

# Directional Statistics, Bayesian Methods of Earthquake Focal Mechanism Estimation, and their Application to New Zealand Seismicity Data

A project  
submitted in fulfillment  
of the requirements for the Degree  
of  
Master of Science in Statistics  
by  
David Walsh

School of Mathematics, Statistics and Computing Sciences  
Victoria University of Wellington  
P.O. Box 600  
Wellington  
New Zealand

Victoria University of Wellington  
2008



# Abstract

A focal mechanism is a geometrical representation of fault slip during an earthquake. Reliable earthquake focal mechanism solutions are used to assess the tectonic characteristics of a region, and are required as inputs to the problem of estimating tectonic stress. We develop a new probabilistic (Bayesian) method for estimating the distribution of focal mechanism parameters based on seismic wave polarity data. Our approach has the advantage of enabling us to incorporate observational errors, particularly those arising from imperfectly known earthquake locations, allowing exploration of the entire parameter space, and leads to natural point estimates of focal mechanism parameters. We investigate the use of generalised Matrix Fisher distributions for parameterising focal mechanism uncertainties by minimising the Kullback-Leibler divergence.

We present here the results of our method in two situations. We first consider the case in which the seismic velocity of the region of interest (described by a velocity model) is presumed to be precisely known, with application to seismic data from the Raukumara Peninsula, New Zealand. We then consider the case in which the velocity model is imperfectly known, with application to data from the Kawerau region, New Zealand.

We find that our estimated focal mechanism solutions for the most part are consistent with all available polarity data, and correspond closely to solutions obtained using established methods. Further, the generalised Matrix Fisher distributions we examine provide a good fit to our Bayesian posterior PDF of the focal mechanism parameters, enabling the posterior PDF to be succinctly summarised by reporting the estimated parameters of the fitted distribution.





# Acknowledgments

Firstly, a huge thank you to my supervisors, Richard Arnold and John Townend, for all their help over the course of this project. I was, it seemed to me, constantly full of questions, and they never failed to use their knowledge and expertise to guide me in the right direction. Cheers guys.

Thanks to those who provided the data used in this project; Martin Reyners, Daniel Clarke, and Kevin Fenaughty through GeoNet. Thank you to the Earthquake Commission, who funded this project (grant UNI/536). Thank you also to my examiners, Euan Smith and Peter Jupp, for their helpful comments and suggestions.

Thanks to all my fellow graduate students who I've had the pleasure of studying with over the last two years. Thank you to the staff in the Statistics and Operations Research group, particularly Megan Clark, who has always been supportive, and helped me pay the rent by employing me as a Teaching Assistant during my course of study. Thanks also to Ray Brownrigg, our Statistical Computing Engineer, who was a great help in what was a computationally intensive project.

Lastly, thanks to my friends and family for all their support throughout.



# Contents

Abstract . . . . .	iii
Acknowledgments . . . . .	v
<b>1 Introduction</b>	<b>1</b>
1.1 Motivation and Objectives . . . . .	1
1.2 Focal Mechanisms . . . . .	2
1.2.1 P-wave First Motions . . . . .	3
1.2.2 S-wave Information . . . . .	6
1.2.3 Focal Mechanism Ambiguities . . . . .	7
1.2.4 Solution Quality and Sources of Error . . . . .	7
1.3 Previous Methods . . . . .	8
1.3.1 Optimisation Methods . . . . .	9
1.3.2 Probabilistic Methods . . . . .	11
1.4 Contribution of this Thesis . . . . .	12
1.5 Outline . . . . .	14
<b>2 Directional Statistics</b>	<b>15</b>
2.1 Circular Data . . . . .	15
2.1.1 von Mises Distribution . . . . .	15
2.2 Spherical Data . . . . .	16
2.2.1 von Mises-Fisher Distribution . . . . .	16
2.2.2 Bingham Distribution . . . . .	18
2.3 Orientation Data . . . . .	19
2.3.1 Stiefel Manifolds . . . . .	20
2.3.2 Matrix Fisher distribution . . . . .	20
2.3.3 8-Mode Matrix Fisher distribution . . . . .	31
2.3.4 Goodness of Fit Testing . . . . .	36
<b>3 Bayesian Methods</b>	<b>39</b>
3.1 Bayes' Theorem . . . . .	39
3.2 Examples . . . . .	40
3.2.1 Lighthouse Problem . . . . .	40
3.2.2 Earthquake Hypocentre Location . . . . .	42
3.3 Application to Focal Mechanisms . . . . .	45
3.3.1 Velocity Model Known . . . . .	45
3.3.2 Velocity Model Unknown . . . . .	49
3.3.3 Probability Density of P- and T-axes . . . . .	51
<b>4 Computing</b>	<b>55</b>
4.1 R . . . . .	55
4.2 NonLinLoc . . . . .	55
4.2.1 Running NonLinLoc . . . . .	55
4.2.2 Programs . . . . .	56

4.2.3	Obtaining Take-off Parameters from NonLinLoc Output . . . . .	58
4.3	Velest . . . . .	60
4.4	Grid Computing . . . . .	61
<b>5</b>	<b>Applications</b>	<b>63</b>
5.1	Velocity model known — Raukumara Peninsula . . . . .	63
5.1.1	Tectonic Setting . . . . .	63
5.1.2	Velocity Model . . . . .	63
5.1.3	Data . . . . .	64
5.1.4	Posterior PDF Particulars . . . . .	64
5.1.5	Results . . . . .	65
5.2	Velocity model unknown — Kawerau . . . . .	92
5.2.1	Tectonic Setting . . . . .	92
5.2.2	Velocity Models . . . . .	92
5.2.3	Data . . . . .	94
5.2.4	Posterior PDF Particulars . . . . .	94
5.2.5	Results . . . . .	95
<b>6</b>	<b>Conclusions</b>	<b>131</b>
	<b>Glossary</b>	<b>133</b>
	<b>References</b>	<b>139</b>
<b>A</b>	<b>Useful Definitions and Results</b>	<b>143</b>
A.1	Rotation . . . . .	143
A.1.1	Euler Angles . . . . .	143
A.1.2	Passive and Active Rotation . . . . .	144
A.1.3	Rotation Angles and Angular Difference . . . . .	144
A.2	Spherical Coordinates . . . . .	145
A.3	Lambert-Schmidt Projection . . . . .	145
A.4	Averaging Angles . . . . .	147
A.4.1	Circular Data . . . . .	148
A.4.2	Spherical Data . . . . .	149
A.5	Change of Variable Technique . . . . .	149
A.5.1	Lighthouse Example . . . . .	150
A.6	Hypergeometric Functions of a Matrix Argument . . . . .	150
A.7	Numerical Integration . . . . .	151
A.7.1	Trapezium Rule . . . . .	151
A.7.2	Monte Carlo Integration . . . . .	152
A.8	Tensors . . . . .	153
A.8.1	Moment Tensor . . . . .	154
A.9	Dirac Delta Function . . . . .	155
<b>B</b>	<b>Raukumara earthquake hypocentres</b>	<b>157</b>

# List of Tables

1.1	Summary of published focal mechanism estimation methods . . . . .	13
2.1	Table of $\sigma_\xi$ (in degrees) for various values of $\mathbf{D}$ . . . . .	24
3.1	Lighthouse example PDFs . . . . .	41
4.1	GeoNet qualities and arrival time errors $\sigma_{t_i}$ . . . . .	56
4.2	Sources of each element of Equation 3.21 and 3.25 . . . . .	60
5.1	Summary of Raukumara Results . . . . .	66
5.2	Mean P- and S-wave velocity models for Kawerau . . . . .	95
5.3	Summary of Kawerau Results . . . . .	98
B.1	Raukumara earthquake hypocentres . . . . .	157



# List of Figures

1.1	Diagram of a focal mechanism . . . . .	3
1.2	P-wave first motions . . . . .	4
1.3	Stereonet showing first motions . . . . .	5
1.4	Beachball diagram . . . . .	5
1.5	Focal mechanisms for various types of fault . . . . .	6
1.6	Diagram illustrating the problem. . . . .	9
2.1	von Mises distribution . . . . .	18
2.2	von Mises-Fisher distribution . . . . .	20
2.3	Simulation from Bingham distribution . . . . .	22
2.4	Matrix Fisher distribution . . . . .	25
2.5	Plot of $\sigma_\xi$ (log scale) versus $k$ . . . . .	28
2.6	Simulation from Matrix Fisher distribution . . . . .	29
2.7	PDF plots comparing $\mathbf{F}$ and $\hat{\mathbf{F}}$ for the Matrix Fisher distribution . . . . .	36
2.8	8-Mode Matrix Fisher distribution . . . . .	38
2.9	Simulation from 8-Mode Matrix Fisher distribution . . . . .	39
2.10	PDF plots comparing $\mathbf{F}$ and $\hat{\mathbf{F}}$ for the 8-Mode Matrix Fisher distribution . .	41
3.1	Lighthouse problem . . . . .	44
3.2	Lighthouse location - posterior PDF . . . . .	46
3.3	T-axis diagram . . . . .	58
4.1	3D linear interpolation . . . . .	63
5.1	The Raukumara Peninsula . . . . .	68
5.2	$V_p$ velocity model for the Raukumara Peninsula . . . . .	69
5.3	$V_s$ velocity model for the Raukumara Peninsula . . . . .	70
5.4	Diagram showing our event selection criteria . . . . .	71
5.5	Stereonet and PT contour plot for Raukumara events . . . . .	73-89
5.6	Marginal PDF plots - empirical and fitted for Raukumara events . . . . .	74-90
5.23	Map of Raukumara showing MAP focal mechanism estimates . . . . .	93
5.24	Map of Raukumara showing Reyners & McGinty (1999) focal mechanisms . . . . .	94
5.25	Histogram of angular differences . . . . .	95
5.26	Plot of angular difference versus distance between epicentre . . . . .	96
5.27	Map of the Kawerau area . . . . .	97
5.28	Plot of Kawerau velocity models . . . . .	98
5.29	Stereonet and PT contour plot for Kawerau events . . . . .	102-128
5.30	Marginal PDF plots - empirical and fitted for Kawerau events . . . . .	103-129
5.57	Map of Kawerau showing MAP focal mechanism estimates . . . . .	132
A.1	Spherical Coordinates . . . . .	148
A.2	Stereographic projection . . . . .	150
A.3	Beachball diagram . . . . .	151
A.4	Trapezium rule . . . . .	155





# Chapter 1

## Introduction

### 1.1 Motivation and Objectives

The focal mechanism of an earthquake describes the geometry of the fault on which the earthquake occurred using three angular parameters: the strike, dip and rake (see Section 1.2). Reliable earthquake focal mechanism solutions can be used for assessing the tectonic characteristics of a region (see e.g. Reyners & McGinty 1999), and are required as inputs to the problem of estimating tectonic stress (see e.g. Arnold & Townend 2007), changes in which have been hypothesised to occur following large earthquakes and volcanic eruptions (for a recent review see Townend 2006).

