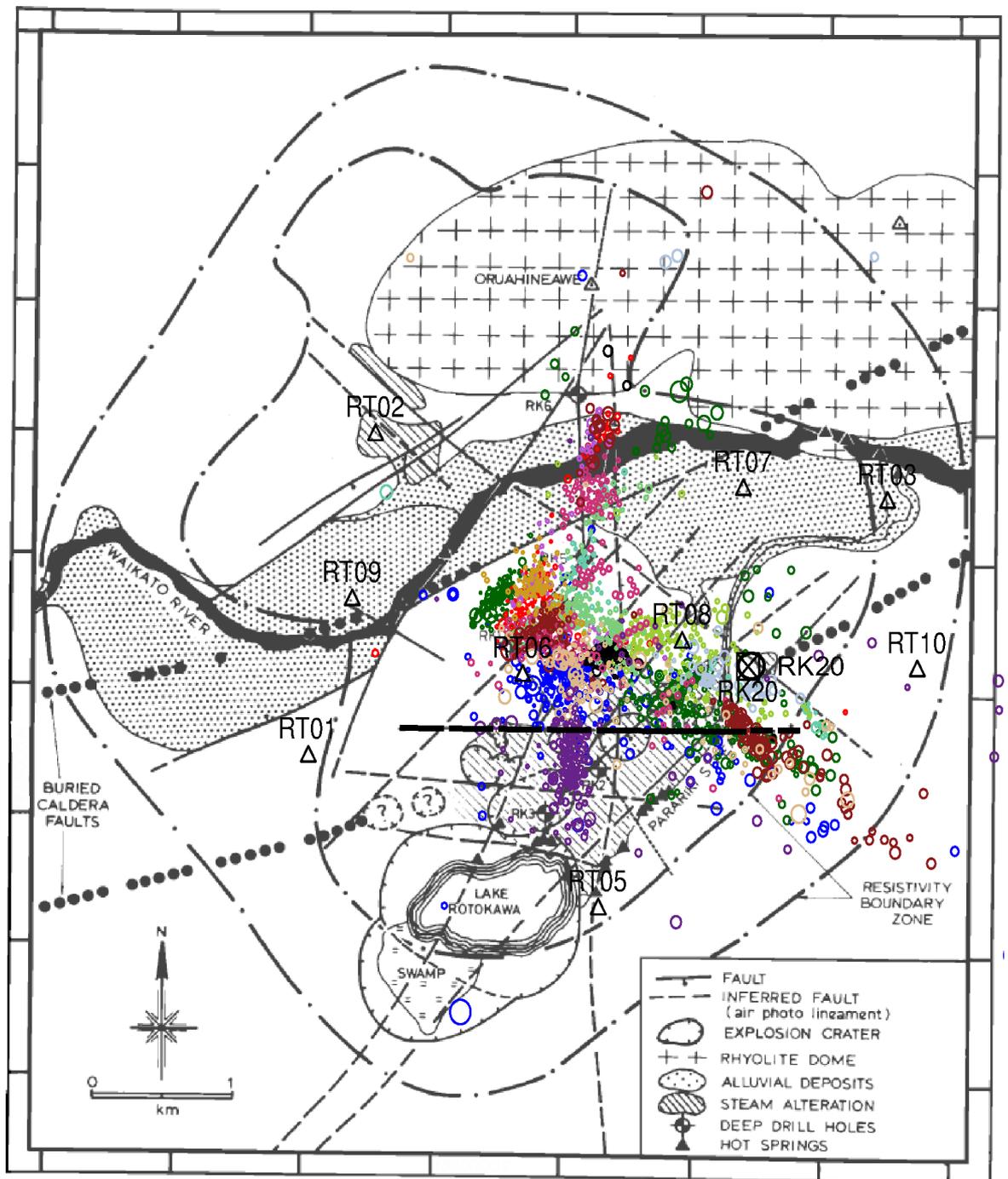


Figure 3.19: Geological map of Rotokawa with overlaid earthquakes. The bold black line is the inferred fault we believe may coincide with the event set of 290. Earthquakes are coloured based on their respective master event and scaled based on uncertainty as per Fig. 3.17. Figure modified from *Krupp and Seward [1987]*.

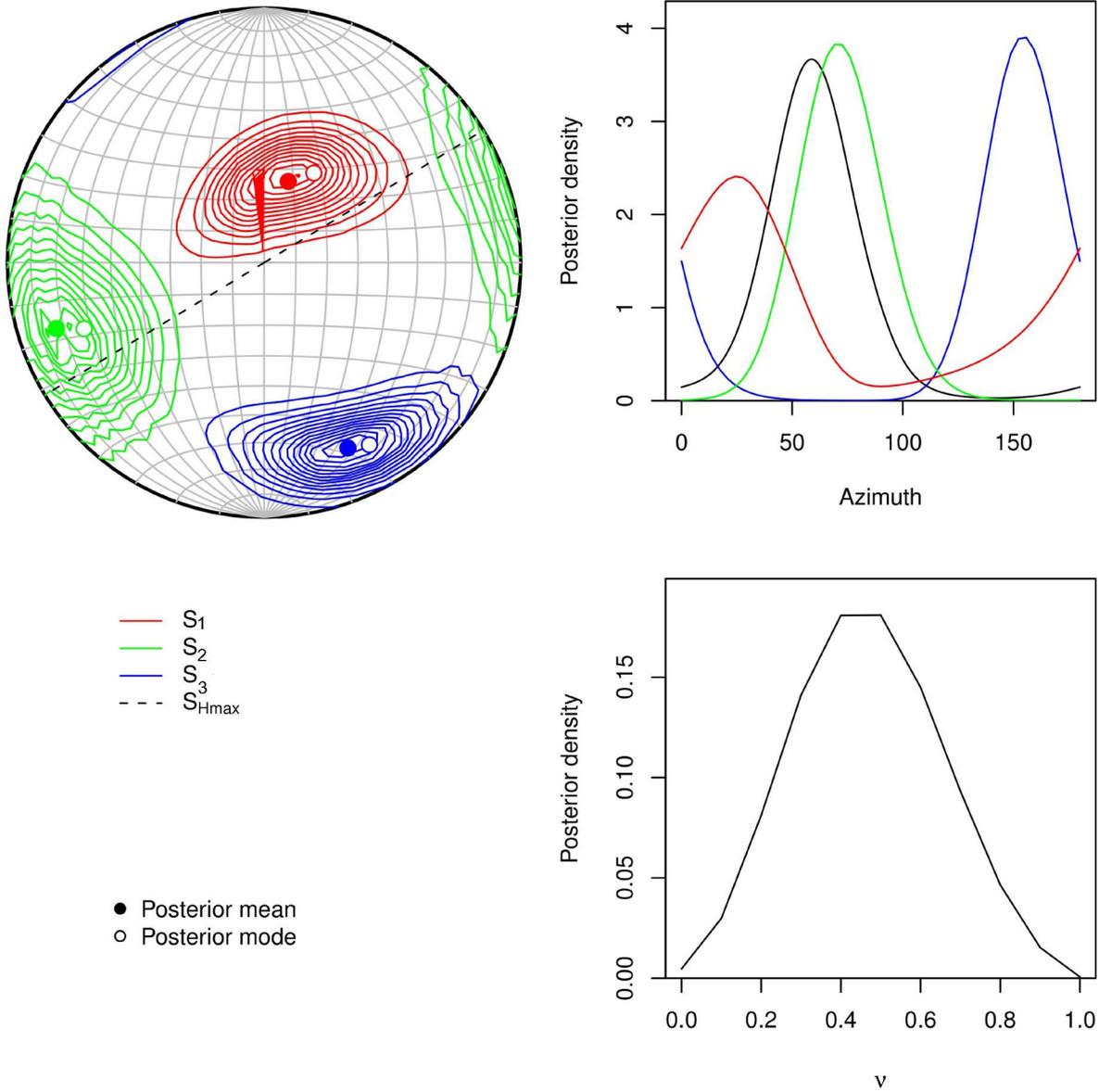


Figure 3.20: Details of the stress inversion results obtained using the method of *Arnold and Townend* [2007] applied to the suite of focal mechanism solutions listed in Table 3.3. v is the stress ratio given by $(S_1 - S_2) / (S_1 - S_3)$.

Details of the stress inversion are displayed in Fig. 3.20 and the prevailing orientation of S_{Hmax} (northeast–southwest) is displayed in the bottom-right corner of Fig. 3.18, with a wedge symbol spanning the 80% confidence interval of the estimate. The TVZ is a region dominated by northwest–southeast extension at an azimuth of $124 \pm 13^\circ$ [Darby and Meertens 1995] and a minimum principal stress trending 148° [Hurst et al. 2002]: these azimuths are consistent, within uncertainties, with the estimate here of northwest–southeast *minimum* principal stress trending 156° [Fig. 3.20].

The focal mechanisms at Rotokawa are consistent with the stress orientation deduced from regional focal mechanisms, indicating that the earthquakes at Rotokawa are occurring in response to regional stresses. The fact that (excluding slave events associated with master event 290) the seismicity pattern largely shows a lack of linear features indicative of earthquakes occurring on pre-existing faults suggests that the actual earthquakes may occur along well-distributed joints and/or fractures rather than recognisable major faults. This has been noted in previous studies of induced seismicity at geothermal fields [Eberhart-Phillips and Oppenheimer 1984; Phillips et al. 1997; Roff et al. 1996].

3.4.2 Relation to injection

The flow rate histories of injection wells at Rotokawa for the period spanned by the seismicity data are displayed in Fig. 3.21 [Steven Sewell, MRP, pers. comm., May 2011], as well as the number of earthquakes we have detected daily. Note that four injection wells were active during this period, and that as injection flow rates are largely constant there is no immediately obvious correlation between injection flow rates and seismicity rates.

Injection well RK16 is located close to station RT09, where little seismicity is observed, and we therefore focus below on the other three active injection wells. Fig. 3.22 displays the locations of the well tracks for RK11, RK12, and RK20. Note that RK20 is a deviated well and its three feed zones are located close to the depth of a large portion of earthquakes. RK11 and RK12 are part of an earlier injection scheme, when reinjection was into the shallow aquifer within the production region [Hunt and Bowyer 2007], they now only inject small

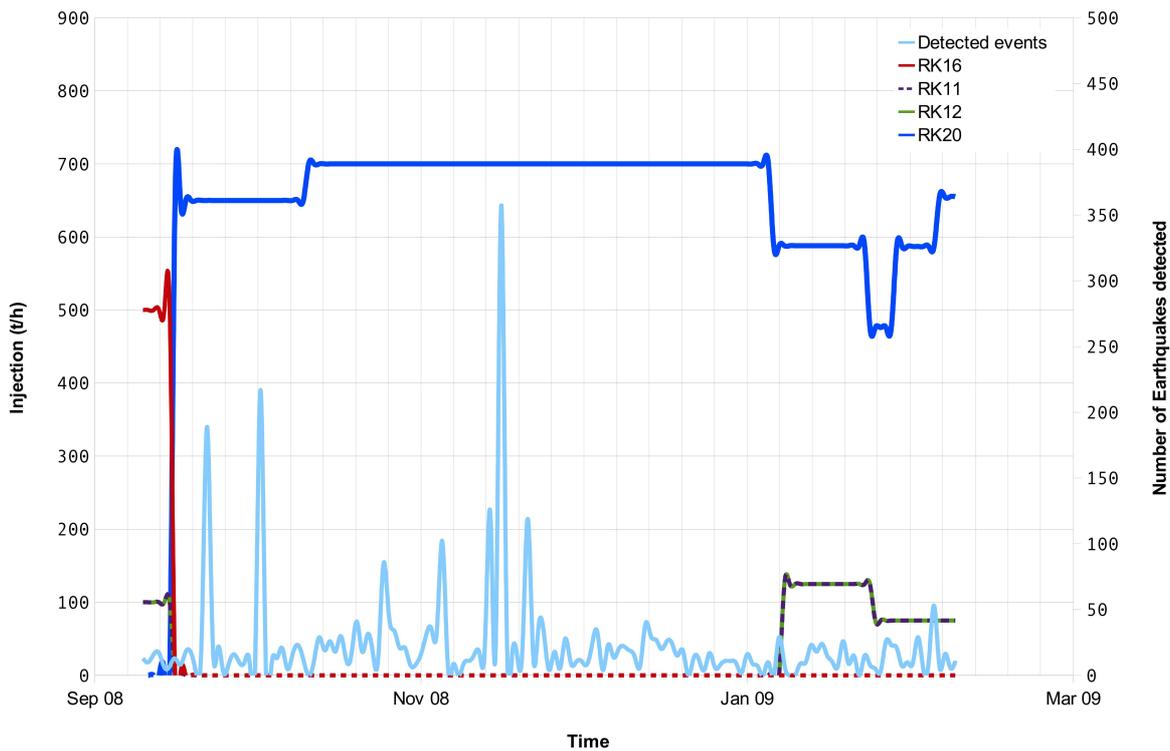


Figure 3.21: Injection flow rate history (in tonnes/hour) and daily number of detected earthquakes (light blue). Note that RK11 and RK12 have an identical flow rate history and so plot over top of each other. Data supplied by MRP [Steven Sewell, MRP, pers. comm., May 2011].

amounts to maintain pressures in this aquifer. They are therefore shallower wells and each have a gap of ~ 450 m between the well bottom and the dense region of earthquakes. In the horizontal plane, all three wells are located close to the dense region of earthquakes.

Earthquake hypocentre plots for different time periods (not shown) highlight different temporal patterns of seismicity between the region near RK11/12 and the region near RK20. Based on this, we focus on two separate regions displayed in Fig. 3.23. Each region is $300 \text{ m} \times 300 \text{ m} \times 5 \text{ km}$, the green region surrounds RK12, and the brown region surrounds the RK20 fluid injection points.

The temporal patterns of seismicity within these two regions are illustrated in Fig. 3.24. It is important to note that these changes are only indicative of the regions we have selected. Four points of change are noted, separating the plot into five time periods (hypocentre plots for each of these periods are presented in Appendix C.4). The first time period is the six

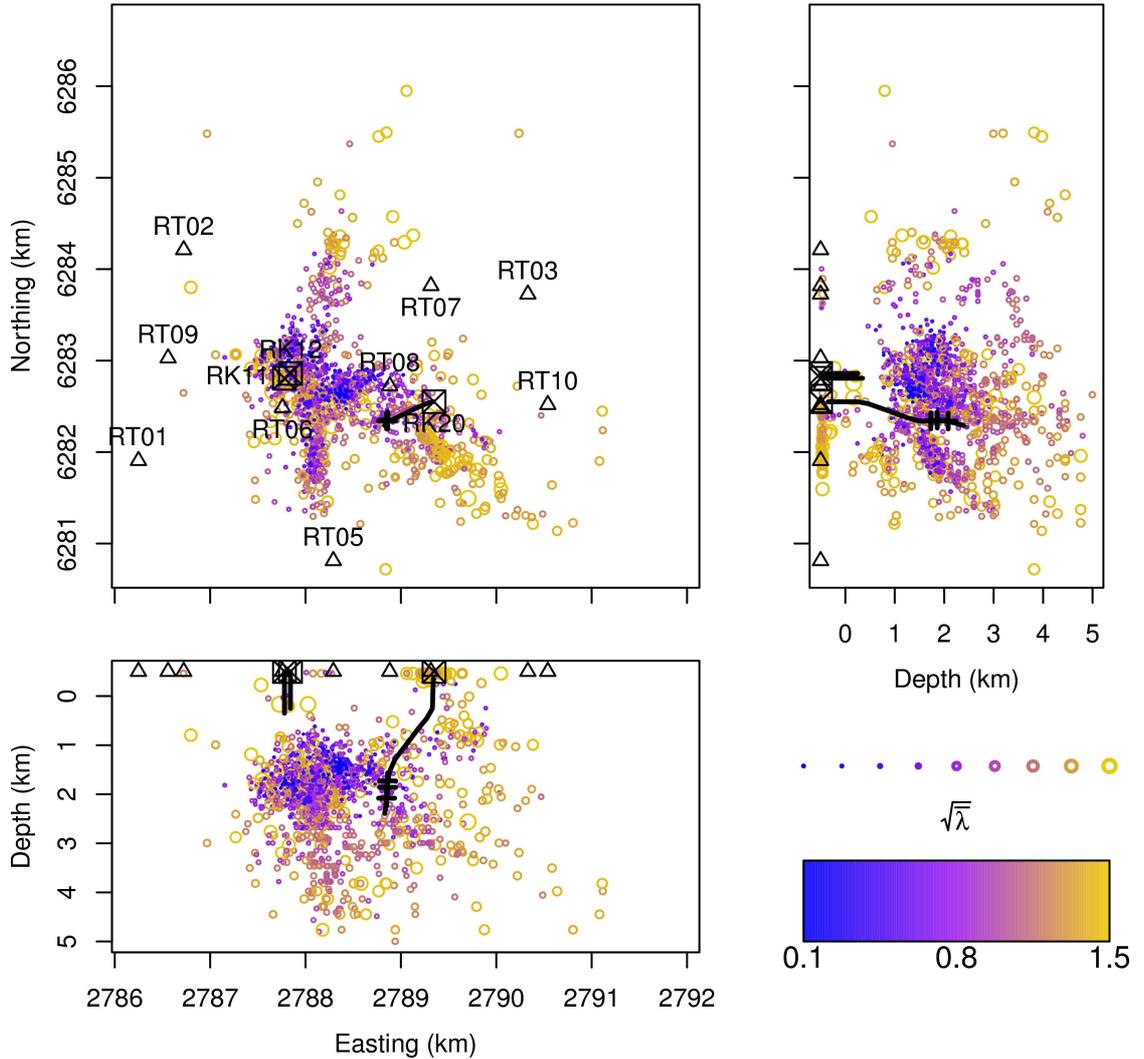


Figure 3.22: Earthquake locations and injection well tracks. Earthquake locations scaled and coloured by location uncertainty $\sqrt{\lambda}$. Black triangles are seismometers, and the black squares are the injection wells RK11, RK12 and RK20. Bold black lines display the well tracks, and three perpendicular lines identify the three feed zones of RK20.

day period before injection commenced at RK20, during which time RK11 and RK12 were both injecting. Note that there was no seismicity around RK20, and a moderate amount of seismicity around RK12. RK11 and RK12 then stopped and RK20 was initiated with a flow rate of 650 t/h (tonnes/hour). Shortly afterwards (~ 2 days), a minor amount of seismicity began in the RK20 region, and seismicity in the RK12 region increased. The next time period occurs after the flow rate at RK20 was increased to 700 t/h, an increase of $\sim 8\%$, and there was a significant increase in seismicity in the RK20 region, again with a delay of ~ 2 days. During this period, seismicity in the RK12 region appears to have decreased very slightly. A

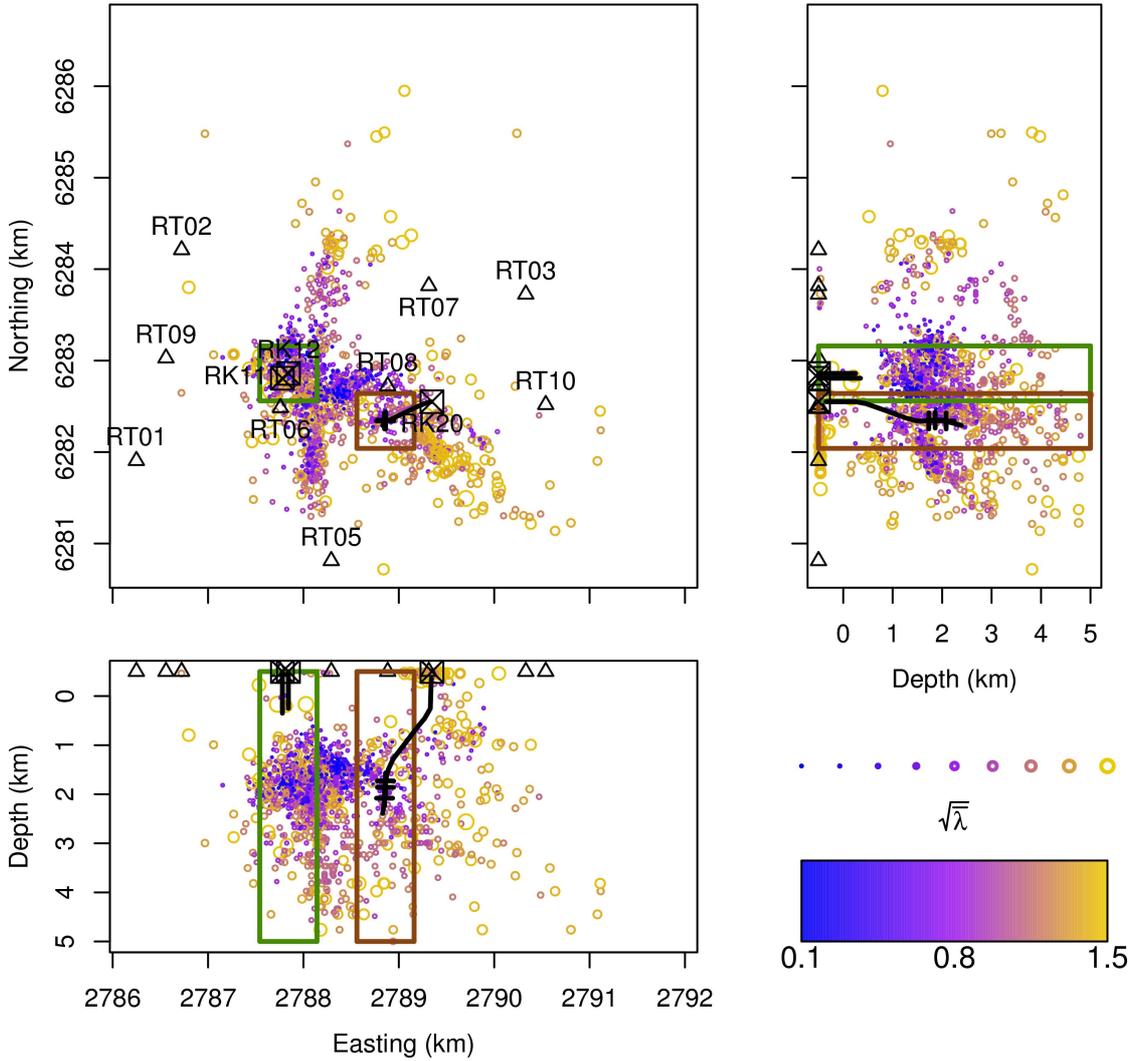


Figure 3.23: Two regions of earthquakes, each $300 \text{ m} \times 300 \text{ m} \times 5 \text{ km}$. The green square surrounds RK12, and the brown square surrounds the RK20 fluid entries. We investigate the different temporal patterns of seismicity in these two regions in Fig. 3.24. Earthquake locations scaled and coloured by location uncertainty $\sqrt{\lambda}$.

sudden change in seismic rate occurred on 11 November 2008 (day 75), ~ 10 weeks after the initiation of RK20, with no discernible cause. The final flow rate change noted occurs when the flow rate at RK20 dropped to 588 t/h (note that there are other changes in RK20's flow rate during this final period, see Fig. 3.24 for details) and RK11 and RK12 were reinitiated. Although the flow rate has dropped, the seismicity rate does not appear to have changed from the previous period's relative inactivity .

This pattern suggests that a flow rate of 650 t/h was not sufficient to cause failure on most fractures in the vicinity of RK20. Conversely, with an increase of only 8%, to a flow rate of

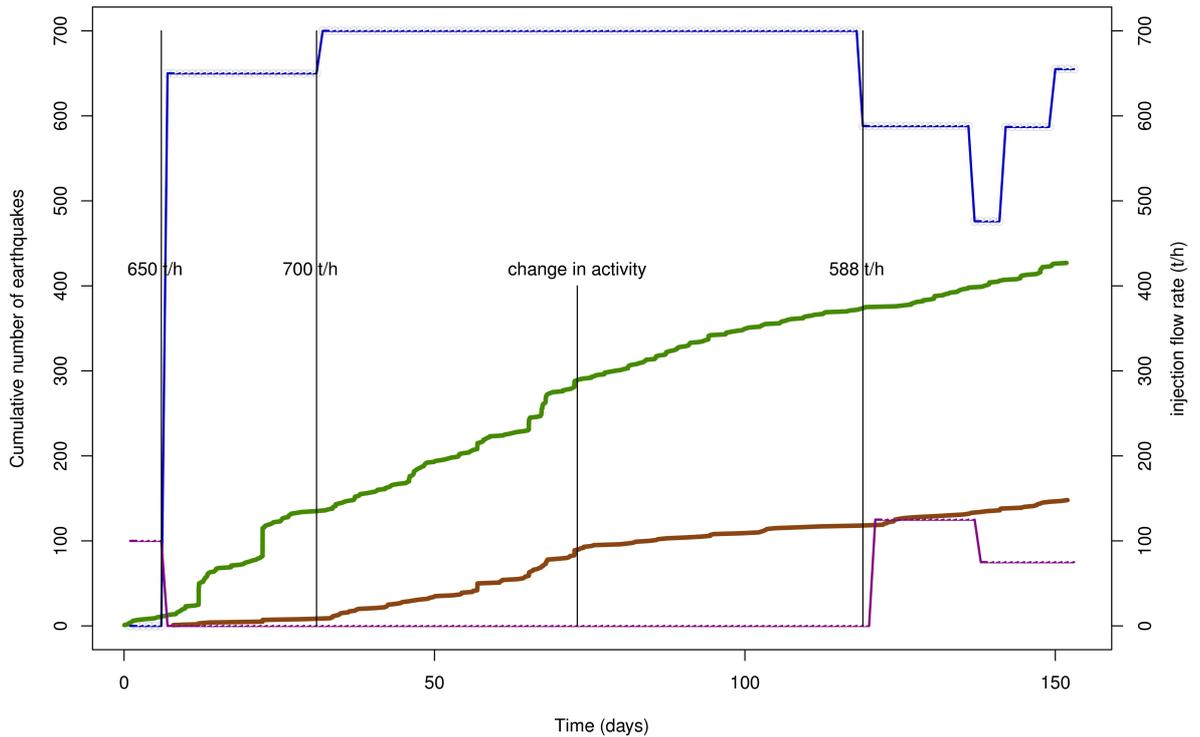


Figure 3.24: Temporal pattern of seismicity for the two regions of earthquakes identified in Fig. 3.23. Green corresponds to the RK12 region, and brown corresponds to the RK20 fluid entry region. The blue line corresponds to RK20 flow rates (in tonnes/hour), and the purple line to RK11/12 flow rates. Four points of change are noted, three of these correspond to changes in injection flow rate, with the flow rate identified. There is no discernible reason for the other changing point noted.

700 t/h, failure occurred on a significant number of fractures in the vicinity of RK20. One interpretation of the sudden decrease in the seismicity rate that occurred on 11 November 2008 is that, after 75 days of injection, the majority of critically-stressed fractures had slipped and the stress in the region had reached a relatively stable state.

We have investigated event distances from the RK20 injection well feed zones versus event times for all earthquakes, as well as for just earthquakes within the RK12 and RK20 regions identified above. We have found no correlation between distance and time (or the square root of time), suggesting there is no discernible diffusive migration of earthquakes as has been found in some other microearthquake studies [*Shapiro et al.* 1997; 2002; *Yukutake et al.* 2011].

One interpretation of this result is that low hydraulic impedance pathways are encountered at the fluid-entry points of RK20, through which fluid flows aseismically, and that

seismicity only occurs when fluid reaches high impedance pathways, which are distributed throughout the vicinity. *Cornet and Scotti* [1993] suggested that near the onset of injection, flow is likely to occur along paths with the lowest hydraulic impedance, *Scotti and Cornet* [1994] suggested that some hydraulically significant fractures are aseismic, and *Baria et al.* [2005] found events located some distance from the injection point after pumping had ceased, suggesting stress propagation mechanisms in relation to pressure gradients are important, not just the magnitude of the pressure.

Fig. 3.25 displays the earthquake locations with respect to the production and injection areas at Rotokawa. Most earthquakes occur within the production region with north–south and northwest–southeast-striking lineations. Part of the northwest–southeast-striking lineation leads southwest out of the injection area, but these earthquakes have large location uncertainties due to the seismic network geometry. If these are true locations it would suggest some of the injectate at RK20 returns to the production region whilst the rest heads southeast out of the field. However, the lineation feature may be a result of the gap in the network providing the location calculations less constraint in this direction and hence allowing spreading of the locations along the azimuth of least constraint. In this case, events may in fact cluster closer to RK20 and remain within the injection region. Although we have investigated seismicity in a box surrounding the RK12 injection well, the shallow depth of this well makes it unclear whether the earthquakes close by and elsewhere in the production region are injection-induced, production-induced, or occurring due to broader scale pore-pressure decreases/increases and temperature alterations within the production region. Nevertheless, it seems likely that the earthquakes in the injection area and the edge of the production region near RK20 are related to injection at RK20.

3.5 Summary

This chapter has outlined the steps used to refine phase-picks and obtain accurate hypocentres. This process has resulted in a final set of 2101 earthquakes with an average RMS time

residual of 0.11 s and a mean hypocentre uncertainty of 0.36 km. The dominant locus of seismicity, which lies at 1.0–2.5 km depth, is consistent with the location of the Rotokawa Andesite forming the Rotokawa reservoir [Fig. 1.3]. Focal mechanisms calculated for the 14 master events used for detection provide a stress estimate consistent with the northeast–southwest S_{Hmax} azimuth expected for the TVZ. Slave events for each respective master event are shown to generally cluster together, as expected for highly-correlated events. One

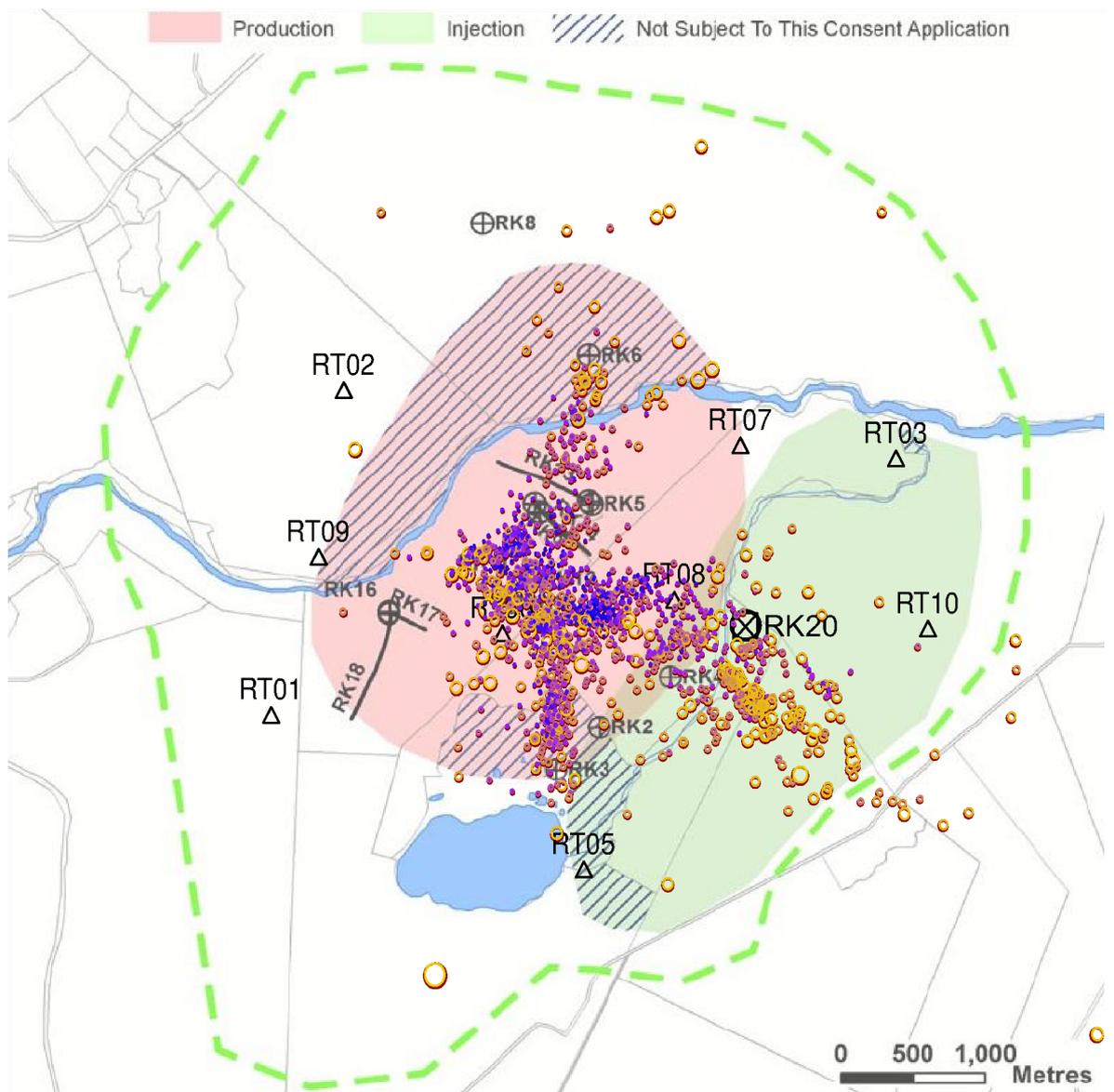


Figure 3.25: Production (pink) and injection (green) regions at Rotokawa with overlaid earthquakes. Earthquake locations are coloured and scaled based on uncertainty as per Fig. 3.23. Figure modified from *Grant* [2007].

cluster of slave events corresponding to master event 290 appears to delineate a previously inferred normal fault. Other seismicity is likely to be occurring on distributed joints and/or fractures within the reservoir. Seismicity in the region near the RK20 injection well feed zones shows a correlation with injection flow rates, with the rate of seismicity increasing as injection flow rates increase: this increase appears to have a ~ 2 day lag. There is a sudden decrease in seismicity after ~ 10 weeks of injection at RK20, perhaps due to the stress in the region reaching a relatively stable state. The seismicity located within the injection region and in the immediate vicinity of RK20 is likely to be related to injection and fluid flow pathways. This suggests some of the injectate at RK20 returns to the production region whilst the rest either heads southeast out of the field or remains to the southeast of RK20 within the injection region. An additional seismic station bridging the gap between RT05 and RT10 would be advantageous for determining RK20 near-vicinity earthquake locations with higher accuracy. The dominant locus of seismicity lies within the production region and it is unclear whether this seismicity is related to injection or production processes. Note that we do not have the production well histories to compare seismicity with.

Chapter 4

Statistical Experimental Design

Accurate earthquake locations underpin many facets of seismological analysis. The resolution obtainable from seismic data, however, is inherently circumscribed by the geometry of the network used to acquire the data. Here, we address the question of how to maximise this resolution by designing an optimal extension of an acquisition network. Given a set of expected earthquake locations and an existing acquisition network, we show how to develop and implement an algorithm that yields an optimal seismic network extension, where optimality is defined in terms of the network configuration that maximises earthquake location resolution.

Note that here, and throughout this chapter, we use the term ‘resolution’ to mean spatial resolution (not the resolution matrix from inverse theory [*Nolet* 2008]), which corresponds to the spatial extent of an earthquake’s location uncertainty. We endeavour to locate earthquakes with the highest accuracy, or lowest spatial uncertainty. This means having the finest spatial resolution over an area of interest in which earthquakes occur. Each earthquake provided as input to our algorithm represents a point in space. We use these point’s spatial resolutions’ as a representative sample to measure the spatial resolution of the entire region of interest.

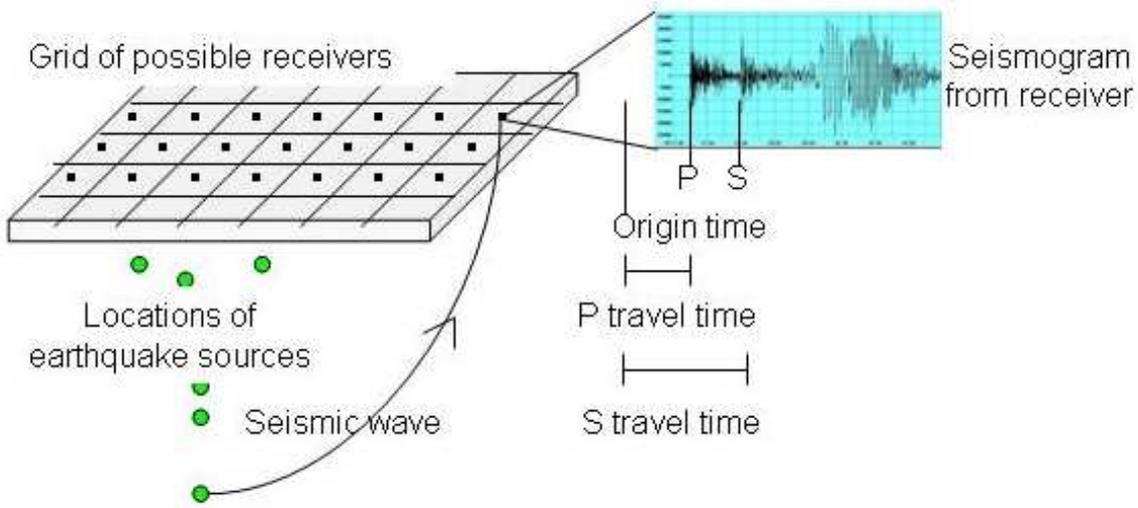


Figure 4.1: Cartoon of our network design problem. Note that *travel time* is the key data link in determining earthquake locations from seismograms provided by receivers.

4.1 Derivation of the design criterion

We begin with a region of possible receiver locations. We are interested in collecting seismic data in order to reduce earthquake location uncertainties. To do this, we need to identify the areas in which earthquakes of interest are expected to occur in order to determine which potential receivers would provide the best resolution for these areas. The network design produced is only optimal for the set of earthquakes provided, or the areas of interest identified. In practice, receivers constrain earthquake locations using P and S wave arrivals. The measurement of arrival times and subsequent estimation of origin times are attempts at measuring travel times accurately. In our application, we treat the input earthquake locations as perfectly known, and hence can ignore origin time errors and deal only with travel times as our data. It is these travel times, that constitute our data, which we wish to measure with the highest resolution [Fig. 4.1]. In this study, we consider both P and S travel times, but for the derivation of our design criterion we first focus on only using P travel times as our data.

Assume a seismic experiment is carried out using a survey design described by the vector ξ (a set of m station locations): travel time data are collected, which we represent by the vector \mathbf{d} (a set of m travel times for a single event), and these data are used to constrain a model of earthquake location, θ . If we let ω represent other earth model parameters for

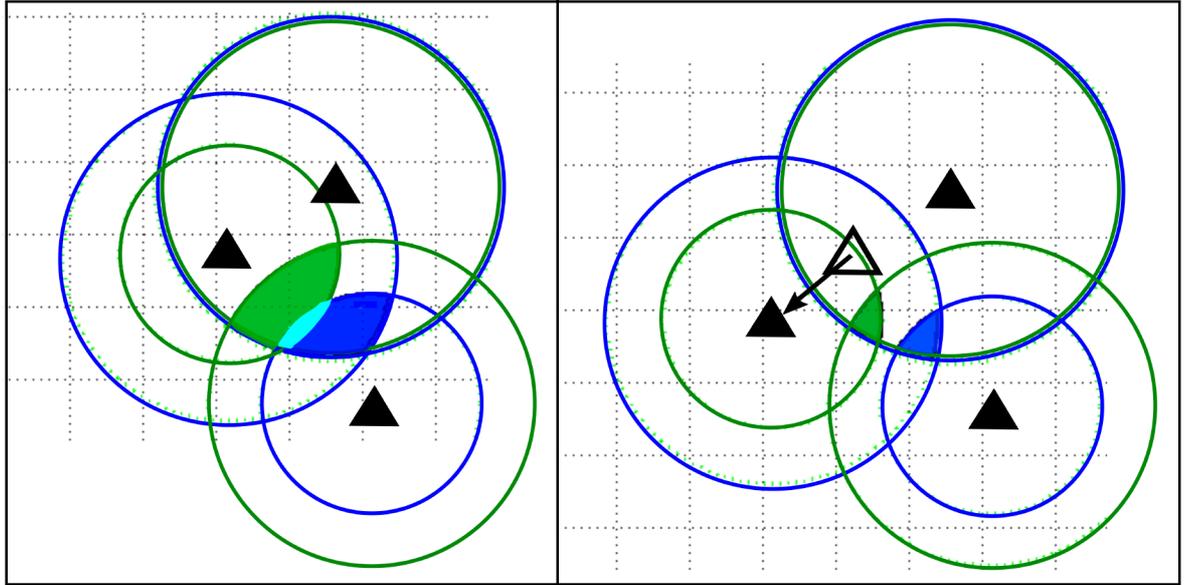


Figure 4.2: Schematic illustration of the epicentral Probability Density Functions (PDFs) ($p(\mathbf{d}|\boldsymbol{\theta}\boldsymbol{\omega}\boldsymbol{\xi})$) of two earthquakes constrained by P-wave travel time measurements ($\mathbf{d}|\boldsymbol{\theta}\boldsymbol{\omega}\boldsymbol{\xi}$) at three stations (black triangles). The blue and green circles display the outer uncertainty bounds of the travel times (the inner uncertainty bounds are omitted for clarity). The left-hand diagram displays a case in which the PDFs of the two earthquakes overlap (light blue region), and the locations are not clearly resolved. The right-hand diagram shows that by shifting only one station — and thereby increasing the variance of the expected PDF values $\Sigma_{g\xi}$ while maintaining constant data uncertainties Σ_{ξ} — the earthquake locations are resolvable. Greater resolution is achieved when the data uncertainty variance is smaller with respect to the location variance ($\Sigma_{g\xi} \gg \Sigma_{\xi}$). Fig. 4.3 represents these two cases in terms of the notation used to derive a design criterion.

which we have *a priori* estimates, such as a velocity model, then the earthquake hypocentre is dependent on $\boldsymbol{\omega}$, and the data \mathbf{d} are dependent on the three variables, $\boldsymbol{\theta}\boldsymbol{\omega}\boldsymbol{\xi}$. Table 4.1 summarises the nomenclature used in this section.

Section 1.3.2 contains an account of previous work in this area. The algorithm we use describes data of the highest resolution as those data that best resolve/constrain earthquake locations: therefore, we wish to minimise the degree to which the data distributions, $p(\mathbf{d}|\boldsymbol{\theta}\boldsymbol{\omega}\boldsymbol{\xi})$, overlap for all $\boldsymbol{\theta}$ [after *Coles and Curtis 2011*]. In other words, given locations $\boldsymbol{\theta}_i$ and $\boldsymbol{\theta}_j$ corresponding to data $p(\mathbf{d}_i|\boldsymbol{\theta}_i\boldsymbol{\omega}\boldsymbol{\xi})$ and $p(\mathbf{d}_j|\boldsymbol{\theta}_j\boldsymbol{\omega}\boldsymbol{\xi})$, we wish these distributions to be sufficiently different that $\boldsymbol{\theta}_i$ and $\boldsymbol{\theta}_j$ can be resolved from each other. See Figs. 4.2 and 4.3 for visual depictions of this idea.

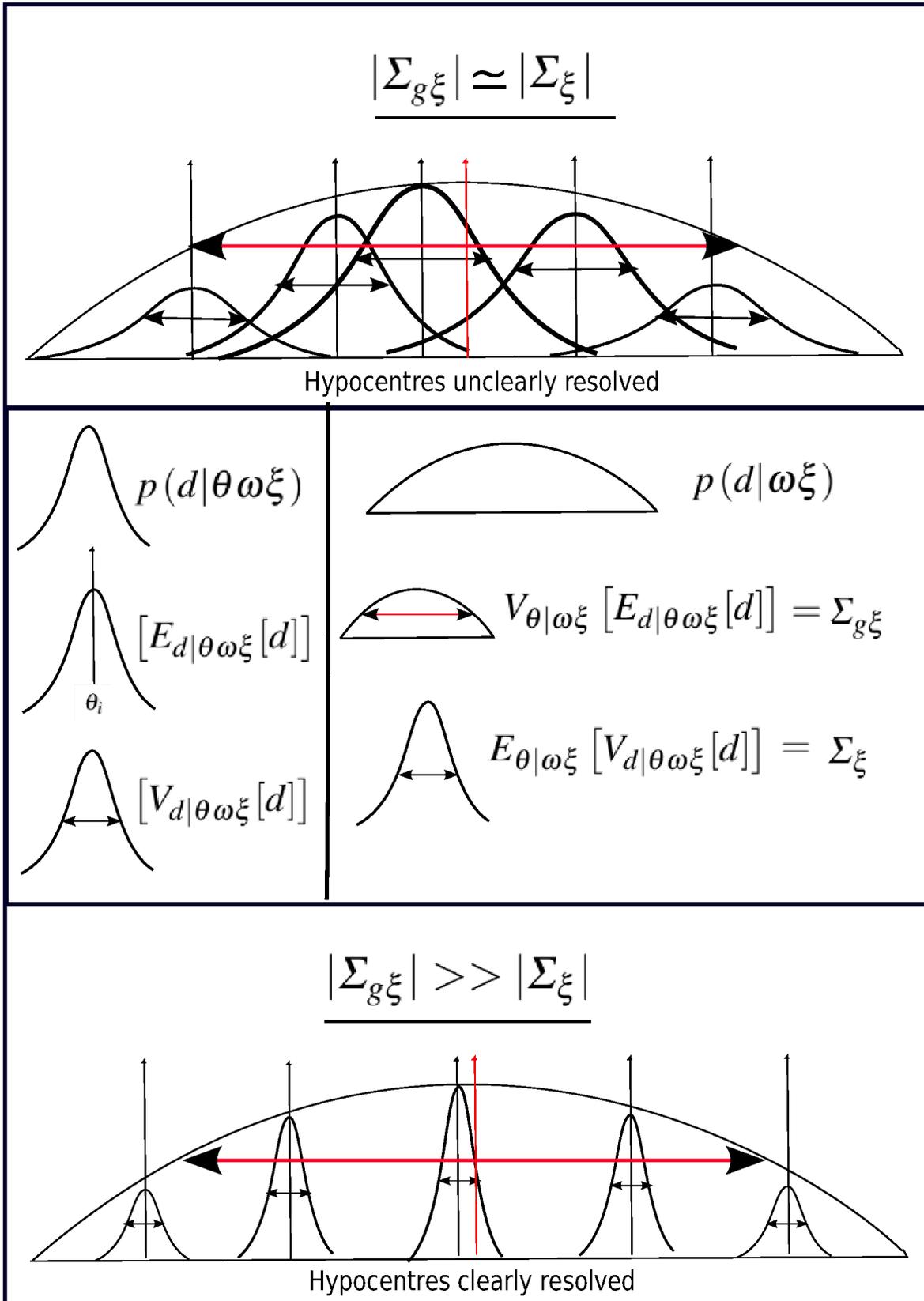


Figure 4.3: Depiction of the basis of *Coles and Curtis* [2011]’s optimal network design criterion, but using the nomenclature employed in this study. The goal of the criterion is to maximise hypocentre resolution (cf. Fig. 4.2).

Our data $\mathbf{d}|\boldsymbol{\theta}\boldsymbol{\omega}\xi$ have an expected value, or mean, represented by¹,

$$E[\mathbf{d}|\boldsymbol{\theta}\boldsymbol{\omega}\xi] = \boldsymbol{\mu}_{\boldsymbol{\theta}\xi} \quad (m \times 1 \text{ vector}) \quad (4.1)$$

and a variance-covariance matrix,

$$V[\mathbf{d}|\boldsymbol{\theta}\boldsymbol{\omega}\xi] = \boldsymbol{\Sigma}_{\boldsymbol{\theta}\xi} \quad (m \times m \text{ matrix}) \quad (4.2)$$

Note that we omit the subscript $\boldsymbol{\omega}$ for clarity.

The experimental design is intended to be optimal for all hypocentres used in its computation. These hypocentres $\boldsymbol{\theta}$ are implemented as perfectly known discrete points. We can integrate over the distribution of earthquake hypocentres $p(\boldsymbol{\theta}|\boldsymbol{\omega})$ in order to gain a distribution independent of $\boldsymbol{\theta}$: this yields the marginal distribution of the data,

$$p(\mathbf{d}|\boldsymbol{\omega}\xi) = \int p(\mathbf{d}|\boldsymbol{\theta}\boldsymbol{\omega}\xi) p(\boldsymbol{\theta}|\boldsymbol{\omega}) d\boldsymbol{\theta} \quad (4.3)$$

¹Note that $\boldsymbol{\mu}_{\boldsymbol{\theta}\xi}$ is equivalent to $g(\boldsymbol{\theta}\boldsymbol{\omega}\xi)$ in the notation used by *Coles and Curtis* [2011].