Existing methods of focal mechanism estimation (see Section 1.3) make use of P-wave first motion polarities (see Section 1.2.1) and/or S-wave information (see Section 1.2.2). These methods can be characterised as either optimisation or probabilistic techniques.

The existing optimisation methods are, for the most part, unable to accommodate all of the relevant sources of uncertainty in the underlying seismological data, although some studies have applied a forward method to this problem (Hardebeck & Shearer 2002). Some of this uncertainty stems from imprecise knowledge of the Earth's seismic velocity structure. The probabilistic methods in the literature (see e.g. De Natale et al. 1991, Zollo & Bernard 1991) take into account P-wave polarity uncertainties, but fail to account for the uncertainty in earthquake hypocentre location, and do not parameterise the resulting probability distribution.

Here we propose a new, probabilistic method of focal mechanism estimation, based on Bayes' rule; a simple probabilistic theorem that can be used to assess the degree to which certain data support certain hypotheses (Sivia 1996). A Bayesian approach allows a complete probabilistic treatment of the problem, and leads naturally to robust point estimates of focal mechanism parameters based on seismological data, taking into account the inherent uncertainties. This is an extension of initial work undertaken by Bouley (2005). Having derived a method for calculating the posterior distribution of the focal mechanism parameters, we investigate the use of directional distributions for representing focal mechanism uncertainties. We attempt to parameterise this uncertainty using generalised Matrix Fisher distributions, fitted by minimising the Kullback-Leibler divergence.

The objectives of this project are, therefore, to:

1. Develop a probabilistic (Bayesian) model of the constraints imposed on focal mechanism parameters by first-motion data given imperfect hypocentre parameters and an error-free velocity model;

2. Generalise the model developed in Objective 1 to situations in which neither the hypocentre nor the velocity model are known perfectly;
3. Investigate the suitability of idealised error distributions (particularly generalised Matrix Fisher distributions) for representing focal mechanism uncertainties.

## 1.2 Focal Mechanisms

A *focal mechanism*<sup>1</sup> describes the geometry of a fault during an earthquake. Although not always the case, it is generally assumed that the fault is a planar surface (known as the *fault plane*) across which respective sides move (Stein & Wyssession 2003). The alternative term fault plane solution is sometimes used (Fowler 1990).

Focal mechanisms can be described by the three angles *strike*, *dip* and *rake*. Strike and dip describe the orientation of the fault plane, subject to an ambiguity which we describe below, while the rake describes the sense of relative motion during the earthquake. Here we represent a focal mechanism by the following:

- The strike direction is the direction of a horizontal line in the fault plane, while the strike angle  $\xi$  is the angle measured clockwise from north to the strike direction ( $0 \leq \xi \leq 360^\circ$ ).
- The dip angle  $\delta$  is the angle that the fault slants downwards from the horizontal to the right of the strike direction ( $0 \leq \delta \leq 90^\circ$ ).
- The rake angle  $\lambda$  (also known as the slip angle) specifies the direction of motion of the upper side of the fault (the hanging wall block) with respect to the lower side of the fault (the foot wall block), measured in the fault plane anti-clockwise from the direction of the strike ( $0 \leq \lambda \leq 360^\circ$ ).

(Stein & Wyssession 2003, Aki & Richards 2002), where  $\Theta = (\xi, \delta, \lambda)$ . This description can be visualised in Figure 1.1.

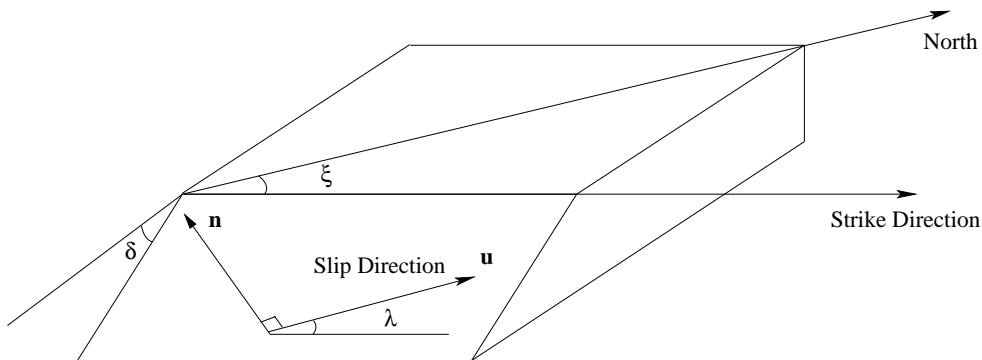


Figure 1.1: Visual representation of fault slip. Figure based on Aki & Richards (2002).

Alternatively the fault plane can equivalently be described by a unit *normal vector*  $\hat{n}$ , and unit *slip vector*  $\hat{u}$ , the direction of motion of the hanging wall block with respect to the foot wall block (Stein & Wyssession 2003). These two vectors are orthogonal, and together with the *null vector*  $\hat{a} = \hat{n} \times \hat{u}$  constitute the orthogonal rotation matrix  $R(\Theta) = [\hat{u} \ \hat{a} \ \hat{n}]$ , which

<sup>1</sup>Italics denote terms explained further in the Glossary, p 133

entirely describes the focal mechanism. These two equivalent definitions are related as described in Appendix A.1.1.

A focal mechanism is the most simple representation of fault slip. Another method of representing a seismic source is by way of the *moment tensor* (see Appendix A.8.1). The advantage of the moment tensor is that it encompasses both fault geometry and earthquake magnitude, and it can describe more complex seismic sources than simply fault slip, such as opening modes. Moment tensor inversion — estimating the moment tensor using the seismograms of stations that recorded an event — is only possible for reasonably large events,  $> 3.5$  local magnitude ( $M_L$ ) (J. Ristau, personal communication, April 4, 2008). These methods are beyond the scope of this project since we assume a double couple source, which can be sufficiently described by the focal mechanism. Focal mechanisms can also be computed for smaller earthquakes than a moment tensor can be computed for, down to around  $2.3 M_L$  in the New Zealand settings considered here.

The following sections outline the basics of focal mechanism estimation.

### 1.2.1 P-wave First Motions

*P-waves* are the seismic waves that travel fastest from the earthquake source (Stein & Wysession 2003). A *first motion* indicates the direction of motion, or polarity, of the first P-wave arrival at a *seismometer* (Stein & Wysession 2003). By observing P-wave polarities at a number of different seismic stations, the focal mechanism can be determined (Fowler 1990, Cox & Hart 1986).

P-waves are an example of a longitudinal wave; the direction of oscillation of the wave is in the same direction as the direction of wave propagation (Stein & Wysession 2003, Fowler 1990, Aki & Richards 2002). If the movement of the material near the fault is toward a certain station (away from the earthquake source), the first motion of the P-wave that arrives at that station will be upwards (termed *compressional*). If the material moves away from the station (toward the earthquake source), the first motion of the P-wave arriving at that station will be downwards (*dilatational*) (Stein & Wysession 2003).

These first motions define four quadrants surrounding the source. As shown in Figure 1.2, the division of these quadrants occurs at the fault plane and a plane perpendicular to the fault plane, known as the *auxiliary plane*, which together are called the *nodal planes* (Cox & Hart 1986, Fowler 1990, Stein & Wysession 2003). As explained in Section 1.2.3, one can generally not distinguish between these two planes.

The *focal sphere* is an imaginary sphere of negligible radius centered at and surrounding the earthquake source (Stein & Wysession 2003, Cox & Hart 1986, Fowler 1990). Locations of compressions and dilatations leaving the earthquake source can be represented as points on the focal sphere. Because a sphere cannot be represented on paper without distortion (Kagan 2005), a 2-dimensional stereographic projection of the lower hemisphere of the focal sphere, known as a *stereonet*, is commonly used (Stein & Wysession 2003, Fowler 1990).

To determine a focal mechanism, the polarities of P-wave first motions at seismometers are first recorded. Each station corresponds to a point on the focal sphere where the P-wave left the earthquake source en route to the seismometer. This point is usually defined in spherical polar coordinates (see Appendix A.2) by its *azimuth*,  $\phi$ , the angle measured clockwise from north to the point, and its *take-off angle*,  $\theta$ , the angle measured from the downward vertical to the point (Stein & Wysession 2003, Fowler 1990, Udias 1999). Alternatively, the take-off

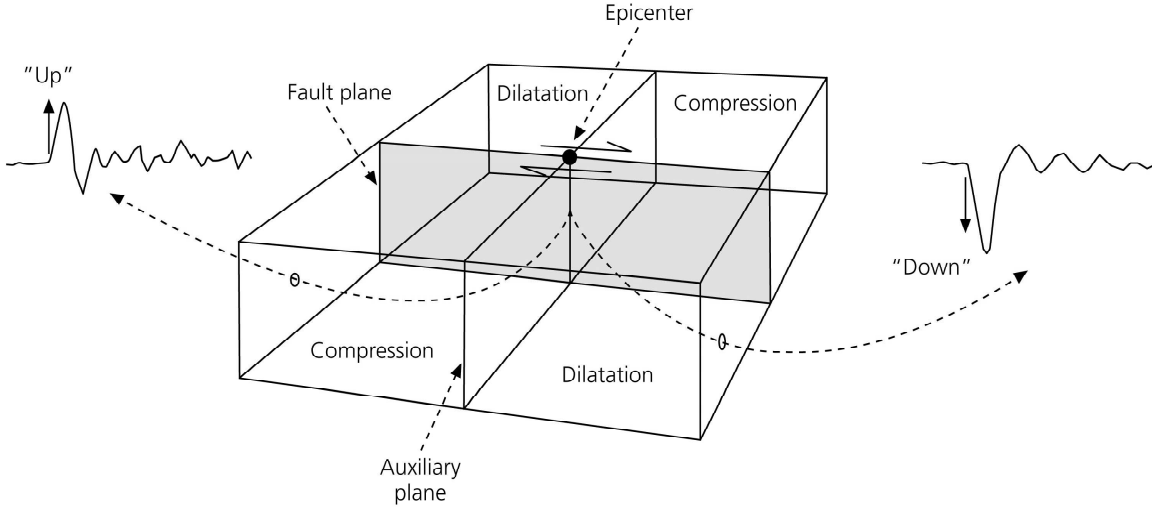


Figure 1.2: First motions of P-waves provide information about the nodal planes (Stein & Wysession 2003).