Table 4.1: Nomenclature for network design

Symbol	Explanation
\mathbf{d}	data ($m \times 1$ vector)
$p(\mathbf{d} \boldsymbol{\theta}\boldsymbol{\omega}\xi)$	probabilistic distribution of data given $\boldsymbol{\theta}$, $\boldsymbol{\omega}$ and ξ
$p(\boldsymbol{\theta} \boldsymbol{\omega})$	<i>a priori</i> probabilistic distribution of hypocentres (given knowledge of $\boldsymbol{\omega}$)
$\boldsymbol{\theta}$	earthquake hypocentre
$\boldsymbol{\Sigma}_{\boldsymbol{\theta}\xi}$	variance of data \mathbf{d} ($m \times m$ matrix)
$\boldsymbol{\Sigma}_{g\xi} = V_{\boldsymbol{\theta} \boldsymbol{\omega}\xi}[\boldsymbol{\mu}_{\boldsymbol{\theta}\xi}]$	hypocentre variance
$\boldsymbol{\Sigma}_{\xi} = E_{\boldsymbol{\theta} \boldsymbol{\omega}\xi}[\boldsymbol{\Sigma}_{\boldsymbol{\theta}\xi}]$	measurement error variance
$\boldsymbol{\mu}_{\boldsymbol{\theta}\xi}$	mean of data \mathbf{d}
ξ	seismic network (fixed for each design iteration)
Φ_{ξ}	design criterion for the network ξ
$\boldsymbol{\omega}$	velocity model and other known parameters (<i>a priori</i> information)

with an expected value,

$$E[\mathbf{d}|\omega\xi] = \int E[\mathbf{d}|\boldsymbol{\theta}\omega\xi] p(\boldsymbol{\theta}|\omega) d\boldsymbol{\theta} \quad (4.4a)$$

$$= \int \boldsymbol{\mu}_{\boldsymbol{\theta}\xi} p(\boldsymbol{\theta}|\omega) d\boldsymbol{\theta} \quad (4.4b)$$

$$= \boldsymbol{\mu}_\xi \quad (4.4c)$$

from Eqs. 4.1 and 4.3. Similarly, the marginal distribution given by Eq. 4.3 has a variance that is equal to the mean of the data variances for all $\boldsymbol{\theta}$ plus the variance of the data means for all $\boldsymbol{\theta}$ (see Appendix D.1):

$$V[\mathbf{d}|\omega\xi] = E_{\boldsymbol{\theta}|\omega\xi}[V[\mathbf{d}|\boldsymbol{\theta}\omega\xi]] + V_{\boldsymbol{\theta}|\omega\xi}[E[\mathbf{d}|\boldsymbol{\theta}\omega\xi]] \quad (4.5a)$$

$$= E_{\boldsymbol{\theta}|\omega\xi}[\boldsymbol{\Sigma}_{\boldsymbol{\theta}\xi}] + V_{\boldsymbol{\theta}|\omega\xi}[\boldsymbol{\mu}_{\boldsymbol{\theta}\xi}] \quad (4.5b)$$

$$= E_{\boldsymbol{\theta}|\omega\xi}[\boldsymbol{\Sigma}_{\boldsymbol{\theta}\xi}] + \boldsymbol{\Sigma}_{g\xi} \quad (4.5c)$$

$$= \boldsymbol{\Sigma}_\xi + \boldsymbol{\Sigma}_{g\xi} \quad (4.5d)$$

Here $\boldsymbol{\Sigma}_\xi$ is the measurement error variance and $\boldsymbol{\Sigma}_{g\xi}$ is the hypocentre variance [Fig. 4.3]. For maximum discrimination between hypocentres we require [Fig. 4.3],

$$|\boldsymbol{\Sigma}_{g\xi}| \gg |\boldsymbol{\Sigma}_\xi| \quad (4.6)$$

and therefore need to maximise,

$$\Phi_\xi = \log\left(\frac{|\boldsymbol{\Sigma}_{g\xi}|}{|\boldsymbol{\Sigma}_\xi|}\right) = \log(|\boldsymbol{\Sigma}_{g\xi}|) - \log(|\boldsymbol{\Sigma}_\xi|) \quad (4.7)$$

This is equivalent to the D_N -Criterion used by *Coles and Curtis* [2011], but we have derived it using a different method: namely, using only the first and second moments of the distributions (the mean and variance), without needing to assume anything about the detail of those distributions, without needing to assume that $V_{\mathbf{d}|\boldsymbol{\theta}\omega\xi}[\mathbf{d}]$ is fixed, and without the explicit use of the Kullback-Leibler divergence [*Cover and Thomas* 1991] — which, as dis-

cussed in Chapter 1, is a measure of the distance between individual hypocentre probability distributions.

The experimental design criterion represented by Eq. 4.7 is intended to ensure that the data produced at all stations by an earthquake location provide the minimum amount of mutual information with other earthquake locations, so that locations are clearly resolved [Figs. 4.2 and 4.3]. This requires the testing of different experimental designs (different configurations of station locations) and choosing the design with the best result.

The rest of this chapter describes our development of an algorithm implementing this criterion, and its application to the Kawerau geothermal field. We also describe a brief application of this algorithm to the 2010 Darfield Earthquake in Appendix D.3. In the Kawerau case, we start with six existing seismometer locations and consider how best to add up to ten more stations [Steven Sewell, MRP, pers. comm. 2010]. We begin with the network of existing stations and test all possible locations for the best location of one additional station. This is iterated until all ten best locations for new stations are found.

This method of iteratively adding one new station at a time is similar to the approach taken by *Curtis et al.* [2004]. Other methods that have been used in statistical seismic network design require initial designs that evolve into improved designs via minor adjustments [*Rabinowitz and Steinberg* 1990; *Steinberg et al.* 1995] or use stochastic optimisation methods such as genetic algorithms or simulated annealing to identify good designs [*Curtis* 1999; *Maurer and Boerner* 1998]. The first method suffers from the final design being strongly dependent on the initial design model, whilst the second method, due to its stochastic nature, does not guarantee a repeatable result. Both methods require substantial computing time and resources to be able to test a sufficiently large number of designs to provide robust results. Though our method does not guarantee a globally optimal result, it does provide a repeatable result, quick processing times, a result influenced by all possible seismometer locations (rather than a subset of locations as in the other two methods), and the production of a network design that always performs well. This is demonstrated with the construction of the expected triangular quadripartite designs during initial testing in Section 4.2. The method

also allows for the case where, due to nancial or logistical constraints for example, fewer seismometers are deployed than anticipated: this scenario does not now require an entire overhaul of the design (as would be the case with the other two methods discussed), but, by simply deploying only the first few stations, network optimality is still approached.

The design criterion can also obviously be used to test for the best network design if a number of different designs (composed of the same number of seismometers) are to be compared, as demonstrated by *Coles and Curtis* [2011].

4.2 Implementation and preliminary testing

We implement the maximisation of the design criterion in an *R* programme [*Becker et al.* 1988]. First, we demonstrate that a generalised example produces results consistent with the findings of [*Uhrhammer* 1980, Fig. 1.9]: specifically, in a simple case the optimal configuration for a seismic network is triangular quadripartite [Fig. 4.4].

Data are produced in this example via the following steps, with the mathematical nomenclature summarised in Table 4.2: We calculate the travel time $t_j(\boldsymbol{\theta})$ from an earthquake $\boldsymbol{\theta}$ to a station j given a uniform velocity model. The signal from this earthquake is observed on a seismogram and its arrival time picks have an uncertainty simulated by a background noise level introducing the pick-time uncertainty ε_j and scaled by the fraction of signal amplitude lost to attenuation from earthquake $\boldsymbol{\theta}$ as discussed below. This implies that the uncertainty in each travel time pick increases as the signal-to-noise ratio (SNR) decreases (attenuation increases).

We calculate this amount of attenuation using the following expression for the amplitude of a wave with a dominant frequency f travelling through a material of velocity v and quality factor Q (the inverse of attenuation), as a function of distance x , and as a fraction of its original amplitude u_0 [*Lay and Wallace* 1995]

$$u(x) = u_0 \exp\left(-\frac{xf\pi}{Qv}\right) \quad (4.8a)$$

Setting $u_0 = 1$ at station j this becomes

$$u_j(\boldsymbol{\theta}) = \exp\left(-\frac{t_j(\boldsymbol{\theta})f\pi}{Q}\right) \quad (4.9)$$

The dominant frequency of the Rotokawa microearthquakes is $f = 10$ Hz, and Q can be varied to investigate different attenuation levels. We model attenuation as the fraction of signal amplitude lost along the event–station path: $a_j(\boldsymbol{\theta}) = 1 - u_j(\boldsymbol{\theta})$, where $a_j(\boldsymbol{\theta}) = 1$ corresponds to complete attenuation, and $a_j(\boldsymbol{\theta}) = 0$ corresponds to zero attenuation.

The background noise process uncertainty ε_j is assumed to be independently and identically distributed (*iid*) with zero mean and constant variance: $\varepsilon_j \sim iid(0, \sigma_\varepsilon^2)$ for all j . As σ_ε is the same constant at every location, it simply acts as a scaling factor, and is set to $\sigma_\varepsilon = 1$ for convenience. The noise process uncertainty is then scaled by the attenuation. For an

Table 4.2: Nomenclature for network design implementation

Symbol	Explanation
$a_j(\boldsymbol{\theta})$	fraction of signal amplitude lost to attenuation from earthquake $\boldsymbol{\theta}$ at station j
$\mathbf{A}(\boldsymbol{\theta})$	diagonal matrix of values $1/(1 - a_j(\boldsymbol{\theta}))$ for $j = 1, 2, \dots, m$
d_j	data at station j
f	frequency
$j = 1, 2, \dots, m$	stations
Q	quality factor
$t_j(\boldsymbol{\theta})$	travel time from earthquake $\boldsymbol{\theta}$ to station j
u_0	initial amplitude of an earthquake signal
$u_j(\boldsymbol{\theta})$	fraction of signal amplitude remaining from earthquake $\boldsymbol{\theta}$ at station j
v	velocity
ε_j	pick time uncertainty at station j
$\boldsymbol{\theta} = 1, 2, \dots, n$	earthquake hypocentres
σ_ε	variance of ε_j
$\Sigma_{\boldsymbol{\theta}\xi}$	variance of data \boldsymbol{d}
$\Sigma_{g\xi} = V_{\boldsymbol{\theta} \omega\xi} [\boldsymbol{\mu}_{\boldsymbol{\theta}\xi}]$	hypocentre variance
$\Sigma_\xi = E_{\boldsymbol{\theta} \omega\xi} [\Sigma_{\boldsymbol{\theta}\xi}]$	measurement error variance
ξ	seismic network (fixed for each design iteration)
Φ_ξ	design criterion for the network ξ
ω	velocity model and other known parameters (<i>a priori</i> information)

earthquake θ at station j the arrival time data are thus

$$d_j = t_j(\theta) + \frac{\varepsilon_j}{1 - a_j(\theta)} \quad (4.10)$$

Here, $t_j(\theta)$ and ε_j are in units of time, whilst $a_j(\theta)$ is a dimensionless fraction which acts as a scaling factor to simulate the effect of amplitude variation on the background noise uncertainty ε_j . For all stations $j = 1, 2, \dots, m$, where $j = 1, 2, \dots, m - 1$ are existing stations, and $j = m$ is a potential station location to be evaluated, this yields the $m \times 1$ vector of travel times,

$$\mathbf{d} = \mathbf{t}(\theta) + \mathbf{A}(\theta)\varepsilon \quad (4.11)$$

where $\mathbf{t}(\theta)$ is the $m \times 1$ vector of (noiseless) travel times and $\mathbf{A}(\theta) = \text{diag}(1/(1 - a_j(\theta)))$ is an $m \times m$ matrix. Since $\varepsilon_j \sim iid(0, \sigma_\varepsilon^2)$, the mean of the data vector is

$$E[\mathbf{d}|\theta] = \mathbf{t}(\theta) + \mathbf{A}(\theta) \times 0 \quad (4.12a)$$

$$= \mathbf{t}(\theta) \quad (4.12b)$$

and the variance of the data is

$$V[\mathbf{d}|\theta] = \mathbf{A}(\theta)V[\varepsilon]\mathbf{A}(\theta)^T \quad (4.13a)$$

$$= \mathbf{A}(\theta)\sigma_\varepsilon^2\mathbf{I}\mathbf{A}(\theta) \quad (4.13b)$$

$$= \sigma_\varepsilon^2\mathbf{A}(\theta)\mathbf{A}(\theta) \quad (4.13c)$$

(noting that $\mathbf{A}(\theta) = \mathbf{A}(\theta)^T$). We are now in a position to calculate the variance of the marginal distribution,

$$V[\mathbf{d}] = E_\theta[V[\mathbf{d}|\theta]] + V_\theta[E[\mathbf{d}|\theta]] \quad (4.14a)$$

$$= E_\theta[\sigma_\varepsilon^2\mathbf{A}(\theta)\mathbf{A}(\theta)] + V_\theta[\mathbf{t}(\theta)] \quad (4.14b)$$

$$= \sigma_\varepsilon^2 E_\theta[\mathbf{A}(\theta)\mathbf{A}(\theta)] + V_\theta[\mathbf{t}(\theta)] \quad (4.14c)$$

Here, the first term yields the measurement error variance Σ_ξ [Eq. 4.5]. Approximating the expectation with an average of $\mathbf{A}(\boldsymbol{\theta}_i)$ over the n earthquake observations we obtain

$$\Sigma_\xi = \sigma_\varepsilon^2 E_\theta [\mathbf{A}(\boldsymbol{\theta})\mathbf{A}(\boldsymbol{\theta})] \quad (4.15a)$$

$$= \frac{\sigma_\varepsilon^2}{n} \sum_i \mathbf{A}(\boldsymbol{\theta}_i)\mathbf{A}(\boldsymbol{\theta}_i) \quad (4.15b)$$

$$= \frac{\sigma_\varepsilon^2}{n} \sum_i \left[\text{diag} \left(\frac{1}{1 - \mathbf{a}_j(\boldsymbol{\theta}_i)} \right) \cdot \text{diag} \left(\frac{1}{1 - \mathbf{a}_j(\boldsymbol{\theta}_i)} \right) \right] \quad (4.15c)$$

$$= \sigma_\varepsilon^2 \text{diag} \left(\frac{1}{n} \sum_i \frac{1}{[1 - \mathbf{a}_j(\boldsymbol{\theta}_i)]^2} \right) \quad (4.15d)$$

where $\mathbf{a}_j(\boldsymbol{\theta}_i)$ is the attenuation of the signal from event i arriving at station j . Being diagonal, the measurement error variance matrix has a determinant given straightforwardly by

$$|\Sigma_\xi| = (\sigma_\varepsilon^2)^m \prod_{j=1}^m \left[\frac{1}{n} \sum_i \frac{1}{[1 - \mathbf{a}_j(\boldsymbol{\theta}_i)]^2} \right] \quad (4.16)$$

For computational purposes — in particular, to ensure than no station sees a (fully attenuated) zero signal from any earthquake — we actually implement $[1 - \mathbf{a}_j(\boldsymbol{\theta}_i)]^2$ as $[1.001 - \mathbf{a}_j(\boldsymbol{\theta}_i)]^2$ in Eq. 4.16.

The second term in Eq. 4.14 is the hypocentre variance (after Eq. 4.5), where $\mathbf{t}(\boldsymbol{\theta})$ is the $m \times 1$ vector of travel times from earthquake $\boldsymbol{\theta}$ to each station,

$$|\Sigma_{g\xi}| = |V_\theta[\mathbf{t}(\boldsymbol{\theta})]| \quad (4.17)$$

We estimate this by calculating the variance-covariance matrix of the observed set of vectors $\mathbf{t}(\boldsymbol{\theta})$.

$$\widehat{\Sigma}_{g\xi} = \frac{1}{n-1} \sum_{i=1}^n [\mathbf{t}(\boldsymbol{\theta}_i) - \bar{\mathbf{t}}(\boldsymbol{\theta})][\mathbf{t}(\boldsymbol{\theta}_i) - \bar{\mathbf{t}}(\boldsymbol{\theta})]^T \quad (4.18a)$$

$$\text{where} \quad \bar{\mathbf{t}}(\boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^n \mathbf{t}(\boldsymbol{\theta}_i) \quad (4.18b)$$

is the vector of mean travel times at each station. We can express the design criterion [Eq.

4.7] in this case as

$$\Phi_{\xi} = \log(|\Sigma_{g\xi}|) - \log(|\Sigma_{\xi}|) \quad (4.19)$$

We evaluate this criterion at each possible station location (every grid point in the model) by changing the location of station $j = m$. The location corresponding to the maximum criterion value is the optimal new station location. This location is stored as a new station and the method repeated until the desired number of new stations have been located.

Three simple examples of this method are shown in Figure 4.4. In each example, the coloured contours demonstrate the distribution of the criterion value Φ , with white corresponding to the best locations for a new station (high Φ), and dark green corresponding to the worst locations (low Φ). In this example, we begin with zero stations and show how the distribution of Φ changes with each additional station, until a triangular quadripartite configuration is formed [cf. *Uhrhammer* 1980]. For each example, 100 earthquake sources are generated in a small region in the centre of the grid. The top and middle panels have earthquake depths of 1–9 km, whilst the bottom panels have earthquake depths of 1–40 km. The top example has $Q = 300$, simulating low attenuation, and as a result the stations are placed at reasonable distances from the earthquakes. The middle example has $Q = 50$, simulating high attenuation, with the stations now restrained to lie close to the earthquake locations. The bottom example also has $Q = 50$, but the geometry imposed by the significantly deeper earthquakes has pushed the stations very far from the earthquakes. Even in such simple cases, it becomes obvious that not only should the horizontal distribution of earthquakes be taken into account when designing a seismic network, but the depth distribution and regional attenuation also have significant impacts on the optimal design.

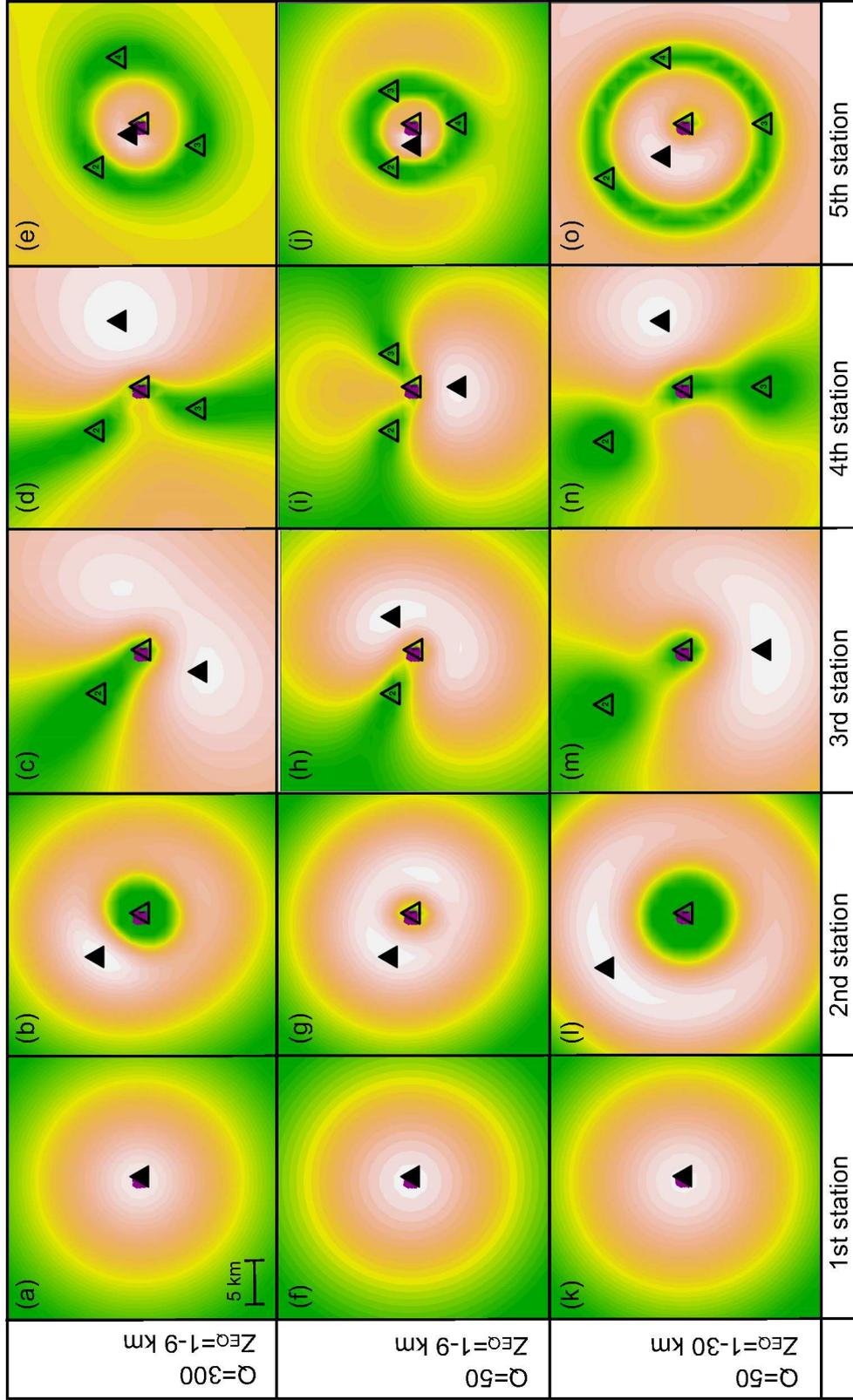


Figure 4.4: Simple synthetic network design testing with coloured contours showing the variation of the optimality factor Φ prior to the addition of the seismic station denoted by the filled triangle. White corresponds to the maximum Φ (best), and dark green corresponds to the minimum Φ (worst). Earthquakes, coloured magenta, are generated in a region 1×1 km in the centre of the $\sim 30 \times 30$ km grid. Existing stations are shown by the outlined triangles and are labelled by order of placement. Note that (e), (j) and (o) all show triangular quadripartite networks.

4.3 Application to Kawerau

We now extend the implementation to consider the existing network geometry at Kawerau. The earthquake catalogue used here contains earthquakes recorded in the two years since commission of the Kawerau power plant in June 2008 (31 June 2008 to 15 July 2010). We have divided this catalogue into three different regions of analysis [Fig. 4.5]. The production region earthquakes are of greatest interest to MRP with respect to production and injection impact and reservoir characterisation. We focus on this set to demonstrate how each addi-

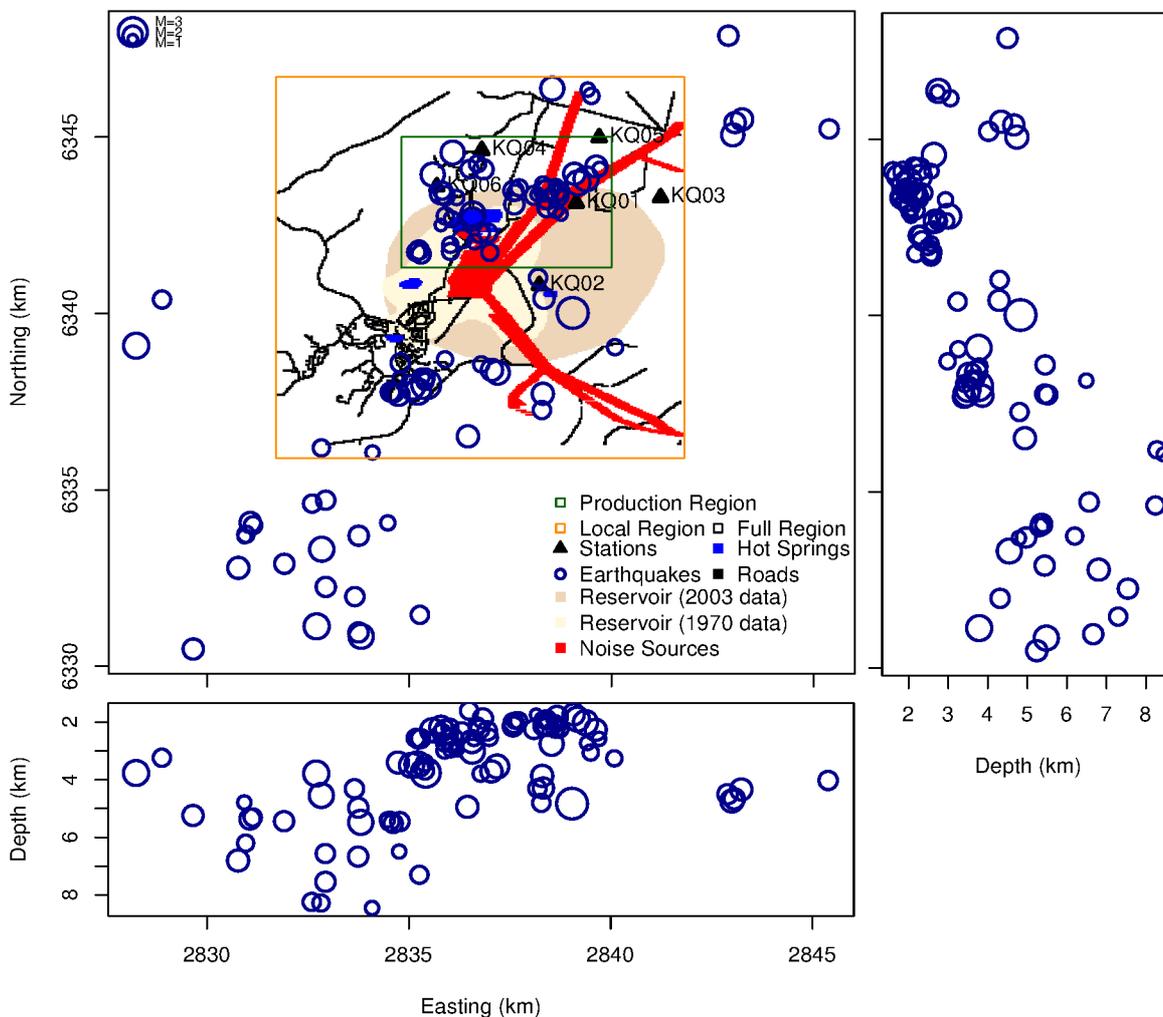


Figure 4.5: Three different earthquake regions used in Kawerau analysis. The 2003 reservoir displayed is the reservoir identified by the distribution of resistivity at a depth of 700 m from 3D-MT modelling in 2003 [Bignall and Harvey 2005], and the 1970 reservoir is that identified by the 1970 Schlumberger survey with AB/2=457 m [Bignall and Harvey 2005].

tional parameter employed in the design model affects the optimal network configuration produced.

The design parameters fall into two categories: 1) the seismological model and 2) surface conditions. The first relates to the path travelled by seismic waves from earthquakes of interest to station locations, whilst the second relates to station location alone. We construct our design model via the following steps: we begin with a uniform velocity and attenuation model [Section 4.3.1]; we then incorporate 3D P-wave velocity and attenuation models [Section 4.3.2]; then 3D S-wave velocity and attenuation models are incorporated [Section 4.3.2.4]; and finally, we introduce the surface conditions created by surface noise sources [Section 4.3.3]. In each case, we aim to calculate the network design ξ which maximises Φ_ξ , and hence maximises hypocentre resolution for the set of earthquakes of interest.

4.3.1 Uniform velocity case

We begin with a uniform velocity model and quality factor Q representing attenuation. The earthquakes in the production region are all less than 3 km deep, and we therefore adopt a P-wave velocity of $V_p=4.6$ km/s based on the results of *Clarke* [2008] and vary Q to investigate the effects of attenuation. In this simple case, the value of the velocity does not in fact matter: only the geometry of the earthquakes and existing stations, and the amount of attenuation affect the station configuration. Two networks produced from different Q values are shown in Fig. 4.6, with a comparison of the results displayed in Fig. 4.7. Fig. 4.6(a) displays the results of a low attenuation model, $Q = 300$ (orange triangles in Fig. 4.7), and (b) displays a high attenuation model, $Q = 50$ (red triangles in Fig. 4.7). The stations are labelled 1–10 in order of importance. Both networks work to surround the earthquake locations, with the configurations stretched in the direction of the long axis of the earthquake domain (northeast–southwest). There are three zones identified as high priority locations which are intuitively expected: as existing stations cluster in the northeast, the best location is to the southwest, which will help constrain earthquake location analysis in the horizontal plane; the second best location is to the north-northwest, as earthquakes to this side are not fully enclosed

by the existing station geometry; thirdly, stations 4–5 in the low attenuation model, and 3–4 in the high attenuation model, cover the centre of the earthquake cluster, which is a good location for constraining earthquake depths. Note that geometrically speaking the two networks evolve along similar lines as stations are added: however, in the low-attenuation model, stations are generally pushed further from the earthquake locations, and there has been some rearrangement in the ordering of importance of similar locations (5 and 3 have swapped places, 10 has shifted west).

It is important when implementing this algorithm that the grid search space used is large enough that edge effects can be avoided, and that the grid sampling rate used is small enough to not create artifacts (we sample every 300 m). The images shown in Fig. 4.6 are actually enlarged sections of the full search space shown in Fig. 4.8(a). If the search space is not set large enough, stations may be pushed towards the edges of the model as in Fig. 4.8(b), and the location of this boundary may influence the network design produced. Another way of determining whether the search space is too small is to examine whether the stations are clustering very close together (generally in the corners of the grid) in the design results.

4.3.2 Seismological model

4.3.2.1 Velocity model

Within the Kawerau region as a whole we identify three different velocity regions. A schematic of these regions is shown in Fig. 4.9: Region 1 lies outside the geothermal reservoir; Region 2 represents the reservoir identified by the distribution of resistivity at a depth of 700 m from 3D-MT modelling in 2003 [Bignall and Harvey 2005]; and Region 3 is the reservoir identified by the 1970 Schlumberger survey with $AB/2=457$ m [Bignall and Harvey 2005].

The velocity models for each region are displayed graphically in Fig. 4.10, with the corresponding values listed in Tables 4.3–4.5. The model for Region 1 is the optimal velocity

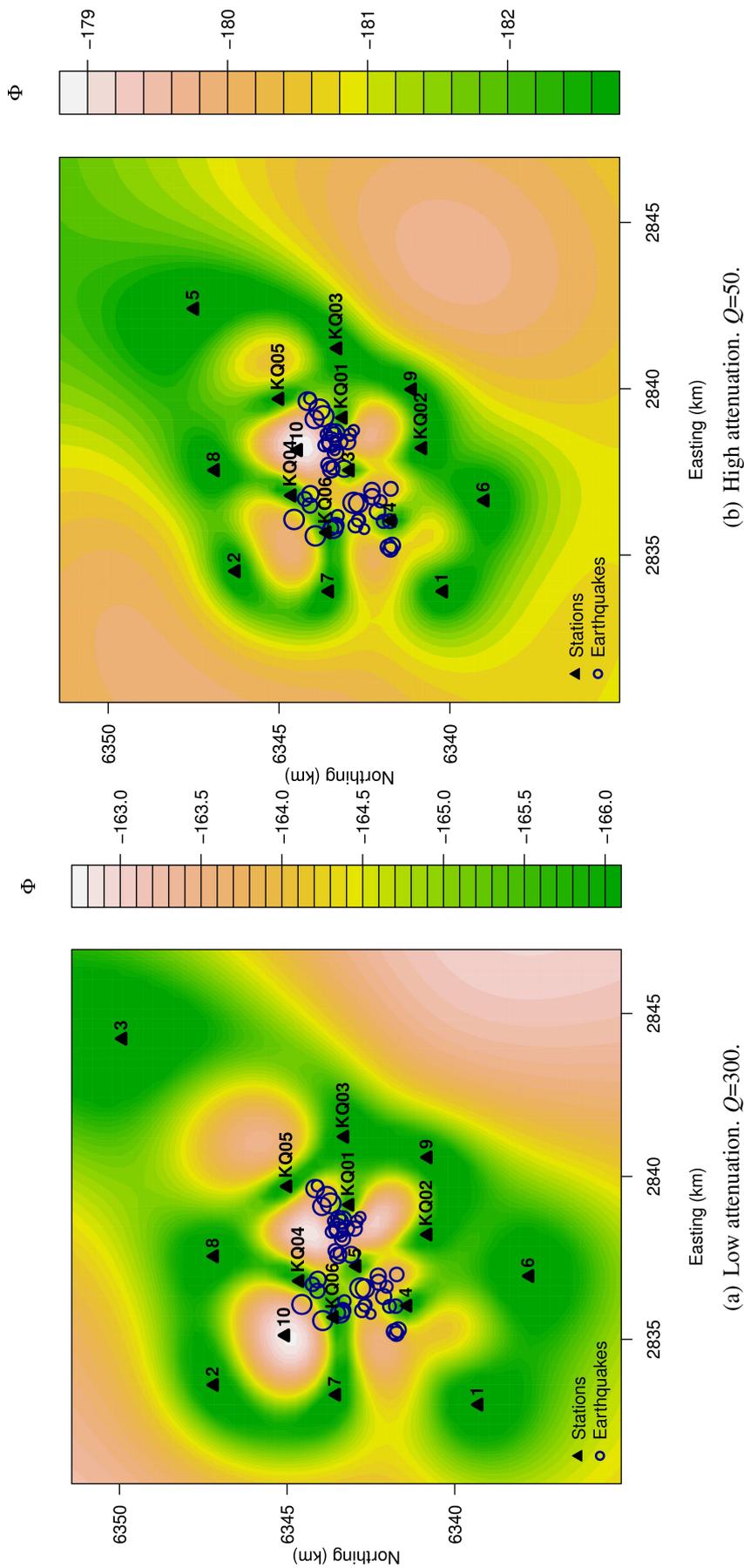


Figure 4.6: Given a uniform velocity model, the optimal location of 10 new stations is shown, labelled 1–10 in order of importance, with coloured contours showing the variation of the optimality factor Φ : white corresponds to the maximum Φ (best), and dark green corresponds to the minimum Φ (worst). We demonstrate the results for two different uniform attenuation models. (a) Low attenuation model, $Q = 300$, (b) High attenuation model, $Q = 50$. As attenuation increases, stations are drawn closer in towards the earthquakes. The two networks are displayed on the same image in Figure 4.7.

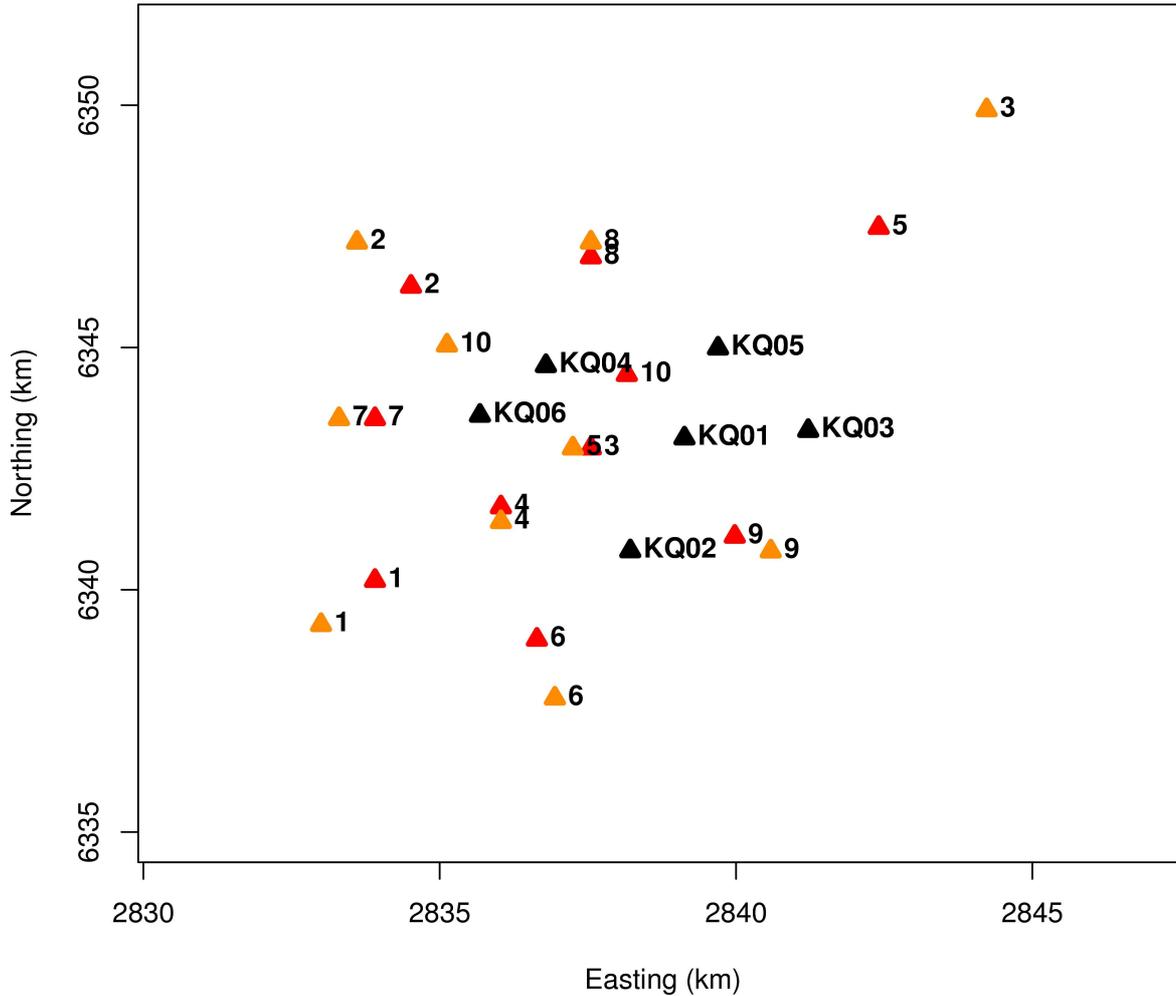


Figure 4.7: Comparison of the two different optimal networks shown in Figure 4.6. Here, the black triangles show the existing 6 station network at Kawerau, the orange triangles display the network configuration produced by the low attenuation model ($Q = 300$), and the red triangles display the network configuration produced by the high attenuation model ($Q = 50$). Stations are labelled 1–10, in order of importance. The Earthquake locations are shown in Figure 4.6.

model for the Kawerau region of *Clarke et al.* [2009]. The models in Region 2 and Region 3 are more complex amalgamations of data. A vertical section showing the upper 2.5 km of the velocity model of Fig. 4.9 is shown in Fig. 4.11 [Steven Sewell, MRP, pers. comm., 14 December 2010]. Fig. 4.11 illustrates four different layers within Regions 2 and 3: a near-surface layer, clay layer, deep clay layer, and the reservoir.

For the reservoir, we use the 1D velocity model (V_p and V_s) calculated for Rotokawa in Chapter 3, with an adjustment for the temperature difference between the two reservoirs.

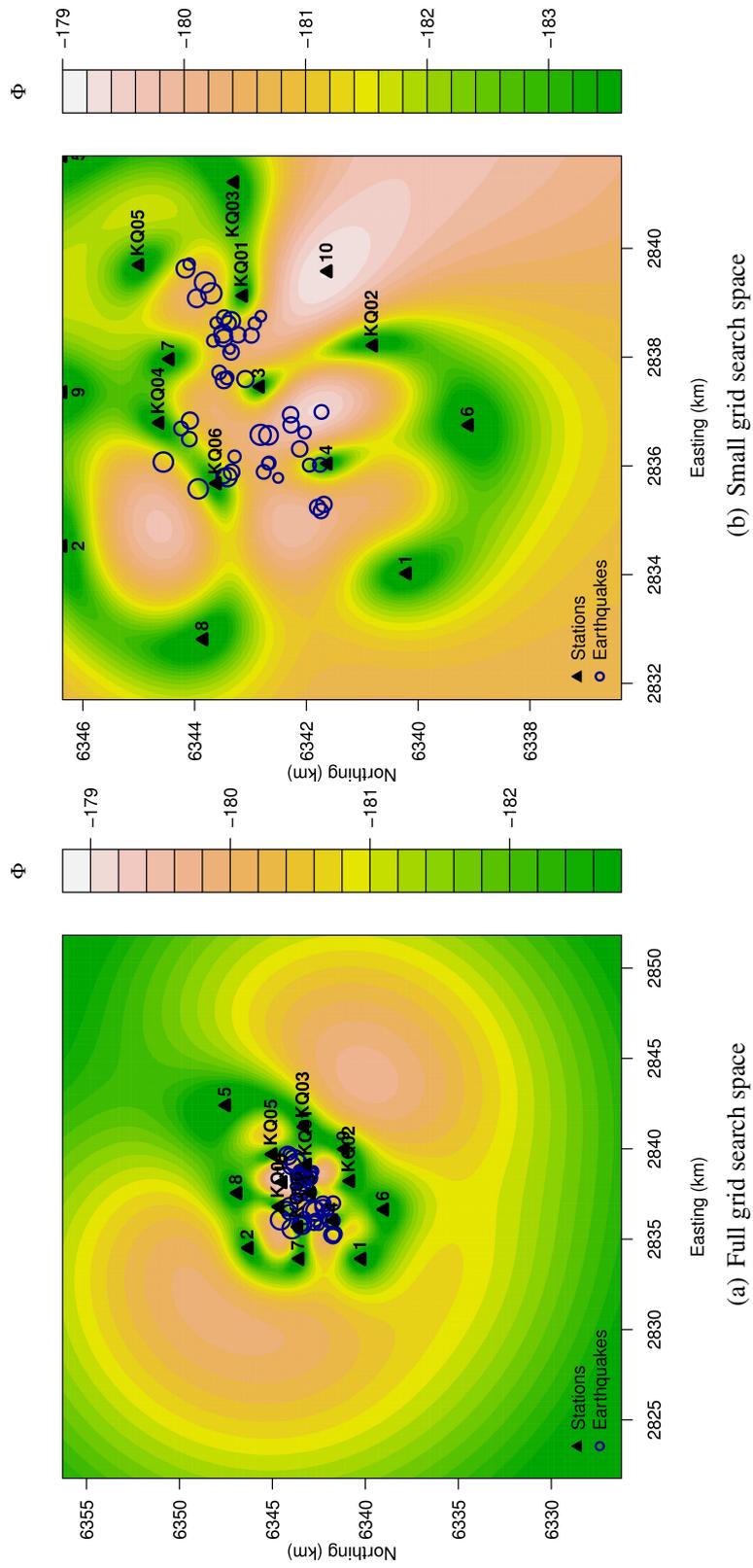


Figure 4-8: Two different design networks produced by varying the size of the search space. (a) The full grid search space used for the designs shown in Fig. 4.6. Note that we ensure the grid is large enough to contain the contours so that edge effects do not influence the network design produced. (b) An example of using a grid search space that is too small. Note how the uppermost three stations (2, 5 and 9) are pushed against the edge of the grid, and the location of this boundary has influenced the network design.

Kawerau is cooler than Rotokawa by $\sim 40^\circ\text{C}$ at the same depth: the production core at Kawerau is $260 - 290^\circ\text{C}$ [Bignall and Harvey 2005], whilst the core at Rotokawa is $300 - 335^\circ\text{C}$ [Heise et al. 2008]. Using this 40°C difference, combined with the variation of P-wave velocity with temperature reported by Jaya et al. [2010] and Kristinsdóttir et al. [2010] for their sample 3A (Hyaloclastite from an in situ temperature of 200°C : both studies used the same rock sample), we find a difference of 0.2 km/s . We add this 0.2 km/s to the Rotokawa P-wave velocities, then use the average V_p/V_s ratio of the Rotokawa data, $V_p/V_s=1.75$, to calculate a corresponding S-wave uniform velocity increase of 0.11 km/s .

For Region 2 we alter this model in the top 1.01 km , setting the deep clay layer to vary between $V_p=3.2 \text{ km/s}$ and $V_p=3.5 \text{ km/s}$ [Stephen Bannister, pers. comm., 10 December 2010, based on exploratory *Velost* runs on Kawerau data in July 2009 using the approach taken by Clarke et al. [2009]].

For Region 3 we alter this model in the top 0.38 km by setting the clay layer to vary between $V_p=2.2 \text{ km/s}$ and $V_p=2.5 \text{ km/s}$ (based on pockets of $V_p < 2.5 \text{ km/s}$ possibly corresponding to clay caps in Stephen Bannister’s experience with 3D modelling of the Rotokawa geothermal field).

For these adjustments we calculate S-wave velocities using the average Rotokawa V_p/V_s ratio in the top 1.01 km , which is also $V_p/V_s=1.75$. For both regions, we set the near-surface layer to the Rotokawa values without adjustment for temperature.

Table 4.3: Region 1 velocity model: Values from the best fit velocity model determined for the Kawerau region in Clarke [2008].

#1: Depth (km)	V_p (km/s)	V_s (km/s)
-3	4.25	2.48
0	4.48	2.62
2	4.97	2.76
4	5.34	3.12
6	5.81	3.40
8	5.92	3.53
10	6.07	3.63

Table 4.4: Region 2 velocity model: 2003 MT reservoir. Values compiled from a variety of sources (detailed in the text).

#2: Depth (km)	Vp (km/s)	Vs (km/s)
-0.40	1.99	1.14
0.10	2.20	1.26
0.16	2.20	1.83
0.30	3.20	2.00
1.01	3.50	1.81
1.30	3.17	1.81
1.60	3.25	1.86
1.80	3.32	1.90
2.00	3.37	1.93
2.20	3.40	1.94
2.60	3.62	2.07
2.90	3.94	2.25
3.40	4.52	2.58
4.00	4.77	2.73
6.00	5.71	3.26
10.00	6.01	3.43

Table 4.5: Region 3 velocity model: 1970s Schlumberger reservoir. Values compiled from a variety of sources (detailed in the text).

#3: Depth (km)	Vp (km/s)	Vs (km/s)
-0.40	1.99	1.14
0.16	2.20	1.26
0.38	2.50	1.43
0.60	3.31	1.89
1.00	3.40	1.94
1.30	3.17	1.81
1.60	3.25	1.86
1.80	3.38	1.90
2.00	3.37	1.93
2.20	3.40	1.94
2.60	3.62	2.07
2.90	3.94	2.25
3.40	4.52	2.58
4.00	4.77	2.73
6.00	5.71	3.26
10.00	6.01	3.43

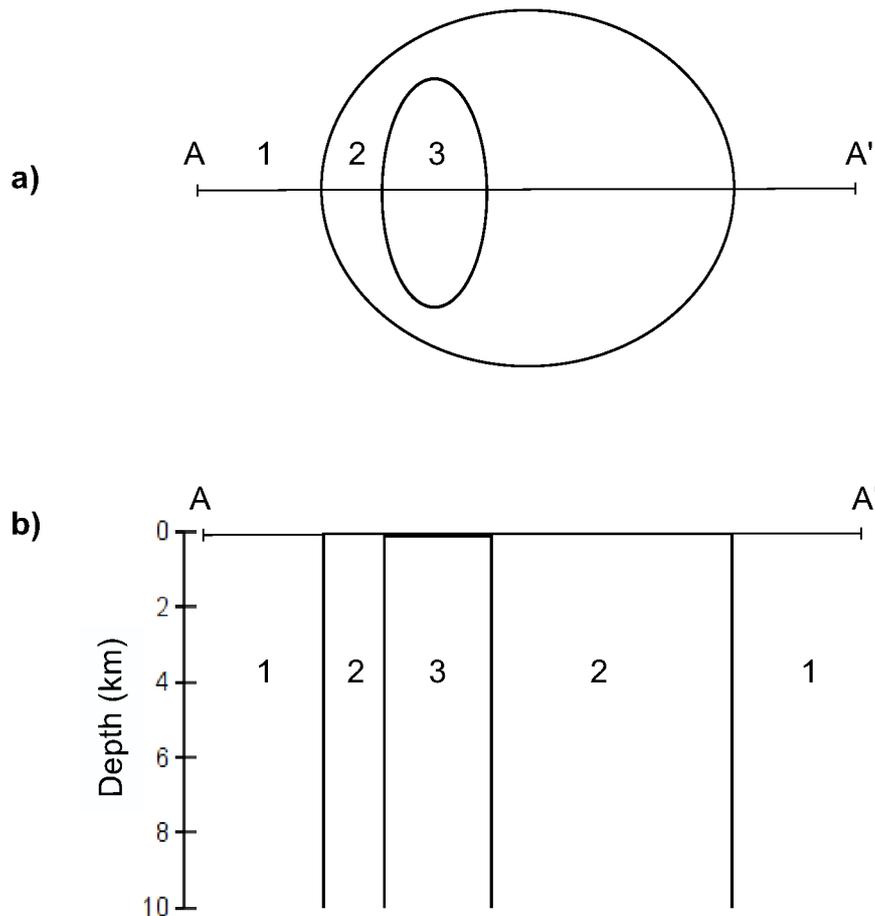


Figure 4.9: Kawerau velocity schematic identifying three different velocity regions. a) Map view. b) Cross section. See text for details.

4.3.2.2 Attenuation model

In a seismological context, attenuation is the decrease of seismic energy with increasing distance from a source due to scattering, spherical divergence, and mechanical energy loss [Gudmundsson 2004; Winkler and Nur 1982]. Attenuation is extremely sensitive to temperature (increasing with increasing temperature), pore saturation (increasing with increasing fluid) and pore geometry (increasing as pores flatten), amongst other factors [Gudmundsson 2004; Gurevich et al. 2007; Jones et al. 1980; Mavko and Nur 1979; Winkler and Nur 1982]. By definition, a geothermal reservoir is a volume of hot water and steam in porous and fractured hot rock: hence, we have the factors required to induce significant scattering

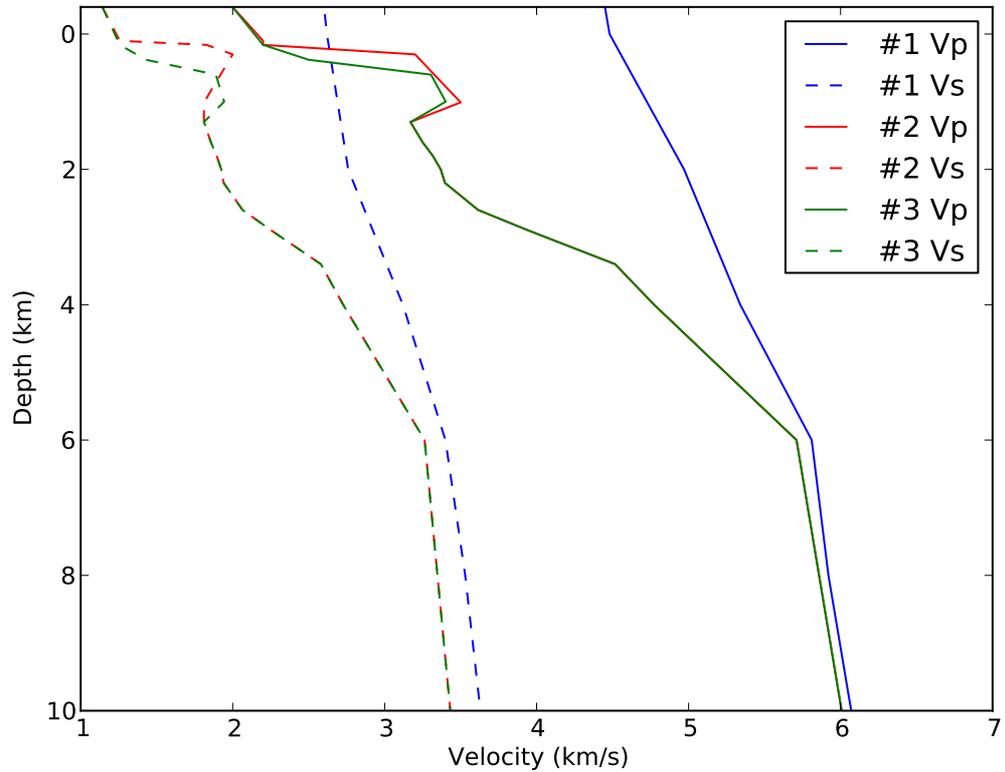


Figure 4.10: Kawerau velocity models for each region identified in Fig. 4.9. The P-wave and S-wave velocity models are labelled Vp and Vs, respectively. Although the velocity models for Regions 2 and 3 are not significantly different, we retain their separation in light of their differing attenuation models, shown in Figure 4.14.

and mechanical energy loss are all present, providing a region of extremely high attenuation.

In this work, attenuation is represented by increased uncertainty in the travel time calculations [Eq. 4.10]. This approach is motivated by the fact that as attenuation increases, the signal amplitude decreases: as a consequence, the SNR decreases and first arrivals become increasingly difficult to pick accurately. We work with the quality factor, Q , which is the inverse of attenuation. Based on Fig. 4.12, which shows as example of the temperature dependence of Q in geothermal rocks at reservoir conditions [Jaya *et al.* 2010], we have

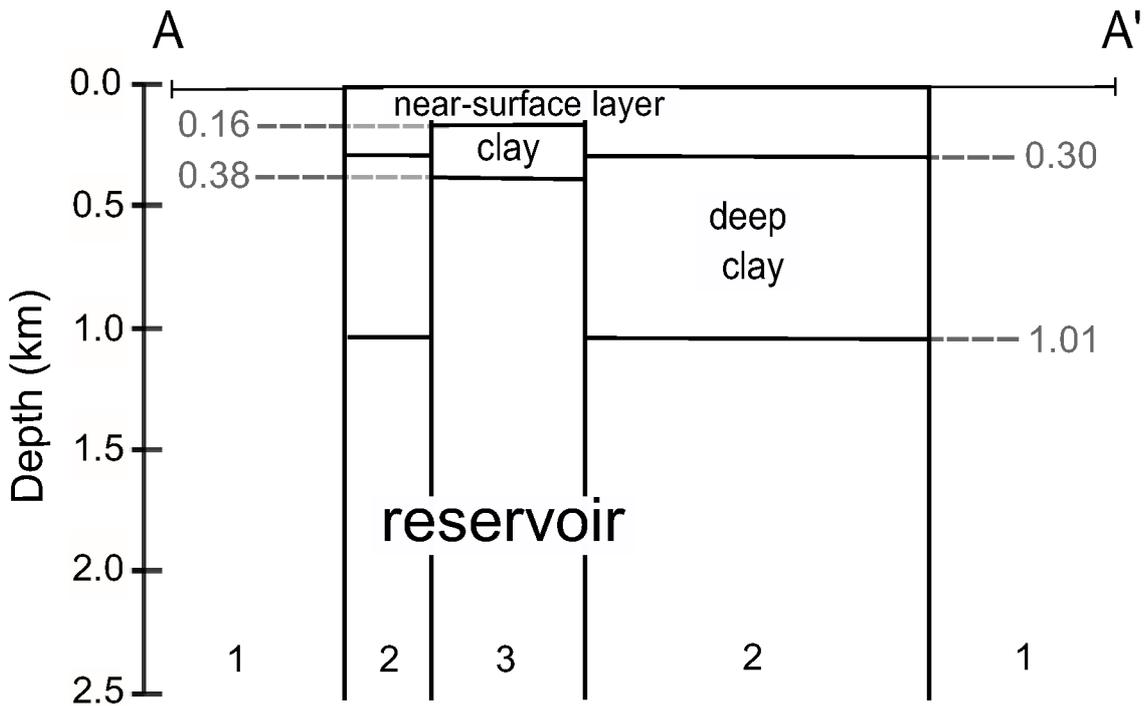


Figure 4.11: A schematic showing the upper 2.5 km of the Kawerau velocity model in Fig. 4.9, identifying four different layers within Regions 2 and 3: Near surface layer, clay layer, deep clay layer, and the reservoir [Steven Sewell, MRP, pers. comm., 14 December 2010].

adopted an empirical relation between Q and temperature T [Fig. 4.13]:

$$Q = \begin{cases} -0.03T^2 + 6T - 22 & \text{if } T < 140^\circ\text{C} \\ 1.2 \times 10^{10}T^{-3.6} & \text{if } T \geq 140^\circ\text{C} \end{cases} \quad (4.20)$$

Temperatures within the reservoir (Regions 2 and 3) have been estimated from the cross section shown in Fig. 4.14. Temperatures outside the reservoir are estimated using a surface temperature of 40°C and a low-temperature gradient of $20^\circ\text{C}/\text{km}$ (as used by *Kissling and Weir* [2005]). This temperature gradient is lower than the standard geothermal gradient of $30^\circ\text{C}/\text{km}$ [*Studt and Thompson* 1969] due to the large amount of downwelling groundwater within the TVZ, which depresses the geothermal gradient to a considerable depth. These downwelling regions of the convective system, which provide the intake for the upwelling geothermal regions, can cause temperature gradients as low as $0^\circ\text{C}/\text{km}$ [*Studt and Thompson* 1969]. It is important to note that the temperature measurements used to calculate these

temperature gradients come from boreholes that only reach depths of 2.5 km, and that at deeper locations the temperature gradient will increase. However, as both the depth of this transition and the form of the temperature gradient following is unknown, modelling this would be purely speculative and beyond the scope of this project. Moreover, the majority of the Kawerau earthquakes lie in the uppermost 4 km. We hence adopt a simple constant gradient of 20°C/km.

4.3.2.3 Travel-time calculations

Given the attenuation and velocity models described above, we now wish to trace ray paths from each earthquake to every point on the surface and assign every surface grid point a travel time and an attenuation value corresponding to each earthquake.

The problem of ray-tracing to provide a fine surface grid of travel times and attenuation values has proved to be more of a challenge than anticipated. We initially performed testing on a uniform velocity model to ensure that we could compare the results with the analytical solution,

$$t_j(\boldsymbol{\theta}) = \frac{\sqrt{(x_j - x_{\boldsymbol{\theta}})^2 + (y_j - y_{\boldsymbol{\theta}})^2 + (z_j - z_{\boldsymbol{\theta}})^2}}{v} \quad (4.21)$$

(where $t_j(\boldsymbol{\theta})$ is the travel time from earthquake $\boldsymbol{\theta}$ to station j) to verify that the ray-tracer was providing accurate results. We attempted to use an initial value and boundary value (two-point) seismic ray tracing programme *TRABOX* [Sambridge and Kennett 1990], but this proved problematic: trying to sample over a sufficiently fine grid resulted in numerous grid points without ray convergence, and hence no assigned values. We next considered the Multistencil Fast Marching Method using the MATLAB function *MSFM* [Dirk-Jan Kroon, University of Twente, Netherlands; Baerentzen 2001; Hassouna and Farag 2007], which provides accurate solutions to the Eikonal equation. To reduce the computational time, we attempted to use the 2D ray-tracing function *MSFM2D* and then translated this to 3D. A variety of tests were performed using both linear and spline interpolation during this process. However, our design algorithm proved too sensitive to even small inconsistencies caused by the 2D to 3D conversion, and required too fine a grid to be computationally feasible. We

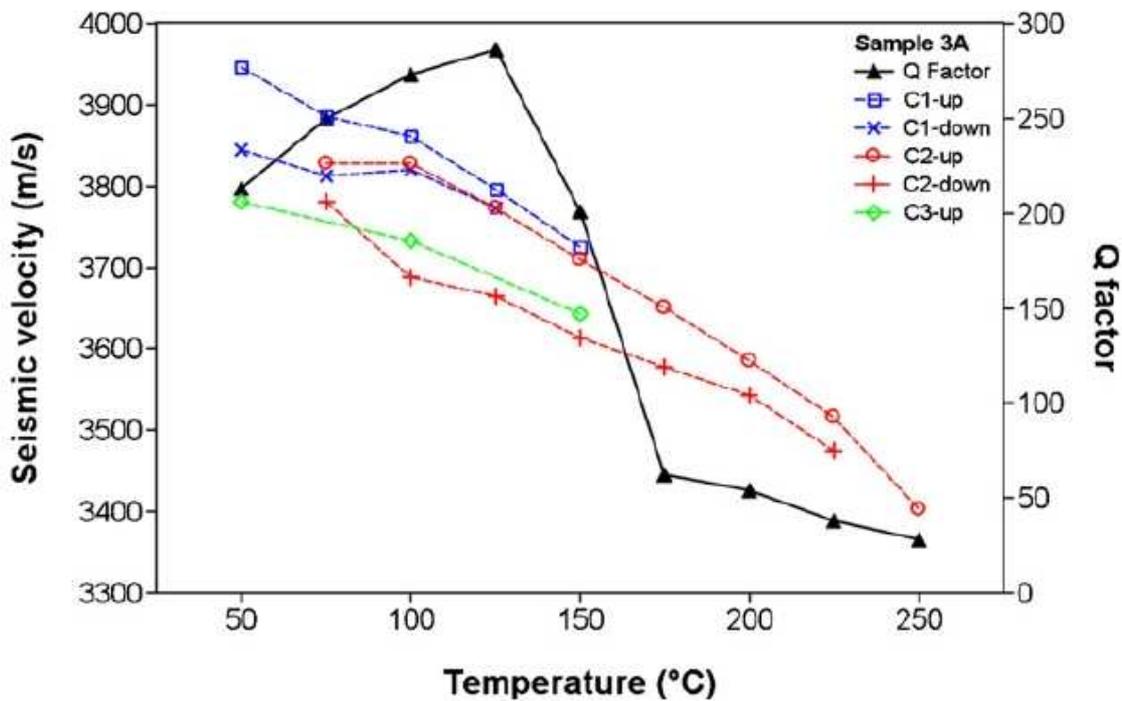


Figure 4.12: Q factor versus temperature. Q is calculated by *Jaya et al.* [2010] from the transmission waveforms of their Sample 3A (Hyaloclastite from an in situ temperature of 200°C). Q is related to the effects of thermophysical characteristics of the fluid, and Q^{-1} is defined as the fractional energy loss per wave cycle [*Jones et al.* 1980]. Note that Q decreases (attenuation increases) rapidly above 125°C. Figure from *Jaya et al.* [2010].

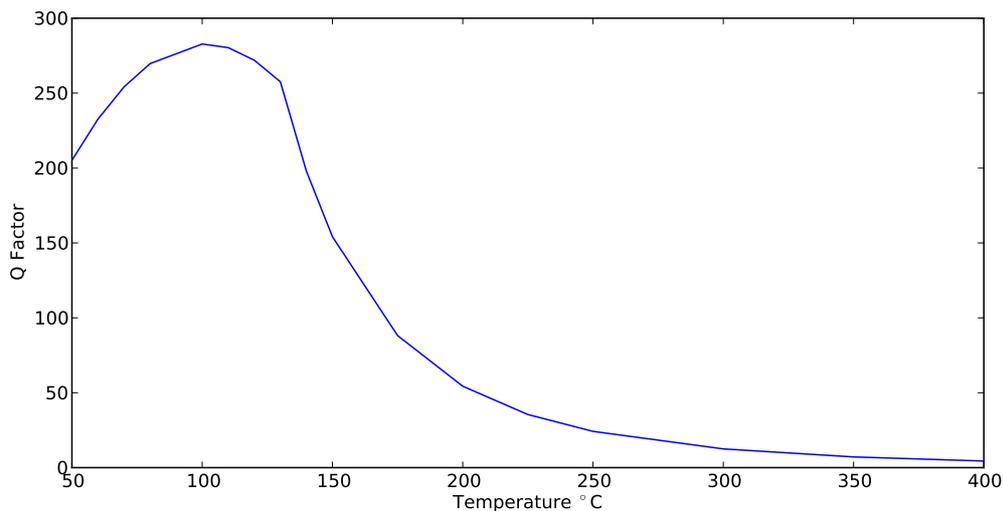


Figure 4.13: Q factor versus temperature calculated using an empirical fit of the black line showing Q factor vs. temperature in Fig. 4.12 [Equation 4.20]. Note that the temperature scale extends further than in Fig. 4.12.

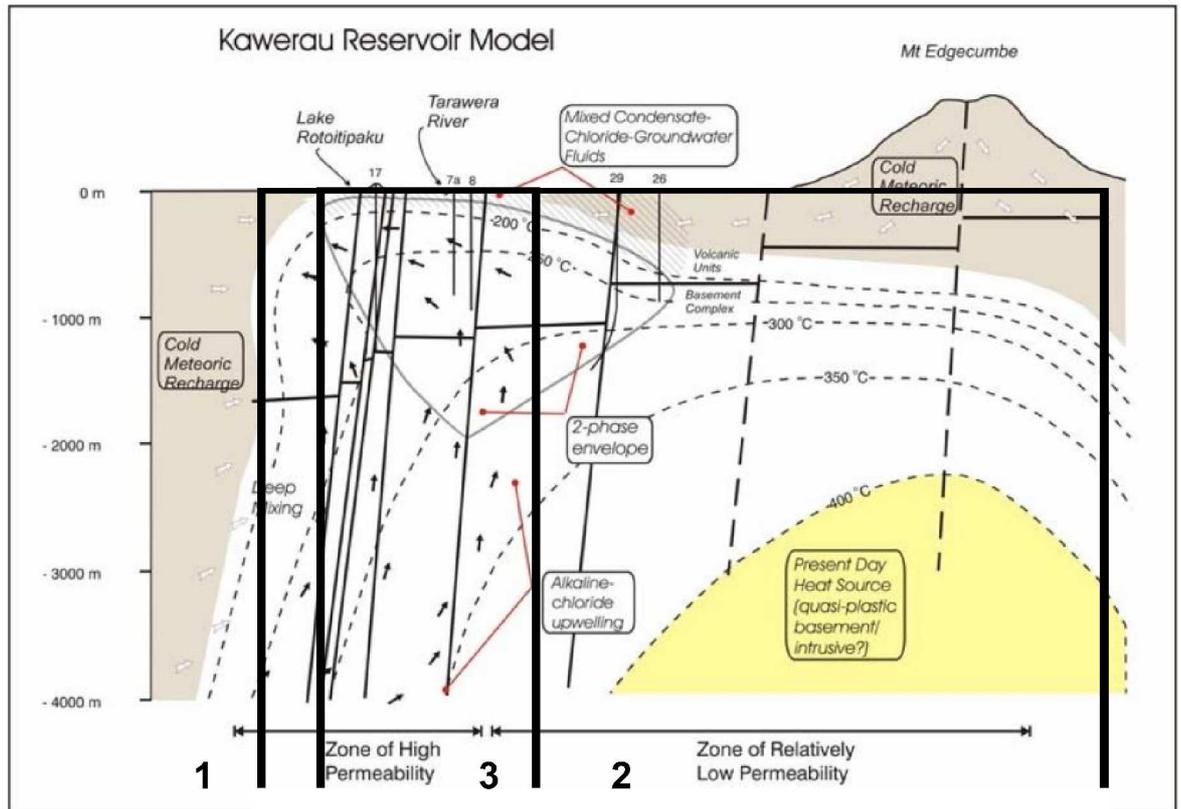


Figure 4.14: Kawerau temperature cross section, from *Bignall and Harvey* [2005]. Overlaid in the thick black lines are the outlines of Regions 2 and 3. The temperatures and corresponding attenuation of the two regions have been estimated from this image.

have therefore chosen to use the 3D ray-tracing function *MSFM3D*, which proves to be more efficient than using *MSFM2D* as a coarser grid can be used without introducing undesirable effects. Care must still be taken in choosing the grid sampling as too coarse a grid sampling will result in artifacts in the design results [see Fig. D.2 in Appendix D]. We have implemented this method using second derivatives, providing a more accurate solution to the Eikonal equation than using first derivatives alone, but we have not used cross-neighbours (see references above), which produced less accurate results. Here, we define the most accurate results as those that provide the most circular travel time contours from a point source in a homogeneous material of uniform velocity.

Our results have shown that the travel times provided by ray tracing differ slightly from those calculated using the analytical solution [Eq. 4.21]. However, if we calculate the average travel time difference (averaged over all earthquakes) between the two, shown in Fig. 4.15,

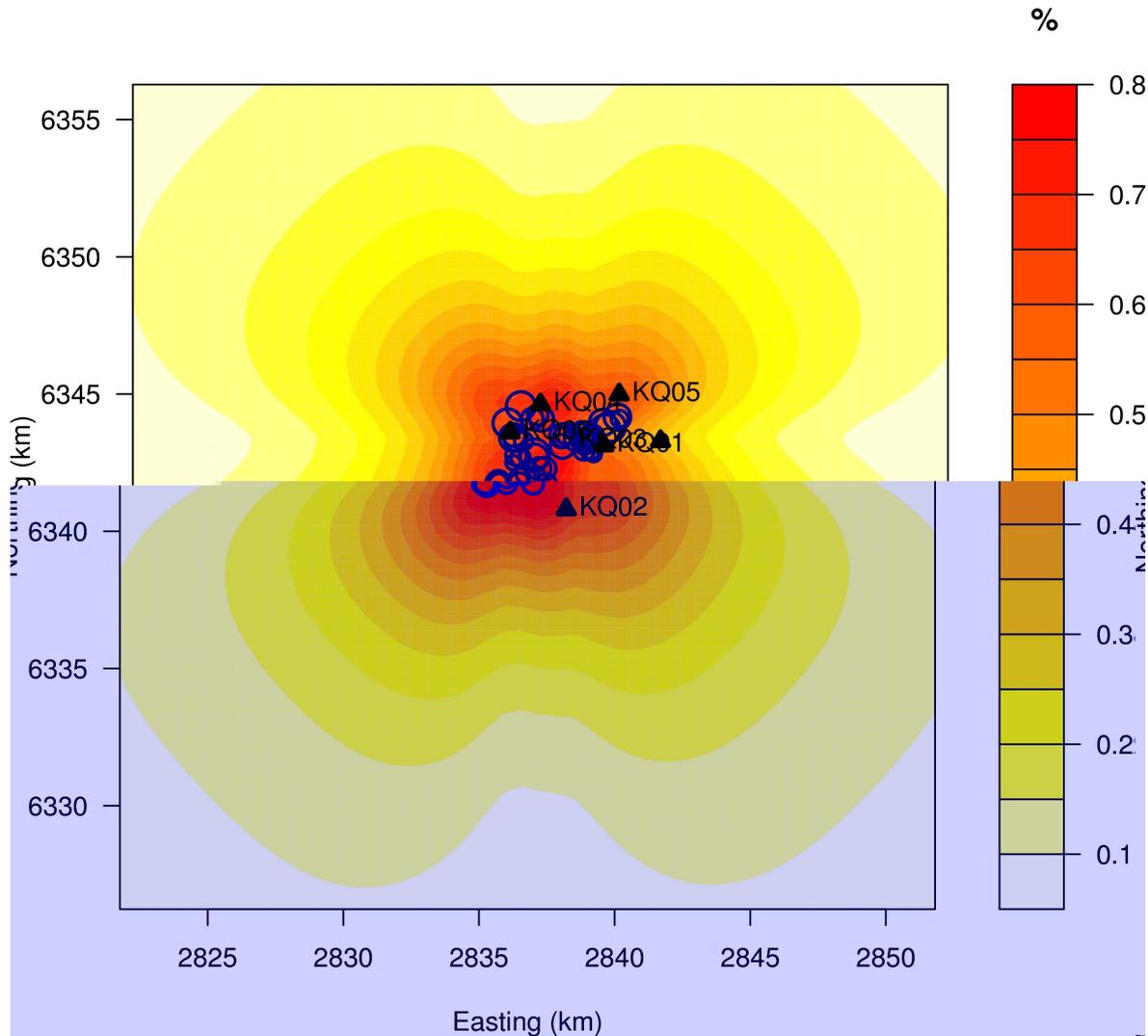


Figure 4.15: Percentage difference between the average travel times, from all the earthquakes shown, calculated via straight line tracing [Eq. 4.21] and ray tracing using *MSFM3D* through a uniform velocity model. The maximum difference is 0.8%, and occurs close to the earthquake cluster. Note that the ray tracing always results in slightly larger travel times.

the maximum difference is found to be less than 0.8%, with ray tracing always supplying a slightly larger travel time.

Note that we have not conducted a full assessment of the ray-tracer's performance, but considered the difference in travel times that the design programme will be 'seeing'. The source of this discrepancy remains to be determined: suffice to say, the difference is small enough (less than 1%) to be acceptable for present purposes. The results of applying this model in our network design does highlight how even subtle velocity changes can result in differing optimal networks, as illustrated by Fig. 4.16, where (a) shows a uniform model

where travel times have been calculated using the analytical solution [Eq. 4.21] and (b) shows the same uniform model, with travel times calculated using *MSFM3D*. The two networks are displayed together for comparison in Fig. 4.17. Examining the two shows that stations 1–4 are placed in very similar locations; station 5 is shifted; most of the subsequent stations are placed in similar locations but numbered differently due to station 5’s shift; and the location of station 9 in Fig. 4.16(a) has no station placed there in (b). In these calculations attenuation is calculated in the same way as for the uniform velocity model [Eq. 4.9] with $Q = 50$, to be sure that the difference we are seeing arises from velocity model ray tracing effects only.

We use this same attenuation model, and *MSFM3D* to ray-trace through the P-wave velocity model determined in Section 4.3.2.1. The optimal network design produced is displayed in Fig. 4.18.

We use *MSFM3D* to model attenuation as follows: we modify Eq. 4.8 again, this time with Q and v as functions of position x ,

$$u(x) = \exp\left(-\frac{xf\pi}{Q(x)v(x)}\right) \quad (4.22)$$

We integrate along the path $S_{j\theta}$, from earthquake θ to station j , determined via ray tracing and set $f = 10$ Hz (the dominant frequency of microearthquakes at Rotokawa),

$$\ln [u_j(\theta)] = -f\pi \int_{S_{j\theta}} \frac{dS_{j\theta}}{Q(S_{j\theta})v(S_{j\theta})} \quad (4.23)$$

The fraction of the original amplitude lost to attenuation along the entire path is $v_j(\theta) = 1 - u_j(\theta)$. This value is scaled depending on the magnitudes of the earthquakes we are interested in locating. For example, the amplitude of a seismic wave from a magnitude 5 earthquake with $v_j(\theta) = 0.2$ would supply sufficient signal for detection and location, however, a seismic wave from a magnitude 0 earthquake with $v_j(\theta) = 0.2$ would not. We therefore need to determine the fraction of amplitude we believe is required for detection and location of our magnitude of interest. Unfortunately the fraction of amplitude that is sufficient also depends on the ambient noise of a region.

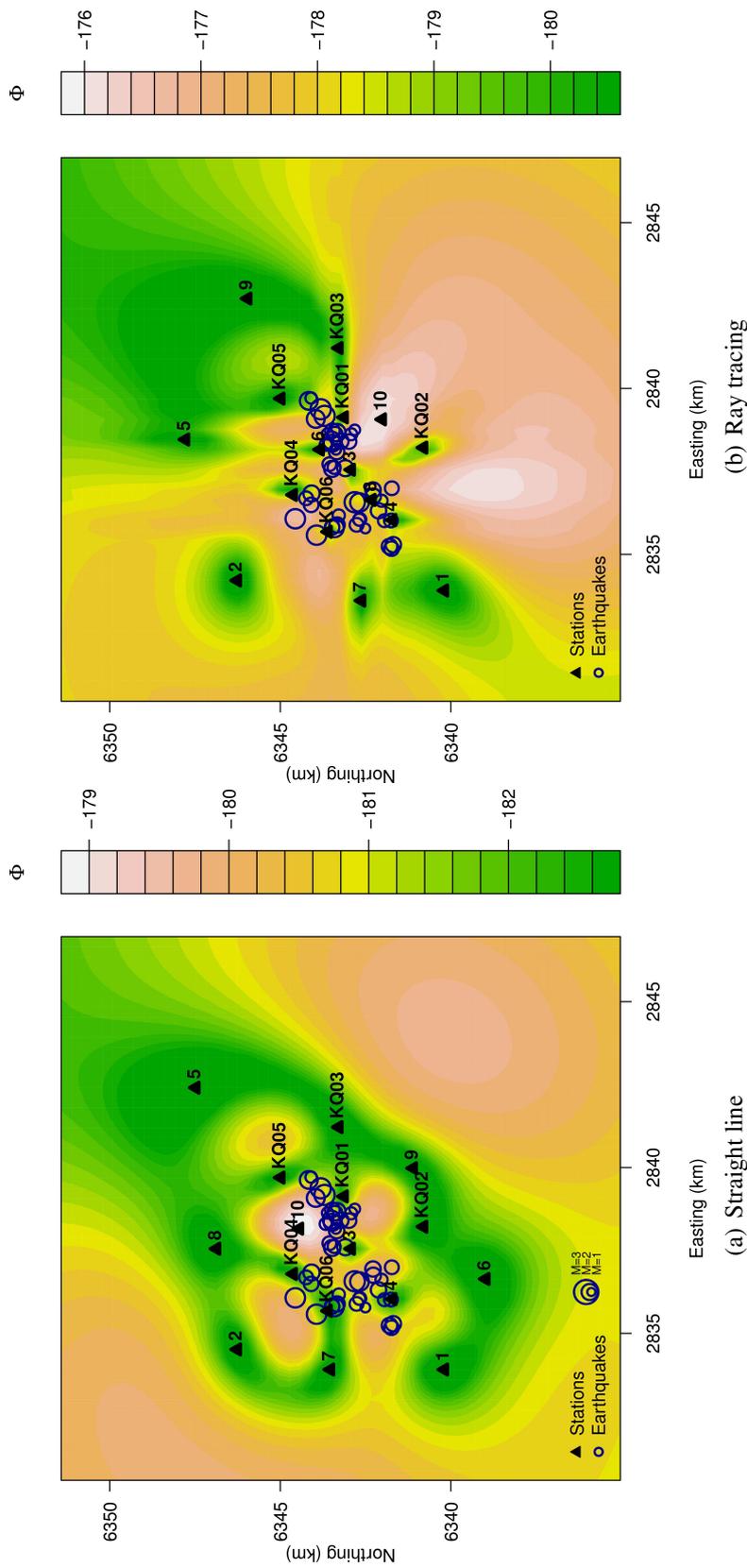


Figure 4.16: Given a uniform velocity model, the optimal location of 10 new stations is shown, labelled 1–10 in order of importance, with contours showing the variation of the optimality factor Φ . (a) Travel times calculated using the analytical solution [Eq. 4.21] (b) travel times calculated using ray tracing with *MSFM3D*. Stations 1–4 are placed in very similar locations; station 5 is shifted; most of the subsequent stations are placed in similar locations but numbered differently due to station 5's shift; and the location of station 9 in (a) has no counterpart in (b). Fig. 4.17 displays these two networks on the same graph.

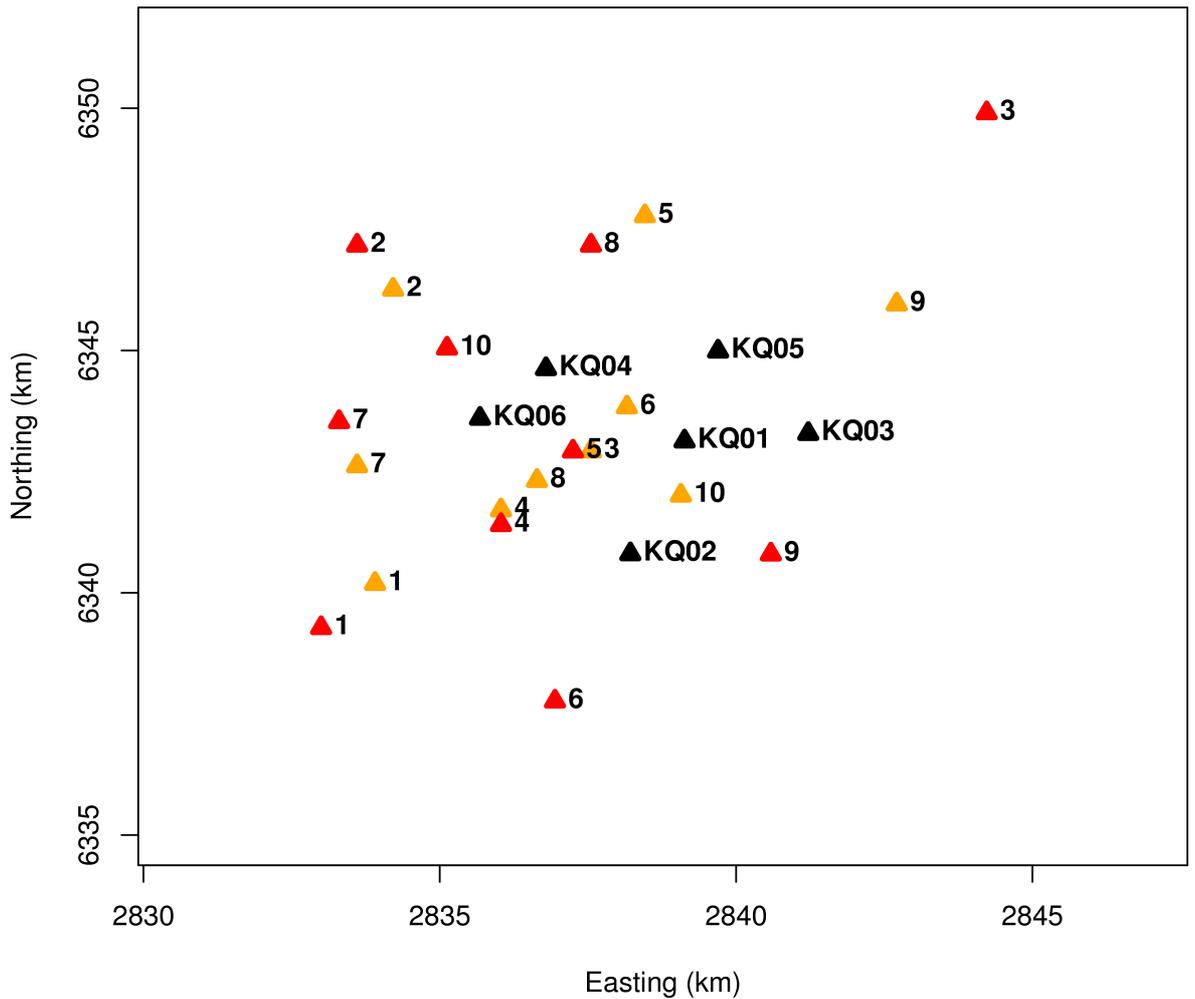


Figure 4.17: Comparison of the two different networks shown in Fig. 4.16. The red triangles display the optimal configuration produced by the analytical solution, and the orange triangles display the optimal configuration produced by ray-tracing. Stations are labelled 1–10, in order of importance. Earthquake locations are displayed in Figure 4.16.

As this value is one of the most influential in the network design model in terms of earthquake–station distance, and can not be easily calculated, experience combined with testing needs to be used to determine the appropriate value. We use the fact that at Rotokawa, in a similar attenuation setting with similar magnitudes of interest, station RT04 was placed ~ 5 km from the earthquake zone of interest, and this distance proved too far away for the station to detect any usable signal from many of the earthquakes in the Rotokawa reservoir that were otherwise detected by closer stations (it detected $M > 2$ events), so was removed [Stephen Bannister, pers. comm., 2010]. We therefore set our scaling value q to ensure stations lie less than ~ 5 km from the earthquake zone. Testing different q values reveals that

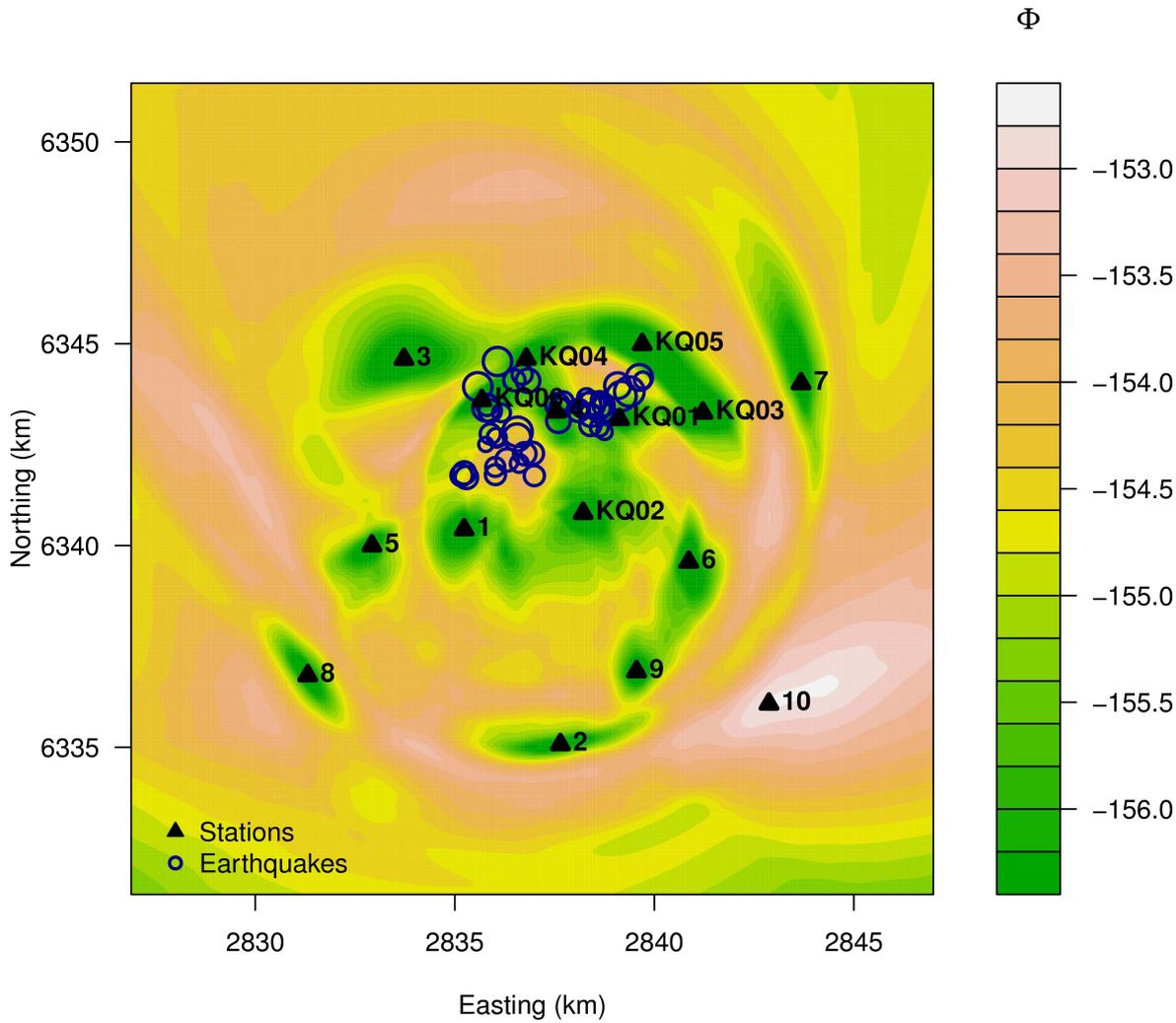


Figure 4.18: Optimal network design created by ray tracing through the 3D P-wave velocity model described in Section 4.3.2.1, using a uniform attenuation model ($Q = 50$).

the q value corresponding to this ~ 5 km constraint is $q = 0.345$ [Fig. 4.19]. Our attenuation value is thus

$$a_j(\boldsymbol{\theta}) = \begin{cases} v_j(\boldsymbol{\theta}) + q & \text{if } v_j(\boldsymbol{\theta}) + q \leq 1 \\ 1 & \text{if } v_j(\boldsymbol{\theta}) + q > 1 \end{cases} \quad (4.24)$$

After ray-tracing, we output two matrices for each earthquake $\boldsymbol{\theta}$: one is a surface grid of travel times, $\mathbf{t}(\boldsymbol{\theta})$, and the other is a surface grid of attenuation before scaling, $\mathbf{v}(\boldsymbol{\theta})$. Our data distribution is the same as in Eq. 4.10, but now $t_j(\boldsymbol{\theta})$ has been extracted from the travel

time grid described above, and $a_j(\boldsymbol{\theta})$ from Eq. 4.24:

$$d_j = t_j(\boldsymbol{\theta}) + \frac{\varepsilon_j}{1 - a_j(\boldsymbol{\theta})} \quad (4.25)$$

It is important to note that during the ray tracing step, the grid sampling rate (the interval between samples from the grid) used is important, and the grid must be sampled finely enough to not create artifacts in the output model.

We calculate an optimal design using the ray tracing described, through the P-wave velocity model described in Section 4.3.2.1 and the attenuation model described in Section 4.3.2.2 [Fig. 4.19(b)]. Fig. 4.19(b) displays an approximately spherical region of radius ~ 5 km from the centre of the earthquake cluster, outside of which Φ is very low due to the strong attenuation. Asymmetry is introduced to the south of the earthquake cluster by the higher attenuation inside the reservoir than outside.

4.3.2.4 P and S waves

In earthquake location, both P and S picks are typically used to locate earthquakes. We therefore extend our algorithm to incorporate P and S travel times together. The difference between the two phases is velocity. We can incorporate P- and S-waves using the same criterion, but the extended data set is now

$$d = \begin{bmatrix} d^P \\ d^S \end{bmatrix} = \begin{bmatrix} \mathbf{t}^P(\boldsymbol{\theta}) + \mathbf{A}^P(\boldsymbol{\theta})\boldsymbol{\varepsilon}^P \\ \mathbf{t}^S(\boldsymbol{\theta}) + \mathbf{A}^S(\boldsymbol{\theta})\boldsymbol{\varepsilon}^S \end{bmatrix} \quad (4.26)$$

As before, the background noise process uncertainties $\boldsymbol{\varepsilon}^P$ and $\boldsymbol{\varepsilon}^S$ are assumed to follow a normal distribution with constant variance, $\boldsymbol{\varepsilon}^P \sim iid(0, (\sigma_\varepsilon^P)^2)$ and $\boldsymbol{\varepsilon}^S \sim iid(0, (\sigma_\varepsilon^S)^2)$, with $\sigma_\varepsilon^P = \sigma_\varepsilon^S = 1$.

We calculate S-wave travel time and attenuation grids as for P-waves, and our application remains the same. Now, however, the S-wave becomes attenuated more quickly than the P-wave due to having a slower velocity, hence longer travel times (we assume the same

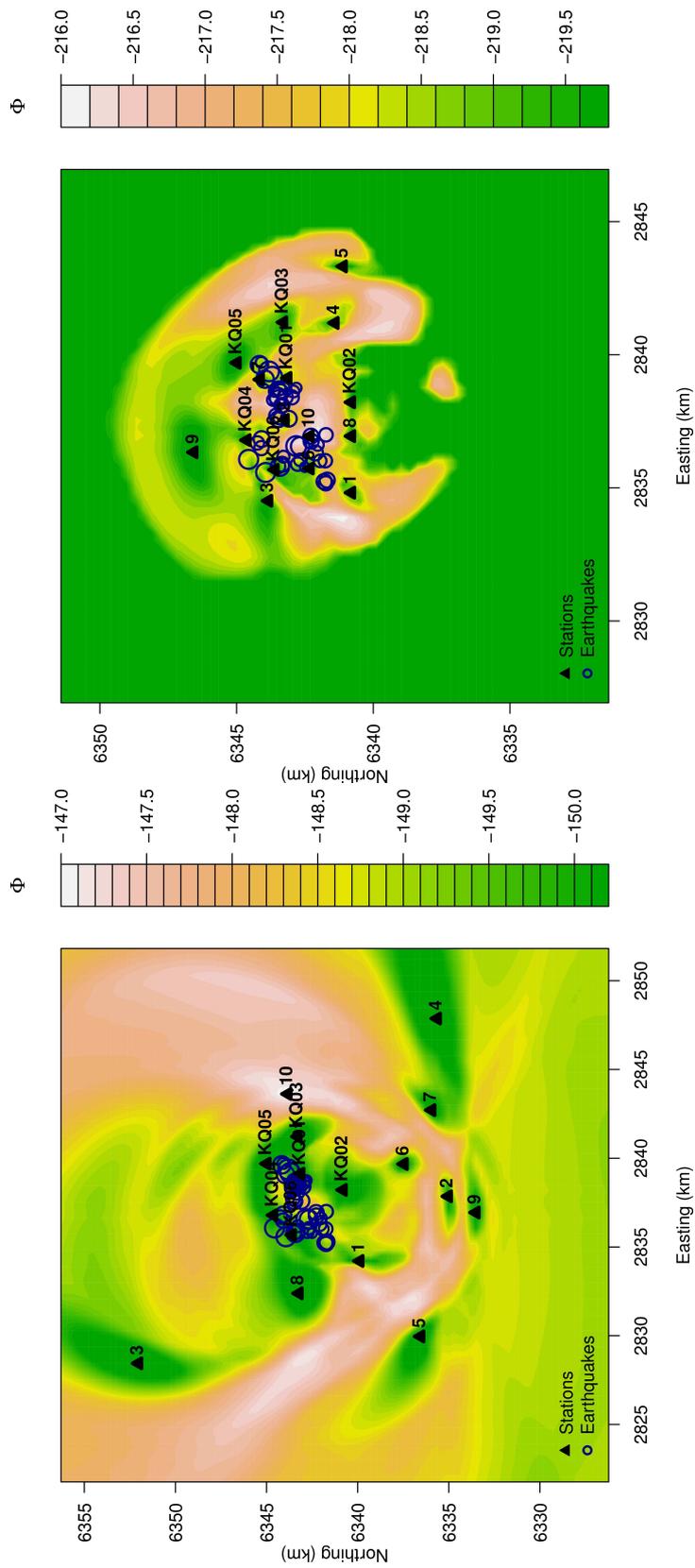


Figure 4.19: Optimal network design produced by ray-tracing through the 3D P-wave velocity model and attenuation model described in Sections 4.3.2.1 and 4.3.2.2 respectively. (a) Attenuation scaling value $q = 0.00$. Note that with no earthquake magnitude constraint imposed, optimal stations are located as far away as ~ 10 km. These locations will be of little use for low-magnitude earthquakes with low initial signal amplitudes. (b) Attenuation scaling value $q = 0.345$. Note the approximately spherical region of radius ~ 5 km from the centre of the earthquake cluster, outside of which Φ is very low due to the strong attenuation. Asymmetry is introduced to the south of the earthquake cluster by the higher attenuation inside the reservoir than outside.

temperature dependence of Q^P and Q^S), and we must reduce our q scaling value or the S-waves will not affect the calculation. By inspecting contour plots of the surface S-wave attenuation values we ascertain that the S-wave attenuates at a rate which is +0.2 that of the P-wave attenuation values at equivalent locations. We therefore drop q by 0.2 to $q = 0.145$ (applied to both the P- and S-wave attenuation models).

We treat the S-wave inclusion relatively simplistically here. Future work is needed for a more sophisticated treatment of S-attenuation that can take into account differences from P-waves in terms of temperature dependence, attenuation rates, amplitudes, and differences in vertical and horizontal noise and amplitudes.

The results of incorporating the S-wave velocity and attenuation model into our design model is shown in Fig. 4.20. In comparison with the model computed using P-waves alone [Fig. 4.19(b)] this model shows a spherical region, outside of which Φ is very low due to the strong attenuation, with a slighter larger radius of ~ 7 km, but stations still lie within ~ 5 km. The asymmetry to the south has been reduced and there have been a number of subtle shifts in station locations.

4.3.3 Noise sources

Seismic station installations are susceptible to several types of noise that can contaminate or mask signals of interest, effectively reducing the operating sensitivity of each instrument. Siting the seismometer on hard rock and away from anthropogenic noise sources is preferable, but not always possible. With respect to choosing locations for seismometers, the Incorporated Research Institutions for Seismology (IRIS) Program for Array Seismic Studies of the Continental Lithosphere (PASSCAL) suggests that “The area within a few kilometres of the chosen site should not have hydroelectric turbines (10 km), frequent train traffic (3 km), irrigation pumping stations (2 km), or other large machinery (12 km) including heavy earth-moving equipment”¹. As we are interested in siting seismometers for the detection of geothermal microearthquakes, in a region of volcanic sediment, we are faced with a situa-

¹<http://www.passcal.nmt.edu/content/station-siting-considerations>. Last accessed 7 June 2010.

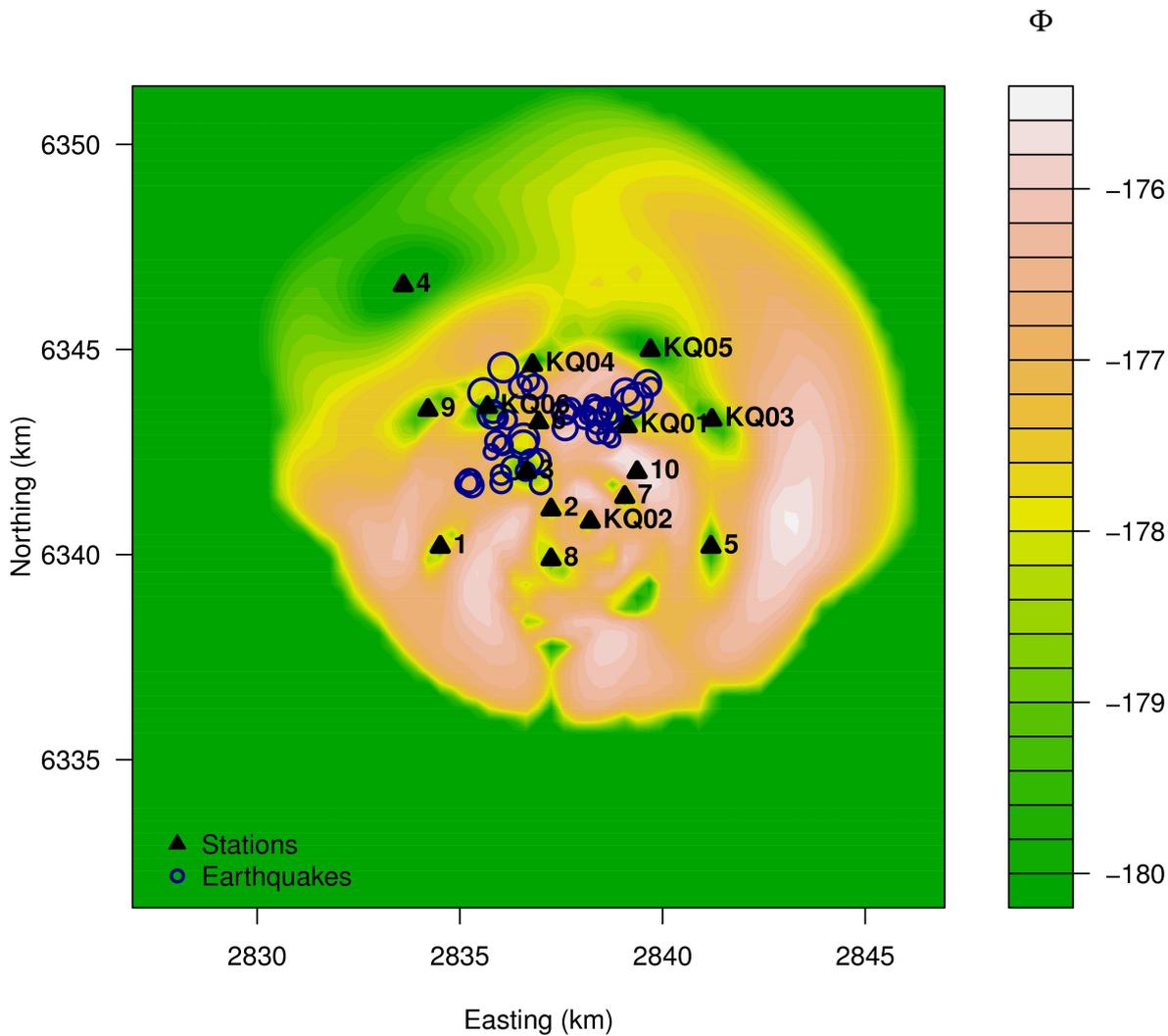


Figure 4.20: Optimal network design produced by ray-tracing through the 3D P-wave and S-wave velocity models and attenuation models described in Sections 4.3.2.1 and 4.3.2.2 respectively. Attenuation scaling value $q = 0.145$ applied to both P- and S-wave attenuation models.

tion in which the attenuation of the signals of interest is sufficiently high that distant station sitings would be ineffectual. Moreover, the locations of interest are centred on a noisy industrial region. Hence, we face an obvious trade-off between background noise and signal attenuation, and the IRIS guidelines can not be strictly adhered to in this situation. We quote these guidelines here to provide a sense of the impact of various surface noise sources on data resolution, and how the ‘optimality’ of the station sites may be compromised by the field conditions and environmental factors.

Anthropogenic noise contaminates seismic recordings through ground vibrations and at-

atmospheric noise. Atmospheric noise affects seismometers via pressure fluctuations applied directly to the recording system and indirectly through the surrounding ground. The distance from a noise source at which a seismometer is useful depends upon the magnitude of the earthquakes of interest, and the magnitude, attenuation and frequency of the noise. To accurately quantify the noise environment and its impact on seismometer performance, a detailed noise study would need to be carried out. Such a study is beyond the scope of our current investigation, so we have used the available literature to estimate the impact of surface noise sources.

Most of the available literature pertains to audible noise, commonly expressed in units of A-Weighted Sound Level (dBA). This corresponds to the sound level in decibels measured on a sound level meter using the A-weighted filter network, which de-emphasises the very low and very high frequency components of the sound in a manner similar to the frequency response of the human ear and correlates well with subjective reactions to noise [AES Highgrove 2006]. We refer below to the L_{10} noise level measurement, which corresponds to the noise level exceeded 10% of the time [Kingett Mitchell Limited 2005].