vector, the unit vector from the origin to the point  $(\phi, \theta)$  on the focal sphere, given by

$$\hat{\mathbf{p}} = (\sin \theta \cos \phi, \sin \theta \sin \phi, \cos \theta)$$

can be used. As P-wave polarity at a station depends on  $\Theta$  and location on the focal sphere  $(\phi, \theta)$ , this relationship can be formalised as follows

$$F^P = 2(\hat{\mathbf{p}} \cdot \hat{\mathbf{n}})(\hat{\mathbf{p}} \cdot \hat{\mathbf{u}}) = \cos \lambda \sin \delta \sin^2 \theta \sin 2(\phi - \xi) - \cos \lambda \cos \delta \sin 2\theta \cos(\phi - \xi) + \quad (1.1)$$

$$+ \sin \lambda \sin 2\delta (\cos^2 \theta - \sin^2 \theta \sin^2(\phi - \xi)) + \sin \lambda \cos 2\delta \sin 2\theta \sin(\phi - \xi) \quad (1.2)$$

(Aki & Richards 2002) where  $F$  denotes the radiation pattern, or normalised amplitude, of the P-wave leaving the focal sphere at point  $(\phi, \theta)$ .

To plot points  $(\phi, \theta)$  onto a stereonet, we use a technique known as the Lambert-Schmidt projection (see Section A.3). Once all points are plotted on the stereonet, the focal sphere is partitioned by two great circles — the nodal planes corresponding to the focal mechanism solution  $\Theta$  — creating four quadrants (see Figure 1.3). Appendix A.3 describes how  $R(\Theta)$  defines the nodal planes. In principle (i.e. given perfectly observed data) each quadrant contains only compressional or only dilatational first motions (Aki & Richards 2002). The quadrants where the first motions are compressional are coloured dark, while the quadrants where the first motions are dilatational are coloured white (Stein & Wysession 2003). This results in a “beachball” appearance (see Figure 1.4); these stereographic projections are sometimes referred to as beachball diagrams.

Beachball diagrams can indicate certain types of faulting. Strike-slip faulting occurs when the two sides of the fault move horizontally past one-another ( $\delta = \pi/2$  and  $\lambda = 0$  or  $\pi$ ) (Aki & Richards 2002). Dip-slip faulting occurs when the movement is vertical ( $\delta = \pi/2$  and  $\lambda = \pi/2$  or  $3\pi/2$ ). There are two types of dip-slip faults; reverse faults, occurring when the hanging wall moves upwards relative to the foot wall ( $\lambda \in (0, \pi)$ ), and normal faults, occurring when the hanging wall moves downwards relative to the foot wall ( $\lambda \in (\pi, 2\pi)$ ) (Aki & Richards 2002, Stein & Wysession 2003). Figure 1.5 shows typical beachball diagrams corresponding to these fault types.

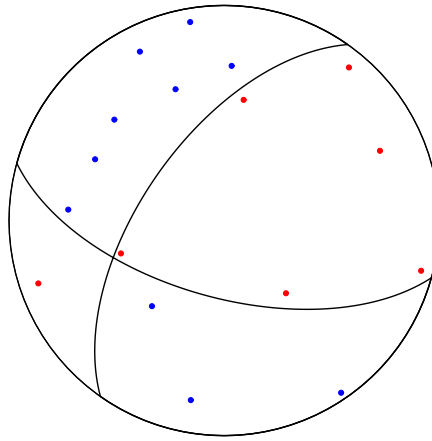


Figure 1.3: Stereonet showing how the nodal planes partition the focal sphere into four quadrants, each quadrant containing only compressional (blue) or dilatational (red) first motions.

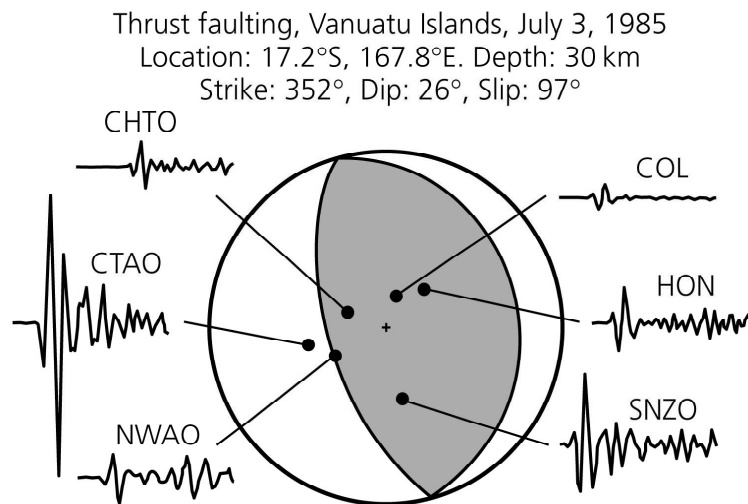


Figure 1.4: Example of a focal mechanism, with P wave first motions shown, represented on a stereonet/beachball diagram (Stein & Wyssession 2003).

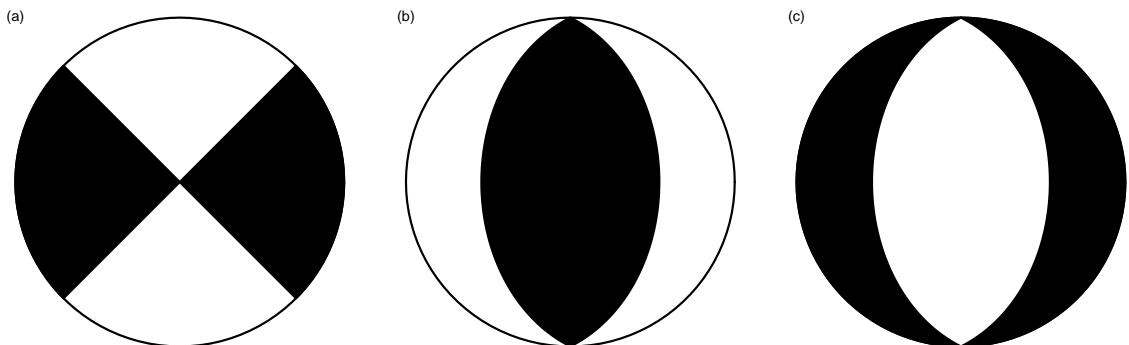


Figure 1.5: Focal mechanisms for various types of fault. (a) shows strike-slip faulting, (b) shows reverse faulting, and (c) shows normal faulting. Figure based on Stein & Wyssession (2003).

The dilatational quadrant of the focal sphere is bisected by the pressure (or P-) axis, which is parallel to  $\hat{\mathbf{u}} - \hat{\mathbf{n}}$ , and the compressional quadrant is bisected by the tensional (or T-) axis, parallel to  $\hat{\mathbf{u}} + \hat{\mathbf{n}}$  (Arnold & Townend 2007, Aki & Richards 2002). We denote the unit vectors in the direction of the P- and T-axes  $\hat{\mathbf{v}}^P$  and  $\hat{\mathbf{v}}^T$  respectively.

### 1.2.2 S-wave Information

*S-waves*, or secondary waves, are so-called because they are the slower of the two types of seismic wave to arrive at seismometers (Udias 1999). These waves are transverse, meaning that the direction of oscillation of the wave is perpendicular to the direction of the wave's propagation. S-waves are commonly used in two ways to supplement P-wave information in determining focal mechanisms; S-wave polarisation and  $S/P$  amplitude ratios.

Determination of S-wave polarisation involves analysing the oscillation geometry of the S-wave. A plane defined by two lines — the vertical and the path connecting an earthquake hypocentre to a seismometer — is identified. The S-wave is split into two perpendicular components, SV and SH, based on its oscillation in relation to this plane. The SV displacement is vertical, in the plane, while the SH displacement is horizontal, normal to the plane (Stein & Wysession 2003). As with P-waves, the radiation patterns  $F$  of SV and SH are directly related to fault geometry  $\Theta$  and position on the focal sphere  $(\phi, \theta)$  (Aki & Richards 2002).

$$F^{SV} = \sin \lambda \cos 2\delta \cos 2\theta \sin(\phi - \xi) - \cos \lambda \cos \delta \cos 2\theta \cos(\phi - \xi) + \frac{1}{2} \cos \lambda \sin \delta \sin 2\theta \sin 2(\phi - \xi) - \frac{1}{2} \sin \lambda \sin 2\delta \sin 2\theta (1 + \sin^2(\phi - \xi)) \quad (1.3)$$

$$F^{SH} = \cos \lambda \cos \delta \cos \theta \sin(\phi - \xi) + \cos \lambda \sin \delta \sin \theta \cos 2(\phi - \xi) + \sin \lambda \cos 2\delta \cos \theta \cos(\phi - \xi) - \frac{1}{2} \sin \lambda \sin 2\delta \sin \theta \sin 2(\phi - \xi) \quad (1.4)$$

Thus S-wave polarisation can be used to help constrain a focal mechanism, particularly by comparing theoretical and observed values of  $F^{SV}$  and  $F^{SH}$ . However, S-wave information can be sparse, since three-component seismometers are required to identify SV and SH components (Zollo & Bernard 1991), and since the S-wave arrives after the P-wave, the S-wave polarisation may be hard to measure.

Because P-wave amplitudes are small near nodal planes, whereas S-wave amplitudes are large,  $S/P$  amplitude ratios can be useful in constraining focal mechanism solutions. Large  $S/P$  amplitude ratios indicate a point near a nodal plane and vice versa (Hardebeck & Shearer 2003). Observed  $S/P$  amplitude ratios can be compared to theoretical ratios and the solution with the minimum misfit selected. Alternatively,  $SV/P$  and/or  $SH/P$  ratios can be used (see, e.g. Rau et al. 1996, Snoke 2003). Log amplitude ratios are often used when comparing observed and theoretical values (see Section 1.3). To see why, consider that if  $A \geq B$ , then  $A/B \in (1, \infty]$ , however if  $A \leq B$  then  $A/B \in (0, 1)$ . This lack of symmetry makes comparing ratios difficult. Taking the log ratio restores the symmetry, i.e.  $\log(A/B) = -\log(B/A)$ .

The advantage of using amplitude data is the increased number of observations over P-wave data alone. One disadvantage is that, along with the focal mechanism, event magnitude, attenuation (the loss of energy, and thus amplitude, from the S-wave as it propagates), geometric spreading (energy dispersion due to the expansion of the wavefront as it propagates), and site effects can all influence the observed amplitude (Hardebeck & Shearer 2003, Rau et al. 1996). Using amplitude ratios eliminates geometric spreading and magnitude effects, however. Another disadvantage is that S-wave arrivals and amplitudes can be difficult to pick

due to noise caused by seismic wave scatter (Hardebeck & Shearer 2003, Nakamura 2002).

Due to the existence of these uncertainties, and the fact that S-wave data are not as commonly available as P-wave data, we restrict our analysis to that of P-wave first motions.

### 1.2.3 Focal Mechanism Ambiguities

There are two ambiguities associated with focal mechanism solutions that must be considered. The first is the fact that a fault plane can be represented equally well by its fault normal in either direction, thus  $[\hat{\mathbf{u}} \ \hat{\mathbf{a}} \ \hat{\mathbf{n}}]$  is equivalent to  $[-\hat{\mathbf{u}} \ +\hat{\mathbf{a}} \ -\hat{\mathbf{n}}] = [\hat{\mathbf{u}} \ \hat{\mathbf{a}} \ \hat{\mathbf{n}}]\mathbf{A}_2$ , where

$$\mathbf{A}_2 = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix}.$$

Adopting the convention that the dip lies in the range  $0 \leq \delta \leq 90^\circ$  dictates which of the two equivalent directions will be referred to as  $\hat{\mathbf{n}}$  for a given focal mechanism. Thus the fault normal is an example of axial data, the treatment of which is discussed further in Chapter 2.

The second ambiguity arises because first motion and other remote observations alone are not sufficient to infer which of the two nodal planes is the fault plane and which is the auxiliary plane (Udias 1999, Stein & Wyssession 2003, Aki & Richards 2002, Fowler 1990). This is because the first motions, for example, produced by slip on either of the two planes would be the same (Stein & Wyssession 2003). This is equivalent to being unable to distinguish between the slip vector  $\hat{\mathbf{u}}$  and the fault normal  $\hat{\mathbf{n}}$ , so  $[\hat{\mathbf{u}} \ \hat{\mathbf{a}} \ \hat{\mathbf{n}}]$  is equivalent to  $[+\hat{\mathbf{n}} \ -\hat{\mathbf{a}} \ +\hat{\mathbf{u}}] = [\hat{\mathbf{u}} \ \hat{\mathbf{a}} \ \hat{\mathbf{n}}]\mathbf{C}_2$ , where

$$\mathbf{C}_2 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & -1 & 0 \\ 1 & 0 & 0 \end{bmatrix}.$$

Lastly, if both ambiguities are combined we have a fourth equivalent representation of the focal mechanism. Hence  $[\hat{\mathbf{u}} \ \hat{\mathbf{a}} \ \hat{\mathbf{n}}]$  is equivalent to  $[-\hat{\mathbf{n}} \ -\hat{\mathbf{a}} \ -\hat{\mathbf{u}}] = [\hat{\mathbf{u}} \ \hat{\mathbf{a}} \ \hat{\mathbf{n}}]\mathbf{C}_2\mathbf{A}_2$  (Arnold & Townend 2007). Additional information can sometimes clarify the second ambiguity, such as when the earthquake breaks the surface of the earth or subsequent seismicity (aftershocks) delineates a planar structure and identifies the fault plane, but this information is usually not available, especially for small earthquakes (Fowler 1990, Stein & Wyssession 2003).

### 1.2.4 Solution Quality and Sources of Error

The quality of a focal mechanism solution  $\Theta$  is affected by a number of factors:

1. The number of seismometer stations with polarity readings. If there are a small number of such stations, the solution is more sensitive to individual station polarities (Rabinowitz & Hofstetter 1992, Zollo & Bernard 1991).
2. Station distribution. If observations are distributed unevenly over azimuth  $\phi$  and take-off angle  $\theta$ , then the focal sphere will not be well covered. This can lead to large uncertainties in the focal mechanism solution  $\Theta$  (Udias 1999, Rabinowitz & Hofstetter 1992, Zollo & Bernard 1991).
3. P-wave polarity readings. Measured polarities are uncertain due to  $\pi'_p$ , the probability of an incorrectly wired seismometer, and  $\sigma_a$ , the standard deviation of the amplitude  $A_i$  of the first motion at station  $i$ . If the ratio of  $A_i$  to  $\sigma_a$  is low, incorrect polarity readings can occur, due to instrumental or human effects (Hardebeck & Shearer 2002).

4. *Hypocentre* location  $\mathbf{x}$ . Focal mechanism estimation depends on the take-off parameters  $\phi$  and  $\theta$  corresponding to the paths to each seismometer, which in turn depend on  $\mathbf{x}$ . The hypocentre location is uncertain due to *seismic noise*, which is here parameterised by  $\sigma_{t_i}$ , the P-wave arrival time error at station  $i$ . Seismic noise is created by fluctuations in temperature and pressure, storms, ocean waves, solid earth tides, and human activity (Aki & Richards 2002, Stein & Wysession 2003). The hypocentre location is also uncertain if the seismic velocity structure — as represented by a *velocity model* — of the region of interest is unknown.

Given a large enough dataset, one can select events with a desired minimum number of polarity readings and good focal sphere coverage, thus mitigating the effects of Factors 1 and 2 above. Factors 3 and 4, however, are sources of error inherent in any earthquake observations. Figure 1.6 shows the effect of uncertainty in hypocentre location — the locations of P-waves leaving the source are uncertain, creating a cloud of corresponding positions on the focal sphere. In this project we address this by constructing a Bayesian probability distribution for the focal mechanism parameters that directly accounts for these errors.

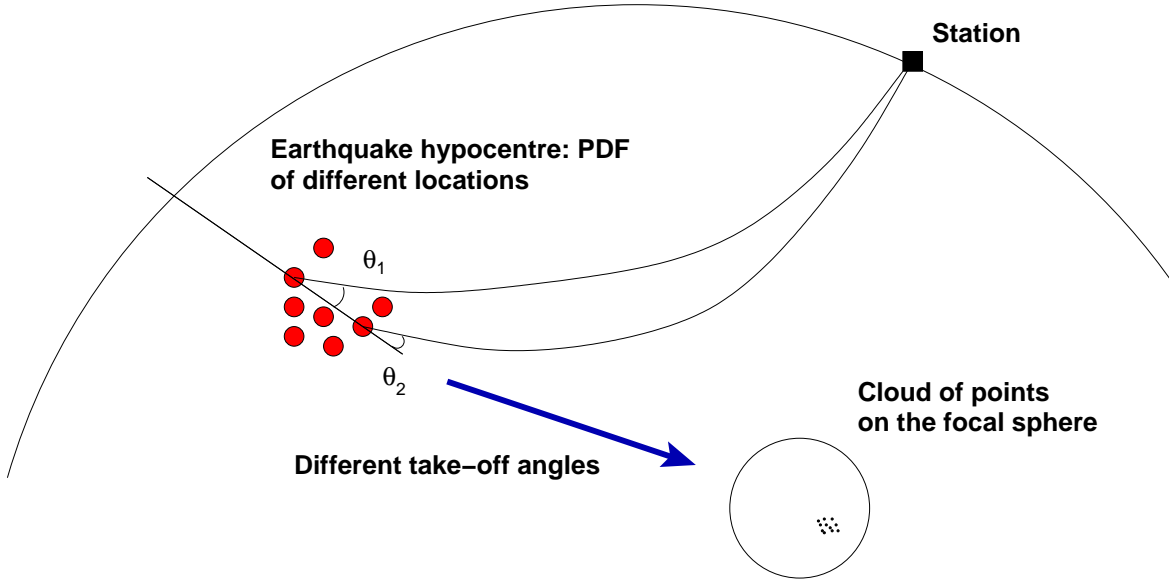


Figure 1.6: Diagram illustrating the problem. Uncertainties in P-wave arrival times at stations lead to different possible earthquake locations, each with its own take-off angle. Hence, when considering take-off angles — an important part of estimating the focal mechanism of an earthquake — we must take into account all possible hypocentre locations.

### 1.3 Previous Methods

In this section we discuss published methods of focal mechanism estimation, which can be divided into two categories. The first category contains optimisation methods, in which the number of discrepancies between the observed and theoretical polarities (given a certain focal mechanism solution  $\Theta$ ) is minimised in some way. The second category contains those methods in which a probability distribution is used for the data, given the parameters, and a maximum likelihood or Bayesian approach is applied to estimate  $\Theta$ . We briefly review each method, the data used, and how uncertainties in the solutions are dealt with. Table 1.1 summarises the methods.



### 1.3.1 Optimisation Methods

#### Reasenber and Oppenheimer (1985) — FPFIT

FPFIT is a Fortran routine developed by Reasenber & Oppenheimer (1985) that finds the focal mechanism  $\Theta$  that best fits P-wave first motion polarities. This involves a two stage (coarse then fine) grid search to find the value of  $\Theta$  that minimises  $F$ , a weighted sum of polarity discrepancies. There are two weights involved:

- The square root of the normalised theoretical amplitude  $A_i^*$  at the  $i^{th}$  station, which down weights observations near the nodal planes.
- A weight based on assessed quality codes for polarities, which are supplied by the seismic analyst.

Uncertainties are parameterised by a one-sided 90% confidence interval for  $F$ , and the values of  $\Theta = (\xi, \delta, \lambda)$  that result in a value for  $F$  inside this confidence interval. Also returned is a value indicating the degree to which the observations lie close to the nodal planes of the solution.

#### Rabinowitz and Hofstetter (1992)

Rabinowitz & Hofstetter (1992) used P-wave polarities and amplitudes to minimise  $F$  in a similar manner to that used by Reasenber & Oppenheimer (1985). The algorithm imposes the constraint that the theoretical P-wave amplitude  $A_i^*$  at station  $i$  resulting from the solution, should be close to the observed amplitude  $A_i$ . This method uses an algorithm called the Flexible Tolerance Method (FTM).