We characterise different noise zones as follows:

1. **Industrial Zone**

Power station: Kingett Mitchell Limited [2005] provided a table of operational noise levels calculated based on field test measurements at the 60 MW Mokai geothermal power station, with adjustments to account for existing structures at Kawerau that result in noise screening. We use these values to determine an equation for noise attenuation as a function of distance d .

$$L_{10} = 129.33 - 27.19 \log d \quad (4.27)$$

Following Kingett Mitchell Limited's [2005] report, the Kawerau facility implemented changes in fan designs to minimise noise and achieve levels of 50 dBA at 500 m. Taking this change into account, we retain the gradient from Eq. 4.27 and decrease the

constant term to obtain a function representing noise levels from the Kawerau power station:

$$L_{10} = 123.39 - 27.19 \log d \quad (4.28)$$

These dBA values measure airborne acoustic noise. To check that this is usefully representative of the noise effects on a seismometer, we consider known seismometer distance restraints. In similar surface conditions to Kawerau (temperature and rock type) at Rotokawa geothermal field, *Bannister et al.* [2008] determined that an instrument placed a few hundred metres from the power station was too noisy: a distance of 300 m from the power station proved too noisy, whilst 550–600 m corresponded to the limit of acceptable noise levels [Stephen Bannister, pers. comm. 2011]. We apply an SNR=3 as a minimum desired noise level and $M \sim 0$ earthquakes as our target magnitudes of interest. Results from Rotokawa show that $M \sim 0$ earthquakes typically have an amplitude of ~ 1000 counts on Taurus recorders within a 3 km radius. A $10\times$ increase in amplitude is equivalent to a 20 dB increase in noise level [Clinton 2007]. As the earthquakes are often just above magnitude zero, we drop the amplitude slightly, to 900, to determine our desired noise level,

$$\frac{\log(900/3)}{\log(10)} 20 = 50 \text{ dBA} \quad (4.29)$$

If we insert this value into Eq. 4.28, we obtain a distance of 520 m (note that for all calculations in this section more significant figures have been used than are displayed). This result is within the uncertainty limits of the 550–600 m estimate from Rotokawa.

We therefore assert that for the purposes of network design calculations: (1) Eq. 4.28 provides a satisfactory function describing noise levels from the Kawerau power station with respect to their affect on seismometers; (2) 50 dBA is the value to be used as the noise limit for seismometer siting; (3) and we assume that audible noise estimates provide satisfactory estimates for the noise impacts on seismometers.

Transmission line: noise is swamped by power station/railway/roads, so we will just

ignore this source.

Tasman Mill: the existing activities in the industrial zone are within the local councils District Plan noise limits [*Kingett Mitchell Limited 2005*] and will be lower than the Power station noise levels, therefore we just apply the Power station noise levels to the entire industrial zone.

2. Railway:

Hanson et al. [2006], conducted a thorough review of the noise impacts of rail traffic, and how to assess this noise. Kawerau railway hosts 12 trains per day [Langley Atkinson, Ontrack, pers. comm., 18 January, 2011], which corresponds to the high-end of a moderately used corridor as defined by *Hanson et al.* [2006] (5–12 trains per day). The impact of the railway noise on seismometers is therefore likely to be significant. We use *Hanson et al.*'s graph displaying the vibrational noise level versus distance from the track centreline, for a “Locomotive powered Passenger or Freight train running at 50 mph” (80.5 km/h) to obtain the following equation,

$$VdB = 104.17 - 18.95 \log d \quad (4.30)$$

VdB denotes vibrational decibels, where $1 \text{ VdB} = 1 \times 10^6 \text{ in/sec}$. As the measurement of VdB is different to dBA, which we have used for our other noise assessments, we need to adjust this to be in dBA. *Hanson et al.* [2006] stated that industry standards for vibration-sensitive equipment regard 66 VdB as the vibrational noise level “adequate for medium- to high-power optical microscopes (400X), microbalances, optical balances, and similar specialized equipment.” We therefore adapt this vibrational noise level as the limit of acceptability for seismometers. Our noise limit in dBA is 50 [Eq. 4.29], we therefore adjust the amplitude of Eq. 4.31 so that now $66 \text{ VdB} \equiv 50 \text{ dBA}$.

$$dBA = 87.71 - 18.95 \log d \quad (4.31)$$

3. **Roads**

Main roads: The Department Of Planning And Community Development, Victoria, Australia [2008] compiled a table of estimated noise levels at varying distances from a road truck. Using these values we have adopted the following gradient for noise attenuation as a function of distance d , with the amplitude adjusted for the fact that receiver E (roughly 25 m from SH 34 [Kingett Mitchell Limited 2005]) is exposed to night-time noise levels of up to 70 dBA, and that SH 34 has an average noise level of 74 dBA [Kingett Mitchell Limited 2005]:

$$L_{10} = 98.36 - 20.29 \log d \quad (4.32)$$

Minor roads: We adjust Eq. 4.32 to account for the lower amplitude of traffic noise produced on minor roads, using as a reference the average of 47 dBA recorded on Onepu Springs Road [Kingett Mitchell Limited 2005]

$$L_{10} = 71.36 - 20.29 \log d \quad (4.33)$$

Lumber Yard: A noise monitoring location within a lumber yard in California recorded an average daily noise level of $L_{10} = 58$ dBA [AES Highgrove 2006]. As lumber yard noise sources will be similar to a road truck, we adapt Eq. 4.32 again,

$$L_{10} = 82.36 - 20.29 \log d \quad (4.34)$$

4. **Power lines:**

Power line noise is very hard to quantify due to a number of influencing factors. Experienced seismometer deployment teams use the rule-of-thumb that seismometers must be at least 50–100 m from power line poles [Stephen Bannister, pers. comm., 2010]. As we do not know the locations of the individual power line poles, we use the location of seismometer station KQ01, which is 50 m away from the power lines as the

acceptable distance, and place a swath of high noise 50 m to either side of the power lines.

5. **River:**

The Tarawera River flows slowly through the region of interest in Kawerau (determined by the meandering curves of the river in this region). It will therefore not create sufficient noise to affect seismometers, and it is ignored as a noise source for present purposes.

Table 4.6 contains a summary of the noise zone characterisations described above.

Table 4.6: Summary of terms used in noise variation with distance equations of the form $L_{10} = \text{Constant} - \text{Gradient} \times \log d$

Source	Constant	Gradient
Power station	123.39	27.19
Railway	87.71	18.95
Main Roads	98.36	20.29
Minor Roads	71.36	20.29
Lumber Yard	82.36	20.29
Power lines	—	— *

* 50 m swath.

We model the surface noise as an increase in the uncertainty of the travel time pick due to the increased background noise reducing the signal-to-noise ratio (SNR). We introduce this additional source of noise by adding a random pick-uncertainty source $\eta_j \sim iid(0, \sigma_j^2)$ where $\sigma_j^2 = \rho 10^{\kappa_j/20}$. Here κ_j is the surface noise in dB at station j (the scaling converts this value to a linear scale), and ρ is an empirical scaling factor determined by trial and error from the data. The value of ρ is chosen so that no stations are placed within any region where the surface noise exceeds 50dB (the noise threshold chosen previously). Eq. 4.10 thus becomes,

$$d_j = t_j(\boldsymbol{\theta}) + \frac{\varepsilon_j}{1 - a_j(\boldsymbol{\theta})} + \eta_j \quad (4.35)$$

For all stations $j = 1, 2, \dots, m$, η_j becomes $\boldsymbol{\eta}$,

$$\mathbf{d} = \mathbf{t}(\boldsymbol{\theta}) + \mathbf{A}(\boldsymbol{\theta})\boldsymbol{\varepsilon} + \boldsymbol{\eta} \quad (4.36)$$

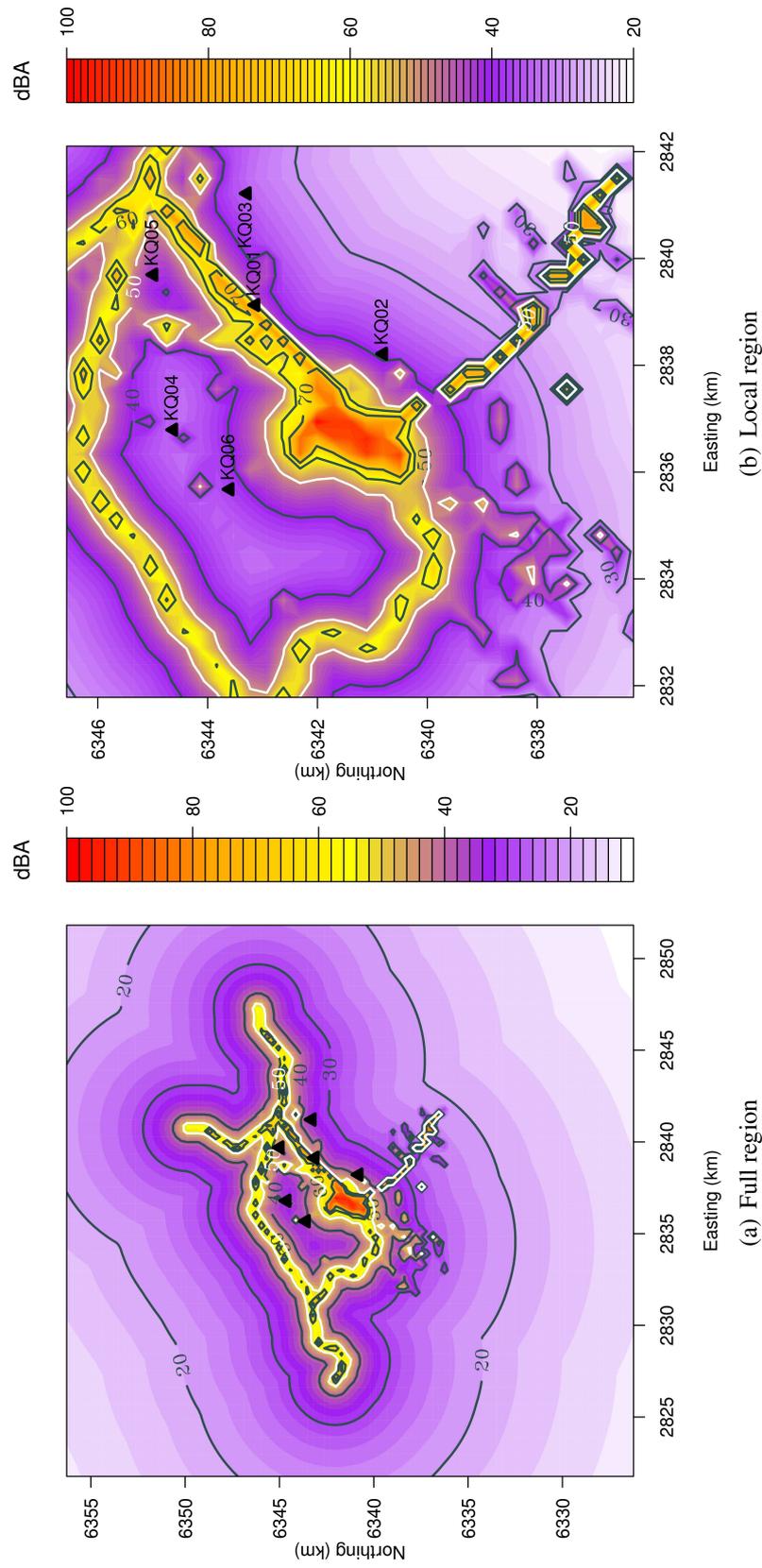


Figure 4.21: Noise levels at Kawerau (dBA). The white line identifies 50 dBA, our determined limit of noise acceptability for a seismometer [Eq. 4.29]. (a) displays the full calculation region we use, and (b) displays a zoom-in of the local region.

As $\eta_j \sim iid(0, \sigma_j^2)$, the data mean is as before [Eq. 4.12],

$$E[\mathbf{d}|\boldsymbol{\theta}] = \mathbf{t}(\boldsymbol{\theta}) + \mathbf{A}(\boldsymbol{\theta}) \times 0 + 0 \quad (4.37a)$$

$$= \mathbf{t}(\boldsymbol{\theta}) \quad (4.37b)$$

Moreover, as $\boldsymbol{\eta}$ and $\varepsilon\mathbf{A}(\boldsymbol{\theta})$ are uncorrelated, the variance of the data is now,

$$V[\mathbf{d}|\boldsymbol{\theta}] = E[(\mathbf{d} - \mathbf{t}(\boldsymbol{\theta}))^2] \quad (4.38a)$$

$$= \sigma_\varepsilon^2 \mathbf{A}(\boldsymbol{\theta})\mathbf{A}(\boldsymbol{\theta}) + \mathbf{D} \quad (4.38b)$$

where \mathbf{D} is the $m \times m$ diagonal matrix $\mathbf{D} = \text{diag}(\sigma_j^2)$, being diagonal because the η_j are independent of $\boldsymbol{\theta}$. We are now in a position to calculate the variance of the marginal distribution:

$$V[\mathbf{d}] = V_\theta[E[\mathbf{d}|\boldsymbol{\theta}]] + E_\theta[V[\mathbf{d}|\boldsymbol{\theta}]] \quad (4.39a)$$

$$= V_\theta[\mathbf{t}(\boldsymbol{\theta})] + E_\theta[\sigma_\varepsilon^2 \mathbf{A}(\boldsymbol{\theta})\mathbf{A}(\boldsymbol{\theta}) + \mathbf{D}] \quad (4.39b)$$

which yields the measurement error variance,

$$\Sigma_\xi = E_\theta[\sigma_\varepsilon^2 \mathbf{A}(\boldsymbol{\theta})\mathbf{A}(\boldsymbol{\theta}) + \mathbf{D}] \quad (4.40a)$$

$$= \frac{1}{n} \sum_i \sigma_\varepsilon^2 \mathbf{A}(\boldsymbol{\theta}_i)\mathbf{A}(\boldsymbol{\theta}_i) + \mathbf{D} \quad (4.40b)$$

$$= \frac{1}{n} \sum_i \left[\sigma_\varepsilon^2 \text{diag}\left(\frac{1}{1 - \mathbf{a}_j(\boldsymbol{\theta}_i)}\right) \cdot \text{diag}\left(\frac{1}{1 - \mathbf{a}_j(\boldsymbol{\theta}_i)}\right) + \text{diag}(\sigma_{ij}^2) \right] \quad (4.40c)$$

$$= \text{diag}\left(\frac{1}{n} \sum_i \frac{\sigma_\varepsilon^2}{[1 - \mathbf{a}_j(\boldsymbol{\theta}_i)]^2} + \sigma_{ij}^2\right) \quad (4.40d)$$

with determinant (cf. Eq. 4.16),

$$|\Sigma_\xi| = \prod_{j=1}^m \left[\frac{1}{n} \sum_i \frac{\sigma_\varepsilon^2}{[1 - \mathbf{a}_j(\boldsymbol{\theta}_i)]^2} + \sigma_{ij}^2 \right] \quad (4.41a)$$

and finally the hypocentre variance (unchanged from Eq. 4.17),

$$|\Sigma_{g\xi}| = |V_{\theta}[\mathbf{t}(\theta)]| \quad (4.42)$$

We can therefore express the design criterion as

$$\Phi_{\xi} = \log(|\Sigma_{g\xi}|) - \log(|\Sigma_{\xi}|) \quad (4.43)$$

The optimal network design result incorporating this surface noise source information applied with $\rho = 1$ is displayed in Fig. 4.22. Note that the presence of high surface noise in the centre of the earthquake cluster now prevents stations from being located there.

4.4 When is enough enough?

It is worth considering the first component of the question posed by *Uhrhammer* [1980] (see Section 1.3.2), that is, what is the minimum number of stations required to achieve a desired precision in locating the earthquakes of interest? Here we outline one method of determining this that could be used in a future study.

The value of the design criterion Φ_{ξ} increases with each additional station in a network as the dimension of the problem increases. To compare two different networks with the same number of stations is the trivial matter of comparing the Φ_{ξ} values of each network: however, to compare two networks of different sizes is more problematic. To determine how much the addition of a station has increased the performance of a network, we need to determine by how much that station's addition has increased the probability that the earthquake set is accurately located.

The probability density of an earthquake's location is

$$P(\theta|d\omega_{\xi}) = \frac{P(d|\theta\omega_{\xi})P(\theta|\omega_{\xi})}{P(d|\omega_{\xi})} \quad (4.44)$$

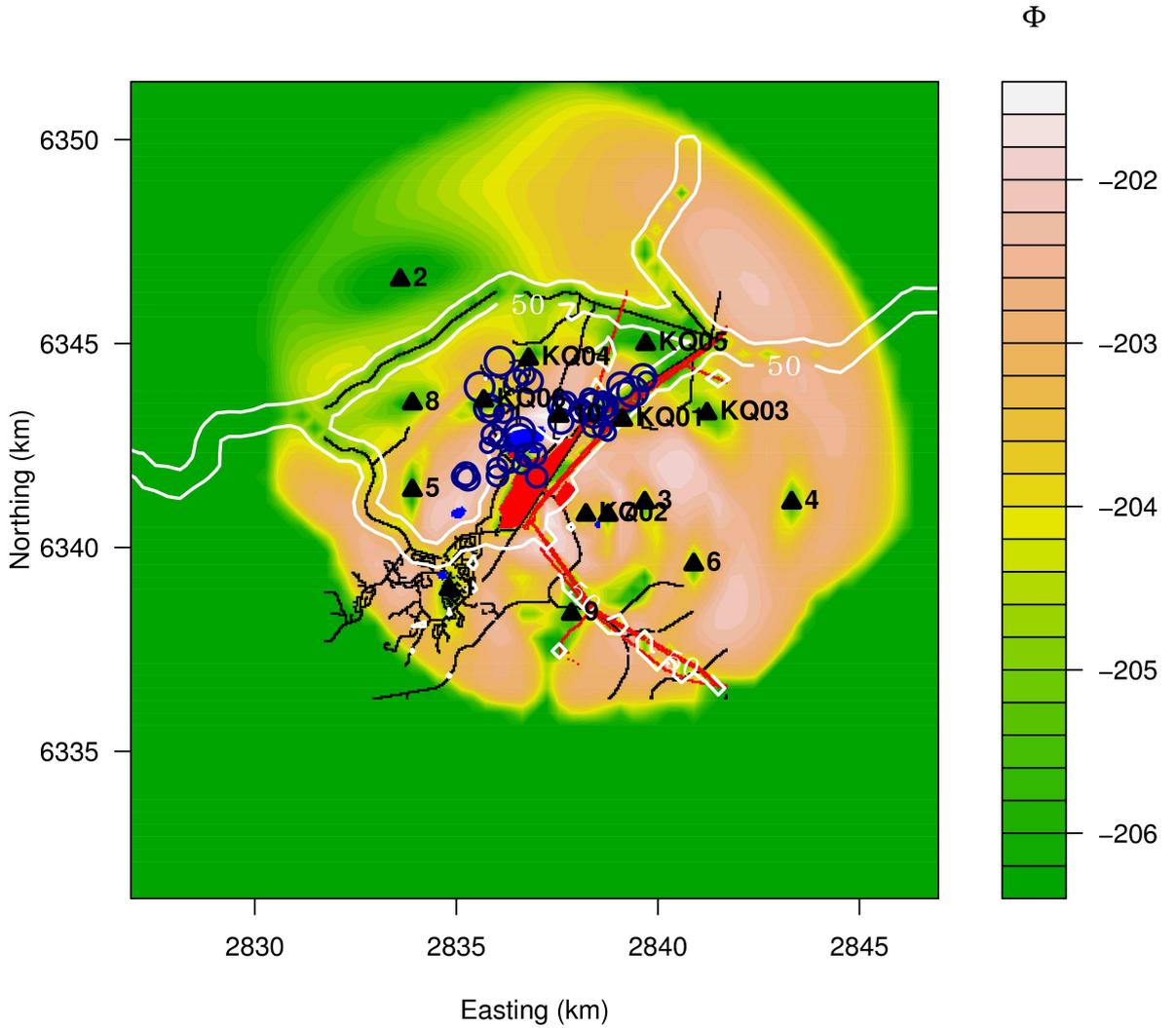


Figure 4.22: Optimal station locations using the surface noise model described in Section 4.3.3 and the P and S-wave velocity and attenuation models as detailed previously. Note that the presence of high surface noise in the centre of the earthquake cluster now restricts stations from being located there.

Based on the derivation of our design criterion we can calculate $P(\mathbf{d}|\boldsymbol{\theta}\omega\xi)$ and $P(\mathbf{d}|\omega\xi)$ straightforwardly, but there is no simple analytical method of calculating $P(\boldsymbol{\theta}|\omega\xi)$. In other words, earthquake location algorithms must be employed to determine $P(\boldsymbol{\theta}|\mathbf{d}\omega\xi)$.

As an example of how this might be done, given the data $\mathbf{d}_{i\xi}$ recorded by the network ξ from earthquake $\boldsymbol{\theta}_i$, the earthquake location software *NonLinLoc* [Lomax et al. 2000; 2009] can be used to provide $P(\boldsymbol{\theta}_i|\mathbf{d}_{i\xi}\omega\xi)$ in the form of a ‘scatter file’ of possible locations for earthquake $\boldsymbol{\theta}_i$, with each location assigned a probability. We can hence calculate the weighted variance-covariance matrix $WV[\boldsymbol{\theta}_i|\mathbf{d}_{i\xi}\omega\xi]$ for earthquake $\boldsymbol{\theta}_i$ using these pos-

sible locations weighted by their probability. This weighted variance-covariance matrix has a determinant

$$\Omega_{i\xi} = |WV [\theta_i | d_{i\xi} \omega \xi]| \quad (4.45)$$

For earthquakes $i = 1, 2, \dots, n$ we form a distribution Ω_ξ . We require our network to be optimal for all the earthquakes — avoiding distributions with large tails — and therefore require a measure of quality that takes into account the entire form of this distribution. We hence calculate the weighted 90% quantile of the distribution Ω_ξ , denoted γ_ξ . The smaller γ_ξ is, the better the earthquake locations are constrained and the better the performance of the network.

To determine when enough stations are enough, and measure the improvement in earthquake location resolution obtained by adding an additional station, we compare γ_{ξ_j} for different values of j corresponding to the addition of stations $j = 1, 2, \dots, m$,

$$r_j = 1 - \frac{\gamma_{\xi_j}}{\gamma_{\xi_{j-1}}} \quad (4.46)$$

Here $100r_j$ corresponds to the percentage improvement obtained by adding station j to a $j - 1$ station network.

A threshold must then be determined, at which we state that the increase in resolution supplied by any additional stations will not be worth the cost of their deployment: for example, a station must provide at least a 10% improvement for the network to be worth the cost of its deployment. Then when $r_j < 0.1$ we can state that the network is of sufficient size and no new stations need be added.

4.5 Final designs

For the three analysis regions identified in Fig. 4.5, we display the optimal network designs in Figures 4.23(a)–4.24(a), with corresponding station coordinates in Table 4.7, and a final figure displaying all three configurations in Fig. 4.24(b). The designs incorporate 3D variations in P- and S-wave velocity, temperature dependent attenuation, and noise sources.

First, we have the optimal network design for the production region earthquakes, the construction of which has been detailed in the preceding sections. A zoom in of the final result is displayed in Figure 4.23(a) and the station coordinates detailed in Table 4.7.

We then display the results of the local region [Fig. 4.23(b)] and full region [Fig. 4.24(a)] with station coordinates detailed in Table 4.7. The surface noise scaling factor must be increased to $\rho = 9$ for these larger data sets. As expected, extending the distribution of the earthquake set causes the stations to spread outwards as we move from the production region, to the local region, to the full region. The most notable feature present in Fig. 4.24(b), with all three networks displayed, is a southwest facing half circle with a south-southwest line extending out from the arc. These define the regions in which a new station would consistently improve earthquake location constraint due to the northeast clustering of the existing stations.

Table 4.7: Optimal seismic stations for locating earthquakes in the three different analysis regions. See Figs. 4.23(a)–4.24(a) for illustrations.

	Production region		Local region		Full region	
Rank	Easting	Northing	Easting	Northing	Easting	Northing
1	2834819	6338977	2821767	6326229	2821767	6326229
2	2833605	6346566	2833605	6339585	2851817	6356279
3	2839676	6341102	2840890	6335942	2833908	6337156
4	2843318	6341102	2830570	6347173	2841193	6335942
5	2833908	6341406	2837247	6338977	2825106	6340192
6	2840890	6339585	2844836	6348994	2841193	6326229
7	2838765	6340799	2832694	6334728	2833908	6341102
8	2833908	6343531	2839979	6338674	2836033	6339585
9	2837854	6338370	2834515	6340495	2832391	6332907
10	2837551	6343227	2837247	6340799	2839372	6338370

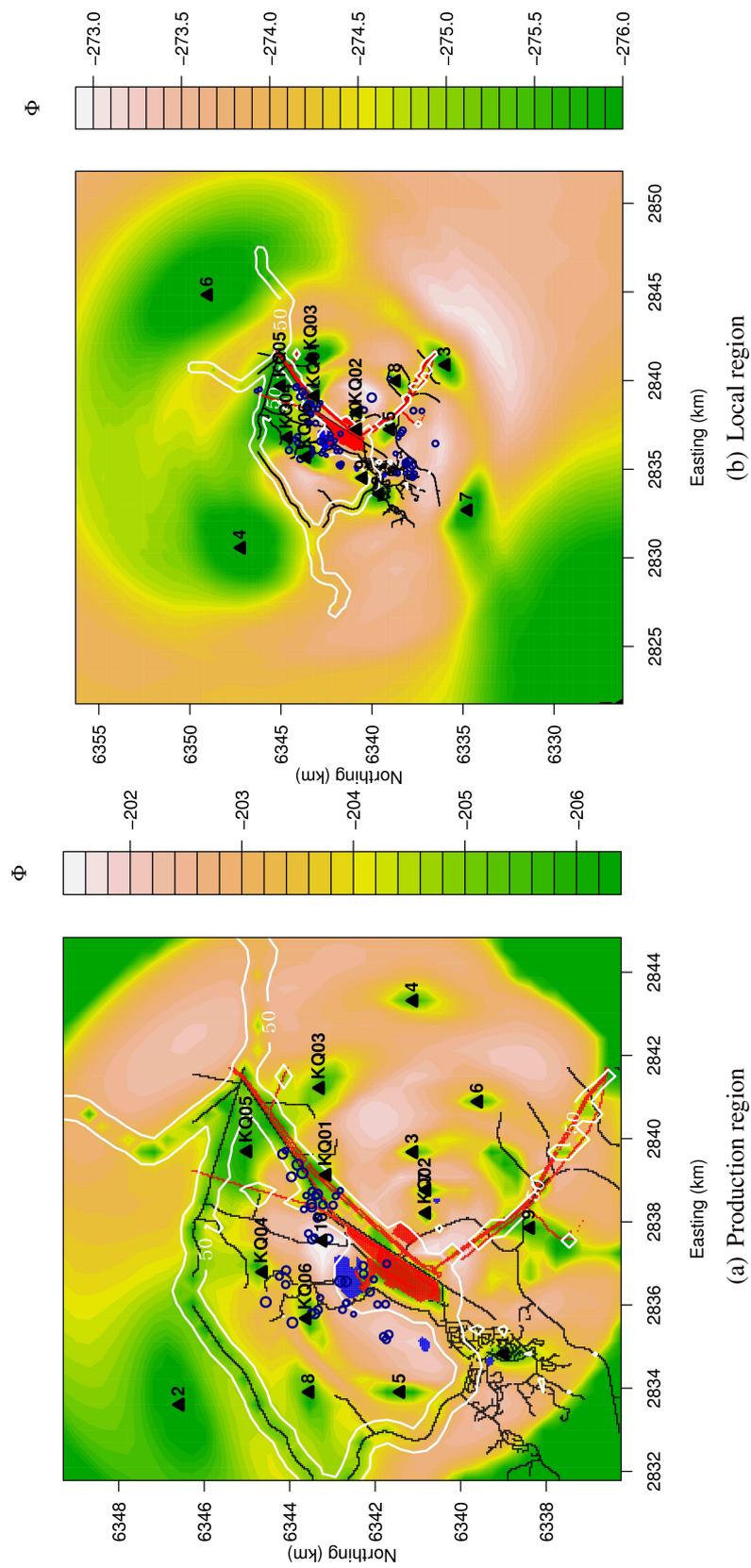


Figure 4.23: Optimal seismic stations to locate a) the production region earthquakes, b) the local region earthquakes. See Table 4.7 for station coordinates.

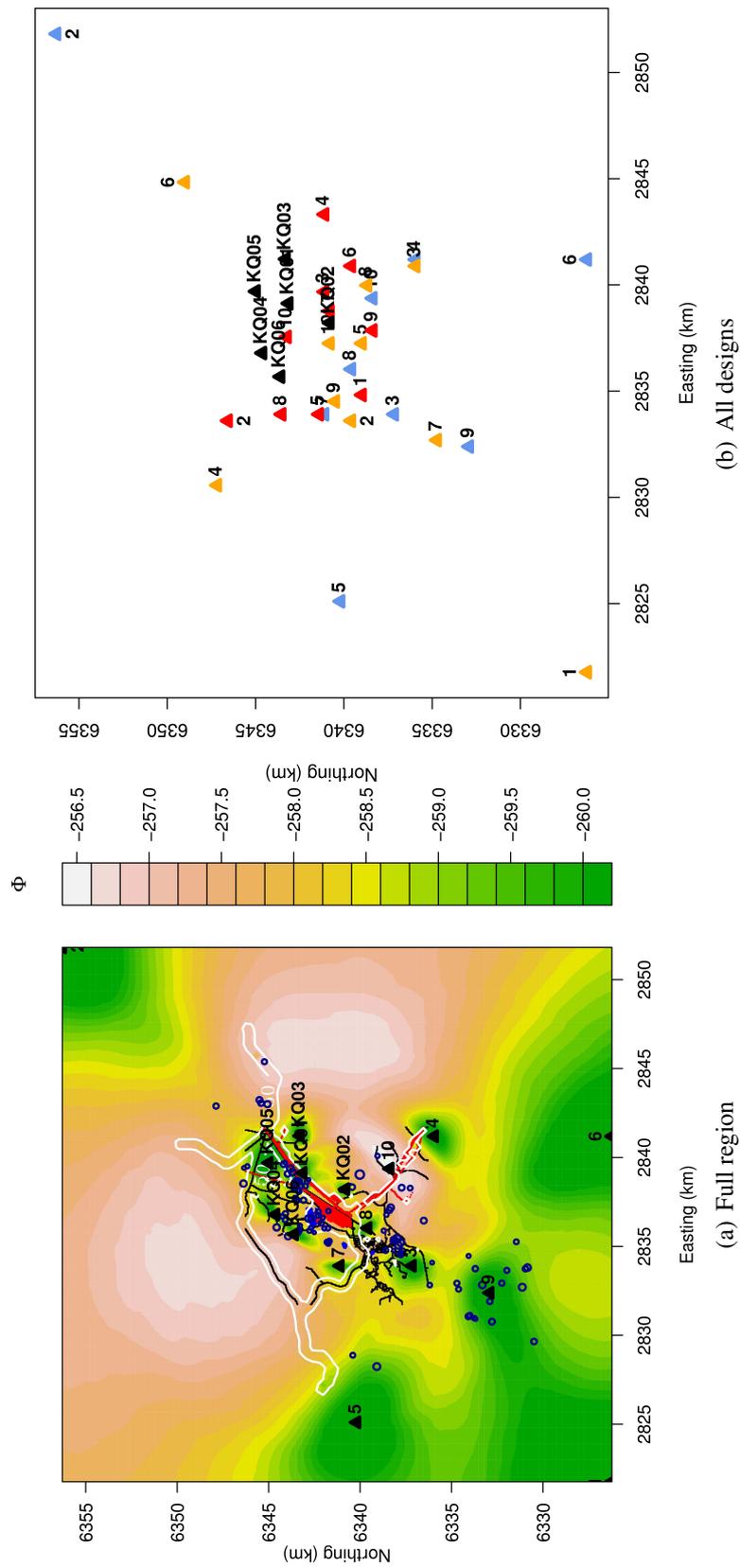


Figure 4.24: Optimal seismic stations to locate a) the full region earthquakes, see Table 4.7 for station coordinates. b) Comparison of the three different networks shown in Figs. 4.23(a)–4.24(a). The red triangles display the optimal configuration for the production region, the orange triangles display the optimal configuration for the local region, and the blue triangles display the optimal configuration for the full region. Note that with the wider region of earthquakes some station locations are now influenced by the edge of our calculation grid [see Fig. 4.8].

As mentioned previously, the production region earthquakes are of particular interest to MRP. We can constrain the network to reflect this fact by assigning a weighting to the different regions of earthquakes. Replacing $p(\boldsymbol{\theta}|\boldsymbol{\omega})$ with a scaled version $w(\boldsymbol{\theta})p(\boldsymbol{\theta}|\boldsymbol{\omega})$ using a set of prior weights $w(\boldsymbol{\theta})$ alters the total variance calculation [Eq. 4.5]:

$$V[\mathbf{d}] = E_{w(\boldsymbol{\theta})} [V[\mathbf{d}|\boldsymbol{\theta}]] + V_{w(\boldsymbol{\theta})} [E[\mathbf{d}|\boldsymbol{\theta}]] \quad (4.47a)$$

$$= \sigma_\varepsilon^2 \text{diag} \left(\frac{\sum_{i=1}^n w(\boldsymbol{\theta}_i) [1 - a_j(\boldsymbol{\theta}_i)]^{-2}}{\sum_{i=1}^n w(\boldsymbol{\theta}_i)} \right) + \frac{\sum_{i=1}^n w(\boldsymbol{\theta}_i) (\mathbf{t}(\boldsymbol{\theta}_i) - \bar{\mathbf{t}}_w) (\mathbf{t}(\boldsymbol{\theta}_i) - \bar{\mathbf{t}}_w)^T}{\sum_{i=1}^n w(\boldsymbol{\theta}_i)} \quad (4.47b)$$

$$= \Sigma_\xi^w + \Sigma_{g\xi}^w \quad (4.47c)$$

where

$$\bar{\mathbf{t}}_w = \frac{\sum_{i=1}^n w(\boldsymbol{\theta}_i) \mathbf{t}(\boldsymbol{\theta}_i)}{\sum_{i=1}^n w(\boldsymbol{\theta}_i)} \quad (4.48)$$

and $w_{\boldsymbol{\theta}_i}$ is the weighting assigned to earthquake $\boldsymbol{\theta}_i$.

We leave the production region earthquakes assigned a weight $w(\boldsymbol{\theta}) = 1$, and investigate the results of changing the weighting of the other regions' earthquakes. In Figs. 4.25(a)–4.26(b) we display weightings for the local region $w(\boldsymbol{\theta}) = 0.5$, full region $w(\boldsymbol{\theta}) = 0.5$; local region $w(\boldsymbol{\theta}) = 0.5$, full region $w(\boldsymbol{\theta}) = 0.2$; and local region $w(\boldsymbol{\theta}) = 0.1$, full region $w(\boldsymbol{\theta}) = 0.1$, respectively. The station coordinates are detailed in Table 4.8, and a figure displaying all three configurations is shown in Fig. 4.26(b). Note that with the wider region of earthquakes some station locations are now influenced by the edge of our calculation grid [see Fig. 4.8], suggesting that a larger calculation grid would be preferable. Due to the greater importance of the production region earthquakes, and the computational time required to perform raytracing to expand this calculation grid, we have chosen to remain with the current grid for our demonstration.

Although we have investigated only simple weighting schemes, as a weighting is assigned to each earthquake individually, a complex weighting scheme may be used if desired.

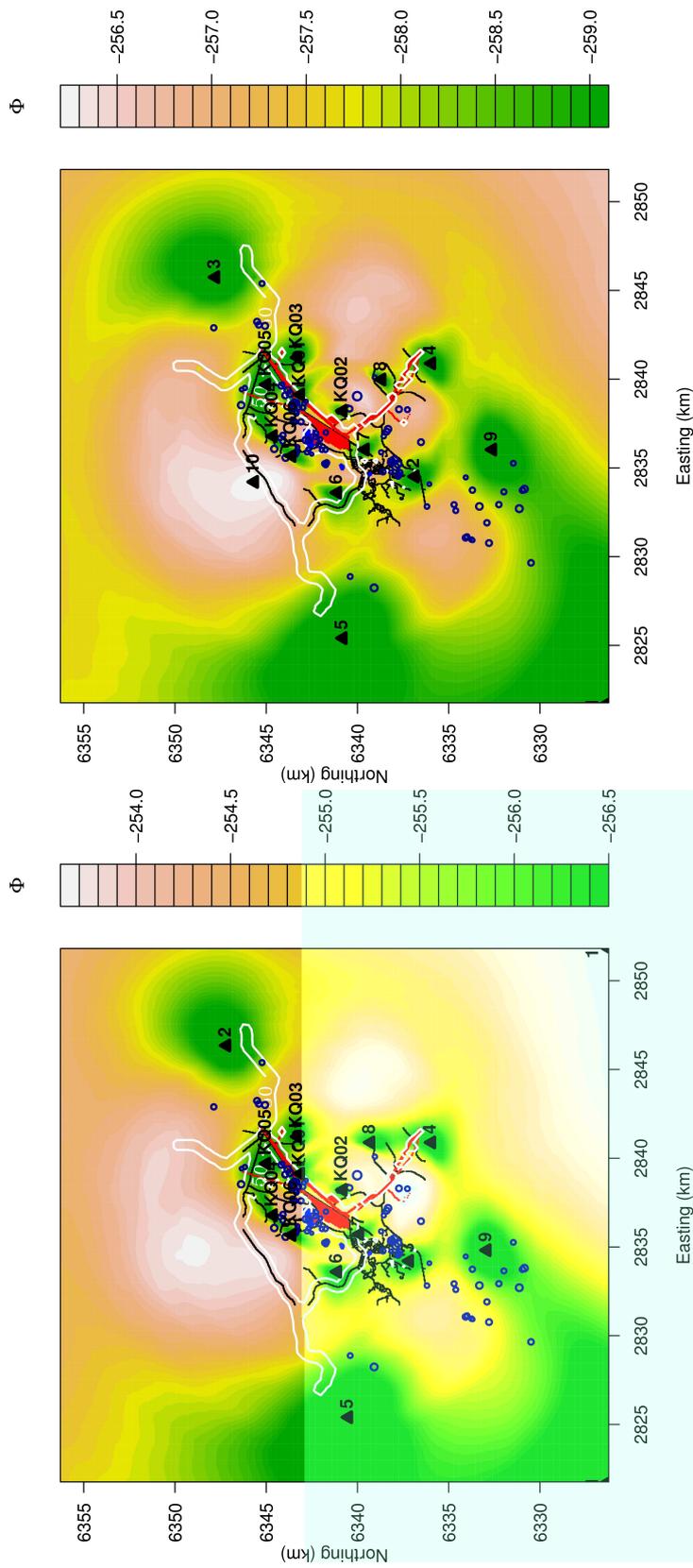
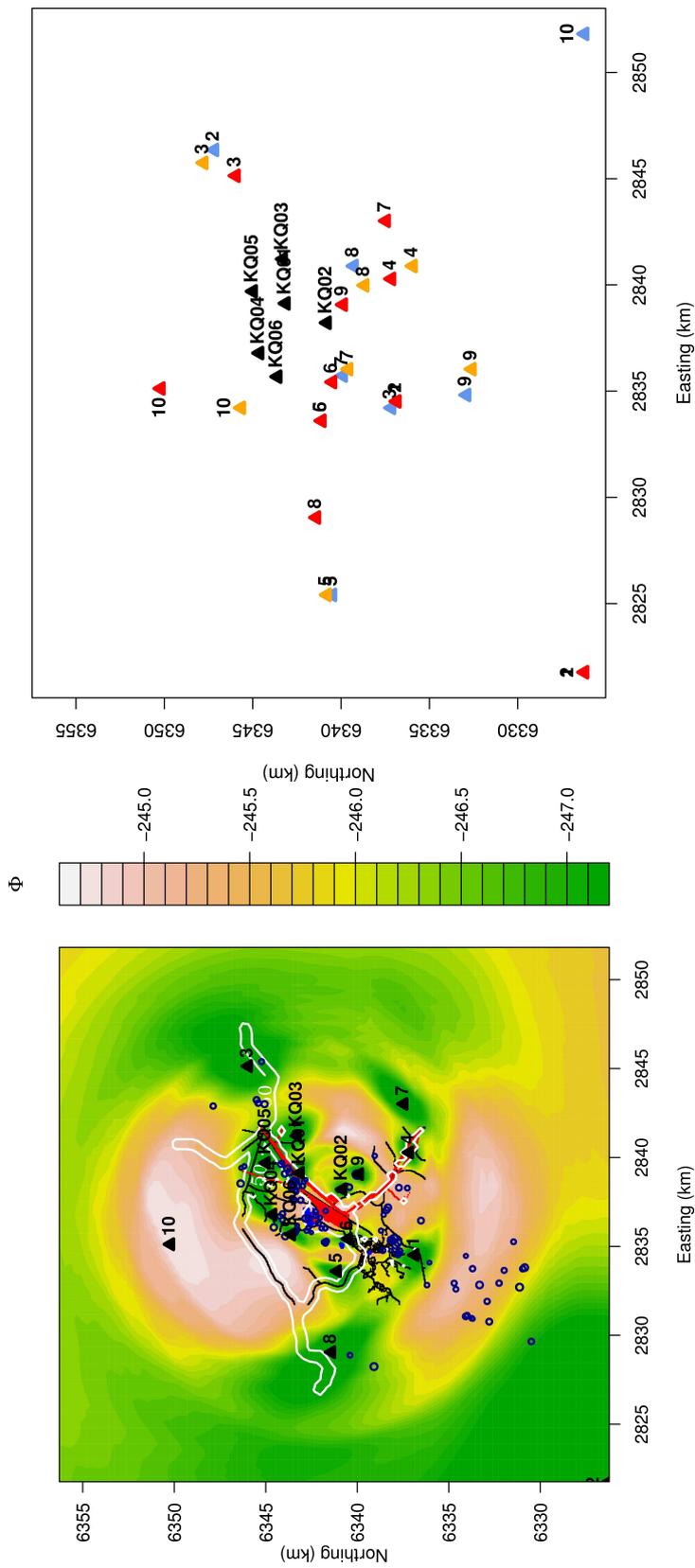


Figure 4.25: Optimal seismic stations to locate the full region earthquakes with production region earthquakes assigned a weight $w(\theta) = 1$, and a) local region $w(\theta) = 0.5$, full region $w(\theta) = 0.5$, b) local region $w(\theta) = 0.2$, full region $w(\theta) = 0.2$. The station coordinates are listed in Table 4.8. Note that with the wider region of earthquakes some station locations are now influenced by the edge of our calculation grid [see Fig. 4.8].



(a) local region $w(\theta) = 0.1$, full region $w(\theta) = 0.1$

(b) All designs

Figure 4.26: Optimal seismic stations to locate the full region earthquakes with production region earthquakes assigned a weight $w(\theta) = 1$, and a) local region $w(\theta) = 0.1$, full region $w(\theta) = 0.1$, the station coordinates are listed in Table 4.8. b) Comparison of the three different networks shown in Figs. 4.25(a)–4.26(b). The red triangles display the optimal configuration for local region $w(\theta) = 0.1$, full region $w(\theta) = 0.1$, the orange triangles display the optimal configuration for local region $w(\theta) = 0.5$, full region $w(\theta) = 0.2$, and the blue triangles display the optimal configuration for local region $w(\theta) = 0.5$, full region $w(\theta) = 0.5$. Note that with the wider region of earthquakes some station locations are now influenced by the edge of our calculation grid [see Fig. 4.8].

Table 4.8: Optimal seismic stations to locate all earthquakes with different weighting schemes applied to each region. Production region earthquakes are assigned a weight $w(\boldsymbol{\theta}) = 1$ for all schemes. $lw(\boldsymbol{\theta})$ is the local region weighting applied, and $fw(\boldsymbol{\theta})$ is the full region weighting applied. See Figs. 4.25(a)–4.26(b) for a graphical comparison.

	$lw(\boldsymbol{\theta}) = 0.5$	$fw(\boldsymbol{\theta}) = 0.5$	$lw(\boldsymbol{\theta}) = 0.5$	$fw(\boldsymbol{\theta}) = 0.2$	$lw(\boldsymbol{\theta}) = 0.1$	$fw(\boldsymbol{\theta}) = 0.1$
Rank	Easting	Northing	Easting	Northing	Easting	Northing
1	2821767	6326229	2821767	6326229	2834515	6336853
2	2846353	6347173	2834515	6336853	2821767	6326229
3	2834212	6337156	2845746	6347780	2845139	6345959
4	2840890	6335942	2840890	6335942	2840283	6337156
5	2825409	6340495	2825409	6340799	2833605	6341102
6	2833605	6341102	2833605	6341102	2835426	6340495
7	2835730	6339888	2836033	6339585	2843014	6337460
8	2840890	6339281	2839979	6338674	2829052	6341406
9	2834819	6332907	2836033	6332603	2839069	6339888
10	2851817	6326229	2834212	6345655	2835123	6350208

4.6 Summary

This chapter has outlined the derivation, implementation, and application of a statistical seismic network design programme. We have demonstrated the evolution of both the implementation and resulting network design, from those based on a uniform velocity model and attenuation model, to those incorporating 3D velocity structure and attenuation models for both P and S travel times, a surface noise model, and sets of earthquakes weighted differently. The algorithm has been implemented in codes that supply rapid results and facilitate easy adaption to other data sets. Computation times depend on the size and number of grid points in the calculation region, and the number of stations and earthquakes; for the Kawerau data set, results are supplied within $\sim 5 - 10$ minutes using a laptop computer.

Sections 4.3.2–4.3.3 demonstrate how each additional parameter incorporated in the design model yields a different optimal network design. These sections highlight the need for: an accurate ray tracer for 3D velocity and attenuation model incorporation; an ambient surface noise study to be performed; and the experience required when assigning the attenuation scaling value depending on the magnitudes of the earthquakes of interest. Section 4.5 demonstrates the need to determine the desired weighting of each earthquake, as

this has significant impacts on the design results. It is also important to ensure that a large enough calculation grid is used so that the grid boundaries do not influence the design results, however, if stations are only desired within a certain area then the design produced will be optimal for the calculation area assigned.

Although we have constructed a complex design model, Sections 4.2 and 4.3.1 highlight that even when a simplistic model is used for velocity and attenuation, our network design algorithm is still of value, since the depths of earthquakes and their epicentres both affect the station geometry that will optimally locate them. When combining a set of earthquakes (or sets) in an irregularly shaped 3D domain, this algorithm can reduce the guess work involved in ensuring the geographic coverage of the domain required to geometrically constrain the earthquakes. Incorporating existing stations in the analysis makes this fact even more apparent.

The network design produced is optimal for the set of target earthquakes provided. By using weightings and assigning the input earthquake locations to cover a sufficient number of spatial samples from a geometric target shape (e.g. a parallelepiped or sphere) or a particular geological domain (e.g. a target reservoir) the network can be straightforwardly customised to a region rather than a set of earthquakes. Moreover, as the algorithm utilises only travel times and attenuation values, a 3D grid could equally be investigated to allow for the possibility of borehole instrumentation.

Chapter 5

Discussion and conclusions

This study has addressed methods of seismologically characterising actively exploited geothermal fields with respect to two principal objectives:

1. Automated microearthquake detection using a matched filter technique and robust hypocenter estimation for earthquakes associated with fluid injection at Rotokawa, to elucidate how spatiotemporal patterns of induced events relate to injection processes;
2. A pilot study of optimal seismic network design that will guide the expansion of the existing seismic monitoring programme at Kawerau geothermal field.

We have made the following contributions towards these two objectives:

Chapter 2: Event Detection

1. We have developed and implemented an algorithm to perform low-magnitude earthquake detection using a matched filter technique and a suite of high-quality master events;
2. The algorithm expands on the preliminary cross-correlation codes supplied by Charlotte Rowe [Rowe 2002] by creating entire network correlation coefficients (NCC) which has the advantage of increasing the gain of small events due to constructive

interference of CCs, and decreasing the gain of spuriously high CCs via destructive interference;

3. Synthetic testing to determine an appropriate detection threshold for events recorded on at least six stations yields a threshold value of $8.72\text{MAD}(\text{NCC}) + \text{median}(\text{NCC})$.
4. Using a suite of 14 master events, application of the algorithm to the three component, nine station network at Rotokawa geothermal field, between 29 September 2008 and 28 February 2009, yields 2461 slave earthquakes spanning a magnitude range $-0.4 \leq M \leq 2.6$ with a mean magnitude of $M = 0.47$;
5. The earthquakes detected with each master event exhibit high waveform similarity over ~ 3 orders of magnitude, and appear to follow a Gutenberg-Richter power law with catalogue completeness down to $M \sim 0$.

The earthquake detection algorithm has been implemented in codes that are easily adaptable to other studies. We have applied this algorithm in a geographically small region so that the premise that our master event set is representative of the locations and focal mechanisms of all events within the region is an acceptable one. For applications to larger regions, care must be taken in choosing master events, and recognising the bias that the master event set creates in the resultant earthquake catalogue. In the Rotokawa case, we are unable to formally confirm whether the detected earthquakes span the entire catalogue of existent seismicity, but in a region with a reasonable number of large events, or in an application to an aftershock sequence, one method of confirming this would be to check that the Gutenberg-Richter law is complied with. Recent successful regional studies using cross-correlation detection include *Gibbons et al.* [2007] and *Schaff* [2009], whilst cross-correlation applied to an aftershock sequence has been performed *Gibbons et al.* [2007].

Any future study will also have to examine carefully the affects of the chosen detection threshold, as the identification of a detected event is highly dependent on this threshold. This study has used only the maximum correlation coefficient within each window, there is the possibility for events to be missed that are closer in time than the window shift step used.

As it is complex to calculate magnitudes for low-amplitude earthquakes, we have calculated magnitudes using an amplitude ratio with the master events whose magnitudes were provided by GNS Science. For future work, the recent paper *Rubinstein and Ellsworth* [2010] offers a more sophisticated approach — SVD of correlated waveforms — to calculating magnitudes for low magnitude events.