The FTM does not provide information on the uncertainty of the solution. The algorithm is robust to changes in polarity, and the addition of amplitude information provides more reliable solutions than those found with FPFIT (Rabinowitz & Hofstetter 1992).

#### Snoke (2003) — FOCMEC

FOCMEC is another Fortran routine, published by Snoke (2003), that uses P- and S-wave polarities and/or amplitude ratios to determine focal mechanisms. FOCMEC reports the set of solutions satisfying a specified number of polarity and/or amplitude ratio misfits.

When using amplitude ratios, FOCMEC selects a best solution based on the minimum root mean square (RMS) error (the square root of the sum of differences squared between the calculated and observed log amplitude ratios).

#### Rau et al. (1996)

Rau et al. (1996) used P-wave polarities and  $SH/P$  amplitude ratios to calculate the focal mechanisms of small to moderate events ( $2.7 \leq M_L \leq 5.7$ ) recorded by the Taiwan Seismic Network. The authors used an early version of FOCMEC (Snoke 2003) and compared 1D and 3D velocity models using only P-wave polarity data, and then using both P-wave polarities and  $SH/P$  amplitude ratios. They found the quality of the solutions to be improved using the 3D model, and that incorporating amplitude ratios allowed them to select the solution that was most consistent with the P-wave polarities, based on the minimum RMS error (Rau et al. 1996).

## Hardebeck and Shearer (2002) — HASH

Hardebeck & Shearer (2002) introduced a method (the HASH algorithm) that accounts for uncertainties in velocity model, event location and P-wave polarities. Using P-wave polarity data from Northridge, California, take-off parameters were computed for a suite of 50 randomly selected combinations of hypocentre locations (varying randomly in depth) and 1D velocity models.

The set of focal mechanism solutions from each of the 50 runs that had less than 10% misfit polarities were identified. This set of acceptable solutions was averaged, and solutions that were greater than  $30^\circ$  from the average were iteratively removed. When all solutions lay within this tolerance, a final average was taken, termed ‘the most preferred solution’ (Hardebeck & Shearer 2002). The quality of the solution obtained in this manner is based on how closely the set of acceptable solutions is clustered around the preferred solution.

The authors elected to deal non-parametrically with uncertainties in solutions, by reporting the set of acceptable solutions that were within  $30^\circ$  of the most preferred solution.

Hardebeck & Shearer (2002) tested their method using the fact that events occurring in spatial clusters should have similar focal mechanisms, as they are likely to originate from the same source. It was found that the method produced similar solutions for tightly spaced events, and performed better in this regard than the FPFIT algorithm discussed above.

## Hardebeck and Shearer (2003)

Hardebeck & Shearer (2003) investigated whether using  $S/P$  wave amplitude ratios could improve their focal mechanism solutions for the Northridge events, using two methods. Their first method was to select the set of acceptable solutions from P-wave data using HASH (Hardebeck & Shearer 2002), and from this choose the mechanism that minimised the misfit of the  $\log(S/P)$  observations. For clusters of similar events, they found that the inclusion of  $S/P$  ratio data reduced the similarity of the solutions; in other words the  $S/P$  ratios actually downgraded the solution quality.

The authors observed a lot of noise in the  $S/P$  amplitude data. Thus their preferred approach is said to account for the uncertainty in  $S/P$  observations, although how this is achieved is omitted in their description. In this case the  $S/P$  data helped constrain solutions that were of poor quality when estimated with P-wave data alone. The authors conclude that  $S/P$  amplitude ratio data can be useful when constraining poor quality solutions if  $S/P$  amplitude ratio noise is accounted for in the estimation procedure (Hardebeck & Shearer 2003).

## Nakamura (2002)

Nakamura (2002) developed a method of using both P and S-wave polarities. The motivation behind this was to use the extra information provided by S-waves, while eliminating the S-wave amplitude noise that scattering creates. S-wave polarity is measured in the North–South and East–West directions of a three component seismometer.

The method compares the observed and theoretical polarities: theoretical polarities for S-waves are taken to be  $\text{sgn}(F^{SH})$  and  $\text{sgn}(F^{SV})$ , where  $F^{SH}$  and  $F^{SV}$  are as defined in Equations 1.4 and 1.3. A grid search over  $\Theta$  is performed by calculating a weighted sum  $Q$  of the difference between observed and theoretical polarities at each value of  $\Theta$ .

Uncertainties in the solution are reported by finding the set of solutions that satisfy  $Q_{min} \leq Q \leq Q_{min} + \epsilon$ , where  $\epsilon$  is said to account for the possibility of incorrect polarity readings. The set of solutions is better constrained when S-wave polarities are included, but the author notes that S-wave polarities can be difficult to measure, and errors can occur in cases in which seismic anisotropy (when wave speed varies with direction) causes the splitting of the wave (Nakamura 2002).

### **Reyners et al. (1997)**

The method used by Reyners et al. (1997) and Reyners & McGinty (1999) places primary importance on P-wave polarities and employs theoretical and observed log amplitude ratios to improve the solution. The mechanism with the lowest number of polarity inconsistencies and the lowest RMS error between the observed and theoretical amplitude ratios is termed the “best” solution. This method uses the computer programs AMPRAT and MECHTOOL by Robinson & Webb (1996).

## **1.3.2 Probabilistic Methods**

### **Brillinger et al. (1980)**

Brillinger et al. (1980) adopted a maximum likelihood estimation (MLE) approach. They assumed that P-wave first motion polarities observed at station  $i$  are Bernoulli random variables

$$Y_i = \begin{cases} +1 & \text{if the first motion is recorded as positive (a compression)} \\ -1 & \text{if the first motion is recorded as negative (a dilatation)} \end{cases}$$

with Bernoulli probability distribution

$$P(Y_i = y_i) = \pi_i^{\frac{1}{2}(1+y_i)} (1 - \pi_i)^{\frac{1}{2}(1-y_i)} \quad (y_i = -1, 1)$$

where

$$\pi_i = P(Y_i = 1) = \pi'_p + (1 - 2\pi'_p) \Phi \left( \frac{A_i^*}{\sigma_a} \right).$$

Here  $\pi'_p$  denotes the probability of an incorrectly wired seismometer,  $A_i^*$  denotes the theoretical amplitude at station  $i$  and  $\sigma_a$  denotes the amplitude noise. This formulation is discussed in more depth in Section 3.3.  $A_i^*$  is a function of  $\Theta$  (see Equation 1.2), which is how the focal mechanism parameters enter the formulation. The values of  $\Theta$  that maximise the log likelihood are found numerically.

The uncertainties in the estimates of  $\Theta$  are quantified by 95% confidence intervals obtained from the asymptotic properties of MLEs — the estimators of  $\Theta$  are asymptotically normal with covariance matrix equal to the inverse of the Fisher information matrix (Brillinger et al. 1980).

### **Zollo and Bernard (1991)**

Zollo & Bernard (1991) used P-wave polarities and S-wave polarization angles in a Bayesian approach (see Chapter 3) to estimate  $\Theta$ . The posterior PDF of  $\Theta$  is given by

$$P(\Theta|\mathbf{d}) \propto P(\mathbf{d}^S|\Theta)P(\Theta)\mu_0(\Theta)$$

where the prior  $P(\Theta)$  is taken to be the PDF of Brillinger et al. (1980), i.e.  $P(\Theta) = P(\mathbf{d}^P|\Theta)$  and  $P(\mathbf{d}^S|\Theta)$  is a conditional probability function for the observed S-wave polarisations.

Also,  $\mu_0(\Theta) = \text{const.} \sin \delta$  is the non-informative PDF representing the state of total ignorance on  $\Theta$ . In this state, the normal vector  $\hat{\mathbf{n}}$  has equal probability in all directions on the focal sphere. The area of any infinitesimal patch on the unit sphere created by the differential angles  $d\xi, d\delta$  is equal to  $\sin \delta d\delta d\xi$ . The  $\sin \delta$  term scales the area of the patch as  $\hat{\mathbf{n}}$  tends towards either pole, where the spacing between  $\xi$  becomes very small.

Rather than giving point estimates and uncertainties for  $\Theta$ , computing the posterior PDF allows the authors to locate regions of high probability in the parameter space of  $\Theta$ , which are displayed graphically. Including S-wave data better constrains the areas of high probability, however one must carefully analyse the stability of the S-wave data before proceeding (Zollo & Bernard 1991).

### De Natale et al. (1991)

De Natale et al. (1991) used P-wave polarities, S-wave polarisations and  $S/P$  amplitude ratios in another Bayesian approach. The posterior PDF of  $\Theta$  is

$$P(\Theta|\mathbf{d}) \propto P(\mathbf{d}_1|\Theta)P(\mathbf{d}_2|\Theta)P(\mathbf{d}_3|\Theta)P(\Theta)\mu_0(\Theta)$$

where

- $P(\mathbf{d}_1|\Theta)$  is Brillinger et al. (1980)'s PDF;
- $P(\mathbf{d}_2|\Theta)$  is a PDF for the observed S-wave polarisations similar to the one used by Zollo & Bernard (1991);
- $P(\mathbf{d}_3|\Theta)$  is a PDF for  $S/P$  amplitude ratios;
- the prior  $P(\Theta)$  modifies the posterior PDF in favour of parts of the parameter space that are likely to produce observable S-waves with stable polarisations, or measurable amplitude ratios, at the stations where the data is available;
- $\mu_0(\Theta)$  is the non-informative PDF as in Zollo & Bernard (1991).

Again, regions of high probability in the parameter space can be displayed.

## 1.4 Contribution of this Thesis

While probabilistic methods of focal mechanism estimation have a clear advantage over optimisation methods, in that the entire parameter space can be explored, the methods in the literature do not account for uncertainty in hypocentre location or velocity model. Additionally, the resulting probability distributions (see e.g. Zollo & Bernard 1991, De Natale et al. 1991) are not parameterised in any way.

In this project we develop a new probabilistic Bayesian method of focal mechanism estimation that directly accounts for the sources of error mentioned in Section 1.2.4. We present the resulting probability distributions visually so one can assess areas of high and low probability in the parameter space. Point estimates of  $\Theta$  result naturally from the distributions.