Chapter 3: Earthquake Location and Spatiotemporal Analysis

1. We have used the probabilistic earthquake location algorithm *NonLinLoc* to locate 2101 earthquakes within the Rotokawa Geothermal field, obtaining hypocentres with an average RMS time residual of 0.11 s, and a mean hypocentre uncertainty of 0.36 km;
2. The dominant locus of seismicity lies between 1.0–2.5 km depth consistent with the location of the Rotokawa Andesite forming the Rotokawa reservoir;
3. Focal mechanism solutions for the 14 master events are predominantly normal, with half displaying a large strike-slip component, and are consistent with the results of previous studies within the TVZ;
4. A stress estimate using the calculated focal mechanism solutions provides a northeast–southwest S_{Hmax} direction consistent with the expected maximum horizontal stress orientation in the TVZ;
5. The location of earthquakes detected with master event 290, and the normal-faulting focal mechanism of the master event are consistent with a previously inferred fault;
6. Seismicity within 300 m horizontally of the RK20 injection well’s feedzones initially shows a correlation with injection flow rates with a ~ 2 day lag, but seismicity reduces ~ 10 weeks after injection, perhaps due to the stress in the region reaching a relatively stable state;

7. Seismicity within the injection region and close to RK20 is likely to be injection-induced, with one portion of the injectate returning to the production region, while the other either migrates southeast out of the field or remains to the south-east of RK20 within the injection region; the origin of seismicity within the production region is unclear.

The highest source of error in the computed hypocentres will likely be induced by inaccurate phase-picks. This is a result of the difficulty of refining phase-picks for a large set of very low-magnitude earthquakes. Methods presented in a number of studies such as those of *Rowe* [2002], *Yang* [2004], *Song et al.* [2010], *Galiana-Merino et al.* [2008], and *Gentili and Michelini* [2006] for low-SNR earthquake phase picking, hold promise for future work.

The other large source of hypocentre error was due to the seismic station geometry. We recommend that the addition of a seismic station bridging the gap between RT05 and RT10 would be advantageous for determining RK20 near-vicinity earthquake locations with higher accuracy.

Due to the low number of observations for each event supplying insufficient constraints for the double-difference location calculations, we were unable to implement *hypoDD* effectively. In this case, a better strategy might be to use the Joint Hypocentre Determination (JHD) method [*Console and Giuntini* 2006] and locate the master events using Geonet stations as additional constraints, and to then locate each cluster of slave events relative to their master event's hypocentre.

Chapter 4: Statistical Experimental Design

1. We have derived and implemented a statistical seismic network design algorithm, and demonstrated its utility in even simplistic cases.
2. The algorithm has been applied to the extension of the seismic network at Kawerau geothermal field and an early version was usefully used during the short-term/rapid-response network design following the Darfield earthquake [Appendix D.3];

3. Unlike previous seismic network design algorithms [e.g. *Coles and Curtis* 2011; *Curtis et al.* 2004; *Rabinowitz and Steinberg* 1990; *Steinberg et al.* 1995], we have presented methods for realistically incorporating 3D velocity structure and attenuation models for both P and S travel times, a surface noise model, and the ability to apply complex weighting functions to the earthquake set, and have shown how each additional parameter incorporated in the design model changes the optimal network design produced;
4. The algorithm has been implemented in codes that supply rapid results and facilitate easy adaption to other data sets.
5. Computation times depend on the size and number of grid points in the calculation region, and the number of stations and earthquakes; for the Kawerau data set, results are supplied within $\sim 5 - 10$ minutes using a laptop computer;
6. We have also provided guidelines for a method to determine how many instruments is sufficient for a given earthquake set.

Preliminary application of the seismic network design algorithm to the Kawerau geothermal field highlighted the need for an accurate ray tracer for 3D velocity and attenuation model incorporation; an ambient surface noise study to be performed; the experience required when assigning the attenuation scaling value depending on the magnitudes of the earthquakes of interest; and the determination of the desired weighting of each earthquake. The earthquake set used is implemented as a perfectly known set of discrete points, but, a surface or a volume could be investigated in future studies by considering a sufficient set of earthquake sample points from the surface/volume. For example, a fault could be investigated by a volume surrounding the fault sampled by a set of Gaussian-weighted earthquakes.

The results of Chapter 3 highlighted the importance of station geometry in constraining earthquake locations — as the worst located earthquakes occurred to the north and southeast, due to gaps in the network between RT02 and RT07, and RT05 and RT10.

Finally, we note that the design criterion derived here can be applied in the design of observational data collection networks constraining *any parameter* of interest. For example,

rather than hypocentre resolution alone, seismic networks could be designed to also produce high quality focal mechanisms by finding the station locations giving optimal coverage of the focal sphere; other observation geophysics such as GPS networks could be designed to optimally resolve earth deformation in areas of interest. In such cases, the design criterion is the same as used here, but the method by which it is implemented will differ depending on the problem at hand.

Appendix A

Signal Processing

A.1 Multi-taper method

To reduce spectral leakage and bias, tapering is applied to a time series before frequency spectrum estimation. We apply the multi-taper method proposed by *Thomson* [1982] using the inverse iteration procedure outlined by *Bell et al.* [1993]. Single taper methods face a trade-off of losing portions of the time series in order to reduce spectral leakage in the frequency domain: the use of multi-tapers alleviates this trade-off. The multi-taper method uses a set of orthogonal tapers and averages the resulting set of spectral estimates. This orthogonal condition ensures that each taper captures new information, and the averaging over their results provides a spectrum with better stability and a lower variance. Also, multi-tapering can be used in an automatic fashion, rather than requiring the design of a filter specific to the situation. A detailed discussion of the advantages of using multitaper spectrum estimation can be found in *Percival and Walden* [1993].

We calculate eigentapers $w^{(r)}$ up to the eighth order, $r = 0, 1, \dots, 8$, with a variable bandwidth $W = 8/N_{ik}$, by discrete prolate spheroidal sequences (DPSS) [*Bell et al.* 1993]. Discrete prolate spheroidal wave functions are eigenfunctions of the Dirichlet kernel. These

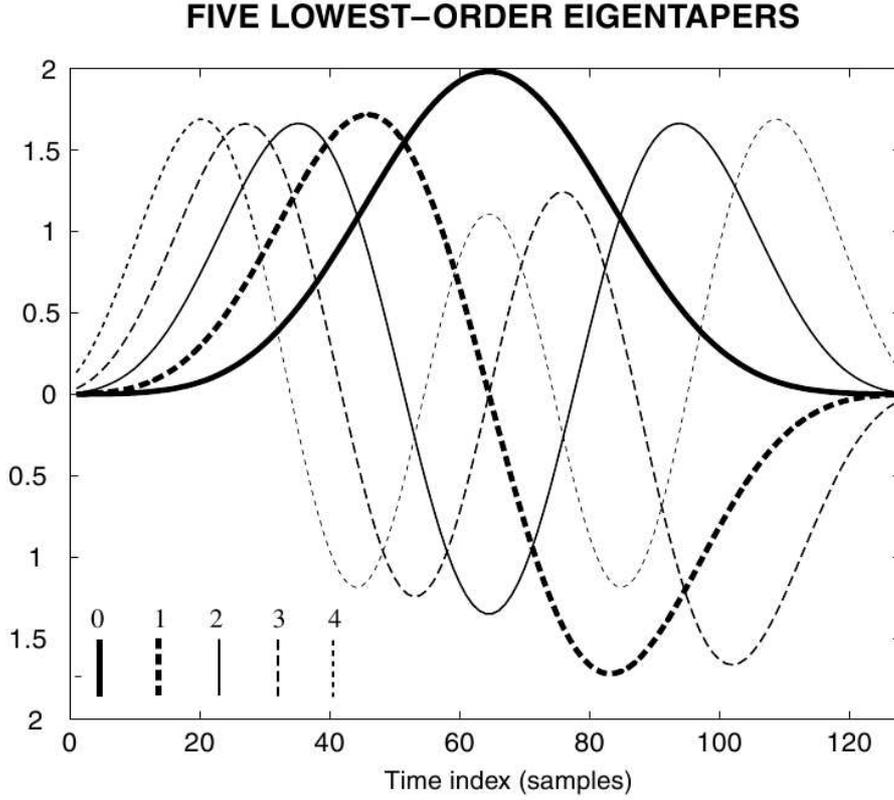


Figure A.1: The five lowest-order multi-taper functions (prolate spheroidal eigentapers) for a 128 sample long window. Higher-order tapers have increasingly steep initial slopes, therefore, spectral leakage becomes greater with higher-order tapers. Figure from *Rowe* [2002]

eigenfunctions, denoted $U_r(R, W; f)$, $r = 0, 1, \dots, R - 1$, are solutions to

$$\int_{-W}^W \frac{\sin R\pi(f - f')}{\sin \pi(f - f')} U_r(R, W; f') df' = \lambda_r(R, W) \cdot U_r(R, W; f) \quad (\text{A.1})$$

where the bandwidth W is $(0 < W < 1/2)$. The functions are ordered by their eigenvalues:

$$1 > \lambda_0(R, W) > \lambda_1(R, W) > \dots > \lambda_{R-1}(R, W) \quad (\text{A.2})$$

The first $2RW$ eigentapers have eigenvalues that are extremely close to 1. The discrete prolate spheroidal wave functions are orthogonal over the interval $(-W, W)$, and their Fourier transforms provide the discrete prolate spheroidal sequences (DPSS) with which we window the time series before estimating its spectrum [Fig. A.1; *Rowe* 2002]. These provide a linearly independent series of eigenspectra for the time series, which may be combined

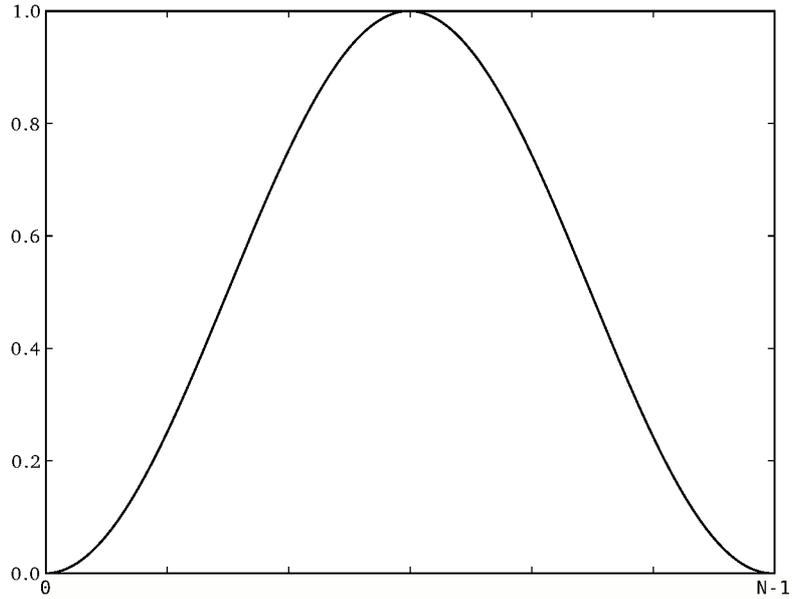


Figure A.2: A Hann window

in a weighted sum to estimate its true spectrum. As spectral leakage becomes greater with higher-order tapers [Fig. A.1], within the first eight eigentapers we only apply those tapers with eigenvalues > 0.5 . During the calculation of eigenvalues, if one order of DPSS does not converge *Bell et al.* [1993] then a Hann window is used instead [Fig. A.2; *Blackman and Tukey* 1959]. A Hann window is a “raised cosine” single taper method [Eq. A.3] to reduce spectral leakage,

$$w[n] = 0.5 \left(1 - \cos \frac{2\pi n}{N_{ik} - 1} \right), 1 \leq n \leq N \quad (\text{A.3})$$

A.2 Coherence filtering

Application of the cross-coherency weighting allows for recovery of a greater portion of the spectrum during cross-correlation than would be achieved with standard bandpass filtering methods, while enhancing the coherent portion of the signals being compared [*Stankova et al.* 2008]. In the absence of highly correlated noise, adaptively filtering on the basis of

high cross-coherency may yield superior correlation results to those that use standard *a priori* filters [DeShon et al. 2007].

SYNTHETIC EXAMPLE - ADAPTIVE PRE-FILTERING

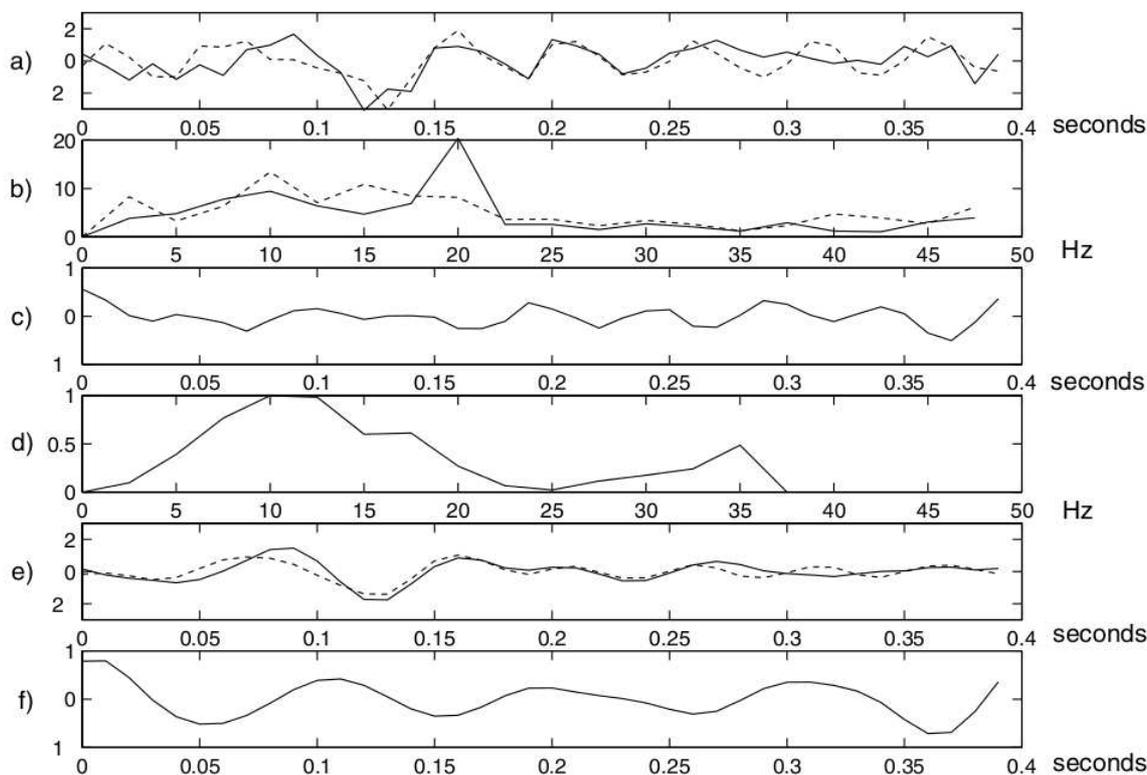


Figure A.3: Illustration of adaptive coherency pre-filtering. a) Two synthetic seismograms lagged by one sample with independent broadband Gaussian noise and a sampling rate of 100 Hz. b) Amplitude-frequency spectra of unfiltered seismograms. c) Cross-correlation function for unfiltered seismograms of (a). The maximum cross-correlation coefficient is approximately 0.5 and the maximum cross-correlation lag is at zero samples. d) Coherency calculated from spectra of 1b. Although most coherent energy resides below 20 Hz, cross-coherency has a small peak at 35 Hz. *A priori* low pass filtering may reject this energy, which could have an important effect on the cross-correlation. e) Adaptively filtered seismograms. Note significant reduction in the random noise constituent and overall increased waveform similarity. f) Re-computed cross-correlation function for the waveform pair, showing a maximum cross-correlation coefficient > 0.9 and a retrieved correct coarse lag of 1 sample. Figure and text from Aster and Rowe [2000].

Appendix B

Code

The preliminary codes provided by Charlotte Rowe (written ca. 2000) were designed to perform the tasks displayed in Fig. B.1. The modifications we have made to this code have resulted in the processes being performed as displayed in Figs. B.2 & B.3. The cross-correlation and network cross-correlation calculations are time intensive, and we have performed these on a computer grid to reduce the processing time. This Appendix describes the details of this procedure.

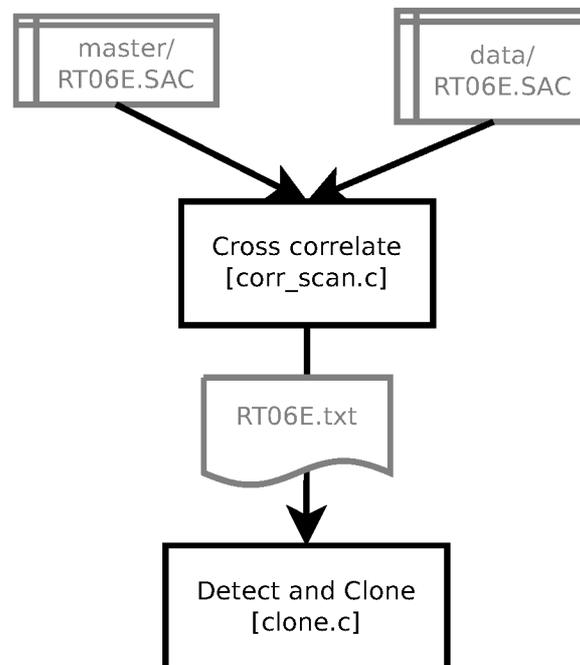


Figure B.1: The code flow from when the correlation code was inherited

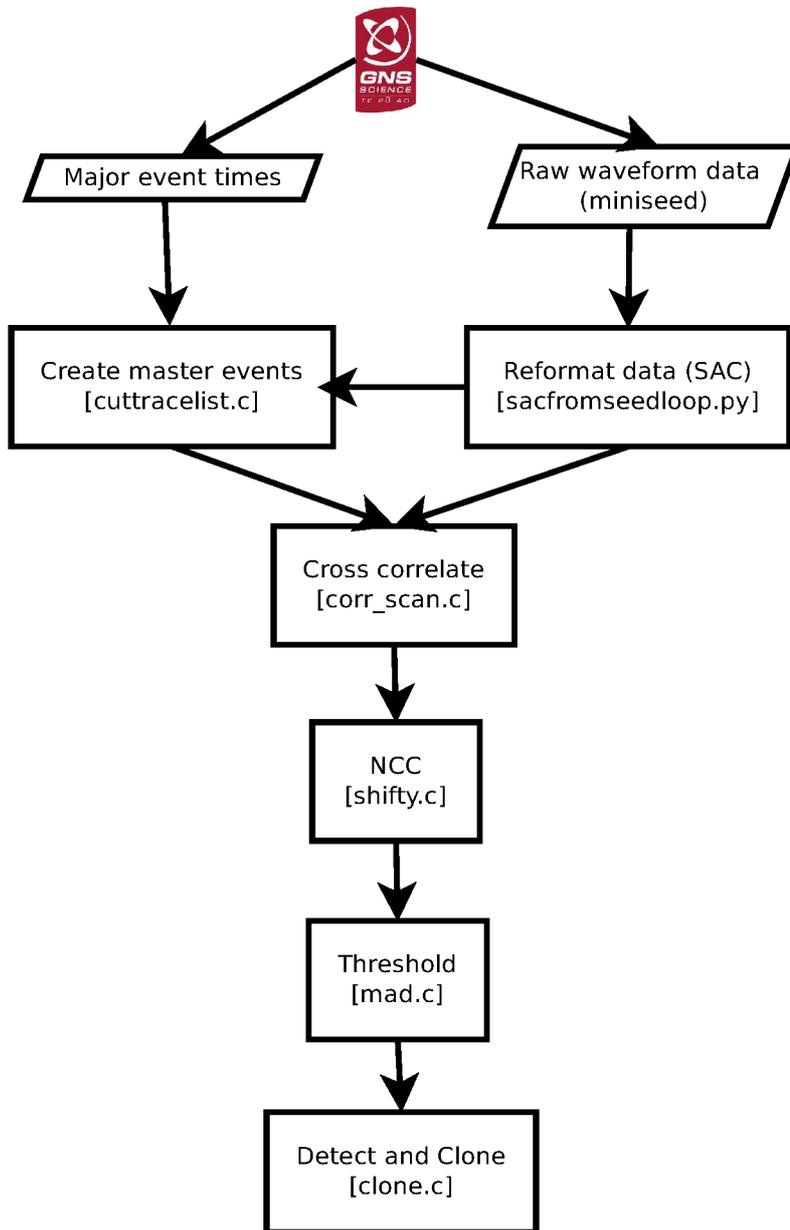


Figure B.2: The code flow we have developed

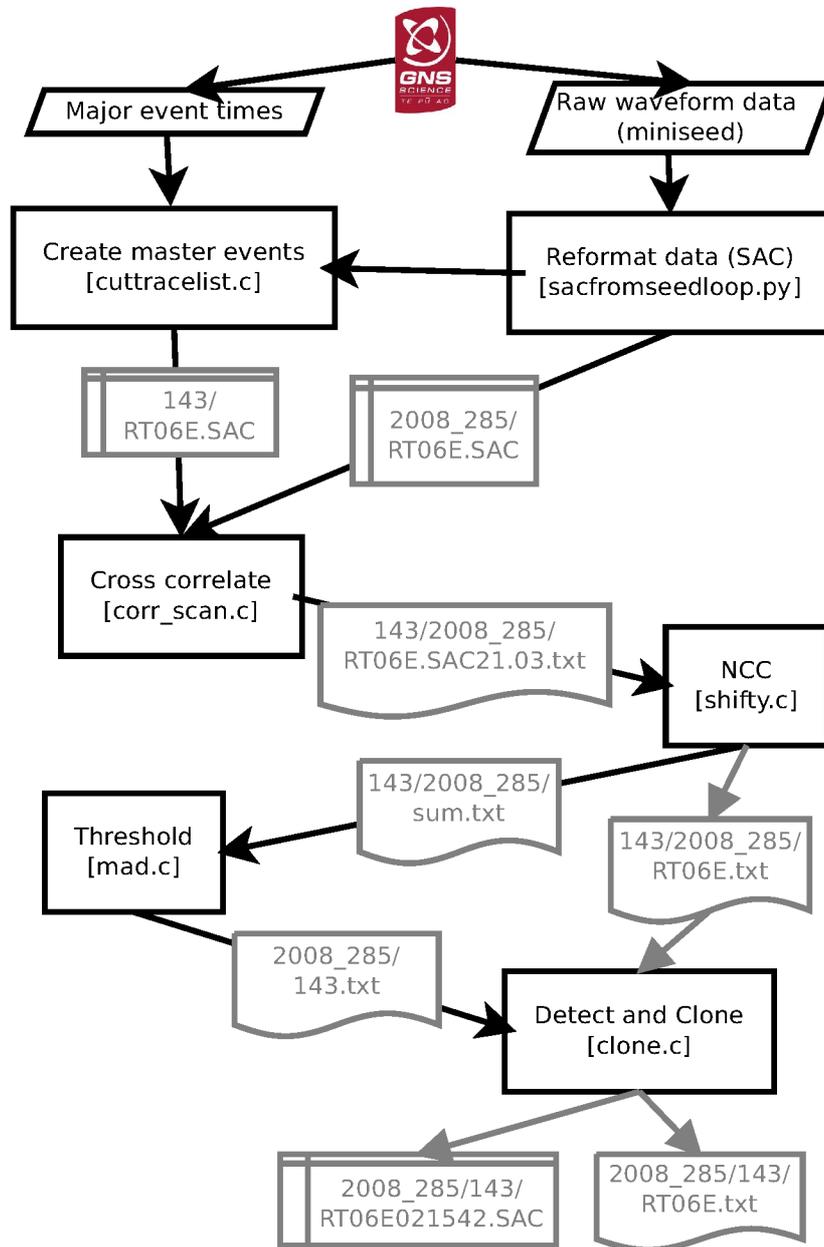


Figure B.3: Our code flow with input and output files

B.1 Convert miniseed to SAC

sacfromseedloop.py: Loops through miniseed data provided by GNS Science, calls **sac_from_mseed.py** (modified from Yannik Behr) to convert data into SAC format and places the SAC data in the desired directory structure [Fig. B.4].

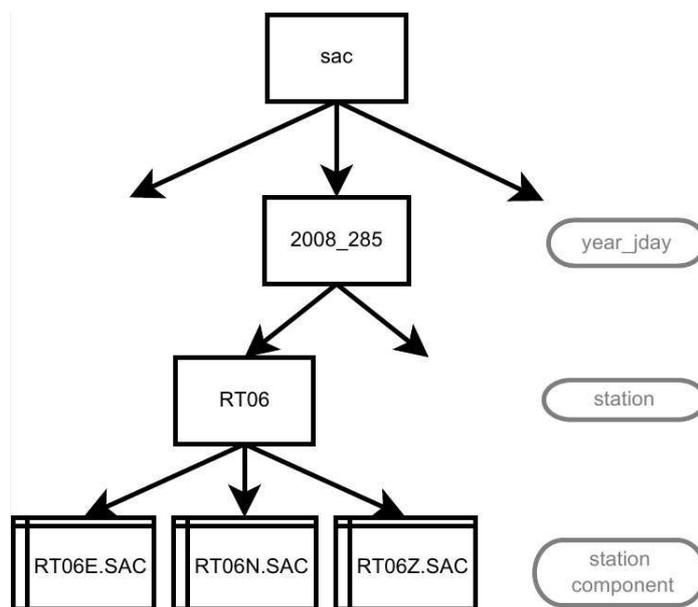


Figure B.4: Data directory structure

B.2 Create master events

cuttracelist.c: Reads an arrival-listing file provided by GNS Science that contains the P and S picks of large events ($M > \sim 1.5$). Using these pick times, Master templates are cut (after *Baisch et al.* [2008], as described in Section 2.3.1) across all components on stations where the event is detected. An ‘event time’ is assigned as the P pick at the closest station to the event, and the time shift from this ‘event time’ to all other stations is calculated [Section 2.3.3]. Output [Fig. B.5] includes text files which contain e.g.:

```
2008_285 RT01 E
Event no. is 143, eventtime is 1228.007160, shift is 0.105160
P-pick is at 1228.450000, S-pick is at 1229.053000
```

start time is 1228.112320, end time is 1229.800720
window length in samples is 337.680000

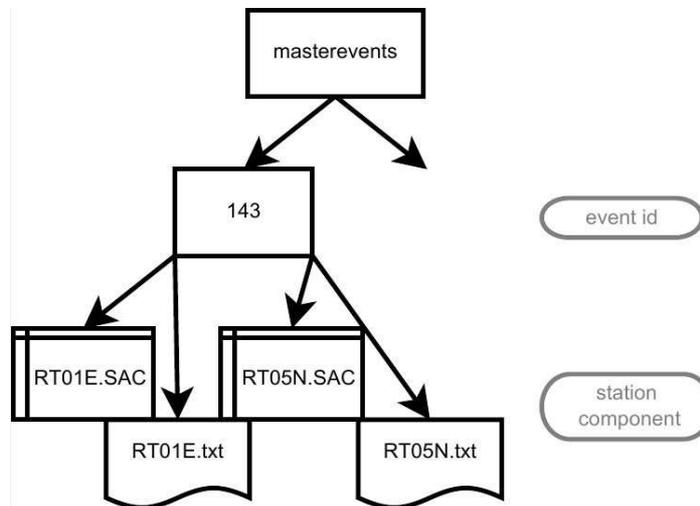


Figure B.5: Master directory structure. SAC files contain header variables *user1*=station shift (in seconds) calculated during creation, and *user2*=event id. **Important:** These header variables are crucial for later directory name creation and the code will not function correctly without them.

B.3 Cross-correlation loop

eventlooptotal.py: Performs cross-correlation on all desired events and data dates. Run by:
`./eventlooptotal.py -c totalcontrol.txt`, where all required control parameters are set in `totalcontrol.txt`.

Explanation of Control Parameters:

`eventfile = YES` (use `eventfilename` to get event list) = `NO` (don't have an event list)

`eventfilename =` event list file (event list has event names as the first word on each line)

`singleevent = YES` (carry out just on event specified in event) = `NO` (don't use single event)

`event =` eventname

`startday =` 2008_273 (year_jday inclusive)

`endday =` 2008_285 (year_jday exclusive)

restricted = YES (restrict calculations to stations (and components) in stationlist) =
NO

stationlist = stationnamecomponent stationname stationname

npts = number of points in data (used in **shifty.c** to determine number of bins and to determine the detection threshold so if number of points varies these will have to be calculated another way)

cohpow = the power of the coherency to use (0 to 1, where 1 applies no shaping) for coherence filtering in **corr_scan.c**

ishift = shift step in samples for **corr_scan.c**

binsz = sample difference less than which two times are considered equivalent in **shifty.c**

false = detection value used to calculate detection threshold

runcontrol = YES (run **controlfile.py** to create control file for **corr_scan.c**) = NO

runcorr = YES (run **corr_scan.c**) = NO

runshift = YES (run **shifty.c**) = NO

runmad = YES (run **mad.c**) = NO

runcontrol2 = YES (run **controlfileclone.py** to create control file for **clone.c**) = NO

runclone = YES (run **clone.c**) = NO

Example of totalcontrol.txt:

```
#####CONTROL-FILE totalcontrol.txt FOR PROGRAMME eventlooptotal.py#####  
#####ensure environment is set up for SAC before running
```

```
[main]
```

```
#####MAIN LOCATION
```

```
maindata=/data/sabine/rotokawa/data/
```

```
maincode=/data/sabine/rotokawa/zara/
```

```
#####python location (needs obspy module installed)
```

```
python=/usr/local/python2.6/bin/python
```

```
[details]
```

```
#####WHAT EVENTS AND DAYS DO YOU WANT TO RUN OVER?
```

```
####Do you have an event file? (YES/NO) (eventfile and singleevent
```

```

cannot have the same value)
eventfile=YES
eventfilename=/data/sabine/rotokawa/zara/mycode/event.listing1
singleevent=NO
event=146

startday=2008_273
endday=2008_285

#####RESTRICT TO CERTAIN STATIONS? (YES/NO)
restricted=NO
stationlist=RT01 RT03Z

#####SOME PARAMATERS
#####number of points in your correlating file
npts=17280000
#####what control file name do you want?
controlname=/data/sabine/rotokawa/zara/detect/control1.dat
control2name=/data/sabine/rotokawa/zara/detect/clonecontrol1.dat

#####CORR_SCAN VARIABLES
#####coherency power for corr_scan
cohpow=0.5
#####shift for corr_scan
ishift=100

#####SHIFTY VARIABLES
#####binsize for calculating network correlation coefficient
binsz=50
#####CLONE VARIABLES
#####detection factor for calculating the detection threshold
false=0.5

[process]
#####WHAT PROCESSES DO YOU WANT TO RUN? (YES/NO)
runcontrol=YES
runcorr=YES
runshift=NO
runmad=NO
runcontrol2=NO
runclone=NO

#####ALLOCATE ALL PATHS USED DURING PROCESSING

[codes]
#####CODE LOCATION AND NAME

```

```

control=/data/sabine/rotokawa/zara/mycode/controlfile.py
corr=/data/sabine/rotokawa/zara/detect/corr_scan
shift=/data/sabine/rotokawa/zara/mycode/shifty
mad=/data/sabine/rotokawa/zara/mycode/mad
control2=/data/sabine/rotokawa/zara/mycode/controlfileclone.py
clone=/data/sabine/rotokawa/zara/detect/snapshot/clone

```

[data]

```

#####DATA LOCATIONS, names "sac" and "masterevents" must be maintained
for naming purposes of output files

```

```

masters=/data/sabine/rotokawa/data/masterevents/
data=/data/sabine/rotokawa/data/sac/

```

```

#####OUPUT LOCATIONS

```

```

corres=/data/sabine/rotokawa/data/corresult/
madres=/data/sabine/rotokawa/data/MAD/
cloneres=/data/sabine/rotokawa/data/cloneresult/
detectres=/data/sabine/rotokawa/data/detections/

```

If all [process] variables in totalcontrol.txt are set to YES, **eventlooptotal.py** performs the following:

1. **controlfile.py**: Creates the control file **control.dat** required for **corr_scan.c**. The first line of **control.dat** is the path to the master template file, and subsequent lines are paths to data the master template is to be cross-correlated with, e.g

```

/data/sabine/rotokawa/data/masterevents/145 RT06E.SAC
/data/sabine/rotokawa/data/sac/2008_273/RT06 RT06E.SAC
/data/sabine/rotokawa/data/sac/2008_274/RT06 RT06E.SAC
/data/sabine/rotokawa/data/sac/2008_275/RT06 RT06E.SAC
/data/sabine/rotokawa/data/sac/2008_276/RT06 RT06E.SAC

```

Important: master templates can only be cross-correlated with matching file names.

2. **corr_scan.c**: (significantly modified from Charlotte Rowe ca. 2000) Carries out the cross-correlation of a master template with data corresponding to the same station and component [Section 2.3.2]. Assumes the station shift (in seconds) is stored in the SAC header user1 and the master event name (a three digit number) is stored in the SAC header user2, this is used for naming purposes (and hence used in **shifty.c**). Output [Fig. B.6] is a text file with: the first line containing the arguments used during

correlation; master window length, coherence power, shift step in samples: Subsequent lines have Column 1 as the number of the file (in **control.dat**) the master template is being correlated with. Column 2 is the mean lag (in samples) weighted by the magnitude of the maximum,

$$\mathbf{cbest} = \frac{\sum(\mathbf{cmax} \times \mathbf{ilag})}{\sum \mathbf{cmax}}$$

where **ilag** is the index of the correlation maximum **cmax** for each narrowband. Ideally want each narrowband to have the correlation maximum at the same location/index. Column 3 is the standard deviation of the lag (in samples), which is

$$\mathbf{stdlag} = \sqrt{\frac{\sum[\mathbf{cmax}(\mathbf{ilag} - \mathbf{cbest})]^2}{\sum \mathbf{cmax}}}$$

stdlg is a weighted standard deviation, with **cmax** providing the weighting. Column 4 is $R_{ab}(h)$, which is the broadband correlation maximum. Column 5 is h , which is the index of the broadband correlation maximum. Column 6 is j , which is the start point of the window (in samples). e.g.

```

704 0.500000 100
00002 153 22.506699 0.1258 068 000000
00002 154 23.050089 0.1475 297 000100
00002 006 7.795477 0.1557 -037 000200

```

3. **shifty.c**: Using the result file from **corr_scan.c** and the calculated time shifts from **cut-tracelist.c**, shifts all results (corresponding to a single event and single day's data) to the common 'event time' and sums time-corresponding cross-correlation coefficients to calculate entire Network Cross-correlation Coefficients (NCC) (described in Section 2.3.3) [Fig. B.7].
4. **mad.c**: Using the results from **shifty.c**, calculates the median and the Median Absolute Deviation (MAD) of the set of NCC's corresponding to a single event and single day's

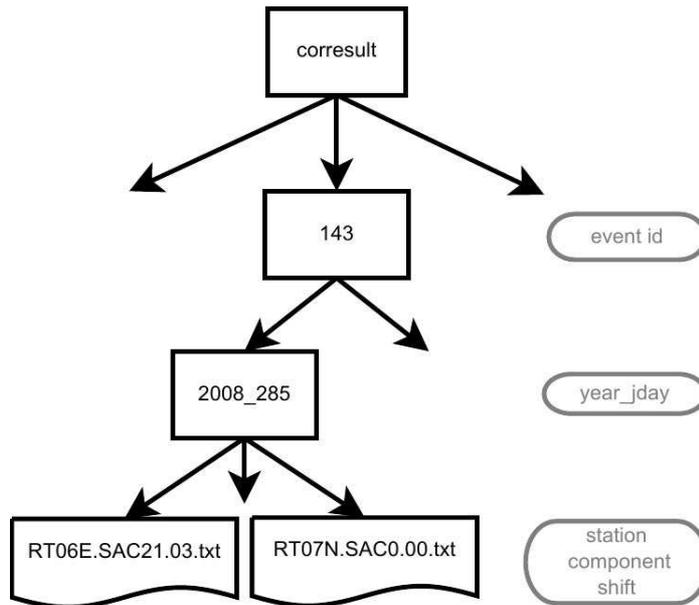


Figure B.6: Corr_scan result directory structure.

data [Section 2.3.4]. Output files [Fig. B.8] have the format:

```

median 1.416000
mad 0.359200

```

5. **controlfileclone.py**: Creates the control file **clonecontrol.dat** required for **clone.c**.

The first line of **clonecontrol.dat** is the path to the master template file, and the next line is the path to the data the master template was cross-correlated with, e.g

```

/data/sabine/rotokawa/data/masterevents/145/RT06E.SAC
/data/sabine/rotokawa/data/sac/2008_273/RT06/RT06E.SAC

```

As master templates can only be cross-correlated with matching file names, and each day has a different threshold, only one data file at a time can be run through **clone.c**.

6. **clone.c**: (modified from Charlotte Rowe ca. 2000) Using the results from **shifty.c** and **mad.c**, points in time with NCC'S over the detection threshold are noted as events, and a new-event SAC-file is cut to the same length as the corresponding master template length (see Section 2.3.4) [Fig. B.9]. The correlation value is stored in the header variable `user2` and the master event name in `kuser2`. A file is also created with three

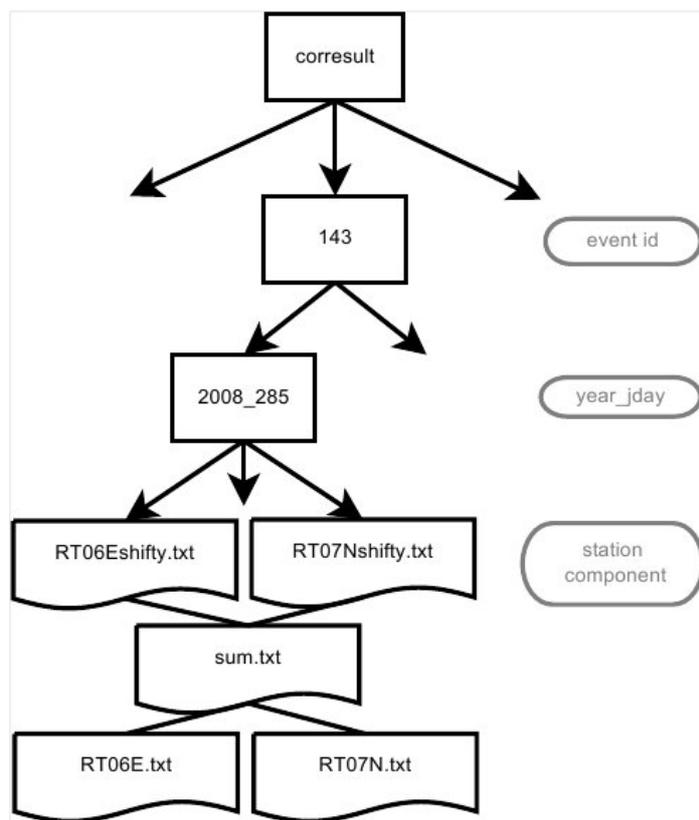


Figure B.7: Shifty result directory structure. Files named *shifty.txt contain two columns: CC's, and shifted times (in samples). sum.txt contains two columns: 50-sample wide bins labelled with their midpoint (Ψ), and NCC's. RT06E.txt (for example) has the same lay out as the corr_scan.c output (apart from the first line): column 1 is given a default value of 1, columns 2, 3 are set to 0, column 4 is set to the NCC, column 5 is set to 0, and for each template the results in sum.txt are shifted forward using that station's shift and written to column 6.

lines for each detection [Fig. B.10]: the NCC value and timing in samples on the first line. The timing in year day hour minutes and seconds on the second line, and the timing in seconds, then day hour minutes seconds and milliseconds on the third line.
e.g.

```

1 max is 8.864901 at 0 + 232830 = 232830
2008 285 0 19 24 0.15
1164.801392 285 00 19 24 150
1 max is 7.576600 at 0 + 233780 = 233780
2008 285 0 19 28 0.90
1169.551392 285 00 19 28 900
  
```

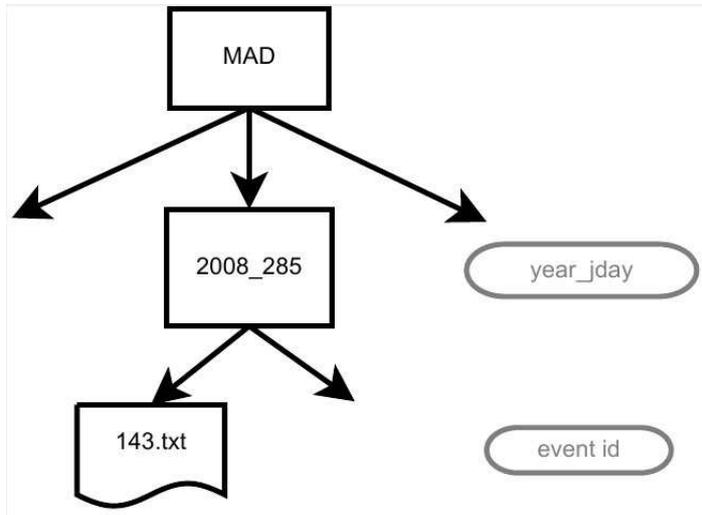


Figure B.8: Mad result directory structure.

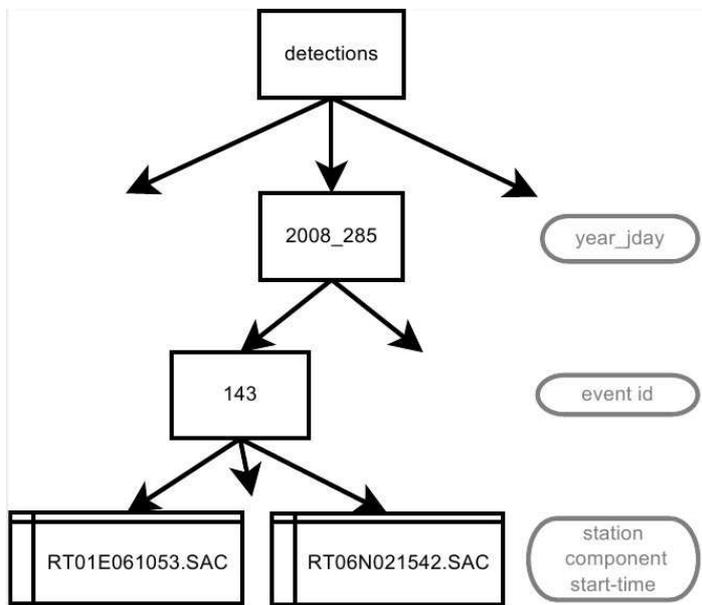


Figure B.9: Clone detection directory structure.

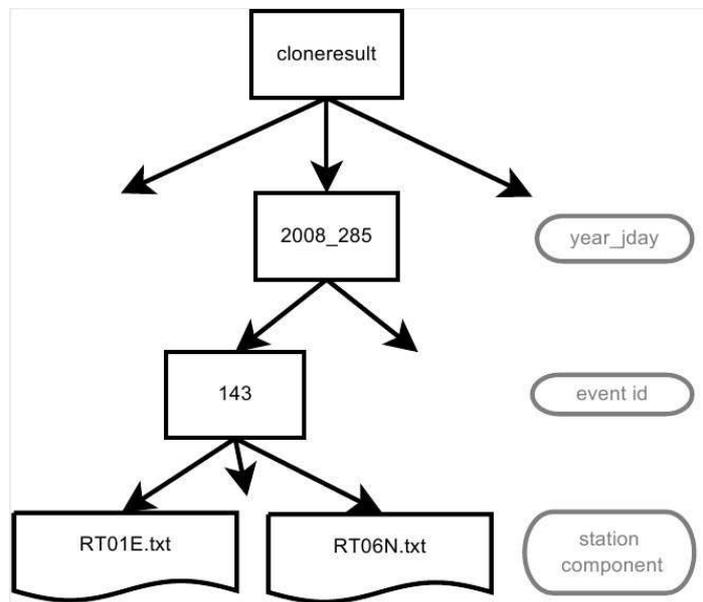


Figure B.10: Clone result file directory structure.

Appendix C

Location

C.1 Double-difference relocation

As described in Section 3.2, we have attempted to apply the double-difference earthquake algorithm *hypoDD* of *Waldhauser* [2001] to the Rotokawa data set, but have obtained unsatisfactory results. The details of the parameters used in these calculations are outlined below. We describe the calculations of cross-correlation differential travel times using *BCSEIS* [*Du et al.* 2004] in Section C.1.1, and present the subsequent double-difference relocation parameters used in *hypoDD* [*Waldhauser* 2001] in Section C.1.2.

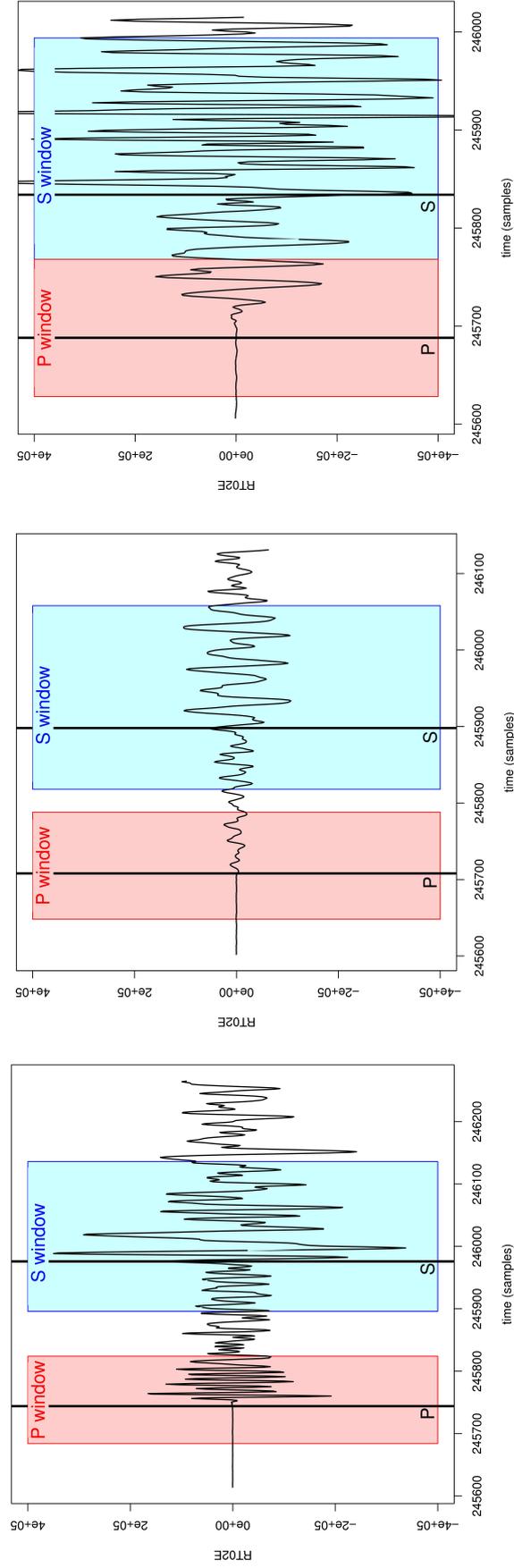
C.1.1 Cross-correlation delay time estimates

To perform double-difference relocation, we must first calculate cross-correlation delay times for each phase of interest. We use *BCSEIS* [*Du et al.* 2004] to perform this calculation. As discussed in Section 3.1.1, *BCSEIS* uses the bispectrum method which is less influenced by correlated noise than standard cross-correlation may be [*Du et al.* 2004; *Statz-Boyer et al.* 2009]. We retain the same filtering of 3–16 Hz, as described in Section 3.1.1, but adjust the window sizes to 0.3 s pre-P pick and 0.4 s post-P pick, and 0.4 s pre-S pick and 0.8 s post-S pick. Window sizes are chosen so that several cycles of seismic energy are included in the window, but the P-window does not overlap into S-phase arrival. Fig. C.1 displays three

examples of our window sizes: note that the same window sizes must be used at all stations, and that our sizes have therefore been influenced by the closer stations with shorter S–P times (e.g. Fig. C.1(c)). As cross-correlating every earthquake against every other earthquake in the data set is a very time-intensive task, we first down-sample all event traces to 50 Hz.

BCSEIS performs three measurements: the first uses time-domain cross-correlation on band-pass filtered waveforms; the second uses the bispectrum method on the same band-pass filtered seismograms; and the third again uses the bispectrum method, but on unfiltered seismograms. After these measurements have been performed, data are selected or rejected based on three thresholds: a cross-correlation coefficient threshold CC^{lim} , an upper limit $CC^{lim(u)}$, and a lower limit $CC^{lim(l)}$. All correlation-derived measurements with coefficients above the threshold CC^{lim} are accepted provided they and their bispectrum method-derived counterparts are close to each other. The estimates are deemed to be “close” when their lag is within a specified number (Δ) of samples of each other. In addition, if an event pair has a correlation coefficient that is greater than the upper limit $CC^{lim(u)}$ at any station, then all correlation-derived measurements with coefficients above the lower limit $CC^{lim(l)}$ are accepted. Again, these measurements are only chosen when the three delay time estimates are consistent. While these criteria provide a stringent screen against inconsistent data, they also offer correlation measurements with relatively low coefficients (between $CC^{lim(l)}$ and CC^{lim}) a second chance at selection. Provided the two events in question are similar enough, and the bispectrum- and correlation-derived delays are consistent, then data that would otherwise be abandoned are used to bolster the resulting set of delay times and improve the quality of the corresponding earthquake locations.

We adopt the following limits: $CC^{lim(l)} = 0.3$, $CC^{lim} = 0.5$, $CC^{lim(u)} = 0.7$, and $\Delta = 1$ sample. These limits are similar to those used by *Du et al.* [2004], apart from CC^{lim} and $CC^{lim(u)}$, where more stringent cut-offs of 0.7 and 0.8 were used. We impose more lenient cut-offs as we are working in a small region, with typically very similar earthquake forms (by virtue of the detection method employed), and we do not wish low-SNR earthquakes to be overly-penalised.



(a) Station RT01

(b) Station RT02

(c) Station RT08

Figure C.1: P and S window sizes used in *BCSEIS*. Window sizes are chosen so that several cycles of seismic energy are included in the window, but the P-window does not overlap into the S-phase even for stations with very short S–P times (e.g. station RT08). Note that time is in samples (200 Hz), and is relative to the start of the day.

C.1.2 *hypoDD*

To perform double-difference earthquake relocation, we use the *hypoDD* software that implements the double-difference earthquake location algorithm of *Waldhauser and Ellsworth* [2000].

The velocity model we use in *hypoDD* is shown in Table C.1. *hypoDD* requires the use of a constant V_p/V_s ratio, and a maximum of 12 layers: we therefore adjust the V_p 1D model used with *Velost* [Table 3.1], and use the average V_p/V_s from this data of $V_p/V_s = 1.75$. As discussed in Section 3.2, we use both catalogue and cross-correlation-derived travel time

Table C.1: 1D V_p velocity model used in *hypoDD* for Rotokawa

Depth (km)	V_p (km/s)
-0.40	2.00
0.10	2.63
1.00	2.90
1.60	3.05
1.80	3.12
2.20	3.20
2.60	3.41
2.90	3.74
3.40	4.32
4.00	4.57
6.00	5.51
10.00	5.81

differences, and as we have more than 100 earthquakes to solve for, we must use *hypoDD*'s LSQR inversion method. Uncertainties reported by the LSQR method are unreliable [*Paige and Saunders* 1982; *Waldhauser* 2001], and we therefore assess earthquake uncertainties based on the conditioning number CND . CND represents how stable the inversion is, and a damping factor is used to damp the hypocentral adjustments if the adjustment vector becomes large or unstable. Generally, a damping factor between 1 and 100 is appropriate, resulting in a condition number that is between about 40 and 80 [*Waldhauser* 2001]. If a very high damping is needed to lower the condition number, or if the condition number remains high, then the system may not be well conditioned.

Phase weights dictate the extent to which each travel time difference is allowed to in-

fluence the final locations obtained with *hypoDD*. A measurement with a high weight has more influence than one with a comparatively low weight. The weighting scheme used by *hypoDD* is summarised as:

$$w_i = w_i^{ap} w_i^{res} w_i^{dist} \quad (\text{C.1})$$

where w_i is the weight applied to the i^{th} datum, and w_i^{ap} is the *a priori* weight for that measurement. w_i^{res} and w_i^{dist} are dynamic weights applied during the inversion to remove data with high residuals and interevent distances. Our *a priori* weights w_i^{ap} are the individual pick weights assigned using Jeffreys' weighting [Section 3.1.2] multiplied by the *a priori* weights summarised in Table C.2. w_i^{res} is a weight that depends on the i^{th} double-difference residual from the preceding iteration [Waldhauser and Ellsworth 2000],

$$w_i^{res} = \max^2 \left(0, 1 - \left(\frac{dr_i}{\alpha \cdot \frac{\mathbf{dr}_{MAD}}{\sigma_{MAD}}} \right)^2 \right) \quad (\text{C.2})$$

where dr_i represents the i^{th} residual difference, \mathbf{dr} is the vector of residuals, $\sigma_{MAD} = 0.67449$ is the MAD for Gaussian noise, and $\mathbf{dr}_{MAD} = \text{median}(|dr_i - \text{median}(\mathbf{dr})|)$. The residual weight is designed to be close to unity for data with small residuals, and to drop rapidly to zero as the residuals approach a given cut-off value of α residual spreads. α is typically set between 3 and 6 to remove non-Gaussian outliers [Waldhauser and Ellsworth 2000]. We set $\alpha = 4$ for both cross correlation and catalogue data after Clarke [2008]. This introduces

Table C.2: Weighting scheme used in *hypoDD*. Note that -9 denotes an unused option, and hence the dynamic reweighting schemes are only used in the last 3 sets of iterations. We set the damping to the limit of the acceptable range of 1–100.

Iterations	Cross correlation data				Catalogue data				Damping
	A priori, P-wave w_i^{ap}	A priori, S-wave w_i^{ap}	Misfit weight α [Eq. C.2]	Dist. Weight c [Eq. C.3]	A priori, P-wave w_i^{ap}	A priori, S-wave w_i^{ap}	Misfit weight α [Eq. C.2]	Dist. Weight c [Eq. C.3]	
1–5	0.001	0.001	–9	–9	1.0	1.0	–9	–9	100
6–10	0.1	0.1	–9	–9	0.9	0.9	–9	–9	100
11–15	0.5	0.5	3	2.5	0.3	0.3	3	10	100
16–20	0.9	0.9	3	2.5	0.2	0.2	3	10	100
20–25	1.0	1.0	3	2.5	0.1	0.1	3	10	100
								CND =	267

a relatively strict cut-off of four standard deviations. w_i^{dist} is a weight that is dependent on the distance between two earthquakes for which the i^{th} travel time difference is calculated [Waldhauser and Ellsworth 2000],

$$w_i^{dist} = \max^b \left(0, 1 - \left(\frac{s_i}{c} \right)^a \right) \quad (\text{C.3})$$

where s is the interevent distance, c is a cut-off value to remove event pairs separated by distances larger than c , and a and b are exponents defining the shape of the weighting curve. Waldhauser and Ellsworth [2000] set $a = b = 5$ for cross-correlation data and $a = b = 3$ for catalogue data. The waveform similarity that cross-correlation utilises is dependant on events being sufficiently close so that they sample the ‘same’ ray path: this means that there is an interevent distance at which these ray paths become significantly different. Schaff *et al.* [2004] showed that at interevent distances of up to 2 km in the Calaveras Fault in northern California correlation measurements provided a significant improvement on phase pick data. Nakahara [2004] extended this testing in Japan, and found improvements using correlation measurements of separations of up to $\sim 4 - 5$ km. Clarke [2008] tested correlation measurements of interevent distances of ~ 10 km in the Rotorua and Kawerau geothermal systems, and found that the correlation coefficients broke down at 2.5 – 3 km. As the distance at which cross-correlation-derived data break down will depend on the velocity heterogeneity of the particular region, we adopt the parameters chosen by Clarke [2008], which were determined within the same region and in a very similar setting to Rotokawa. We therefore set our distance cut-off for dynamic cross-correlation weights to 2.5 km, and our distance cut-off for dynamic catalogue weights to 10 km. These are also similar to the cut-off distances used by Waldhauser and Ellsworth [2000] of 2 km and 10 km respectively. The variation of w_i^{dist} with interevent distance is plotted in Fig. C.2. At distances less than 2.5 km, cross-correlation-derived measurements are given precedence. As the interevent distance increases to 2.5 km, the cross correlation weights drop abruptly to zero. Beyond this distance, only catalogue-based travel time differences are incorporated into the inversion.

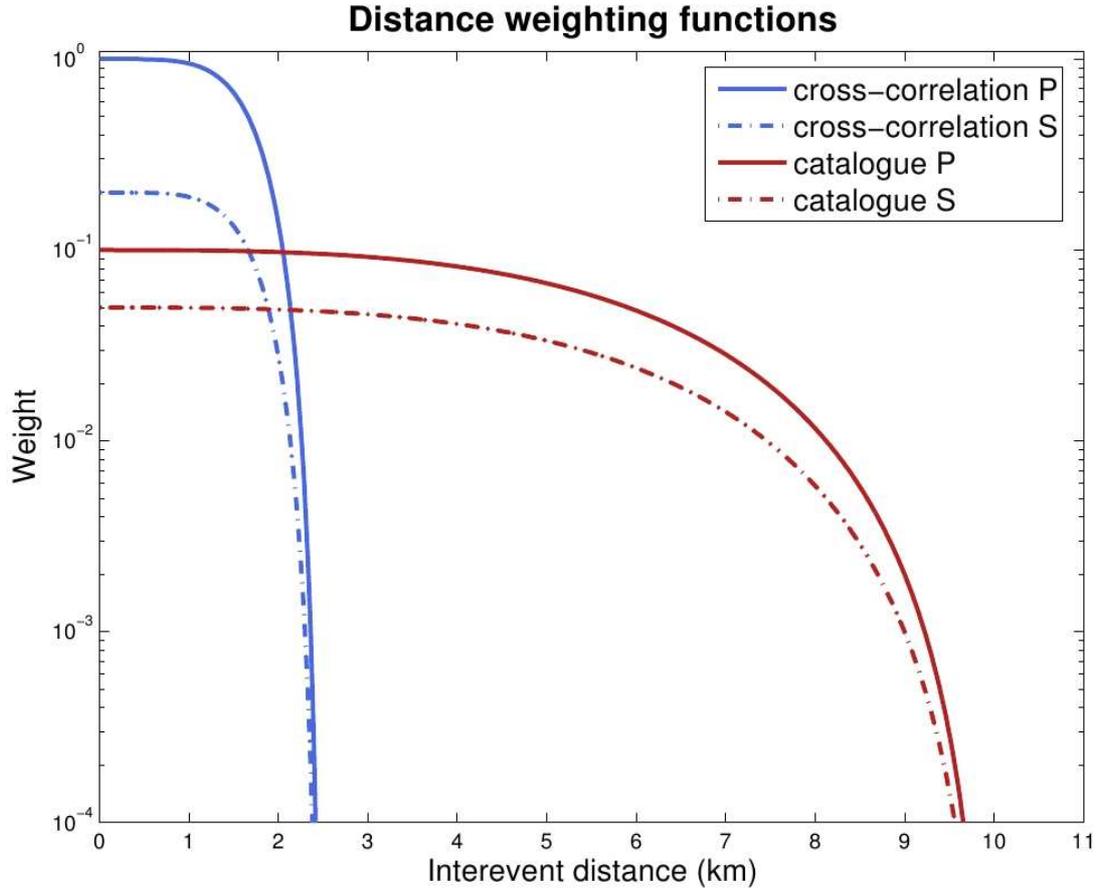


Figure C.2: Intervent distance reweighting functions for *hypoDD*. Each function is scaled by the *a priori* weight for that data type (see text). Further weighting is applied to account for data quality and residual size. Figure from Clarke [2008].

Table C.2 displays our weighting scheme, with damping set to 100: the limit of acceptable values. Our CND, however, is 267. Table C.3 displays the same weighting scheme but

Table C.3: Weighting scheme used in *hypoDD*. Note that -9 denotes an unused option, and hence the dynamic reweighting schemes are only used in the last 3 sets of iterations. 155 is the lowest damping that produces $CND < 80$.

Iterations	Cross correlation data				Catalogue data				Damping
	A priori, P-wave w_i^{ap}	A priori, S-wave w_i^{sp}	Misfit weight α [Eq. C.2]	Dist. Weight c [Eq. C.3]	A priori, P-wave w_i^{ap}	A priori, S-wave w_i^{sp}	Misfit weight α [Eq. C.2]	Dist. Weight c [Eq. C.3]	
1–5	0.001	0.001	–9	–9	1.0	1.0	–9	–9	270
6–10	0.1	0.1	–9	–9	0.9	0.9	–9	–9	270
11–15	0.5	0.5	3	2.5	0.3	0.3	3	10	270
16–20	0.9	0.9	3	2.5	0.2	0.2	3	10	270
20–25	1.0	1.0	3	2.5	0.1	0.1	3	10	270
								CND =	78

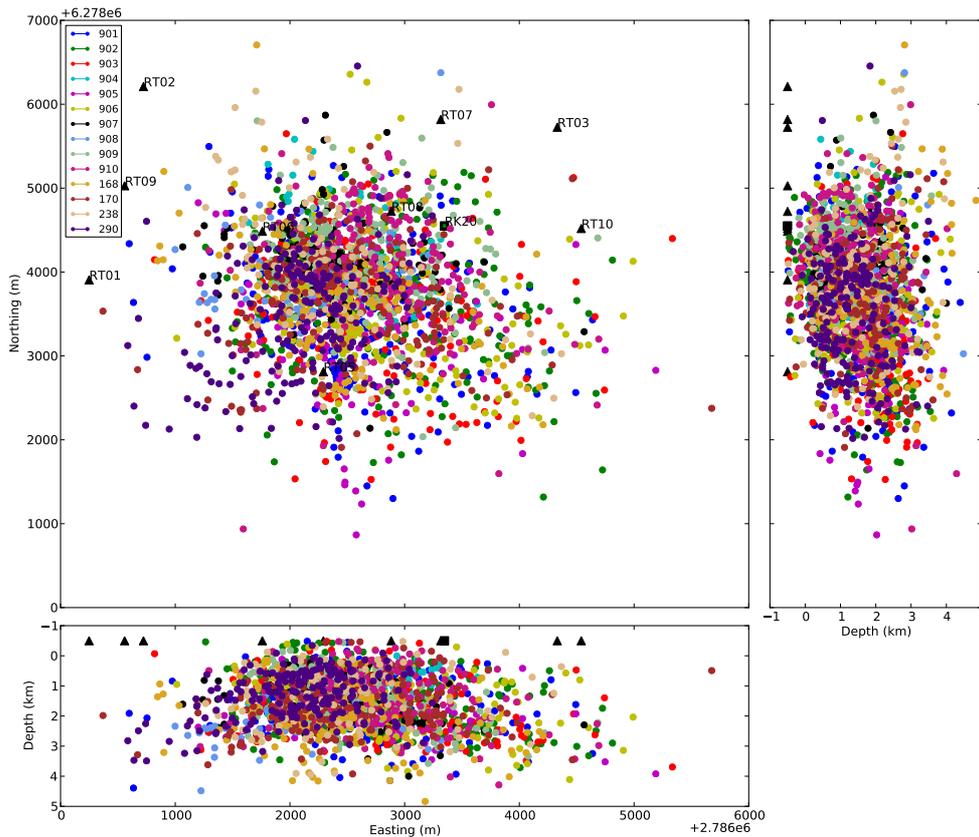


Figure C.3: *hypoDD* locations with slave-events colour coded based on their master event. Note the lack of slave-event clustering and overall lack of any spatial structure.

with a damping of 270, which is the lowest damping that achieves a CND of 40–80. This is an extremely high damping, and reflects the fact that the inversion performed on the data set is unstable. We believe this is due to our low number of stations and correspondingly low number of observations associated with each low-magnitude earthquake. We display the locations determined using the parameters displayed in Table C.3 in Fig. C.3, but as discussed in Section 3.2 we do not have confidence in these locations.

C.2 *NonLinLoc* cluster locations

Each location plot below shows a separate cluster of slave events coloured according to their common master event. The locations are those obtained by using *NonLinLoc* [Section 3.3].

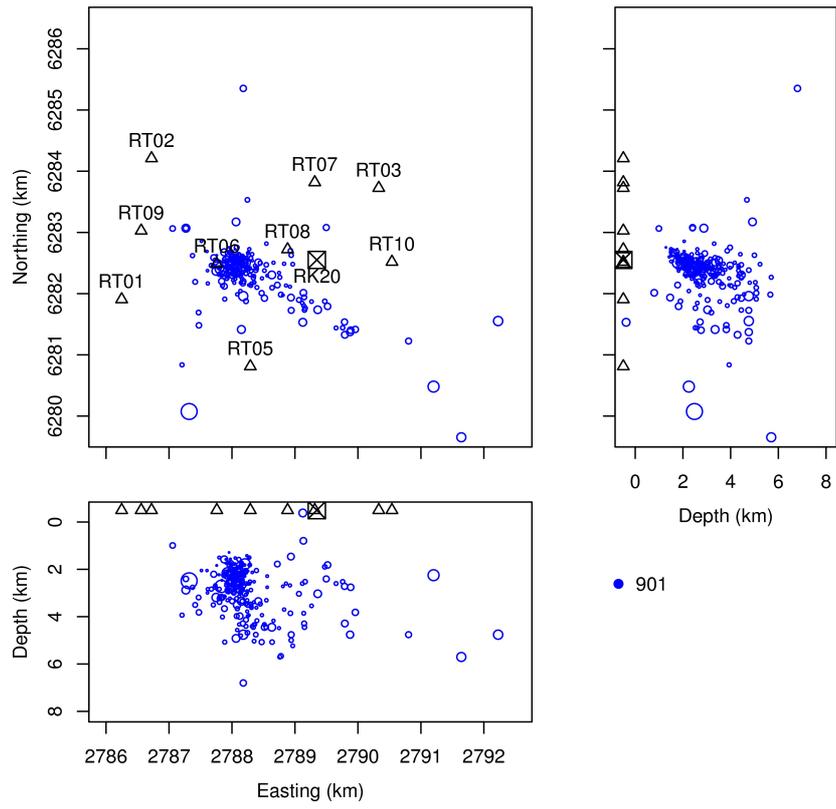


Figure C.4: 901 cluster

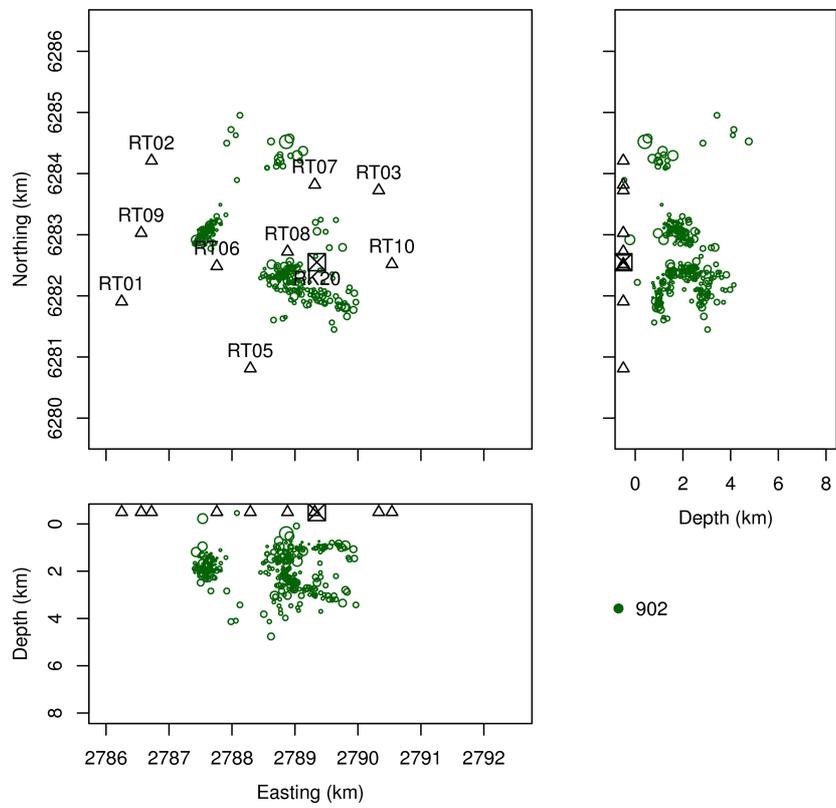


Figure C.5: 902 cluster

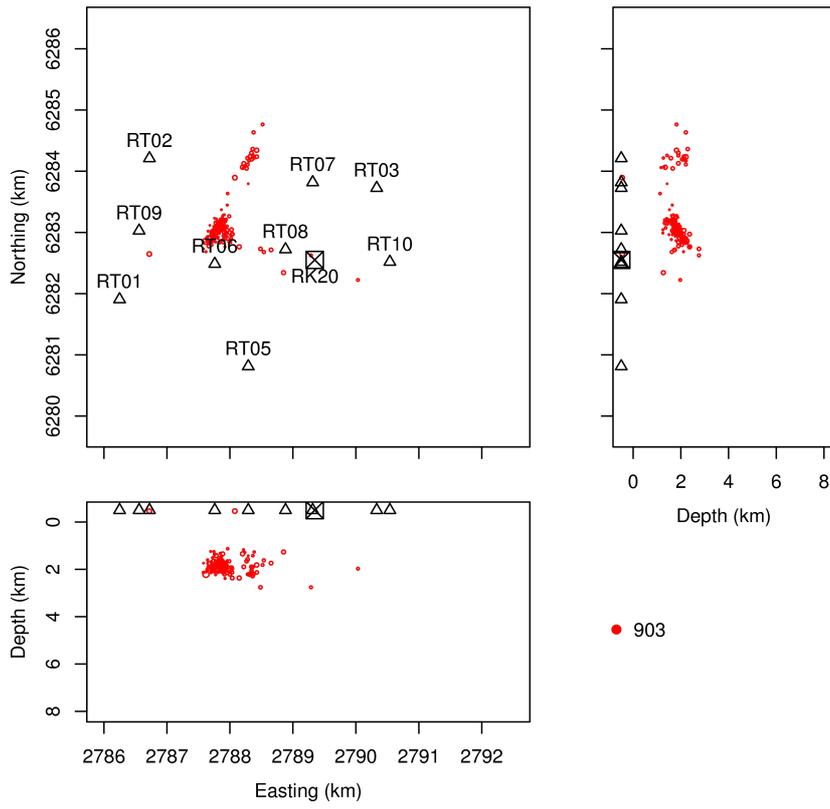


Figure C.6: 903 cluster

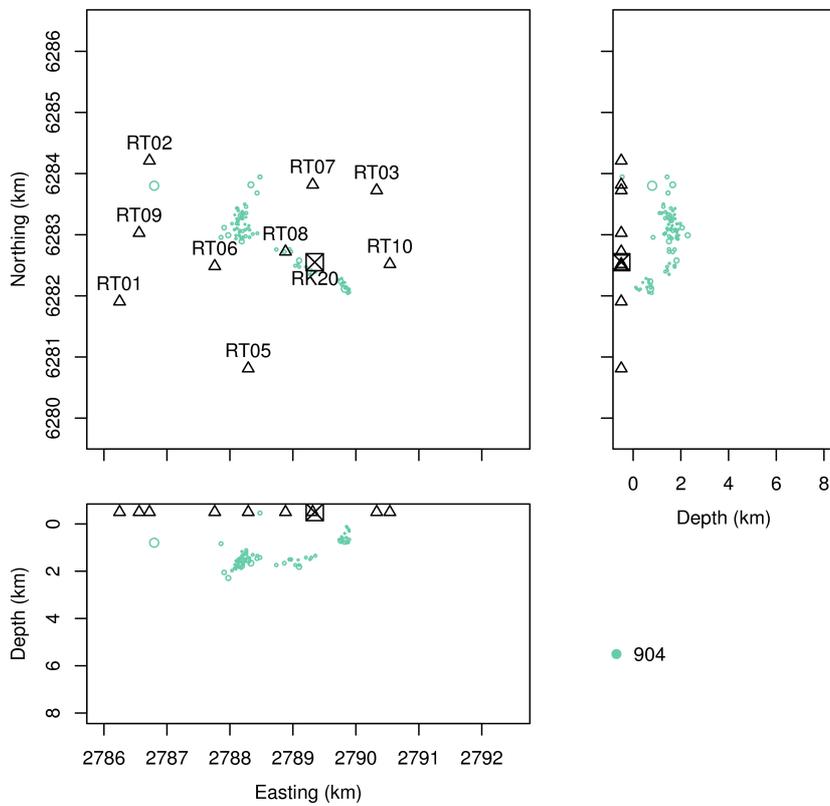


Figure C.7: 904 cluster

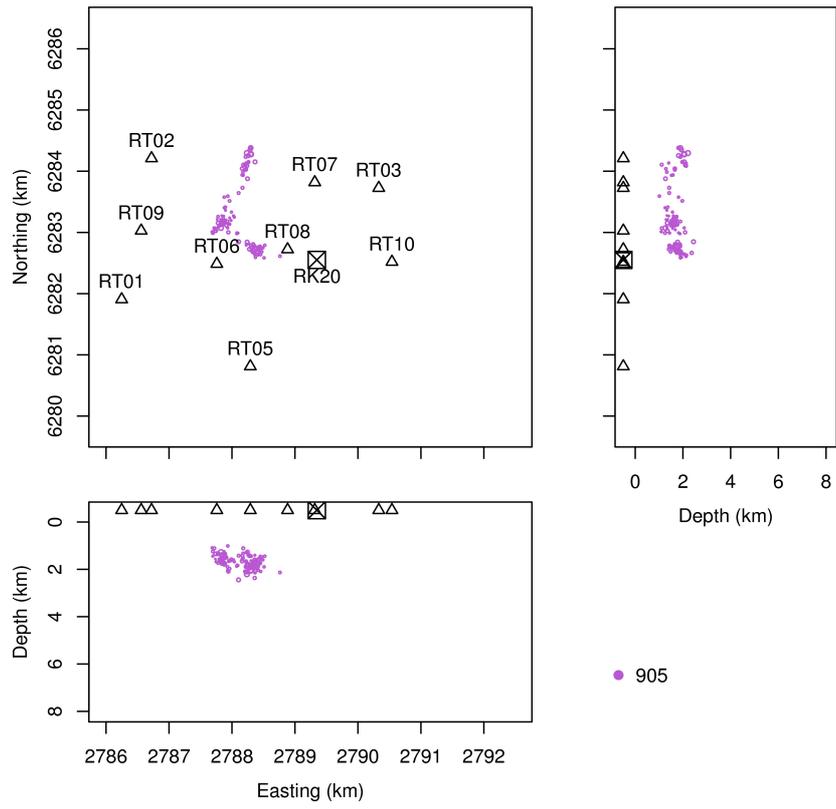


Figure C.8: 905 cluster

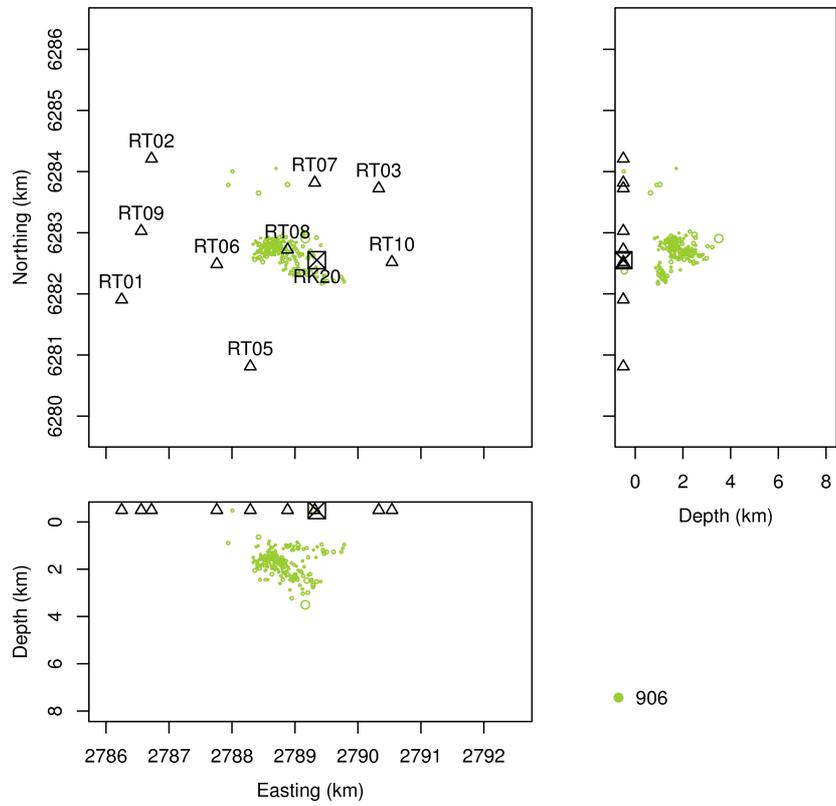


Figure C.9: 906 cluster

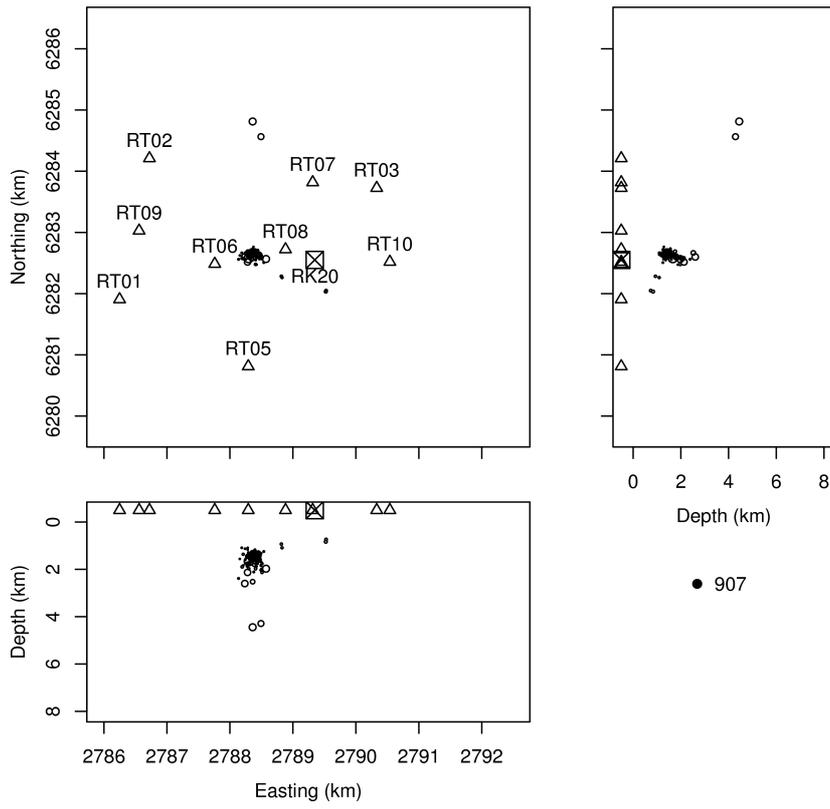


Figure C.10: 907 cluster

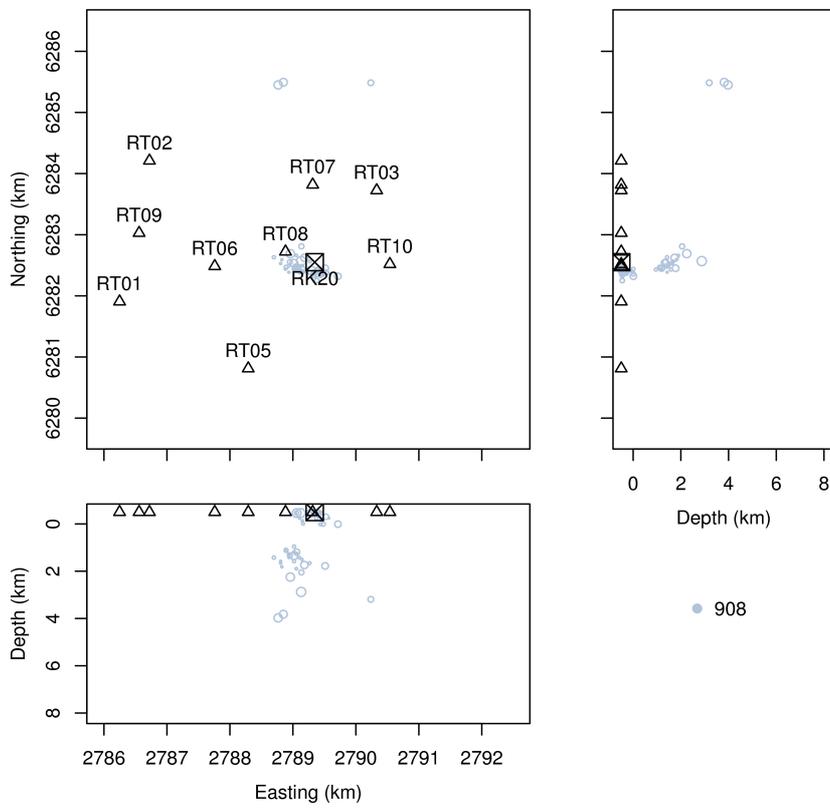


Figure C.11: 908 cluster

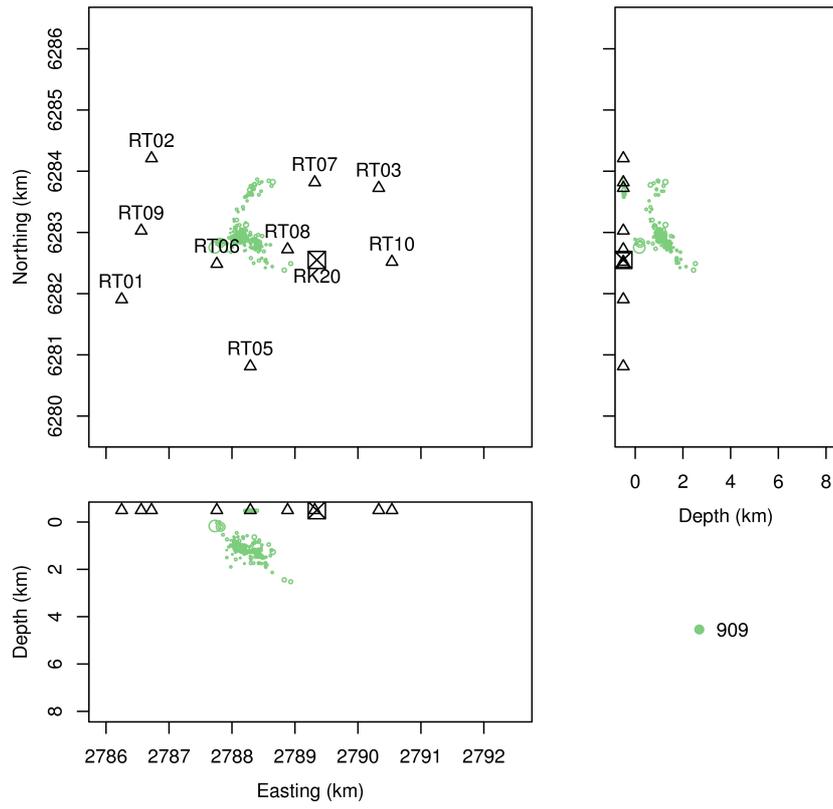


Figure C.12: 909 cluster

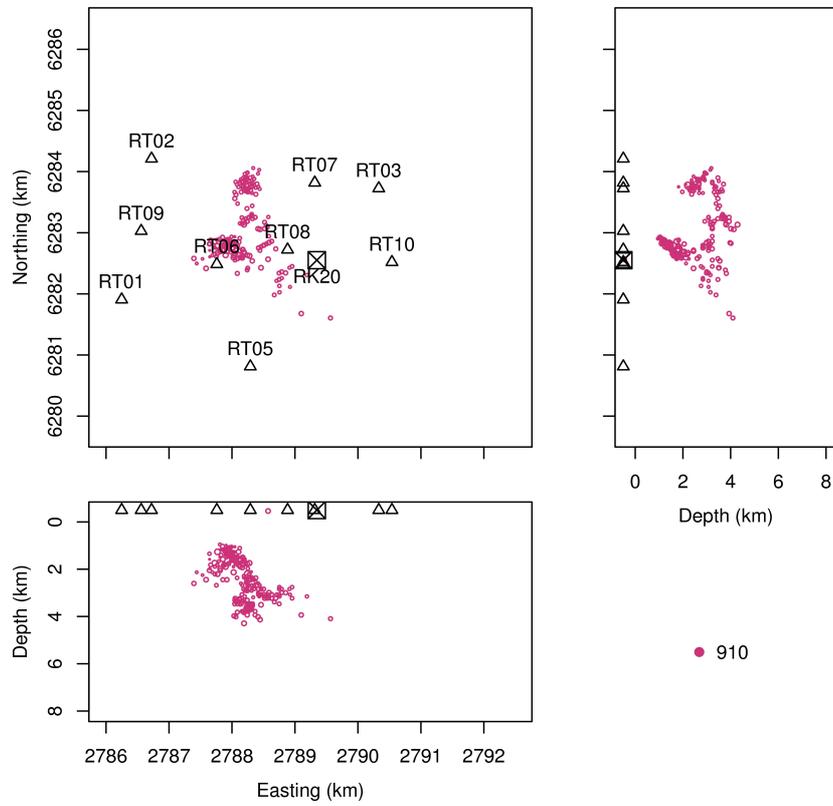


Figure C.13: 910 cluster

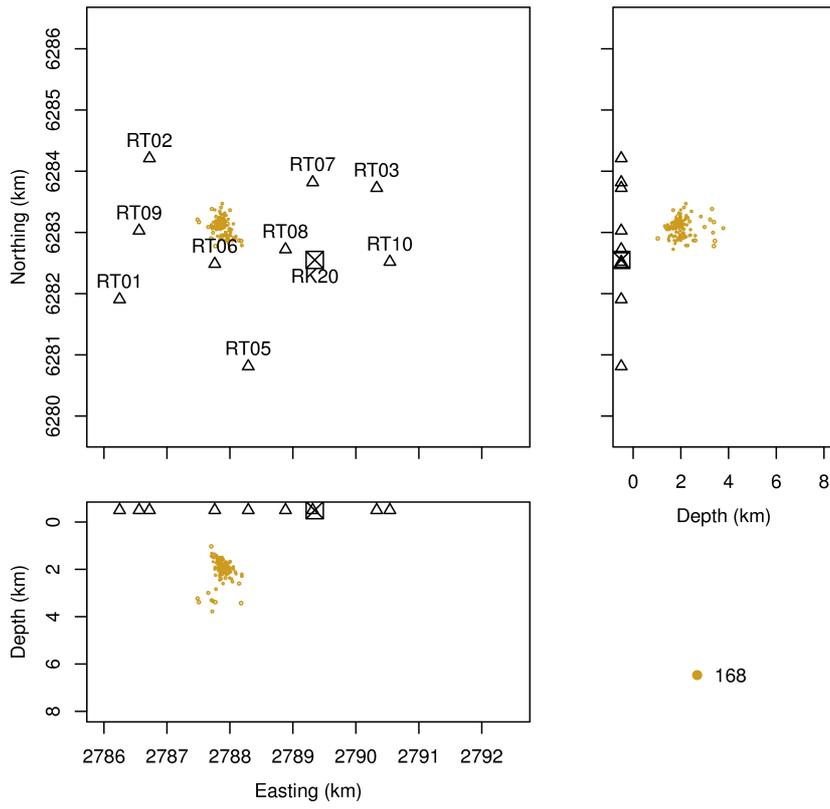


Figure C.14: 168 cluster

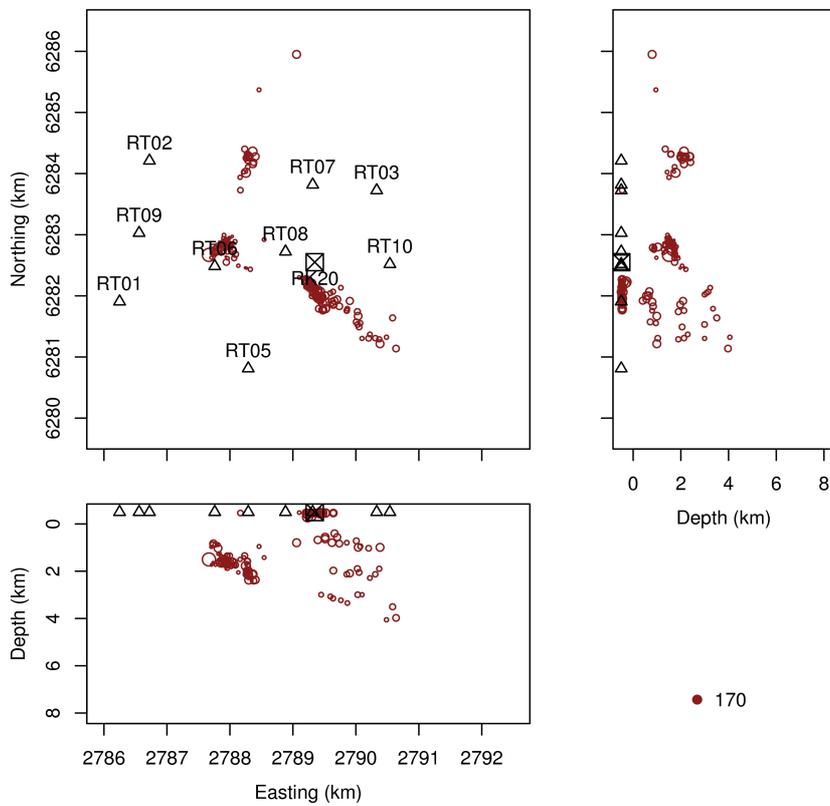


Figure C.15: 170 cluster

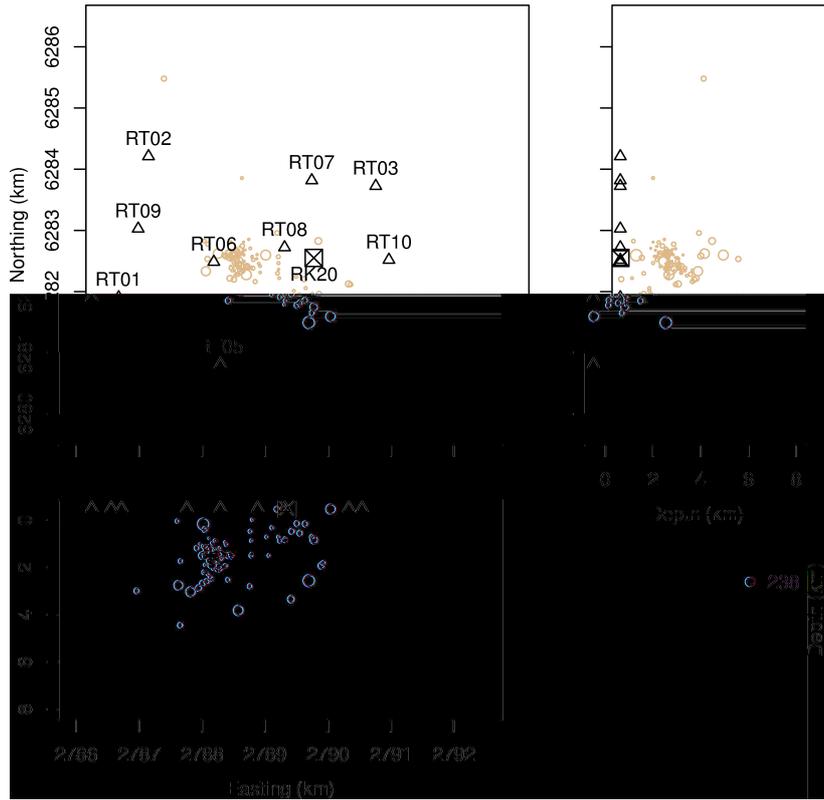


Figure C.16: 238 cluster

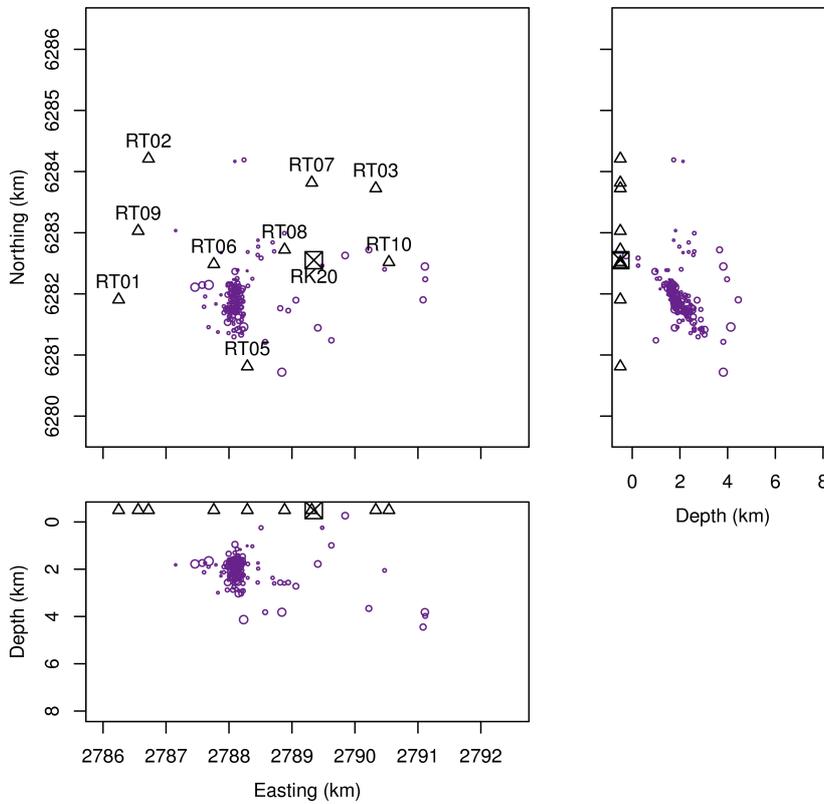


Figure C.17: 290 cluster

C.3 Focal mechanisms

Focal mechanisms are displayed below for each of the 14 master events: the left hand panels show PT contour plots, with orange denoting the P-axis and green the T-axis; the right hand panels show the locations of P-wave compressions (blue points) and dilatations (red points). Errors are calculated using the Matrix Fisher Distribution with scalar concentration parameter after *Walsh et al.* [2009] and *Arnold and Townend* [2007] (see Section 3.4). The focal mechanisms are illustrated in map view in Fig. 3.18 and the strike, dip and rake parameters summarised in Table 3.3.

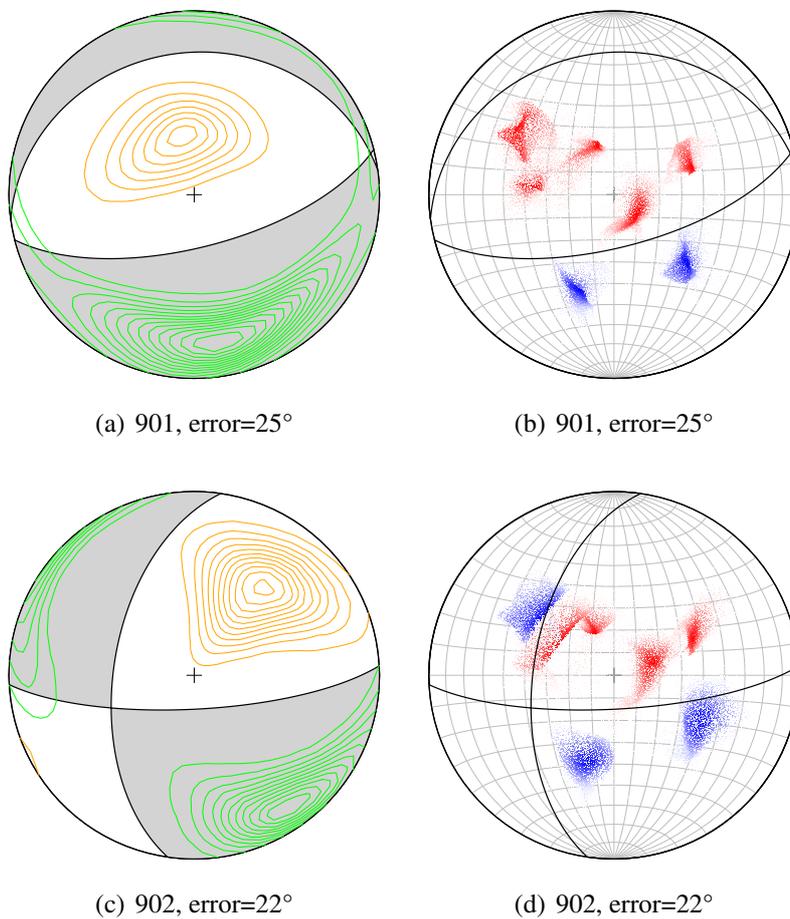
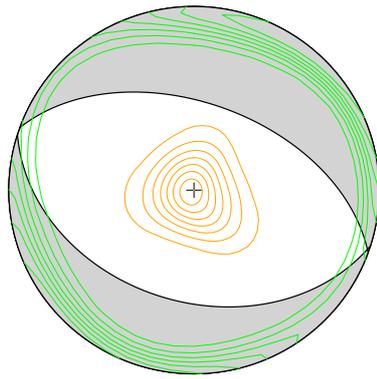
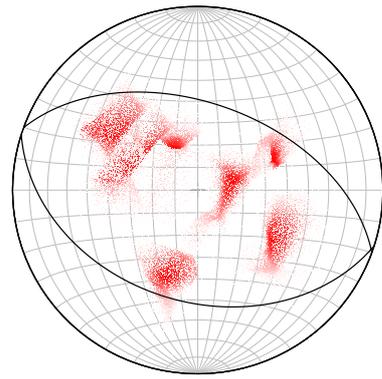


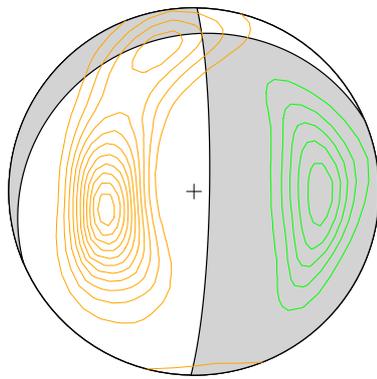
Figure C.18: Focal mechanisms for 901 and 902.