Little research has been carried out into the nature of focal mechanism error distributions. We explore directional distributions, in particular generalised Matrix Fisher distributions, that can be used to parameterise the resulting probability distribution of  $\Theta$ . We investigate whether these directional distributions provide an appropriate fit to the distribution of  $\Theta$ , in the hope of succinctly summarising the distribution of  $\Theta$  by reporting the estimated parameter(s) of the fitted distribution.

Table 1.1: Summary of focal mechanism estimation methods

Optimisation Methods				
Author	Data	Data errors	Method	Uncertainties in $\Theta$
Reasenbergs & Oppenheimer (1985)	P-wave polarities	Allows for uncertainty in polarity via quality codes	Minimises $F$ , a weighted sum of polarity discrepancies	Reports $\Theta$ that fit inside a 90% CI for $F$
Rabinowitz & Hofstetter (1992)	P-wave polarities and amplitudes	Allows for polarity errors	Minimises $F$ while trying to match theoretical and observed amplitudes	None
Snoke (2003)	P- and S-wave polarities and amplitude ratios	Allows for polarity and/or amplitude ratio errors	Minimises RMS error of log amplitude ratios	Reports set of acceptable solutions
Rau et al. (1996)	P-wave polarities and $SH/P$ amplitude ratios	Allows for polarity and/or amplitude ratio errors	FOCMEC	None
Hardebeck & Shearer (2002)	P-wave polarities	Allows for uncertainty in $\mathbf{x}$ , velocity model, and polarity	Find set of acceptable solutions averaged over suite of 50 random combinations of hypocentre locations and 1D velocity models	Reports set of acceptable solutions
Nakamura (2002)	P and S-wave polarities	Allows for polarity errors	Minimises $Q$ , a weighted sum of polarity discrepancies	Reports $\Theta$ that fit within $Q_{min} \leq Q \leq Q_{min} + \epsilon$
Reyners et al. (1997) and Reyners & McGinty (1999)	P-wave polarities and $P/S$ amplitude ratios	Allows for polarity and/or amplitude ratio errors	Calculates best value of $\Theta$ - that consistent with the lowest number of polarity errors and lowest amplitude ratio error	Reports set of acceptable solutions
Probabilistic Methods				
Author	Data	Data errors	Method	Uncertainties in $\Theta$
Brillinger et al. (1980)	P-wave polarities	Accounts for uncertainty in polarity, and amplitude noise	MLE approach, assuming P-wave polarities are Bernoulli random variables	95% CIs for $\Theta$ using properties of MLEs
Zollo & Bernard (1991)	P-wave polarities and S-wave polarisations	Accounts for uncertainty in polarity, and amplitude noise	Bayesian approach	A posterior PDF over $\Theta$
De Natale et al. (1991)	P-wave polarities, S-wave polarisations and amplitude ratios	Accounts for uncertainty in polarity, and amplitude noise	Bayesian approach	A posterior PDF over $\Theta$

## 1.5 Outline

The structure of this thesis is as follows. Directional statistical methods are introduced in Chapter 2, which explains why and how directional methods are used, and introduces various directional probability distributions, from simple to complex cases. Chapter 3 introduces Bayes' rule, the probabilistic theorem behind the models used in this thesis, with some examples of its use. We then describe the formulation of our Bayesian probability distribution for the focal mechanism parameters in two specific cases; when the seismic velocity model is precisely known, and when it is imperfectly known. Chapter 4 describes the various computational techniques used in this project, and how they are used to compute the distributions introduced in the Chapter 2. In Chapter 5 we apply the probabilistic methods outlined in Chapter 3, and estimation techniques described in Chapter 2, to seismic data from New Zealand. We first consider the velocity model known case, with application to data from the Raukumara Peninsula. This is followed by the velocity model unknown case, applied to data from the Kawerau region. Chapter 6 contains some concluding statements. Appendix A contains some useful techniques, definitions and mathematical results that are used throughout this project, while Appendix B contains hypocentre summary information for the earthquakes examined in Chapter 5.

# Chapter 2

## Directional Statistics

In this project, we are interested in various *directional data* (directions, or unit vectors, in space) and *axial data* (directions in which the positive and negative directions are equivalent). These observations are directions or axes in space, in which the sample space is often a circle or sphere, and hence special directional methods are used to analyse them (Mardia & Jupp 2000). To see why such methods are needed, consider the case of averaging angles, as described in Appendix A.4. This chapter examines such directional methods, and introduces probability distributions used to describe the data.

### 2.1 Circular Data

We start with the most simple directional case — that of circular data; directional data in 2D space. An example of a circular datum is a compass bearing. We can represent such data as points on the unit circle, or equivalently, unit vectors  $\mathbf{x}$  in the plane. Alternatively,  $\mathbf{x}$  can be represented by an angle  $\phi$  rotated from a reference direction on the unit circle (Mardia & Jupp 2000), in which case  $\mathbf{x} = (\cos \phi, \sin \phi)$ .

#### 2.1.1 von Mises Distribution

The *von Mises distribution* is a probability distribution used to describe the distribution of directions on circles. It is analogous to the Normal distribution used on lines (Mardia & Jupp 2000).

The von Mises distribution  $M(\phi_0, \kappa)$  has PDF

$$f(\phi|\phi_0, \kappa) = \frac{1}{2\pi I_0(\kappa)} \exp [\kappa \cos(\phi - \phi_0)] \quad (2.1)$$

where  $\phi_0$  is the mean direction,  $\kappa$  is the concentration parameter and  $I_0(\kappa)$  is the modified Bessel function of the first kind

$$I_0(\kappa) = \frac{1}{2\pi} \int_0^{2\pi} \exp(\kappa \cos \phi) d\phi \quad (2.2)$$

(Mardia & Jupp 2000). The larger  $\kappa$  is, the more concentrated the distribution is around the mean direction. When  $\kappa = 0$  the distribution is uniform (Mardia & Jupp 2000).

We can simulate from a von Mises distribution using the technique described by Best & Fisher (1979), which uses a wrapped Cauchy distribution (the Cauchy distribution on the line ‘wrapped’ onto the unit circle) as an envelope for an acceptance-rejection sampling

method. Firstly, we set

$$a = 1 + \sqrt{1 + 4\kappa^2}, \quad b = \frac{a - \sqrt{2a}}{2\kappa}, \quad r = \frac{1 + b^2}{2b}$$

then repeat the following steps until the necessary sample size is obtained.

1. Draw a sample  $U_1, U_2, U_3$  from a Uniform(0,1) distribution;
2. Put  $z = \cos(\pi U_1)$ ,  $f = (1 + rz)/(r + z)$ ,  $c = \kappa(r - f)$ ;
3. If  $c(2 - c) - U_2 > 0$  go to step 5, else go to step 4;
4. If  $\ln(c/U_2) + 1 - c \geq 0$  go to step 5, else return to step 1;
5.  $\phi = \phi_0 + \text{sgn}(U_3 - 0.5) \cos^{-1}(f)$ .

Figure 2.1 shows a sample of points, drawn in the above way using the computer software R, marked on the unit circle. The bell-curved shape of the density illustrates the analogy of the von Mises distribution to the Normal distribution.

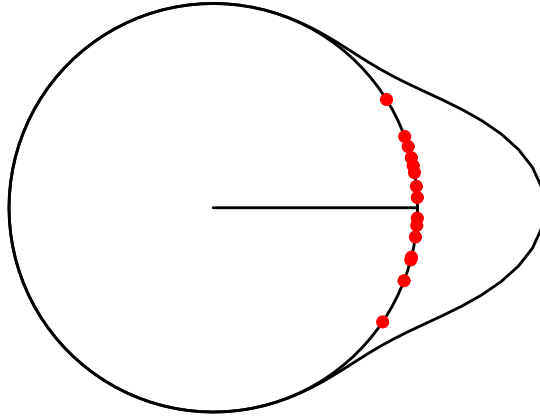


Figure 2.1: A random sample of size 15 drawn from a von Mises distribution with  $\phi_0 = 0$  and  $\kappa = 10$ . The density is also shown.

## 2.2 Spherical Data

Spherical data arise when the observations are directions in 3D space. Again, the data can be directional or axial. In the directional case, these data can be represented as points on the unit sphere, or as unit vectors  $\mathbf{x}$ . In the axial case, these data can be represented as antipodal points on the unit sphere (Mardia & Jupp 2000). The unit vector  $\mathbf{x}$  can alternatively be represented in spherical polar coordinates as  $\mathbf{x} = (\sin \theta \cos \phi, \sin \theta \sin \phi, \cos \theta)$  (see Appendix A.2).

### 2.2.1 von Mises-Fisher Distribution

A *von Mises-Fisher distribution*  $F(\mu, \kappa)$  describes the distribution of a random unit vector  $\mathbf{x}$  on the surface of a sphere (Mardia & Jupp 2000). As in the circular case, the parameter



$\mu$  is the mean direction, while  $\kappa$  is the concentration parameter (Mardia & Jupp 2000). The probability density is given by:

$$f(\mathbf{x}|\mu, \kappa) = \frac{\kappa}{\sinh \kappa} \exp(\kappa \mu^T \mathbf{x})$$

Alternatively, if  $\mathbf{x}$  and  $\mu$  are written in spherical coordinates

$$\begin{aligned}\mathbf{x} &= (\sin \theta \cos \phi, \sin \theta \sin \phi, \cos \theta)^T \\ \mu &= (\sin \theta_0 \cos \phi_0, \sin \theta_0 \sin \phi_0, \cos \theta_0)^T\end{aligned}$$

then  $d\mathbf{x} = \frac{1}{4\pi} \sin \theta d\theta d\phi$  and the probability density can be rewritten using the change of variable technique (see Appendix A.5):

$$\begin{aligned}f(\phi, \theta|\mu, \kappa) &= f(\mathbf{x}|\mu, \kappa) \left| \frac{d\mathbf{x}}{d\phi d\theta} \right| \\ &= \frac{\kappa}{4\pi \sinh \kappa} \exp(\kappa [\cos \theta \cos \theta_0 + \sin \theta \sin \theta_0 \cos(\phi - \phi_0)]) \sin \theta \\ &\quad (0 \leq \phi \leq 2\pi, 0 \leq \theta \leq \pi)\end{aligned}$$

The larger  $\kappa$  is, the more concentrated the distribution is around the mean direction. When  $\kappa = 0$  the distribution is again uniform (Mardia & Jupp 2000).

We can simulate from the von Mises-Fisher distribution using the algorithm described by Wood (1994). It uses the fact that the unit 3D vector  $\mathbf{x}$  has von Mises-Fisher distribution with  $\mu = (0, 0, 1)^T$  if  $\mathbf{x}^T = (\mathbf{v}\sqrt{1 - W^2}, W)$  where  $\mathbf{v}$  is a unit 2D vector which is uniformly distributed, and  $W$  is a random variable on  $[-1, 1]$  with density

$$f(w) = \frac{\exp(\kappa w)}{\sqrt{\frac{\pi\kappa}{2}} I_{\frac{1}{2}}(\kappa)}.$$

Here  $I_{\frac{1}{2}}(\kappa)$  is the modified Bessel function of the first kind and degree 1/2 (Wood 1994). Firstly, we set

$$b = -\kappa + \sqrt{\kappa^2 + 1}$$

then repeat the following steps until the desired sample size is obtained:

1. Put  $x_0 = (1 - b)/(1 + b)$  and  $c = \kappa x_0 + 2 \log(1 - x_0^2)$ ;
2. Generate  $U_1$  and  $U_2$  from a Uniform(0,1) distribution and calculate

$$W = \frac{1 - (1 + b)U_1}{1 - (1 - b)U_1};$$

3. If  $\kappa W + 2 \log(1 - x_0 W) - c < \log(U_2)$  then go to Step 1, else go to Step 4;
4. Generate the uniform 2D vector  $V$  and set  $\mathbf{x}^T = (\mathbf{v}\sqrt{1 - W^2}, W)$ .

Then  $\mathbf{x}$  has a  $F((0, 0, 1)^T, \kappa)$  distribution. We can convert this to a draw from a general  $F(\mu, \kappa)$  by using a rotation matrix (see Appendix A.1). In general, if  $\mathbf{x}$  is a draw from  $F(\mu, \kappa)$  then  $\mathbf{x}' = R\mathbf{x}$  is a draw from  $F(\mu', \kappa)$ , where  $\mu' = R\mu$ . So in this case, where  $\mu = (0, 0, 1)^T$ , any rotation matrix in which the third column is equal to the desired  $\mu$  will convert  $\mathbf{x}$  from  $F((0, 0, 1)^T, \kappa)$  to  $F(\mu, \kappa)$ . This can be done by taking Euler angles  $(\phi_0, \theta_0, 0)$  and setting  $R = R(\phi_0, \theta_0, 0)$ , where  $(\phi_0, \theta_0)$  are the spherical coordinates of  $\mu$ .

Figure 2.2 shows a sample of points from the von Mises-Fisher distribution, drawn in the above way using R, marked on the unit sphere, and shown in stereographic projection.

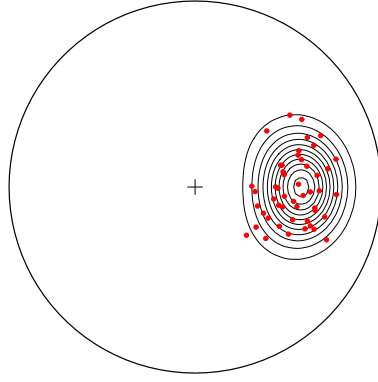


Figure 2.2: A random sample of size 50 drawn from a von Mises-Fisher distribution with  $\mu = (\phi_0, \theta_0) = (\frac{\pi}{2}, \frac{\pi}{4})$  and  $\kappa = 20$ . The contours of the density are shown.

### 2.2.2 Bingham Distribution

The *Bingham distribution* is used to describe axial data on the sphere. In the case of axial data, angles opposite each other are equivalent. Hence, vectors  $\mathbf{x}$  and  $-\mathbf{x}$  cannot be distinguished. The 3-dimensional Bingham distribution has density

$$f(\pm \mathbf{x} | \mathbf{A}) = {}_1F_1 \left( \frac{1}{2}, \frac{3}{2}, \mathbf{A} \right)^{-1} \exp(\mathbf{x}^T \mathbf{A} \mathbf{x})$$

where  $\mathbf{A}$  is a symmetric, traceless  $3 \times 3$  matrix and  ${}_1F_1 \left( \frac{1}{2}, \frac{3}{2}, \mathbf{A} \right)$  is the hypergeometric function given by

$$\begin{aligned} {}_1F_1 \left( \frac{1}{2}, \frac{3}{2}, \mathbf{A} \right) &= \sum_{k=0}^{\infty} \sum_{\kappa} \frac{\left( \frac{1}{2} \right)_{\kappa}}{\left( \frac{3}{2} \right)_{\kappa}} \frac{C_{\kappa}(\mathbf{A})}{k!} \\ &= \int_{S^2} \exp(\mathbf{x}^T \mathbf{A} \mathbf{x}) d\mathbf{x} \end{aligned}$$

Here the integral is obtained over the surface of the sphere (Mardia & Jupp 2000). For more details on hypergeometric functions refer to Appendix A.6. The distribution can be obtained by conditioning the trivariate normal distribution on  $\|\mathbf{x}\| = 1$ . Hence if  $\mathbf{x} \sim N_3(\mathbf{0}, \Sigma)$  then  $\mathbf{x} | (\|\mathbf{x}\| = 1)$  has Bingham distribution with  $\mathbf{A} = -\frac{1}{2}\Sigma^{-1}$  (Mardia & Jupp 2000). When the Bingham distribution has rotational symmetry about some axis, it is known as the Watson distribution (Mardia & Jupp 2000).

The behaviour of the distribution can be assessed by using an eigenvalue decomposition of the parameter matrix  $\mathbf{A}$ .

**Theorem 2.1.** [Eigenvalue Decomposition] A symmetric matrix  $\mathbf{A}$  can be decomposed as

$$\mathbf{A} = \mathbf{U} \mathbf{D} \mathbf{U}^T$$

where  $\mathbf{D}$  is a diagonal matrix of eigenvalues of  $\mathbf{A}$  and the columns of  $\mathbf{U}$  are the eigenvectors of  $\mathbf{A}$ , with  $\mathbf{U}^T \mathbf{U} = \mathbf{I}$ . This is also known as diagonalisation.

For a proof of this theorem see Anton (2000).

Theorem 2.1 implies  $\mathbf{A}$  can be decomposed into  $\mathbf{A} = \mathbf{U} \mathbf{K} \mathbf{U}^T$ . Varying the values in  $\mathbf{K}$  results in various different shapes of the distribution (Mardia & Jupp 2000).

Wood (1987) describes a method with which we can simulate from the density

$$f(\mathbf{x}|\kappa, \beta, \gamma) = [2\pi c]^{-1} \exp [\kappa x_3 + \gamma x_3^2 + \beta(x_1^2 - x_2^2)]$$

which, if we set  $\kappa = 0$ , is a Bingham distribution of the form

$$f(\pm\mathbf{x}|\mathbf{A}) \propto \exp \left( \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \begin{bmatrix} \beta & 0 & 0 \\ 0 & -\beta & 0 \\ 0 & 0 & \gamma \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \right).$$

where  $\beta \geq 0$  and  $\gamma \in (-\infty, \infty)$  (Wood 1987). (This algorithm is very lengthy; refer to Wood (1987) for a description). After implementing the algorithm in R, we can simulate from a Bingham distribution with mean direction  $(\phi_0, \theta_0) = (0, 0)$ , from which a simulation from a distribution with arbitrary mean can be straightforwardly obtained by applying a rotation (as was the case for simulating from the von-Mises Fisher distribution).

We can assess the behaviour of the density under various values of  $\gamma$  and  $\beta$  (see Figure 2.3). It can be seen that when  $\gamma = \beta = 0$ , the distribution is uniform. Meanwhile,  $\gamma$  acts as a concentration parameter — the higher  $\gamma$ , the more concentration of points at the mean, while with  $\gamma$  negative, the points tend to be situated near the equator. We observe a girdle shape when  $\gamma \approx \beta \neq 0$ . Meanwhile, as  $\beta$  increases, points are drawn away from the mean, to two antipodal groupings on the equator.

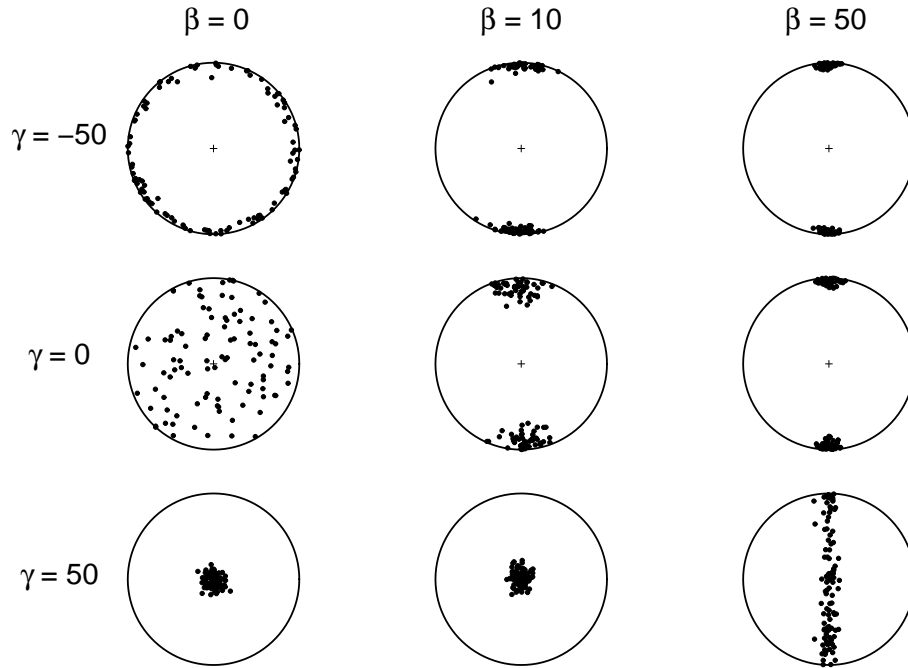


Figure 2.3: Random samples of size 100 drawn from a Bingham distribution with mean  $(\phi, \theta) = (0, 0)$  (the center of the stereonet) and various values of parameters  $\gamma$  and  $\beta$ . Values on the upper sphere have been projected onto the lower sphere.