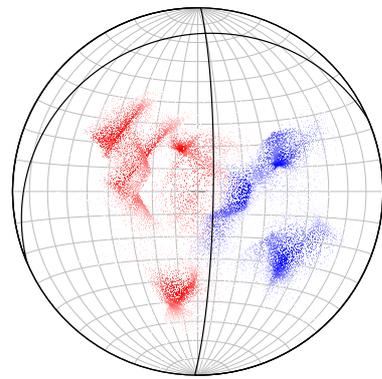
(a) 903, error=26°



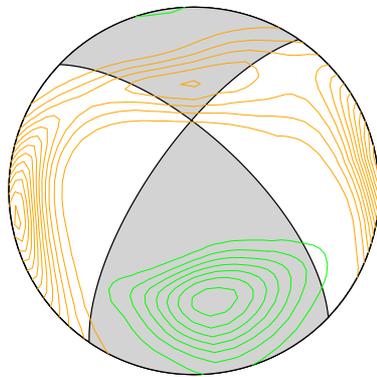
(b) 903, error=26°



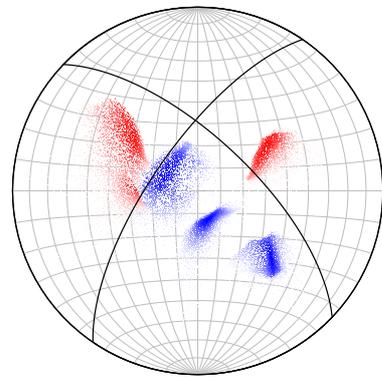
(c) 904, error=27°



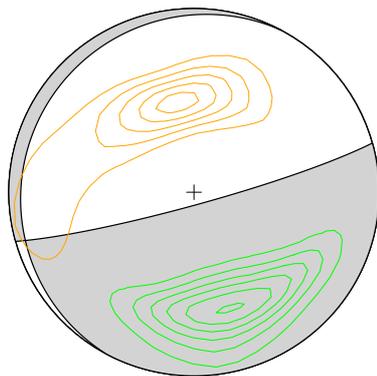
(d) 904, error=27°



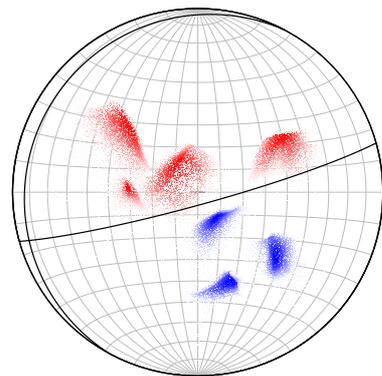
(e) 905, error=26°



(f) 905, error=26°

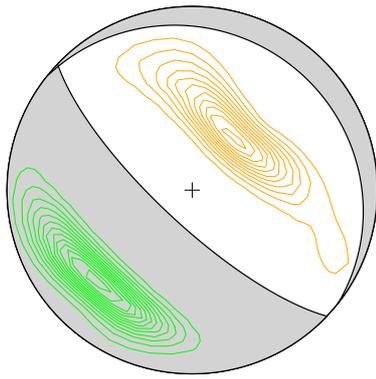


(g) 906, error=27°

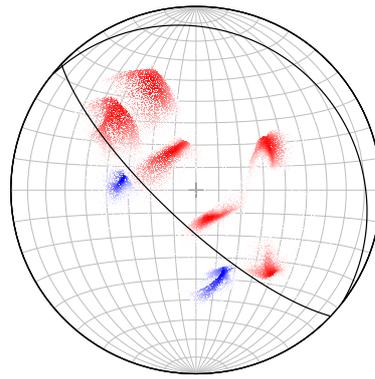


(h) 906, error=27°

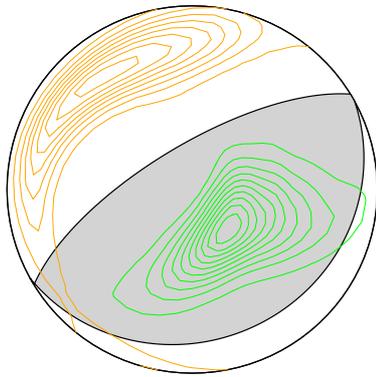
Figure C.19: Focal mechanisms for 903, 904, 905 and 906.



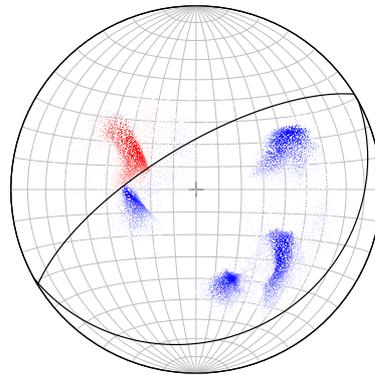
(a) 907, error=23°



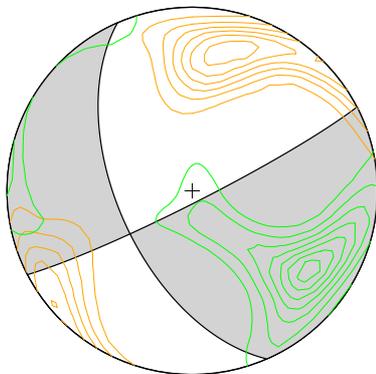
(b) 907, error=23°



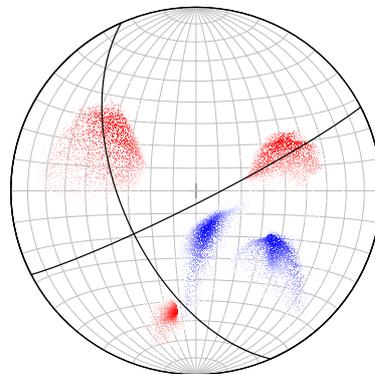
(c) 908, error=27°



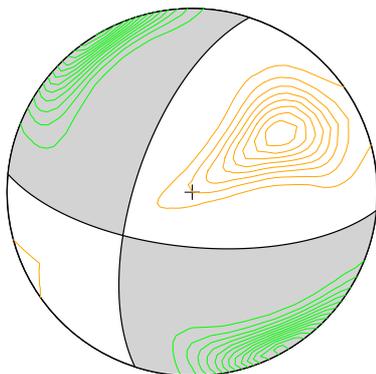
(d) 908, error=27°



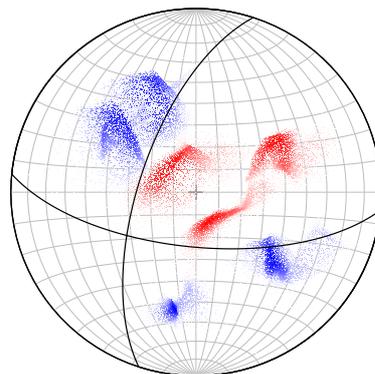
(e) 909, error=27°



(f) 909, error=27°

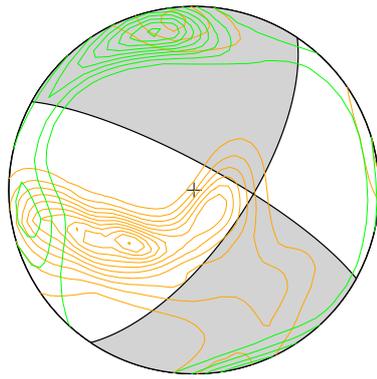


(g) 910, error=21°

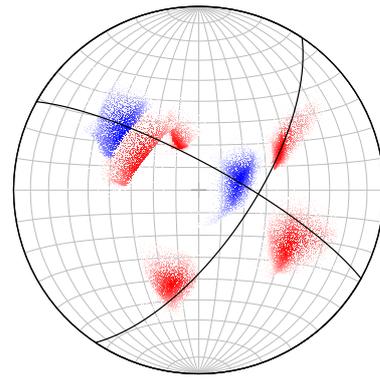


(h) 910, error=21°

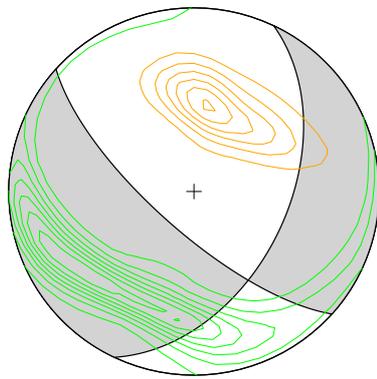
Figure C.20: Focal mechanisms for 907, 908, 909 and 910.



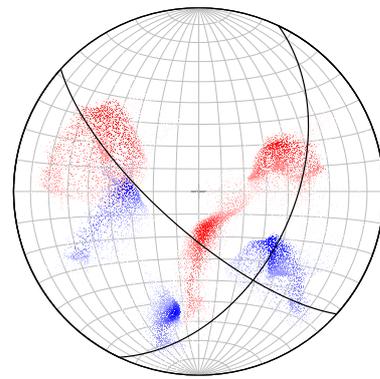
(a) 168, error=30°



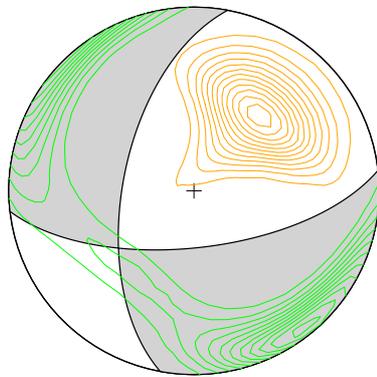
(b) 168, error=30°



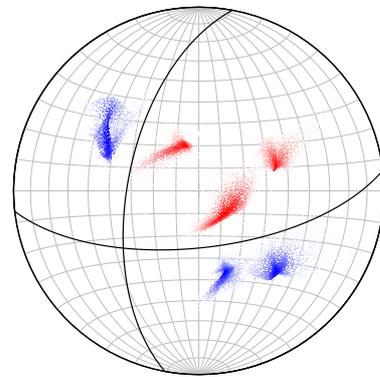
(c) 170, error=29°



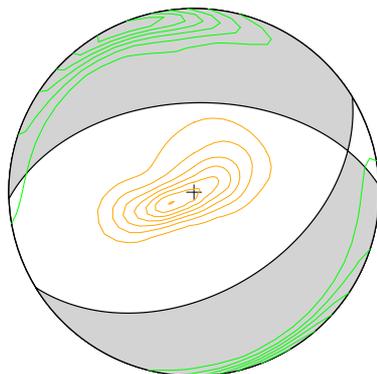
(d) 170, error=29°



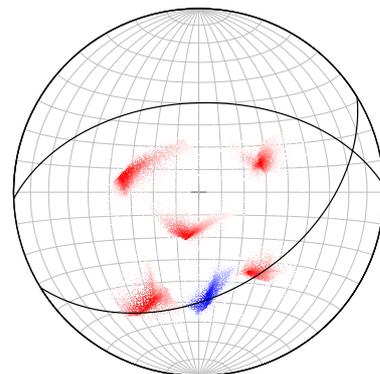
(e) 238, error=25°



(f) 238, error=25°



(g) 290, error=27°



(h) 290, error=27°

Figure C.21: Focal mechanisms 168, 170, 238 and 290.

C.4 Time slices

Location plots for five different time periods are shown below. These correspond to the five time periods identified in Fig. 3.24.

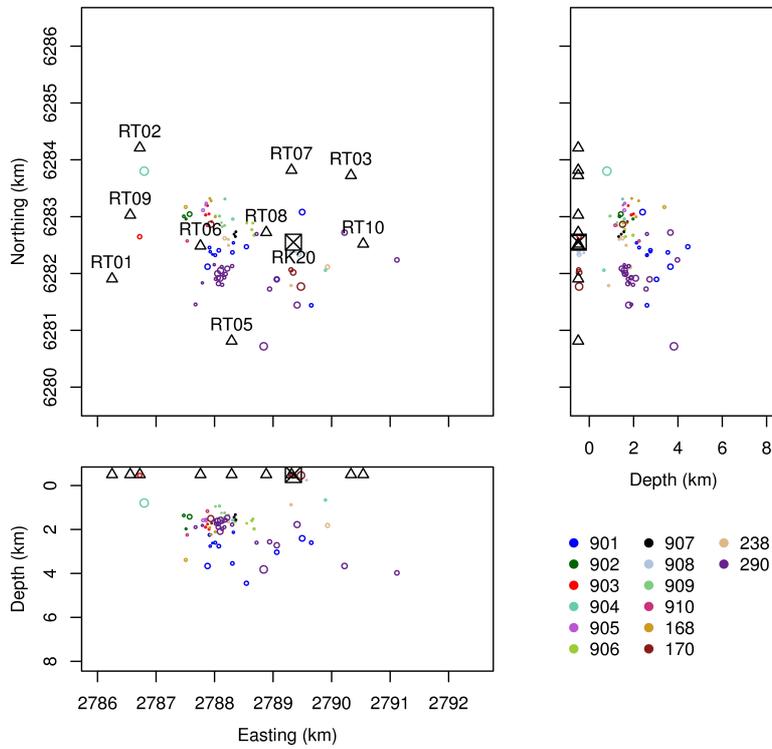


Figure C.22: Pre-RK20 injection, 29 September to 4 October 2008.

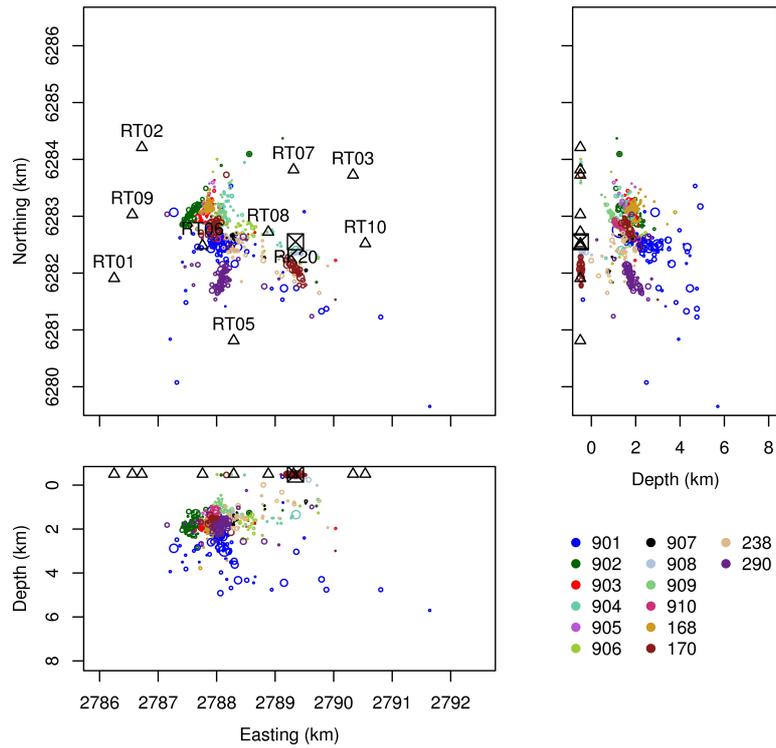


Figure C.23: First 3.5 weeks of RK20 injection, 5 October to 29 October 2008. Flow rate is 650 t/h.

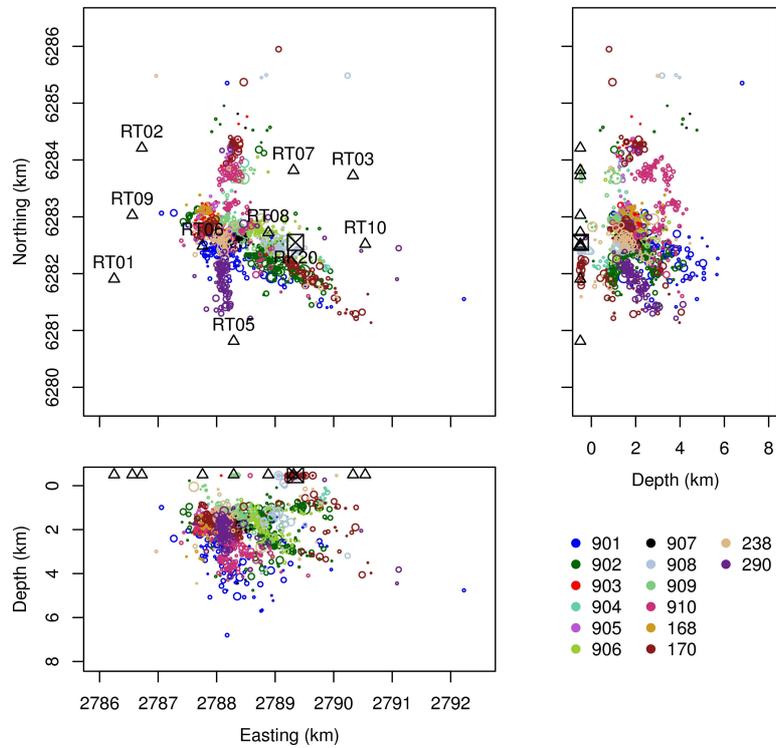


Figure C.24: Very seismically active period, 30 October to 10 November 2008. Flow rate has increased by 8% to 700t/h.

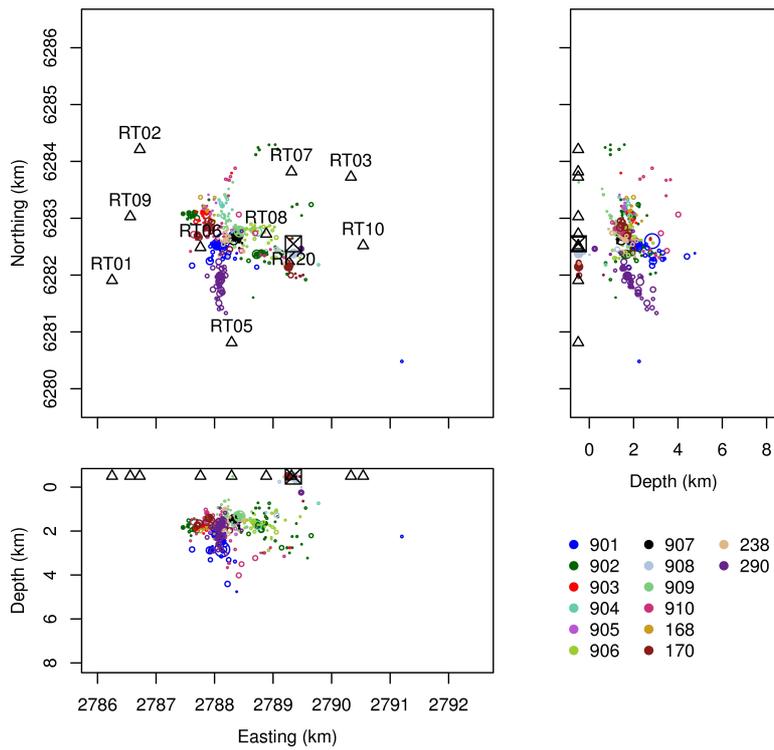


Figure C.25: Seismically inactive period (compared to Fig. C.24), 11 November 2008 to 26 January 2009.

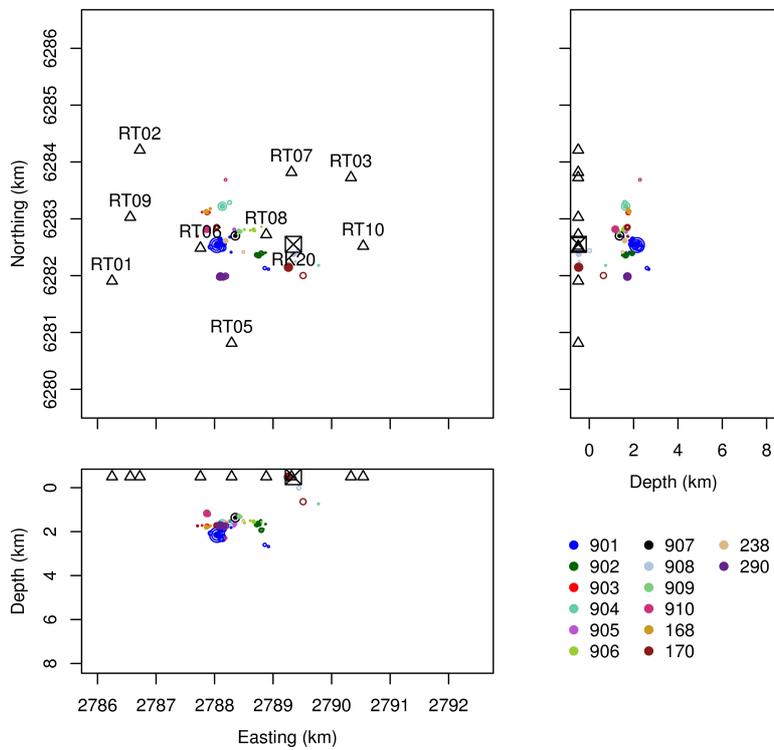


Figure C.26: Last month of data, 27 January to 28 February 2009. RK20 reduces flow rate and RK11/12 begin pumping again.

Appendix D

Network design

D.1 Marginal and conditional means and variances

The marginal (unconditional) variance of a random variable y is defined as

$$V[y] = E[(y - E[y])^2] \quad (\text{D.1})$$

$$= E[y^2 - 2yE[y] + (E[y])^2] \quad (\text{D.2})$$

$$= E[y^2] - E[y]^2 \quad (\text{D.3})$$

Similarly, the variance of y conditional on a fixed value of another (random) variable x is

$$V[y|x] = E[(y - E[y|x])^2 | x] \quad (\text{D.4})$$

If x is a random variable then so is $V[y|x]$, and it has an expected value

$$E[V[y|x]] = E[y^2] - E[E[y|x]^2] \quad (\text{D.5})$$

Similarly, we can treat the conditional expected value $E[y|x]$ as a random variable with variance

$$V[E[y|x]] = E[E[y|x]^2] - E[E[y|x]]^2 \quad (\text{D.6})$$

$$= E[E[y|x]^2] - E[y]^2 \quad (\text{D.7})$$

If we now add $E[V[y|x]]$ and $V[E[y|x]]$ we get

$$E[V[y|x]] + V[E[y|x]] = E[y^2] - E[E[y|x]^2] + E[E[y|x]^2] - E[y]^2 \quad (\text{D.8})$$

$$= E[y^2] - E[y]^2 \quad \text{c.f. Eq. D.3} \quad (\text{D.9})$$

$$= V[y] \quad (\text{D.10})$$

In other words, the marginal variance $V[y]$ is the sum of the expected value of the conditional variance $E[V[y|x]]$ and the variance of the conditional mean $V[E[y|x]]$.

D.2 Code implementation

This Appendix outlines the details of implementing the network design code described in Chapter 4. The functional programmes have been written in *R* [Becker *et al.* 1988], and are called from a *Python* script. *R* and *Python* are both open source programming languages.

The main *Python* script **netdesign.py** is run using the command ‘python netdesign.py -c controlfile.txt’. All necessary parameters are set in controlfile.txt, and the output is a text file containing new station locations, and a PDF file containing images of each new station determination [Fig. D.1].

The following code is an example of controlfile.txt:

```
#####CONTROL-FILE controlfile.txt FOR PROGRAMME runnetdesign.py#####
#####ensure Python and R are installed before running#####
[basic]
#####BASIC PARAMETERS
# # #CALCULATION GRID
```

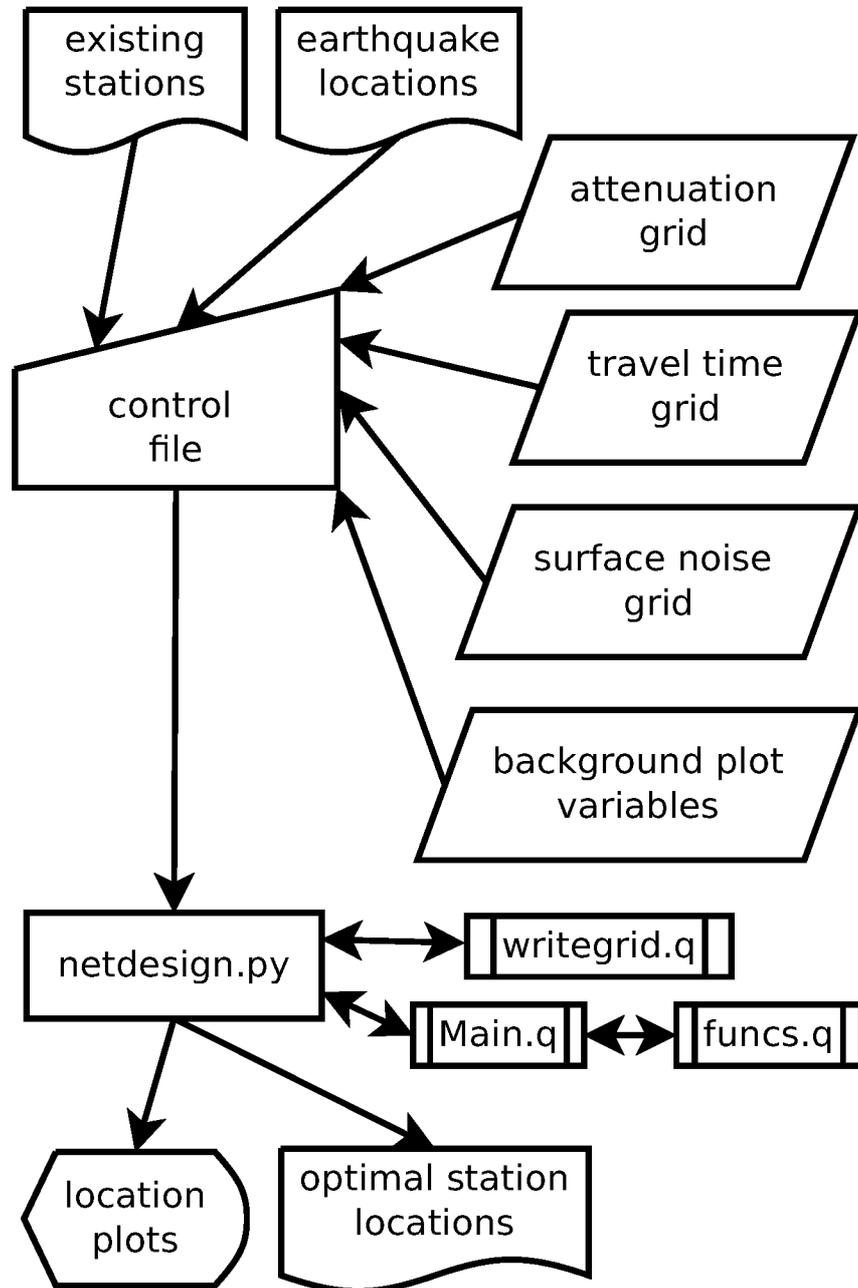


Figure D.1: Network design code flow. All necessary parameters are set in the control file. **netdesign.py** reads in all of these parameters and runs the appropriate *R* programmes (*.q). The output consists of a file with images of each new station determination and a text file containing the new station locations.

```

# min and max of grid, where x is in Eastings and y in Northings (m)
xmin=2831700
xmax=2841700
ymin=6336362
ymax=6346362

# number of grid points wanted along each axis
nxgrid= 25
nygrid =25
# # # #

# # #STATION INFORMATION
#number of existing stations to use
nexist=6
# file with existing stations, in format: Easting Northing
stexistfile=../staexist.txt
# file with existing stations names
stexistlabels=../stalabel.txt

# number of new stations wanted?
newst=10
# # # #

# # # EARTHQUAKE INFORMATION
### Do you have an earthquake file with earthquake locations? (YES/NO)
eqf=YES
# file with earthquake data: Easting Northing Depth Magnitude Weight
# Weight=1 for most important, Weight=0 for the least important
eqfile=../seismall.txt
# if eqf=NO
# Generate earthquakes within a volume specified by the limits below
# x and y in Eastings and Northings, z in km
eqxmin=2837000
eqxmax=2838000
eqymin=6341000
eqymax=6342000
eqzmin=1
eqzmax=40
# how many earthquakes do you want to generate?
eqn=100
# # # #

# # # OUTPUT FILES
# Name Easting Northing of best new stations
bestlocations= bestlocations.txt
# image of results, pdf file

```

```

locpic= newstationpic.pdf
# use image of singular results, higher quality ps file? (YES/NO)
useps=NO
stpic=Pvelmodell
# save data in R format? (YES/NO) (will use name specified by stpic)
Rdata=YES
# # # #

#####
[model]
#####VELOCITY AND ATTENUATION MODELS

# # # UNIFORM VELOCITY
# Use a uniform model? (YES/NO)
unif=NO
# if unif==YES:
# velocity in km/s
velocity=4.6
# attenuation parameters
## Dominant Frequency
freq=10
## Average Quality Factor (for attenuation)
Qunif=50
### scaling factor (q in text) 0 to 1
ascale=0.345
# # # #

# # # 3D VELOCITY MODEL
# if unif==NO:
#have you already created time and attenuation grids?
#these should be named from the variable below:
#ptimefilesnxgridxnygridGRID.txt etc
gridexist=YES
#You need an x,y grid for each earthquake with the name in
#the format "ptimesfiles%4.3f-%4.3f-%1.3f.dat",seisx,seisy,seisz
#format= .dat or .txt
#binary format .dat will be quicker than .txt files
format=.dat
ptimefiles=../Markray/NEWdist2grid
pattenfiles=../Markray/NEWdist2grid

#use P and S? (YES/NO)
PS=NO
#if PS==YES:
#provide S travel time and attenuation grids:
stimefiles=../Markray/NEWdist2grid

```

```

sattenfiles=./Markray/NEWdist2grid

# # # VELOCITY MODEL GRID
#This grid will be projected onto the calculation grid:
#min and max of time and attenuation grid,
#where x is in Eastings and y in Northings
sxmin=2828817
sxmax=2844817
symin=6333279
symax=6349279

#grid sizes
kx=801
ky=801
# # # #

#####
[extras]
#####OPTIONAL EXTRAS
#Use a surface noise model? (YES/NO)
surfn=NO

#if surfn==YES: file with noise data
#currently must load a saved R matrix named 'noise'
noisef=./pics/noise.RData
##noise scaling value, the higher namp the more effect the noise has
namp=100

#Do you want any background information in the images? (YES/NO)
back=YES

#currently loads saved R variables e.g roads, reservoir, buildings etc
# You'll have to change the parameters that are plotted in plot.tabgridback()
#if you want to change these
backg=backg.RData

#####

```

Explanation of Control Parameters:

xmin, xmax, ymin, ymax: the geographic limits, in Eastings and Northings, of the grid on which to search for optimal station locations.

nxgrid, nygrid: number of grid squares in the East and North directions, respectively.

nexist: number of existing stations to use (this doesn't have to be the same number of stations that are in stexistfile: the algorithm can start with 0 or with only a few stations if preferred)

stexistfile: name of the file with the existing station locations, file has two columns: Easting Northing.

stexistlabels: name of the file with a column of existing station names.

newst: number of new stations wanted.

eqf: =YES, uses earthquake locations specified in eqfile. =NO, generate earthquakes within a specified region

eqfile: name of the file with the earthquake locations that the network is desired to locate accurately, file has five columns: Easting Northing Depth Magnitude Weight. Weight is used if some earthquake locations are of greater import than others. Set Weight=0-1, where 0 is the least important and 1 the most.

eqxmin, eqxmax, eqymin, eqymax, eqzmin, eqzmax: if eqf=NO, the limits, in Eastings, Northings, and km, of the region in which to generate earthquakes.

eqn: if eqf=NO, the number of earthquakes to generate.

bestlocations: output file with the names and locations of existing stations and best new stations, file has three columns: Name Easting Northing. Name of new stations are 1-newst, with 1 the most important, and newst the least important.

locpic: output pdf file with images showing the initial set up of earthquakes and stations, and an image showing the design criterion contours for each new station added.

useps:=YES (output results to singular postscript files with the prefix specified by stpic)
= NO (output to pdf file)

stpic: prefix of postscript files and Rdata file

Rdata =YES (save data in R format with the name stpic.RData) = NO (don't save in R format)

unif=YES (use a uniform velocity model - travel times and attenuation are calculated internally) = NO (use a different velocity model - the user needs to provide travel times from each earthquake to every grid point.)

velocity: velocity to be used by the uniform velocity model, in km/s.

freq: dominant frequency, used for attenuation.

Qunif: average quality factor, used for attenuation

ascale: attenuation scaling value, q in the text. Increase with decreasing magnitudes of earthquakes of interest.

gridexist=YES (user has a file containing an $neq \times nxgrid \times nygrid$ grid with travel times from every earthquake (neq) to every grid point. This grid is named

ptimesfilesnxgridxnygridGRID.txt) = NO (the grid will be calculated using **write-grid.q** and neq files (one for every earthquake) containing an $nxgrid \times nygrid$ grid with travel

times from the earthquake to every grid point. These files need to be named `ptimefiles4.3f-%4.3f-%1.3f.dat.txt`, `seisx`, `seisy`, `seisz`. Here `seisx`=Easting (km), `seisy`=Northing (km), `seisz`=Depth (km).)

`format`: file extension of the earthquake traveltime files `ptimefiles`, = `.dat` (binary format, quickest.) = `.txt` (text file).

`ptimefiles`: prefix of the file names for earthquake files and grid files for travel times.

`pattenfiles`: prefix of the file names for earthquake files and grid files for attenuation values.

`PS=YES` (S-wave times are used as well as P-wave, requires S travel time and attenuation files as with P) =`NO` (just P-wave used.) Note, this is not an option with the uniform velocity model as it will not change the results in that case (change will be constant for all grids).

`stimefiles`: prefix of the file names for earthquake files and grid files for travel times for S-wave.

`sattenfiles`: prefix of the file names for earthquake files and grid files for attenuation values for S-wave.

`sxmin`, `sxmax`, `symin`, `symax`: the limits, in Eastings and Northings, of the grid on which the travel time and attenuation calculations have been performed.

`kx`, `ky`: number of grid squares in the East and North directions for the grid on which the travel time and attenuation calculations have been performed.

surfn=YES (use a surface noise model) =NO (don't use a surface noise model).

noisef: file containing the surface noise information, currently uses the R function load() to load a saved $nxgrid \times nygrid$ matrix named 'noise' with surface noise values in dBA.

namp: value to scale the surface noise, the greater namp the more impact the surface noise has.

back=YES (use background information in the plots) =NO (don't use any extra information in the plots).

backg: file containing the extra variables to be plotted. Currently uses the R function load() to load the variables. If these need to be changed, edits must be made within **funcs.q** in the function plot.tabgridback() for the different variable names/colours/symbols.

runnetdesign.py extracts all the parameters provided by `controlfile.txt` and exports these as environment variables. This exportation is required for *R* to find the variables, and they are all removed at the end of **runnetdesign.py** to ensure a clean slate at the start of every run. **runnetdesign.py** then runs **writegrid.q** if required. This function converts files with travel time data for a single earthquake into a file containing a grid with travel time data for all earthquakes. Each earthquake file contains an evenly spaced x by y grid of values at $z=0$. Each row consists of the values at each x location, along a specific y value. Lines step through y values. The first line begins at x_{min} , y_{min} . This is the most time consuming part of the process, and will depend on the number of grid points and the number of earthquakes, as well as the file format. If the file format is `.txt` these files will first be converted into binary format, remember to set `fformat=.dat` for any subsequent runs to speed things up. **writegrid.q** need only be run once for an earthquake set if you are using the same calculation grid. In `controlfile.txt` set `gridexist=YES` for any subsequent runs of the same data

set. **Main.q** is then run, it first calls **first.q**, which just ensures that functions are called from the correct file, **funcs.q**. **Main.q** will calculate the optimal locations of new seismic stations using the algorithm as described in Chapter 4. The output will be a text file of optimal locations specified by `bestlocations` and either a pdf file specified by `locpic` with all images, or individual higher quality postscript files specified by `stpik`. These images show the initial set up of earthquakes and stations, the mean travel time contours, mean attenuation contours, and the design criterion contours for each new station added.

At the most basic level, the code can be run with no other input required: `set unif=YES`, `eqf=NO`, `nexist=0`, and ensure all other optional parameters are set to NO. Now simply specify your grid boundaries, earthquake zone boundaries, number of earthquakes, number of new stations wanted, and attenuation parameters. An optimal network will be designed for the region you have specified. The control file shown below can be used for such a basic computation. Note that `nxgrid` and `nygrid` are set to low numbers: this will speed up the computation during the set up and checking process. Once you are ready to calculate your final designs, these numbers should be increased to provide a more thorough sampling of the grid space.

Example of `controlfile.txt` for basic run:

```
#####CONTROL-FILE controlfile.txt FOR PROGRAMME runnetdesign.py#####
#####ensure Python and R are installed before running#####
[basic]
#####BASIC PARAMETERS
# # #CALCULATION GRID
# min and max of grid, where x is in Eastings and y in Northings (m)
xmin=2831700
xmax=2841700
ymin=6336362
ymax=6346362

# number of grid points wanted along each axis
nxgrid= 25
nygrid =25
# # # #

# # #STATION INFORMATION
```

```

#number of exiting stations to use
nexist=0
# file with existing stations, in format: Easting Northing
stexistfile=./staexist.txt
# file with existing stations names
stexistlabels=./stalabel.txt

# number of new stations wanted?
newst=10
# # # #

# # # EARTHQUAKE INFORMATION
### Do you have an earthquake file with earthquake locations? (YES/NO)
eqf=NO
# file with earthquake data: Easting Northing Depth Magnitude Weight
# Weight=1 for most important, Weight=0 for the least important
eqfile=./seismall.txt
# if eqf=NO
# Generate earthquakes within a volume specified by the limits below
# x and y in Eastings and Northings, z in km
eqxmin=2837000
eqxmax=2838000
eqymin=6341000
eqymax=6342000
eqzmin=1
eqzmax=40
# how many earthquakes do you want to generate?
eqn=100
# # # #

# # # OUTPUT FILES
# Name Easting Northing of best new stations
bestlocations= bestlocations.txt
# image of results, pdf file
locpic= newstationpic.pdf
# use image of singular results, higher quality ps file? (YES/NO)
useps=NO
stpic=Pvelmodel1
# save data in R format? (YES/NO) (will use name specified by stpic)
Rdata=NO
# # # #

#####
[model]
#####VELOCITY AND ATTENUATION MODELS

```

```

# # # UNIFORM VELOCITY
# Use a uniform model? (YES/NO)
unif=YES
# if unif==YES:
# velocity in km/s
velocity=4.6
# attenuation parameters
## Dominant Frequency
freq=10
## Average Quality Factor (for attenuation)
Qunif=50
### scaling factor (q in text) 0 to 1
ascale=0.0
# # # #

# # # 3D VELOCITY MODEL
# if unif==NO:
#have you already created time and attenuation grids?
#these should be named from the variable below:
#ptimefilesnxgridxnygridGRID.txt etc
gridexist=YES
#You need an x,y grid for each earthquake with the name in
#the format "ptimesfiles%4.3f-%4.3f-%1.3f.dat",seisx,seisy,seisz
#format= .dat or .txt
#binary format .dat will be quicker than .txt files
format=.dat
ptimefiles=../Markray/NEWdist2grid
pattenfiles=../Markray/NEWdist2grid

#use P and S? (YES/NO)
PS=NO
#if PS==YES:
#provide S travel time and attenuation grids:
stimefiles=../Markray/NEWdist2grid
sattenfiles=../Markray/NEWdist2grid

# # # VELOCITY MODEL GRID
#This grid will be projected onto the calculation grid:
#min and max of time and attenuation grid,
#where x is in Eastings and y in Northings
sxmin=2828817
sxmax=2844817
symin=6333279
symax=6349279

#grid sizes

```

```

kx=801
ky=801
# # # #

#####
[extras]
#####OPTIONAL EXTRAS
#Use a surface noise model? (YES/NO)
surfn=NO

#if surfn==YES: file with noise data
#currently must load a saved R matrix named 'noise'
noisef=./pics/noise.RData
##noise scaling value, the higher namp the more effect the noise has
namp=100

#Do you want any background information in the images? (YES/NO)
back=NO

#currently loads saved R variables e.g roads, reservoir, buildings etc
# You'll have to change the parameters that are plotted in plot.tabgridback()
#if you want to change these
backg=backg.RData

#####

```

We recommend to next input your specific earthquake locations of interest and existing station locations, changing eqf=YES, and nexist=number of existing stations, and setting the relevant file names. Check the first plot output to be sure earthquakes and stations are being read in and plotted correctly.

Now input your travel time and attenuation grids, which are evenly spaced x by y grids of values at z=0 for every earthquake. Each row consists of the values at each x location, along a specific y value. Lines step through y values. The first line begins at xmin, ymin. These grids should cover a large enough area that no boundary artifacts will be created [see Fig. 4.8]. Set the grid boundaries and sizes, and gridexist=NO. For the first run set fformat=.txt, these will be converted to binary files with the extension .dat, as these are much faster for the conversion to the full earthquake grid. For any subsequent runs set fformat=.dat. These will be projected onto the desired calculation grid and stored in a 3D

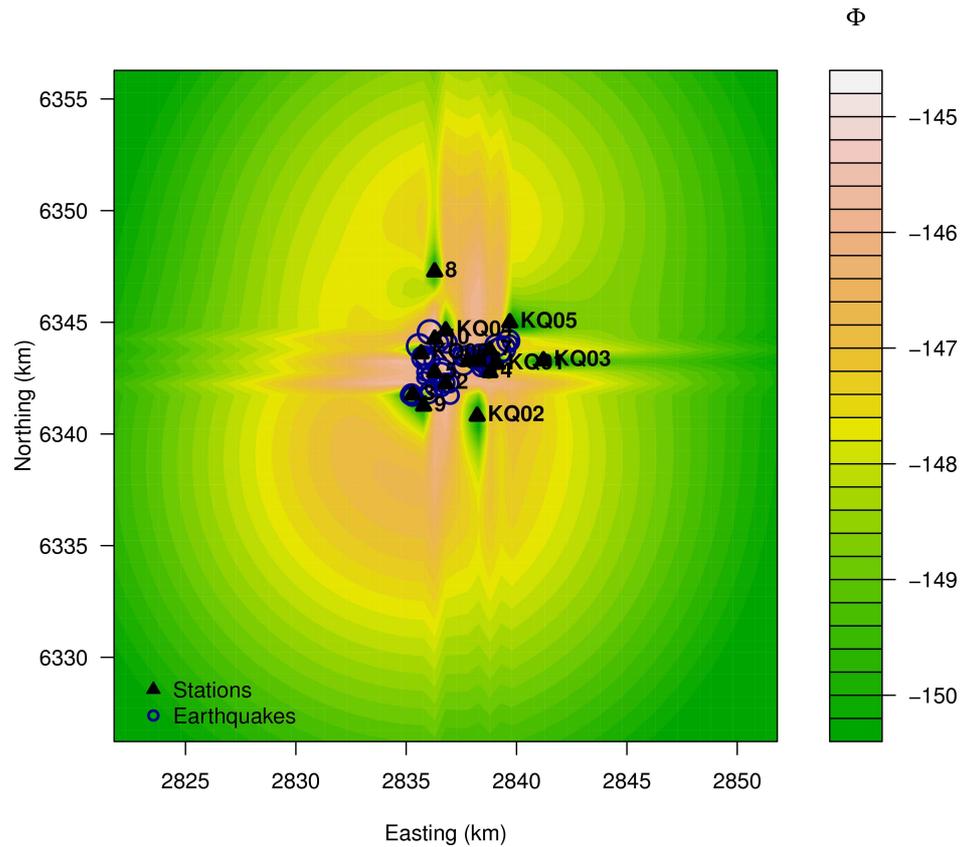


Figure D.2: Example of insufficient sampling during ray tracing resulting in the design criterion ‘seeing’ some of the grid lines.

grid, with the 3rd dimension being the number of earthquakes. Both of these conversions are carried out within **writegrid.q**. The second and third plots in your output file are the mean travel time contours, and the mean attenuation contours. It is important to check that these are centered as expected over your earthquake locations, and that both increase with distance from your earthquake locations. If they behave in this manner then your grids should have been converted correctly. If after running **netdesign.py**, in the plotted images you can ‘see’ the grid as lines through the criterion contours rather than it smoothly varying [see Fig. D.2] this could be due to either an insufficient sampling size during criterion determination (try increasing `nxgrid` and `nygrid` and see if this affects the outcome), or artifacts from your 3D travel time grid calculation, due to e.g. insufficient sampling size during ray tracing (try increasing your grid sampling size), the second of these is the most likely.

D.3 Darfield earthquake

In the aftermath of the magnitude 7.1 earthquake that hit Christchurch, New Zealand at 4:35 am (NZST) on 4th September 2010, there was a joint effort between Victoria University of Wellington, The University of Auckland, University of Wisconsin and GNS Science to deploy seismometers around the affected area. Our network design programme was still in its early stages, but was adapted to the Christchurch region using a uniform velocity model of 5.5 km/s (the top 0–12 km) and attenuation $Q=50$. At the same time, potential sites for new stations were decided upon using standard techniques by seismologists. Due to the sudden nature of the event and the desired quick seismometer deployment time, our design programme was being adapted and run at the same time as the first few stations were deployed. Fig. D.3 displays two seismometers that were deployed concurrently with our first application to the earthquake set. The location of these seismometers matches the region identified as best for a new station by our design programme (white/light-pink region in Fig. D.3). The comparison of locations chosen demonstrates that in some cases seismologists intuition (based on the fault location and existing station locations) matched the programmes' design very well [Figs. D.3 & D.4], whilst in other cases sites that seismologists had not considered were highlighted by the programme [Fig. D.4]. Fig. D.4 shows that the design's most important site was a location not considered by VUW/GNS seismologists. Design results combined with the sites identified by seismologists were used to guide the placement of nine new stations, displayed in Fig. D.5. Note that the five yellow stations from Fig. D.4 have been included in the final station placement set, and have shifted somewhat as these were the decided upon deployment locations, but were shifted slightly due to land access/permissions etc. A few days after this application aftershocks began to migrate eastward; increasing the importance of our site #1.

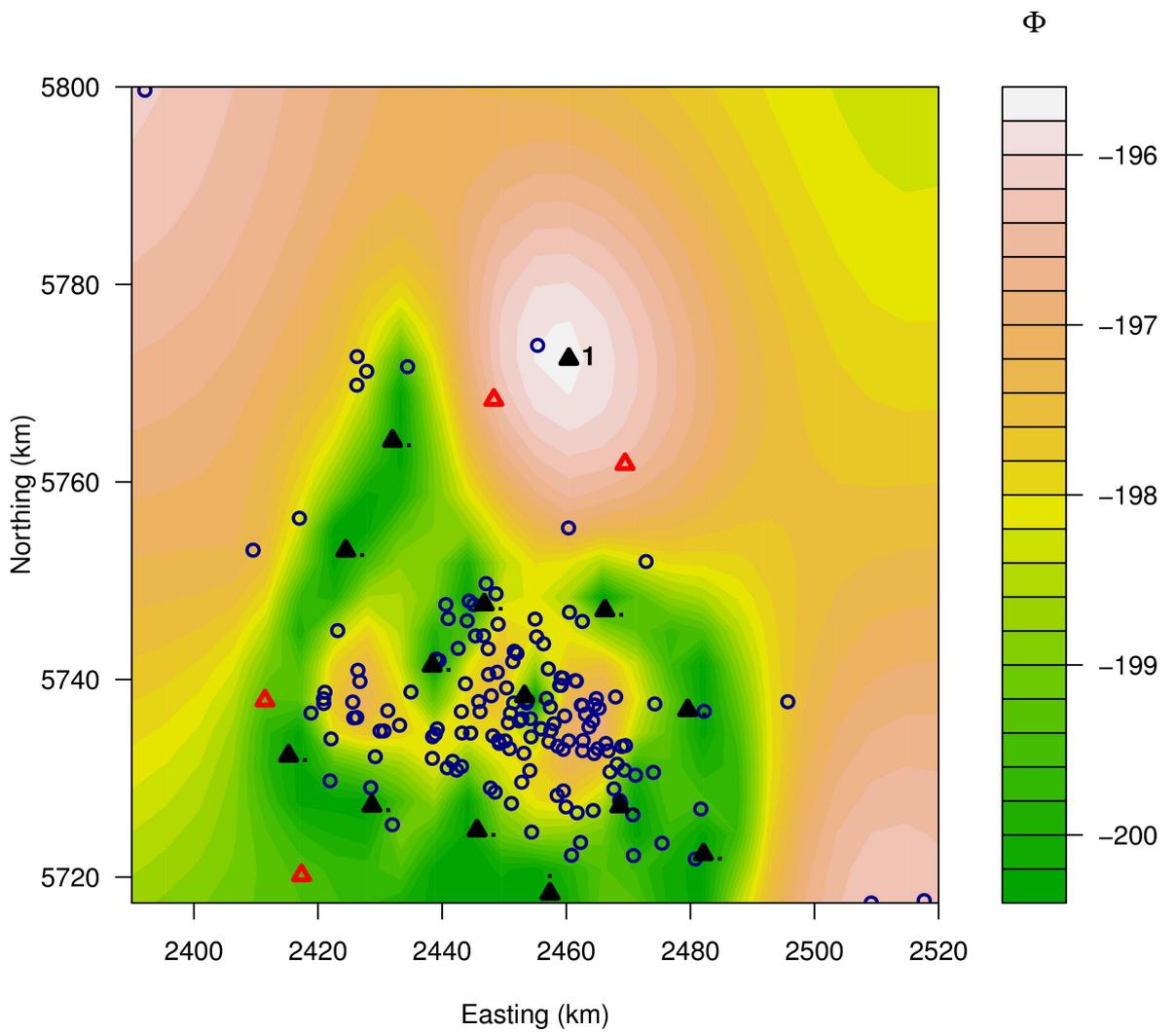


Figure D.3: First application to the Darfield earthquake. Black triangles = permanent Geonet stations and the Geonet stations deployed in the two days immediately succeeding the Christchurch earthquake, red triangles = VUW/UA/UW/GNS sites deployed concurrently with the running of my programme. Note that the location indicated as best for the first new location, white region with black triangle labelled 1, matches two of the sites decided by VUW/GNS seismologists in the field.

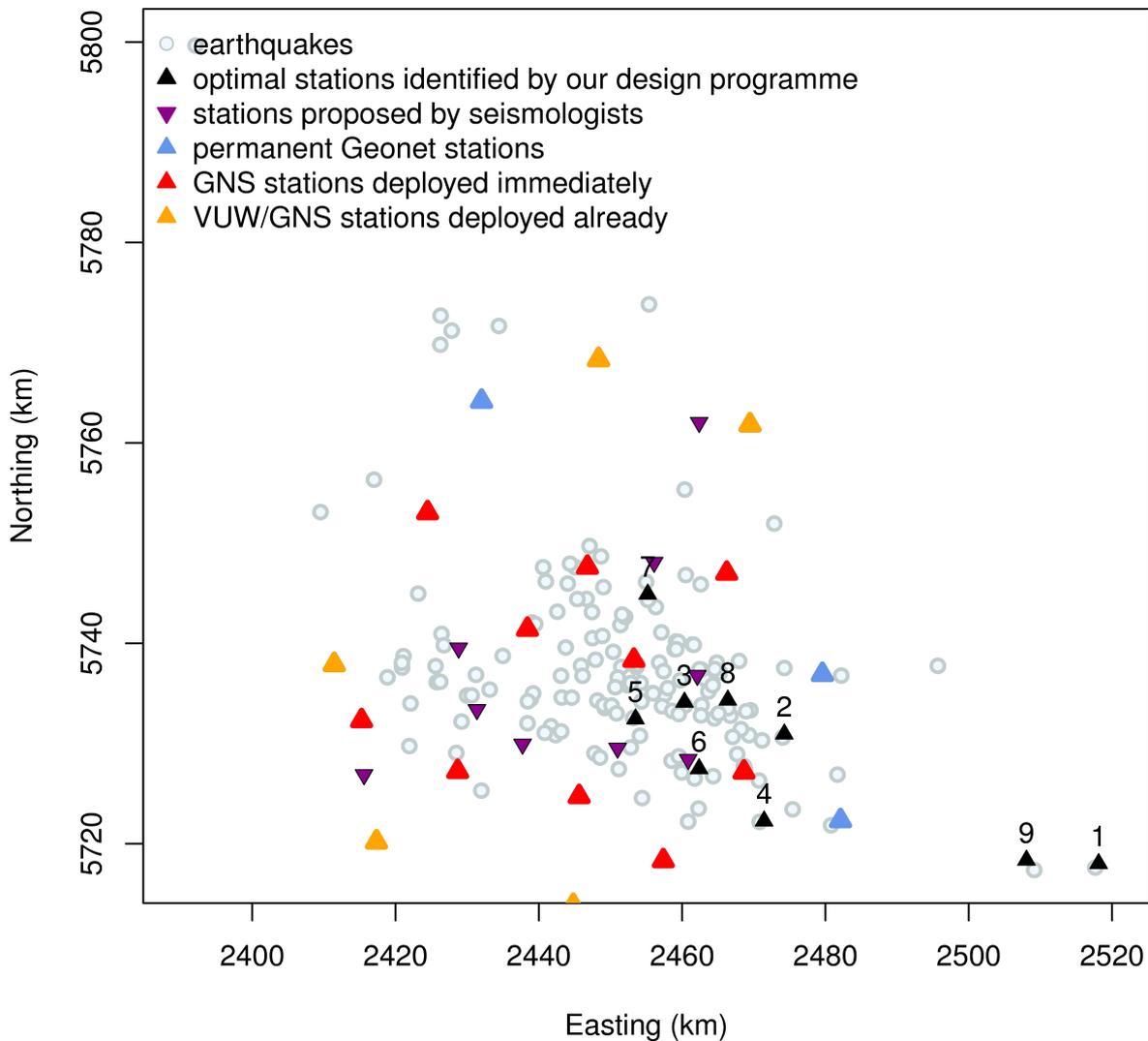


Figure D.4: Second application to the Darfield earthquake. Gray circles = earthquakes, blue triangles = permanent Geonet stations, red triangles = Geonet stations deployed in the two days immediately succeeding the Christchurch earthquake, orange triangles = VUW/UA/UW/GNS sites deployed concurrently with the first run of our design programme (see Fig. D.3), purple triangles = proposed sites chosen by seismologists, black triangles = best new sites identified by design programme: where 1 is most important and 9 is least important. Note the design's stations ranked 3 and 5-8 match well with four proposed stations. The design's stations don't cluster to the west at all, as the proposed stations do. Also, the design's most important site was a location not considered by VUW/GNS seismologists. A few days after this application aftershocks began to migrate eastward; confirming site #1 was indeed a good location.

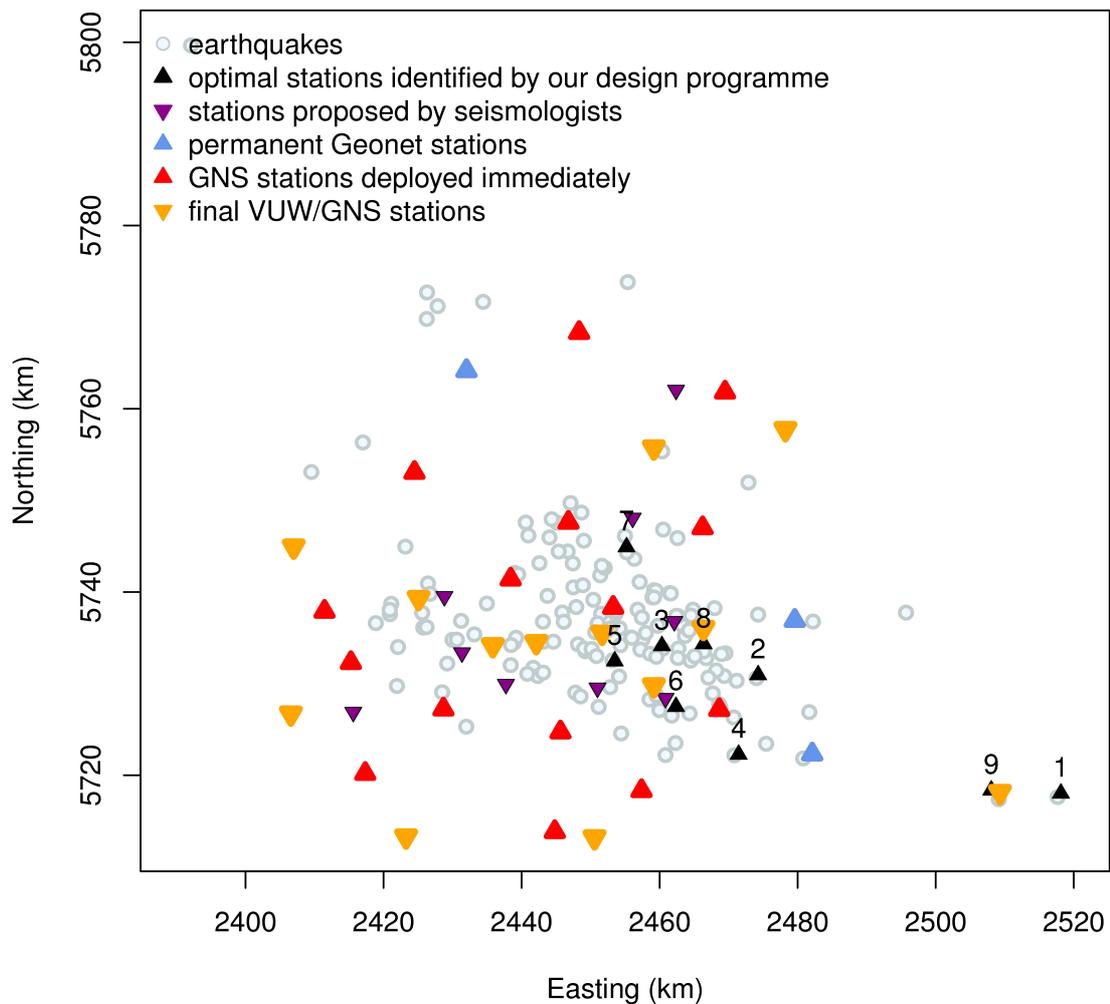


Figure D.5: Final stations deployed after the Darfield earthquake. Gray circles = earthquakes, blue triangles = permanent Geonet stations, red triangles = Geonet stations deployed in the two days immediately succeeding the Christchurch earthquake, orange triangles = VUW/UA/UW/GNS sites deployed, purple triangles = proposed sites chosen by seismologists, black triangles = best new sites identified by design programme: where 1 is most important and 9 is least important. Note that the orange triangles include the previous five orange stations [Fig. D.4], but these locations have shifted somewhat.

Bibliography

Acuna, J. A., J. Stimac, L. Siradazwar, and R. Pasikki (2008), Reservoir management at Awibengkok geothermal field, West Java, Indonesia, *Geothermics*, 37(3), 332–346, doi: 10.1016/j.geothermics.2008.02.005.

AES Highgrove (2006), Application for Certification: AES Highgrove Project: Section 8.5 Noise, viewed July 7 2011, http://www.energy.ca.gov/sitingcases/highgrove/documents/applicant/AFC_CDRM/Volume_01_AES_Highgrove_Project_AFC/.

Alexander, D. E. (1993), *Natural disasters*, part 2 ed., Springer Science and Business.

Allis, R. G. (1997), Natural state and response to development of Kawerau geothermal field, New Zealand, in *Transactions - Geothermal Resources Council*, vol. 21, pp. 3–10.

Anstey, N. (1966), Correlation techniques - a review, *Canadian Journal of Exploration Geophysics*, 2(1), 55–82.

Arehart, G. (2002), Timing of volcanic, plutonic and geothermal activity at Ngatamariki, New Zealand, *Journal of Volcanology and Geothermal Research*, 116(3-4), 201–214, doi: 10.1016/S0377-0273(01)00315-8.

Armannsson, H., and H. Kristmannsdottir (1992), Geothermal environmental impact, *Geothermics*, 21(5-6), 869–880, doi:10.1016/0375-6505(92)90038-B.

Arnold, R., and J. Townend (2007), A Bayesian approach to estimating tectonic stress from seismological data, *Geophysical Journal International*, 170(3), 1336–1356.

- Aster, R., and C. Rowe (2000), Automatic phase pick refinement and similar event association in large seismic data sets, *Advances in Seismic Event Location*, *chapt. 9*, 231–263.
- Aster, R., and J. Scott (1993), Comprehensive characterization of waveform similarity in microearthquake data sets, *Bulletin of the Seimological Society of America*, 83(4), 1307–1307.
- Baerentzen, J. (2001), *On the implementation of fast marching methods for 3D lattices*, IMM, Informatics and Mathematical Modelling, The Technical University of Denmark.
- Baisch, S., L. Ceranna, and H. Harjes (2008), Earthquake Cluster: What Can We Learn from Waveform Similarity?, *Bulletin of the Seismological Society of America*, 98(6), 2806–2814, doi:10.1785/0120080018.
- Bannister, S., S. Sherburn, T. Powell, and D. Bowyer (2008), Microearthquakes at the Rotorokawa Geothermal Field, New Zealand, *GRC Transactions*, 32.
- Barani, S., G. Ferretti, M. Massa, and D. Spallarossa (2007), The waveform similarity approach to identify dependent events in instrumental seismic catalogues, *Geophysical Journal International*, 168(1), 100–108, doi:10.1111/j.1365-246X.2006.03207.x.
- Baria, R., J. Michelet, S. Baumgartner, J. Nicholls, T. Hettkamp, D. Teza, N. Soma, A. H., J. Garnish, and T. Megel (2005), Creation and Mapping of 5000 m deep HDR/HFR Reservoir to produce electricity, in *Proc. World Geothermal Congress, Antalya, Turkey*.
- Beavan, J., P. Tregoning, M. Bevis, T. Kato, and C. Meertens (2002), Motion and rigidity of the Pacific Plate and implications for plate boundary deformation, *Journal of Geophysical Research B: Solid Earth*, 107(10), ETG 19–1 – 19–15.
- Becker, R. A., J. M. Chambers, and A. R. Wilks (1988), *The New S Language*, Wadsworth & Brooks/Cole.
- Behr, Y., J. Townend, S. Bannister, and M. K. Savage (2011), Crustal shear wave tomography of the Taupo Volcanic Zone, New Zealand, via ambient noise correlation between multiple

- three-component networks, *Geochemistry Geophysics Geosystems*, 12(3), 1–18, doi:10.1029/2010GC003385.
- Bell, B., D. Percival, and A. Walden (1993), Calculating Thomson's Spectral Multitapers by Inverse Iteration, *Journal of Computational and Graphical Statistics*, 2(1), 119–130.
- Bibby, H. M., T. G. Caldwell, F. J. Davey, and T. H. Webb (1995), Geophysical evidence on the structure of the Taupo Volcanic Zone and its hydrothermal circulation, *Journal of Volcanology and Geothermal Research*, 68(1-3), 29–58.
- Bibby, H. M., G. F. Risk, and T. G. Caldwell (1999), Resistivity structure of the central Taupo Volcanic Zone, New Zealand, *Journal of Volcanology and Geothermal Research*, pp. 163–181.
- Bibby, H. M., G. Risk, and T. G. Caldwell (2002), Long offset tensor apparent resistivity surveys of the Taupo Volcanic Zone, New Zealand, *Journal of Applied Geophysics*, 49(1-2), 17–32.
- Bignall, G., and C. C. Harvey (2005), Geoscientific Review of the Kawerau Geothermal Field, *GNS Science Client Report*, (520W1048).
- Bjornsson, G., A. Hjartarson, G. S. Bodvarsson, and B. Steingrímsson (2003), Development of a 3-D geothermal reservoir model for the greater Hengill volcano in SW-Iceland, in *TOUGH Symposium*, pp. 1–11.
- Blackett, R. E., G. M. Sowards, and E. Trimmer (2004), Utah's high temperature geothermal resource potential analysis of selected sites, *Utah Geological Survey, Utah Department of Natural Resources*.
- Blackman, R. B., and J. Tukey (1959), *Particular Pairs of Windows*, pp. 98–99, New York: Dover.
- Bodvarsson, G. (1972), Thermal problems in the siting of reinjection wells, *Geothermics*, 1(2), 63–66, doi:10.1016/0375-6505(72)90013-2.

- Brown, J. R., G. C. Beroza, and D. R. Shelly (2008), An autocorrelation method to detect low frequency earthquakes within tremor, *Geophysical Research Letters*, 35(16), doi:10.1029/2008GL034560.
- Browne, P., I. Graham, R. Parker, and C. Wood (1992), Subsurface andesite lavas and plutonic rocks in the Rotokawa and Ngatamariki geothermal systems, Taupo Volcanic Zone, New Zealand, *Journal of Volcanology and Geothermal Research*, 51, 199–215.
- Bryan, C. J., and S. Sherburn (1999), Seismicity associated with the 1995-1996 eruptions of Ruapehu volcano, New Zealand: Narrative and insights into physical processes, *Journal of Volcanology and Geothermal Research*, 90(1-2), 1–18.
- Burnham, L., T. B. Johansson, H. Kelly, A. K. Reddy, and R. H. Williams (Eds.) (1993), *Renewable Energy: Sources for Fuels and Electricity*, Island Press.
- Clarke, D. (2008), Velocity Modelling and Earthquake Relocation in the Rotorua and Kawerau Geothermal Areas, Taupo Volcanic Zone, New Zealand, Ph.D. thesis, Victoria University of Wellington.
- Clarke, D., J. Townend, M. K. Savage, and S. Bannister (2009), Seismicity in the Rotorua and Kawerau geothermal systems, Taupo Volcanic Zone, New Zealand, based on improved velocity models and cross-correlation measurements, *Journal of Volcanology and Geothermal Research*, 180(1), 50–66.
- Clinton, J. (2007), Geophysical Data Processing 2, Chapter 1 : Digitisation, viewed March 10 2010, <http://www.seismo.ethz.ch/staff/jclinton/GDP07/GDP07.html>.
- Coles, D., and A. Curtis (2011), Efficient nonlinear Bayesian survey design using Dn optimization, *Geophysics*, 76(2), Q1.
- Coles, D. A., and F. D. Morgan (2009), A method of fast, sequential experimental design for linearized geophysical inverse problems, *Geophysics Journal International*, 178, 145–158.

- Console, R., and A. Giuntini (2006), An algorithm for double difference joint hypocenter determination: application to the 2002 Molise (Central Italy) earthquake sequence, *Annals of Geophysics*, 49(April/June).
- Cornet, F., and O. Scotti (1993), Analysis of induced seismicity for fault zone identification, *International Journal of Rock Mechanics and Mining Sciences & Geomechanics Abstracts*, 30(7), 789–795, doi:10.1016/0148-9062(93)90024-8.
- Cover, T., and J. Thomas (1991), *Elements of Information Theory*, Wiley-Interscience.
- Cuenot, N., J. Charley, L. Dorbath, and H. Haessler (2006), Faulting mechanisms and stress regime at the European HDR site of Soultz-sous-Forêts, France, *Geothermics*, 35(5-6), 561–575, doi:10.1016/j.geothermics.2006.11.007.
- Curtis, A. (1999), Optimal Design of focused experiments and surveys, *Geophysics Journal International*, 139, 205–215.
- Curtis, A., and H. Maurer (2000), Optimizing the design of geophysical experiments: Is it worthwhile?, *EOS Forum*, 81(20).
- Curtis, A., A. Michelini, D. Leslie, and A. Lomax (2004), A deterministic algorithm for experimental design applied to tomographic and microseismic monitoring surveys, *Geophysics Journal International*, 157, 595–606.
- Darby, D. J., and C. M. Meertens (1995), Terrestrial and GPS measurements of deformation across the Taupo back arc and Hikurangi forearc regions in New Zealand, *Journal of Geophysical Research*, 100(B5), 8221–8232.
- Davey, F., S. Henrys, and E. Lodolo (1995), Asymmetric rifting in a continental back-arc environment, North Island, New Zealand, *Journal of Volcanology and Geothermal Research*, 68(1-3), 209–238, doi:10.1016/0377-0273(95)00014-L.
- Denlinger, R., and C. Bufe (1982), Reservoir conditions related to induced seismicity at

- the Geysers steam reservoir, northern California, *Bulletin of the Seismological Society of America*, 72(4), 1317.
- DeShon, H. R., C. H. Thurber, and C. Rowe (2007), High-precision earthquake location and three-dimensional P wave velocity determination at Redoubt Volcano, Alaska, *Journal of Geophysical Research*, 112(B7), doi:10.1029/2006JB004751.
- Di Stefano, R., F. Aldersons, E. Kissling, P. Baccheschi, C. Chiarabba, and D. Giardini (2006), Automatic seismic phase picking and consistent observation error assessment: application to the Italian seismicity, *Geophysical Journal International*, 165(1), 121–134, doi:10.1111/j.1365-246X.2005.02799.x.
- Du, W. X., C. H. Thurber, and D. Eberhart-Phillips (2004), Earthquake relocation using cross-correlation time delay estimates verified with the bispectrum method, *Bulletin of the Seismological Society of America*, 94(3), 856–866.
- Duncan, P. M. (2007), Microseismic monitoring - coming of age, *E&P*, September.
- Dyer, B. C., U. Schanz, F. Ladner, M. O. Haring, and T. Spillman (2008a), Microseismic imaging of a geothermal reservoir stimulation, *The Leading Edge*, (July 2008), 856–869.
- Dyer, B. C., U. Schanz, F. Ladner, M. O. Haring, and T. Spillman (2008b), Microseismic imaging of a geothermal reservoir stimulation, *The Leading Edge*, 27(7), 856, doi:10.1190/1.2954024.
- Eberhart-Phillips, D., and D. H. Oppenheimer (1984), Induced Seismicity in The Geysers Geothermal Area, California, *Journal of Geophysical Research*, 89(B2), 1191–1207, doi: 10.1029/JB089iB02p01191.
- Eisner, L., D. Abbott, W. B. Barker, J. Lakings, and M. P. Thornton (2008), Noise suppression for detection and location of microseismic events using a matched filter, in *SEG Las Vegas 2008 Annual Meeting*, 2, pp. 1431–1435, doi:10.1190/1.3059184.

- El Hariri, M., R. E. Abercrombie, C. A. Rowe, and A. F. do Nascimento (2010), The role of fluids in triggering earthquakes: observations from reservoir induced seismicity in Brazil, *Geophysical Journal International*, doi:10.1111/j.1365-246X.2010.04554.x.
- Ellis, A. J. (1999), *The Development of Geothermal Power*, Clerestory Press.
- Ferretti, G., M. Massa, and S. Solarino (2005), An Improved Method for the Recognition of Seismic Families: Application to the Garfagnana-Lunigiana Area, Italy, *Bulletin of the Seismological Society of America*, 95(5), 1903–1915, doi:10.1785/0120040078.
- Freiberger, W. (1963), An approximate method in signal detection, *Quart. J. App. Math.*, 20, 373–378.
- Fridleifsson, I. B., R. Bertani, E. Huenges, J. W. Lund, A. Ragnarsson, and L. Rybach (2008), The possible role and contribution of geothermal energy to the mitigation of climate change, in *IPCC Scoping Meeting on Renewable Energy Sources*, pp. 59–80.
- Galiana-Merino, J., J. Rosa-Herranz, and S. Parolai (2008), Seismic P Phase Picking Using a Kurtosis-Based Criterion in the Stationary Wavelet Domain, *Geoscience and Remote Sensing, IEEE Transactions on*, 46(11), 3815–3826.
- Gentili, S., and a. Michelini (2006), Automatic picking of P and S phases using a neural tree, *Journal of Seismology*, 10(1), 39–63, doi:10.1007/s10950-006-2296-6.
- Ghassemi, a., and G. Sureshkumar (2007), Changes in fracture aperture and fluid pressure due to thermal stress and silica dissolution/precipitation induced by heat extraction from subsurface rocks, *Geothermics*, 36(2), 115–140, doi:10.1016/j.geothermics.2006.10.001.
- Ghassemi, a., S. Tarasovs, and a. Cheng (2007), A 3-D study of the effects of thermomechanical loads on fracture slip in enhanced geothermal reservoirs, *International Journal of Rock Mechanics and Mining Sciences*, 44(8), 1132–1148, doi:10.1016/j.ijrmms.2007.07.016.
- Giardini, D. (2009), Geothermal quake risks must be faced, *Nature*, 462(December).

- Gibbons, S., M. Bottgersorensen, D. Harris, and F. Ringdal (2007), The detection and location of low magnitude earthquakes in northern Norway using multi-channel waveform correlation at regional distances, *Physics of The Earth and Planetary Interiors*, 160(3-4), 285–309, doi:10.1016/j.pepi.2006.11.008.
- Gibbons, S. J., and F. Ringdal (2006), The detection of low magnitude seismic events using array-based waveform correlation, *Geophysical Journal International*, 165(1), 149–166, doi:10.1111/j.1365-246X.2006.02865.x.
- Grant, M. A. (2007), Rotokawa reservoir and response to production, *MAGAK Quantitative decisions*.
- Gudmundsson, O. (2004), Seismic attenuation at Rabaul volcano, Papua New Guinea, *Journal of Volcanology and Geothermal Research*, 130(1-2), 77–92, doi:10.1016/S0377-0273(03)00282-8.
- Gurevich, B., R. J. Galvin, M. Brajanovski, T. M. Muller, and G. Lambert (2007), Fluid substitution, dispersion, and attenuation in fractured and porous reservoirs - insights from new rock physics models, *The Leading Edge*, 26(9), 1162, doi:10.1190/1.2780787.
- Gutenberg, B., and C. Richter (1944), Frequency of earthquakes in California, *Bulletin of the Seismological Society of America*, 34, 185–188.
- Hanson, C. E., D. A. Towers, and L. D. Meister (2006), Transit Noise and Vibration Impact Assessment, *Office of Planning and Environment, Federal Transit Administration, U.S. Department of Transportation*.
- Hardebeck, J. L. (2002), A New Method for Determining First-Motion Focal Mechanisms, *Bulletin of the Seismological Society of America*, 92(6), 2264–2276, doi:10.1785/0120010200.
- Harvey, C. C., B. R. White, J. V. Lawless, and M. G. Dunstall (2010), 2005 - 2010 New

- Zealand Country Update, in *Proceedings World Geothermal Congress 2010, Bali, Indonesia*, April, pp. 25–29.
- Hassouna, M. S., and A. A. Farag (2007), Multistencils Fast Marching Methods: A Highly Accurate Solution to the Eikonal Equation on Cartesian Domains, *IEEE Transactions on pattern analysis and machine intelligence*, 29(9), 1563–1574.
- Hayashi, K. (1999), Numerical models of HDR geothermal reservoirs - a review of current thinking and progress, *Geothermics*, 28(4-5), 507–518, doi:10.1016/S0375-6505(99)00026-7.
- Healy, J., R. Hamilton, and C. Raleigh (1970), Earthquakes induced by fluid injection and explosion, *Tectonophysics*, 9(2-3), 205–214, doi:10.1016/0040-1951(70)90017-X.
- Healy, J. H., W. W. Rubey, D. T. Griggs, and C. B. Raleigh (1968), The Denver Earthquakes, *Science*, 161, 1301–1310.
- Heise, W., T. G. Caldwell, H. M. Bibby, and S. C. Bannister (2008), Three-dimensional modelling of magnetotelluric data from the Rotokawa geothermal field, Taupo Volcanic Zone, New Zealand, *Geophysical Journal International*, 173(2), 740–750, doi:10.1111/j.1365-246X.2008.03737.x.
- Hochstein, M., and T. Hunt (1970), Seismic, gravity and magnetic studies, Broadlands geothermal field, New Zealand, *Geothermics*, 2(2), 333–346, doi:10.1016/0375-6505(70)90032-5.
- Hochstein, M., I. Smith, K. Regenauerlieb, and S. Ehara (1993), Geochemistry and heat transfer processes in Quaternary rhyolitic systems of The Taupo Volcanic Zone, New Zealand, *Tectonophysics*, 223(3-4), 213–235, doi:10.1016/0040-1951(93)90139-B.
- Hochstein, M. P. (1995), Crustal heat transfer in the Taupo Volcanic Zone (New Zealand) comparison with other volcanic arcs and explanatory heat source models, *Journal of Volcanology and Geothermal Research*, 68, 117–151.

- Holt, R. J. (2007), Response to Rotokawa I and II Production, *SymGeo Engineering, Inc.*
- Hoover, D., and J. A. Dietrich (1969), Seismic activity during the 1968 test pumping at the Rocky Mountain arsenal disposal well, *Geological Survey Circular*, 613.
- Houghton, B. F., C. J. N. Wilson, M. O. McWilliams, M. a. Lanphere, S. D. Weaver, R. M. Briggs, and M. S. Pringle (1995), Chronology and dynamics of a large silicic magmatic system: Central Taupo Volcanic Zone, New Zealand, *Geology*, 23(1), 13, doi:10.1130/0091-7613(1995)023<0013:CADOAL>2.3.CO;2.
- Hubbert, M. K., and W. W. Rubey (1959), Role of Fluid Pressure in Mechanics of Overthrust Faulting, *Geological Society of America Bulletin*, 70(2), 115, doi:10.1130/0016-7606(1959)70[115:ROFPIM]2.0.CO;2.
- Hunt, T., and D. Bowyer (2007), Reinjection and gravity changes at Rotokawa geothermal field, New Zealand, *Geothermics*, 36, 421–435.
- Hurst, A. W., H. M. Bibby, and R. R. Robinson (2002), Earthquake focal mechanisms in the central Taupo Volcanic Zone and their relation to faulting and deformation, *New Zealand Journal of Geology and Geophysics*, 45(4), 527–536.
- Husen, S. (2003), Probabilistic earthquake location in complex three-dimensional velocity models: Application to Switzerland, *Journal of Geophysical Research*, 108(B2), doi:10.1029/2002JB001778.
- Jaya, M. S., S. a. Shapiro, L. H. Kristinsdóttir, D. Bruhn, H. Milsch, and E. Spangenberg (2010), Temperature dependence of seismic properties in geothermal rocks at reservoir conditions, *Geothermics*, 39(1), 115–123, doi:10.1016/j.geothermics.2009.12.002.
- Jeffreys, H. (1973), *On travel times in seismology*, pp. 36–120, Gordon and Breach Sci. Publ., London.
- Jones, T., W. Murphy, and A. Nur (1980), Effects of temperature and saturation on the velocity and attenuation of seismic waves in rocks: applications to geothermal reservoir

- evaluation, in *PROCEEDINGS, Sixth Workshop on Geothermal Reservoir Engineering, December 16-18, 1980*, Stanford.
- Jupe, A., R. Jones, S. Wilson, and J. Cowles (2000), The role of microearthquake monitoring in hydrocarbon reservoir management, in *Proceedings of SPE Annual Technical Conference and Exhibition*, Society of Petroleum Engineers, doi:10.2118/63131-MS.
- Kingett Mitchell Limited (2005), Kawerau Geothermal Noise Assessment, *Kingett Mitchell Limited*.
- Kissling, E., W. L. Ellsworth, D. Eberhart-Phillips, and U. Kradolfer (1994), Initial reference models in local earthquake tomography, *Journal of Geophysical Research*, 99(B10), 19,635–19,646, doi:10.1029/93JB03138.
- Kissling, W., and G. Weir (2005), The spatial distribution of the geothermal fields in the Taupo Volcanic Zone, New Zealand, *Journal of Volcanology and Geothermal Research*, 145(1-2), 136–150, doi:10.1016/j.jvolgeores.2005.01.006.
- Kisslinger, C. (1976), A review of theories of mechanisms of induced seismicity, *Engineering Geology*, 10(2-4), 85–98.
- Kohl, T., and T. Megel (2007), Predictive modeling of reservoir response to hydraulic stimulations at the European EGS site Soultz-sous-Forets, *International Journal of Rock Mechanics and Mining Sciences*, 44, 1118–1131, doi:10.1016/j.ijrmms.2007.07.022|.
- Kristinsdóttir, L. H., O. G. Flóvenz, K. Árnason, D. Bruhn, H. Milsch, E. Spangenberg, and J. Kulenkampff (2010), Electrical conductivity and P-wave velocity in rock samples from high-temperature Icelandic geothermal fields, *Geothermics*, 39(1), 94–105, doi:10.1016/j.geothermics.2009.12.001.
- Krupp, R. E., and T. M. Seward (1987), The Rotokawa geothermal system, New Zealand; an active epithermal gold-depositing environment, *Economic Geology*, 82(5), 1109–1129, doi:10.2113/gsecongeo.82.5.1109.

Kumano, Y., H. Asanuma, A. Hotta, H. Niitsuma, U. Schanz, and M. Haring (2006), Reservoir structure delineation by microseismic multiplet analysis at Basel, Switzerland 2006, in *SEG Annual Meeting*, pp. SGP-TR-179.

Lay, T., and T. C. Wallace (1995), *Modern Global Seismology*, Academic Press, INC.

Le Calvez, J., M. Craven, R. Klem, J. Baihly, L. Bennett, and K. Brook (2007), Real-Time Microseismic Monitoring of Hydraulic Fracture Treatment: A Tool To Improve Completion and Reservoir Management, in *Proceedings of SPE Hydraulic Fracturing Technology Conference*, Society of Petroleum Engineers, doi:10.2118/106159-MS.

Lees, J. M. (1998), Multiplet Analysis at Coso Geothermal, *Bulletin of the Seismological Society of America*, 88(5), 1127–1143.

Leonard, M. (2000), Comparison of Manual and Automatic Onset Time Picking, *Bulletin of the Seismological Society of America*, 90(6), 1384–1390, doi:10.1785/0120000026.

Lippitsch, R., R. S. White, and H. Soosalu (2005), Precise hypocentre relocation of microearthquakes in a high-temperature geothermal field: the Torfajokull central volcano, Iceland, *Geophysics Journal International*, 160, 370–387.

Lomax, A., J. Virieux, P. Volant, and C. Berge (2000), Probabilistic earthquake location in 3D and layered models: introduction of a Metropolis-Gibbs method and comparison with linear locations, in *Advances in Seismic Event Location*, edited by C. H. Thurber and N. Rabinowitz, pp. 101–134, Kluwer, Amsterdam.

Lomax, A., A. Michelini, and A. Curtis (2009), Earthquake Location, in *Encyclopedia of Complexity and Systems Science*, edited by R. A. Meyers, pp. 2449–2473, Springer, New York.

Lund, B., and J. Townend (2007), Calculating horizontal stress orientations with full or partial knowledge of the tectonic stress tensor, *Geophysical Journal International*, 170(3), 1328–1335.

- Majer, E., R. Baria, M. Stark, S. Oates, J. Bommer, B. Smith, and H. Asanuma (2007), Induced seismicity associated with Enhanced Geothermal Systems, *Geothermics*, 36(3), 185–222, doi:10.1016/j.geothermics.2007.03.003.
- Majer, E., R. Baria, and M. Stark (2008), Protocol for induced seismicity associated with enhanced geothermal systems, *Report produced in Task D Annex I (9 April 2008), International Energy Agency-Geothermal Implementing Agreement (incorporating comments by: C. Bromley, W. Cumming, A. Jelacic and L. Rybach). Available at: <http://www.iea-gea.org/publications.asp>).*
- Maurer, H., and D. E. Boerner (1998), Optimized and robust experimental design: a non-linear application to EM sounding, *Geophysics Journal International*, 132, 458–468.
- Maurer, H., and N. Deichmann (1995), Microearthquake cluster detection based on waveform similarities, with an application to the western Swiss Alps, *Geophysical Journal International*, 123(2), 588–600, doi:10.1111/j.1365-246X.1995.tb06873.x.
- Mavko, G. M., and A. Nur (1979), Wave attenuation in partially saturated rocks, *Geophysics*, 44(2), 161–178.
- Mcloughlin, K., A. Campbell, and G. Ussher (2010), The Nga Awa Purua Geothermal Project, Rotokawa, New Zealand, in *World Geothermal Congress*, April, pp. 25–29.
- Menke, W. (1984), *Geophysical Data Analysis: Discrete Inverse Theory*, Academic Press, INC.
- Menke, W., and D. Schaff (2004), Absolute Earthquake Locations with Differential Data, *Bulletin of the Seismological Society of America*, 94, 2254–2264.
- Mighty River Power (2007), Rotokawa Geothermal Development: Proposed System Management Plan, viewed July 7 2011, <http://www.mightyriverpower.co.nz/ReportsDocuments/Default.aspx?type=Other+Reports+and+Documents>.

- Mighty River Power (2008a), Mighty River Power: Kawerau Geothermal power station, viewed July 7 2011, <http://www.mightyriverpower.co.nz/Generation/PowerStations/Geothermal/Kawerau/>.
- Mighty River Power (2008b), Rotokawa, viewed July 7 2011, <http://www.mightyriverpower.co.nz/Generation/PowerStations/Geothermal/Rotokawa/>.
- Mighty River Power (2010), Geothermal generation, viewed July 7 2011, <http://www.mightyriverpower.co.nz/Generation/PowerStations/Geothermal/Default.aspx>.
- Ministry Of Economic Development (2010), New Zealand Energy Data File, viewed July 7 2011, http://www.med.govt.nz/templates/StandardSummary_15169.aspx.
- Mitchell, T. J. (1974), An Algorithm for the Construction of "D-Optimal" Experimental Designs, *Technometrics*, 16(2), 203–210.
- Nairn, I., and S. Beanland (1989), Geological setting of the 1987 Edgecumbe earthquake, New Zealand, *New Zealand Journal of Geology & Geophysics*, 32, 1–13.
- Nakahara, H. (2004), Correlation distance of waveforms for closely located events. Implication of the heterogeneous structure around the source region of the 1995 Hyogo-Ken Nanbu, Japan, earthquake ($M_w = 6.9$), *Geophysical Journal International*, 157(3), 1255–1268, doi:10.1111/j.1365-246X.2004.02278.x.
- Nanjo, K. Z., T. Ishibe, H. Tsuruoka, D. Schorlemmer, Y. Ishigaki, and N. Hirata (2010), Analysis of the Completeness Magnitude and Seismic Network Coverage of Japan, *Bulletin of the Seismological Society of America*, 100(6), 3261–3268, doi:10.1785/0120100077.
- Nemoto, K., M. Hirokazu, H. Niitsuma, and N. Tsuchiya (2008), Mechanical and hydraulic coupling of injection-induced slip along pre-existing fractures, *Geothermics*, 37, 157–172.
- Niitsuma, H. (1999), Current status of seismic and borehole measurements for HDR/HWR development, *Geothermics*, 28(4-5), 475–490, doi:10.1016/S0375-6505(99)00024-3.

- Nolet, G. (2008), *A Breviary of Seismic Tomography: Imaging the interior of the Earth and Sun*, Cambridge University Press.
- O’Sullivan, M. (2001), State of the art of geothermal reservoir simulation, *Geothermics*, 30(4), 395–429, doi:10.1016/S0375-6505(01)00005-0.
- O’Sullivan, M. J., A. Yeh, and W. I. Mannington (2009), A history of numerical modelling of the Wairakei geothermal field, *Geothermics*, (38), 155–168.
- Paige, C. C., and M. A. Saunders (1982), LSQR: An algorithm for sparse linear equations and sparse least squares, *ACM Transactions on Mathematical Software*, 8(1), 43–71.
- Peng, Z., and P. Zhao (2009), Migration of early aftershocks following the 2004 Parkfield earthquake, *Nature Geoscience*, 2(12), 877–881, doi:10.1038/ngeo697.
- Percival, D., and A. Walden (1993), *Spectral analysis for physical applications: multitaper and conventional univariate techniques*, Cambridge University Press.
- Phillips, C., J. Parr, and E. Riskin (2007), *Signals, systems, and transforms*, Prentice Hall.
- Phillips, W. S. (2000), Precise Microearthquake Locations and Fluid Flow in the Geothermal Reservoir at Soultz-sous-Forets, France, *Bulletin of the Seismological Society of America*, 90(1), 212–228, doi:10.1785/0119990047.
- Phillips, W. S., L. S. House, and M. C. Fehler (1997), Detailed joint structure in a geothermal reservoir from studies of induced microearthquake clusters, *Journal of Geophysical Research*, 102(B6), 11,745–11,763, doi:10.1029/97JB00762.
- Prejean, S., W. Ellsworth, M. Zoback, and F. Waldhauser (2002), Fault structure and kinematics of the Long Valley Caldera region, California, revealed by high-accuracy earthquake hypocenters and focal mechanism stress inversions, *Journal of Geophysical Research B: Solid Earth*, 107(12), ESE 9–1 – 9–19.
- Pukelsheim, F. (2006), *Optimal design of experiments*, SIAM/Society for Industrial and Applied Mathematics.

- Rabinowitz, N., and D. M. Steinberg (1990), Optimal configuration of a seismographic network: A statistical approach, *Bulletin of the Seismological Society of America*, 80(1), 187–196.
- Rae, A. (2007), Rotokawa Geology and Geophysics, *GNS Science Consultancy Report*, (520W1111).
- Raleigh, C. B., J. H. Healy, and J. Bredehoeft (1976), An Experiment in Earthquake Control at Rangely, Colorado, *Science*, 191(26 March), 1230–1237.
- Ripperger, J., P. Kastli, D. Fah, and D. Giardini (2009), Ground motion and macroseismic intensities of a seismic event related to geothermal reservoir stimulation below the city of Basel observations and modelling, *Geophysical Journal International*, 179(3), 1757–1771, doi:10.1111/j.1365-246X.2009.04374.x.
- Risk, G. (2000), Electrical Resistivity Surveys Of The Rotokawa Geothermal Field, New Zealand, in *Proceedings of the 22nd New Zealand Geothermal Workshop*, pp. 121–126.
- Roff, A., W. Phillips, and D. Brown (1996), Joint structures determined by clustering microearthquakes using waveform amplitude ratios, *International Journal of Rock Mechanics and Mining Sciences & Geomechanics Abstracts*, 33(6), 627–639, doi:10.1016/0148-9062(95)00077-1.
- Rothe, J. (1969), Seismes artificiels, *Tectonophysics*, 9(2-3), 215–238, doi:10.1016/0040-1951(70)90018-1.
- Rowe, C. A. (2002), An Automatic, Adaptive Algorithm for Refining Phase Picks in Large Seismic Data Sets, *Bulletin of the Seismological Society of America*, 92(5), 1660–1674, doi:10.1785/0120010224.
- Rowland, J. V., and R. H. Sibson (2004), Structural controls on hydrothermal flow in a segmented rift system, Taupo Volcanic Zone, New Zealand, *Geofluids*, 4, 259–283.

- Rowland, J. V., C. J. Wilson, and D. M. Gravley (2010), Spatial and temporal variations in magma-assisted rifting, Taupo Volcanic Zone, New Zealand, *Journal of Volcanology and Geothermal Research*, 190(1-2), 89–108, doi:10.1016/j.jvolgeores.2009.05.004.
- Rubinstein, J. L., and W. L. Ellsworth (2010), Precise Estimation of Repeating Earthquake Moment: Example from Parkfield, California, *Bulletin of the Seismological Society of America*, 100(5A), 1952–1961, doi:10.1785/0120100007.
- Rutqvist, J., and C. Oldenburg (2007), Analysis of cause and mechanism for injection-induced seismicity at the Geysers Geothermal Field, California, *Lawrence Berkeley National Laboratory*.
- Sambridge, M., and Kennett (1990), Boundary value ray-tracing in a heterogeneous medium: a simple and versatile algorithm, *Geophysical Journal International*, 101, 157–168.
- Sarmiento, Z. (1986), Waste water reinjection at tongonan geothermal field: Results and implications, *Geothermics*, 15(3), 295–308, doi:10.1016/0375-6505(86)90106-9.
- Schaff, D., and F. Waldhauser (2008), Improving Magnitude Detection Thresholds Using Multi-Station, Multi-Event, and Multi-Phase Methods, *Air Force Research Laboratory, Air Force Materiel Command*, (AFRL-RV-HA-TR-2008-1121).
- Schaff, D. P. (2008), Semiempirical Statistics of Correlation-Detector Performance, *Bulletin of the Seismological Society of America*, 98(3), 1495–1507, doi:10.1785/0120060263.
- Schaff, D. P. (2009), Broad-scale applicability of correlation detectors to China seismicity, *Geophysical Research Letters*, 36(11), doi:10.1029/2009GL038179.
- Schaff, D. P., G. H. R. Bokelmann, W. L. Ellsworth, E. Zankerka, F. Waldhauser, and G. C. Beroza (2004), Optimizing correlation techniques for improved earthquake location, *Bulletin of the Seismological Society of America*, 94(2), 705–721.
- Schorlemmer, D., and J. Woessner (2008), Probability of Detecting an Earthquake, *Bulletin of the Seismological Society of America*, 98(5), 2103–2117, doi:10.1785/0120070105.

- Scotti, O., and F. Cornet (1994), In situ evidence for fluid-induced aseismic slip events along fault zones, in *International Journal of Rock Mechanics and Mining Sciences & Geomechanics Abstracts*, vol. 31, pp. 347–358, Elsevier.
- Segall, P. (1989), Earthquakes triggered by fluid extraction, *Geology*, 17(10), 942.
- Segall, P. (1997), Final Report: Induced Seismicity, *Office of Basic Energy Sciences, U.S. Department of Energy*, (DE-FG03-90).
- Shapiro, S. A., E. Huenges, and G. Borm (1997), Estimating the crust permeability from fluid-injection-induced seismic emission at the KTB site, *Geophysical Journal International*, 131(2), F15–F18, doi:10.1111/j.1365-246X.1997.tb01215.x.
- Shapiro, S. a., E. Rothert, V. Rath, and J. Rindschwentner (2002), Characterization of fluid transport properties of reservoirs using induced microseismicity, *Geophysics*, 67(1), 212, doi:10.1190/1.1451597.
- Shapiro, S. a., C. Dinske, C. Langenbruch, and F. Wenzel (2010), Seismogenic index and magnitude probability of earthquakes induced during reservoir fluid stimulations, *The Leading Edge*, 29(3), 304, doi:10.1190/1.3353727.
- Shelly, D. R., and J. L. Hardebeck (2010), Precise tremor source locations and amplitude variations along the lower-crustal central San Andreas Fault, *Geophysical Research Letters*, 37(14), 1–5, doi:10.1029/2010GL043672.
- Shelly, D. R., G. C. Beroza, and S. Ide (2007), Non-volcanic tremor and low-frequency earthquake swarms., *Nature*, 446(7133), 305–7, doi:10.1038/nature05666.
- Shelly, D. R., W. L. Ellsworth, T. Ryberg, C. Haberland, G. S. Fuis, J. Murphy, R. M. Nadeau, and R. Bürgmann (2009), Precise location of San Andreas Fault tremors near Cholame, California using seismometer clusters: Slip on the deep extension of the fault?, *Geophysical Research Letters*, 36(1), doi:10.1029/2008GL036367.

- Shemeta, J., and P. Anderson (2010), It's a matter of size: Magnitude and moment estimates for microseismic data, *The Leading Edge*, (March).
- Sherburn, S., C. J. Bryan, A. W. Hurst, J. H. Latter, and B. J. Scott (1999), Seismicity of Ruapehu volcano, New Zealand, 1971-1996: A review, *Journal of Volcanology and Geothermal Research*, 88(4), 255–278.
- Simiyu, S. (1999), Induced micro-seismicity during well discharge: OW-719, Olkaria, Kenya rift, *Geothermics*, 28(6), 785–802, doi:10.1016/S0375-6505(99)00043-7.
- Song, F., H. Kuleli, and M. Toksoz (2010), An improved method for hydrofracture-induced microseismic event detection and phase picking, *Geophysics*, 75(6), 6–11.
- Stankova, J., S. L. Bilek, C. A. Rowe, and R. C. Aster (2008), Characteristics of the October 2005 Microearthquake Swarm and Reactivation of Similar Event Seismic Swarms over Decadal Time Periods near Socorro, New Mexico, *Bulletin of the Seismological Society of America*, 98(1), 93–105, doi:10.1785/0120070108.
- Statz-Boyer, P., C. Thurber, J. Pesicek, and S. Prejean (2009), High precision relocation of earthquakes at Iliamna Volcano, Alaska, *Journal of Volcanology and Geothermal Research*, 184(3-4), 323–332, doi:10.1016/j.jvolgeores.2009.04.016.
- Stein, S., and M. Wysession (2003), *An Introduction to Seismology, Earthquakes, and Earth Structure*, Blackwell Publishing.
- Steinberg, D. M., N. Rabinowitz, Y. Shimshoni, and D. Mizrachi (1995), Configuring a Seismographic Network for Optimal Monitoring of Fault Lines and Multiple Sources, *Bulletin of the Seismological Society of America*, 85(6), 1847–1857.
- Stern, T. (1987), Asymmetric back-arc spreading, heat flux and structure associated with the Central Volcanic Region of New Zealand, *Earth and planetary science letters*, 85(1-3), 265–276.

- Stern, T. A. (1985), A Back-arc basin formed within continental lithosphere: The Central Volcanic Region of New Zealand, *Tectonophysics*, 112, 385–409.
- Stopa, J., and P. Wojnarowski (2006), Analytical model of cold water front movement in a geothermal reservoir, *Geothermics*, 35(1), 59–69, doi:10.1016/j.geothermics.2005.11.002.
- Stratford, W. R., and T. A. Stern (2006), Crust and upper mantle structure of a continental backarc: Central North Island, New Zealand, *Geophysical Journal International*, 166(1), 469–484.
- Studt, F. E., and G. E. K. Thompson (1969), Geothermal heat flow in the North Island of New Zealand, *New Zealand Journal of Geology & Geophysics*, 12(4), 673–683.
- Tarantola, A., and B. Valette (1982), Inverse problems = quest for information, *Journal of Geophysics*, 50, 150–170.
- The Department Of Planning And Community Development, Victoria, Australia, A. (2008), Victorian Desalination Project Environmental Effects Statement. Volume 5: Environmental effects of power supply. Chapter 10: Noise and vibration.
- Thomson, D. (1982), Spectrum estimation and harmonic analysis, *Proceedings of the IEEE*, 70(9), 1055–1096.
- Tsang, C. (1980), Some aspects of the response of geothermal reservoirs to brine reinjection with application to the Cerro Prieto field, *Geothermics*, 9(1-2), 213–220, doi:10.1016/0375-6505(80)90035-8.
- Uhrhammer, R. A. (1980), Analysis of small seismographic station networks, *Bulletin of the Seismological Society of America*, 70(4), 1369–1379.
- Ungemach, M., M. Antics, and P. Lalosklos (2008), Sustainable geothermal reservoir management practice, *GPC Instrumentation Process*.

- van Trees, H. (1968), *Detection, Estimation and Modulation Theory*, John Wiley & Sons.
- Waldhauser, F. (2001), hypoDD – A Program to Compute Double-Difference Hypocenter Locations, *U.S. Geol. Surv. Open-File Rep., 01-113, Menlo Park, California*.
- Waldhauser, F., and W. L. Ellsworth (2000), A Double-Difference earthquake location algorithm: method and application to the Northern Hayward Fault, California, *Bulletin of the Seismological Society of America*, 90(December), 1353–1368.
- Wallace, L. M., J. Beavan, R. McCaffrey, and D. Darby (2004), Subduction zone coupling and tectonic block rotations in the North Island, New Zealand, *Journal of Geophysical Research B: Solid Earth*, 109(12), 1–21.
- Walsh, D. (2008), Directional Statistics, Bayesian Methods of Earthquake Focal Mechanism Estimation, and their Application to New Zealand Seismicity Data, MSc Thesis, *Victoria University of Wellington*.
- Walsh, D., R. Arnold, and J. Townend (2009), A Bayesian approach to determining and parametrizing earthquake focal mechanisms, *Geophysical Journal International*, 176(1), 235–255.
- Wiemer, S., and M. Wyss (2000), Minimum Magnitude of Completeness in Earthquake Catalogs: Examples from Alaska, the Western United States, and Japan, *Bulletin of the Seismological Society of America*, 90(4), 859–869, doi:10.1785/0120020035.
- Wilson, C. J. N., B. F. Houghton, M. O. McWilliams, M. A. Lanphere, S. D. Weaver, and R. M. Briggs (1995), Volcanic and structural evolution of Taupo Volcanic Zone, New Zealand: a review, *Journal of Volcanology and Geothermal Research*, 68(1-3), 1–28.
- Winkler, K. W., and A. Nur (1982), Seismic attenuation: Effects of pore fluids and frictional sliding, *Geophysics*, 47(1).
- Wood, C. P., R. L. Brathwaite, and M. D. Rosenberg (2001), Basement structure, lithology

and permeability at Kawerau and Ohaaki geothermal fields, New Zealand, *Geothermics*, 30(4), 461–481.

Yang, W. (2004), An adaptive, automatic phase-picking and epicenter locating program based on waveform cross-correlation and its application for processing, Msc thesis, New Mexico Institute of Mining and Technology.

Yukutake, Y., H. Ito, R. Honda, M. Harada, T. Tanada, and A. Yoshida (2011), Fluid-induced swarm earthquake sequence revealed by precisely determined hypocenters and focal mechanisms in the 2009 activity at Hakone volcano, Japan, *Journal of Geophysical Research*, 116(B4), 1–13, doi:10.1029/2010JB008036.

Zhang, H., C. Thurber, and P. Bedrosian (2009), Joint inversion for Vp, Vs, and Vp/Vs at SAFOD, Parkfield, California, *Geochemistry Geophysics Geosystems*, 10(11), 1–17, doi:10.1029/2009GC002709.

Zoback, M. D., and H.-P. Harjes (1997), Injection-induced earthquakes and crustal stress at 9 km depth at the KTB deep drilling site, Germany, *Journal of Geophysical Research*, 102(B8), 18,477–18,491, doi:10.1029/96JB02814.