## 2.3 Orientation Data

The spatial orientation of an object in  $p$  dimensions can be defined by  $n$  distinguishable directions (Downs 1972). An example is an L-shaped object in three dimensions, the orientation of which is defined by two orthogonal unit vectors giving the directions of its two arms (Downs 1972). If  $\mathbf{x}_1$  and  $\mathbf{x}_2$  are the two orthogonal vectors that describe these directions,

then the  $3 \times 2$  matrix  $\mathbf{X}$ , with columns  $\mathbf{x}_1$  and  $\mathbf{x}_2$ , describes the orientation completely, and  $\mathbf{X}^T \mathbf{X} = \mathbf{I}_2$  (Downs 1972, Khatri & Mardia 1977).

In this section we will discuss the Matrix Fisher distribution, which is an extension of the von Mises distribution, and is used to deal with data of this kind. Before discussing the Matrix Fisher distribution, however, we introduce the concept of a Stiefel Manifold.

### 2.3.1 Stiefel Manifolds

An *orthonormal  $n$ -frame* in  $\mathbb{R}^p$  is a set of  $n$  vectors  $(\mathbf{x}_1, \dots, \mathbf{x}_n)$  in  $\mathbb{R}^p$  that are orthonormal, i.e.  $\mathbf{x}_i^T \mathbf{x}_j = 0, i \neq j$  and each vector has length 1 (Mardia & Jupp 2000).

The set of all orthonormal  $n$ -frames in  $\mathbb{R}^p$  is known as the *Stiefel Manifold*  $V_n(\mathbb{R}^p)$  (Mardia & Jupp 2000). For our purposes we can think of the Stiefel Manifold as the set of all  $p \times n$  matrices that describe the orientation of an object in  $p$ -dimensions, defined by  $n$  directions, and for which  $\mathbf{X}^T \mathbf{X} = \mathbf{I}_n$ , i.e.

$$V_n(\mathbb{R}^p) = \{\mathbf{X} : \mathbf{X}^T \mathbf{X} = \mathbf{I}_n\} \quad (2.3)$$

(Khatri & Mardia 1977, Mardia & Jupp 2000).

A useful tool for considering distributions on  $V_n(\mathbb{R}^p)$  is the polar decomposition of matrices.

**Theorem 2.2.** [Polar Decomposition] Any  $p \times n$  matrix  $\mathbf{X}$  can be decomposed as

$$\mathbf{X} = \mathbf{M}\mathbf{K}$$

where  $\mathbf{M}$  is the polar part, in  $V_n(\mathbb{R}^p)$ , and  $\mathbf{K}$  is the elliptical part, an  $n \times n$  symmetric positive semi-definite matrix. We can obtain  $\mathbf{M}$  and  $\mathbf{K}$  uniquely when  $\mathbf{X}$  is invertible.  $\mathbf{K}$  is given by

$$\mathbf{K} = (\mathbf{X}^T \mathbf{X})^{\frac{1}{2}}$$

and

$$\mathbf{M} = \mathbf{X}\mathbf{K}^{-1}$$

(Mardia & Jupp 2000).

To prove the uniqueness of the decomposition it suffices to show that a positive definite matrix  $\mathbf{A}$  has a unique positive definite square root  $\mathbf{B} = \mathbf{A}^{\frac{1}{2}}$ . If this is the case then  $\mathbf{K}$  is unique. For a proof of this see Abadir & Magnus (2005, p 220).

### 2.3.2 Matrix Fisher distribution

The *Matrix Fisher distribution* describes the distribution of matrices  $\mathbf{X}$  on the Stiefel Manifold  $V_n(\mathbb{R}^p)$ . It has PDF

$$f(\mathbf{X}; \mathbf{F}) = \left[ {}_0F_1 \left( \frac{p}{2}; \frac{1}{4} \mathbf{F}^T \mathbf{F} \right) \right]^{-1} \exp [\text{tr}(\mathbf{F}^T \mathbf{X})] \quad (2.4)$$

where  $\mathbf{F}$  is a  $p \times n$  parameter matrix, and  ${}_0F_1 \left( \frac{p}{2}; \frac{1}{4} \mathbf{F}^T \mathbf{F} \right)$  is the hypergeometric function given by

$$\begin{aligned} {}_0F_1 \left( \frac{p}{2}; \frac{1}{4} \mathbf{F}^T \mathbf{F} \right) &= \sum_{k=0}^{\infty} \sum_{\boldsymbol{\kappa}} \frac{1}{\left( \frac{p}{2} \right)_{\boldsymbol{\kappa}}} \frac{C_{\boldsymbol{\kappa}}(\mathbf{F}^T \mathbf{F})}{k!} \\ &= \int_{V_n(\mathbb{R}^p)} \exp (\text{tr}(\mathbf{F}^T \mathbf{X})) d\mathbf{X} \end{aligned}$$

(Muirhead 1982, Mardia & Jupp 2000). Here  $C_\kappa$  and  $(a)_\kappa$  are as defined in Appendix A.6.

Since  $\mathbf{F}$  has polar decomposition  $\mathbf{F} = \mathbf{M}\mathbf{K}$  and  $\mathbf{K}$ , being symmetric, can be diagonalised as  $\mathbf{K} = \mathbf{U}\mathbf{D}\mathbf{U}^T$  for orthogonal  $\mathbf{U}$  and diagonal  $\mathbf{D}$ , then

$$\mathbf{F}^T\mathbf{F} = \mathbf{K}^T\mathbf{M}^T\mathbf{M}\mathbf{K} = \mathbf{K}^T\mathbf{K} = \mathbf{K}\mathbf{K} = \mathbf{U}\mathbf{D}\mathbf{U}^T\mathbf{U}\mathbf{D}\mathbf{U}^T = \mathbf{U}\mathbf{D}^2\mathbf{U}^T.$$

Now, since the zonal polynomial  $C_\kappa$  is a function of the eigenvalues of its argument, and since for any two matrices  $\mathbf{A}$  and  $\mathbf{B}$ ,  $\mathbf{AB}$  and  $\mathbf{BA}$  have the same eigenvalues, we find

$$\begin{aligned} C_\kappa(\mathbf{U}\mathbf{D}^2\mathbf{U}^T) &= C_\kappa(\mathbf{D}^2\mathbf{U}\mathbf{U}^T) \\ &= C_\kappa(\mathbf{D}^2) \end{aligned}$$

and hence  ${}_0F_1\left(\frac{p}{2}; \frac{1}{4}\mathbf{F}^T\mathbf{F}\right) = {}_0F_1\left(\frac{p}{2}; \frac{1}{4}\mathbf{D}^2\right) = c(\mathbf{D})$ . In other words, the hypergeometric function depends only on  $\mathbf{D}$ , the eigenvalues of  $\mathbf{K}$  (Khatri & Mardia 1977).

The focal mechanism of an earthquake is an example of a 3-dimensional object whose orientation can be described by three directions (the fault normal  $\hat{\mathbf{n}}$ , the slip vector  $\hat{\mathbf{u}}$  and null vector  $\hat{\mathbf{a}}$ , which together form the matrix  $R(\Theta) = [\hat{\mathbf{u}} \ \hat{\mathbf{a}} \ \hat{\mathbf{n}}]$ ). Since focal mechanism estimation is the central focus of this project, we will consider from here on only the case where  $p = 3$  and  $n = 3$ , i.e. a 3-dimensional object whose orientation can be described by three directions. The relevant Stiefel Manifold is  $V_3(\mathbb{R}^3)$ .

A Stiefel Manifold where  $n = p$ , as in this case, is equivalent to the *orthogonal group* of  $p \times p$  orthogonal matrices, denoted  $O(p)$  (Downs 1972, Khatri & Mardia 1977). If we add the further restriction that  $\det R = +1$  then this becomes the *special orthogonal group*  $SO(p)$ . Special orthogonal group  $SO(3)$  contains the  $3 \times 3$  matrices  $R(\Theta) = [\hat{\mathbf{u}} \ \hat{\mathbf{a}} \ \hat{\mathbf{n}}]$  that describe all possible orientations of the fault plane in 3 dimensions.

The shape of the distribution is controlled by  $\mathbf{M}$ ,  $\mathbf{U}$  and  $\mathbf{D}$ , where  $\mathbf{K} = \mathbf{U}\mathbf{D}\mathbf{U}^T$ . The density has a mode at the value of  $\mathbf{X}$  where  $\text{tr}(\mathbf{F}^T\mathbf{X})$  is maximised. This occurs when  $\mathbf{X} = \mathbf{M}$ , the polar part of  $\mathbf{F}$  (Downs 1972). The larger the entries of the diagonal matrix  $\mathbf{D}$ , the greater the concentration of the distribution about the vectors defined by the columns of  $\mathbf{M}$ . Meanwhile,  $\mathbf{U}$  twists the shape of the distribution at the modes.  $\mathbf{U}$  is a rotation matrix, and can hence be defined by three Euler angles  $(\phi, \theta, \psi)$  where  $\mathbf{U} = B_z(\phi)B_y(\theta)B_z(\psi)$ . The angles  $\theta$  and  $\psi$  directly twist the distribution at the third of the modes (the normal vector  $\hat{\mathbf{n}}$ ), while the twisting of the other two modes is more complex.  $\mathbf{M}$  is also a rotation matrix; overall therefore each of the three matrices  $\mathbf{M}$ ,  $\mathbf{U}$  and  $\mathbf{D}$  is defined by 3 parameters, which together constitute the 9 parameters of the parameter matrix  $\mathbf{F} = \mathbf{M}\mathbf{U}\mathbf{D}\mathbf{U}^T$ .

Due to their aforementioned properties,  $\mathbf{M}$ ,  $\mathbf{D}$  and  $\mathbf{U}$  are called the *modal matrix*, *concentration matrix* (Downs 1972) and *spin matrix* respectively. When  $\mathbf{D}$  is the zero matrix the distribution is uniform. Figure 2.4 shows the shape of the distribution for various values of these parameters.

A special case of the distribution, considered by Arnold & Townend (2007), occurs when  $\mathbf{D} = \kappa\mathbf{I}$ . In this case  $\mathbf{F} = \mathbf{M}\mathbf{U}\mathbf{D}\mathbf{U}^T = \kappa\mathbf{M}\mathbf{U}\mathbf{U}^T = \kappa\mathbf{M}$ , and so the distribution can be described as the Matrix Fisher distribution with scalar concentration parameter  $\kappa$ . The concentrations around each of the three modes are equal. Arnold & Townend (2007) assumed that the distribution of focal mechanism parameters  $\Theta$  could be fitted by this distribution. Later we explore the validity of this assumption.

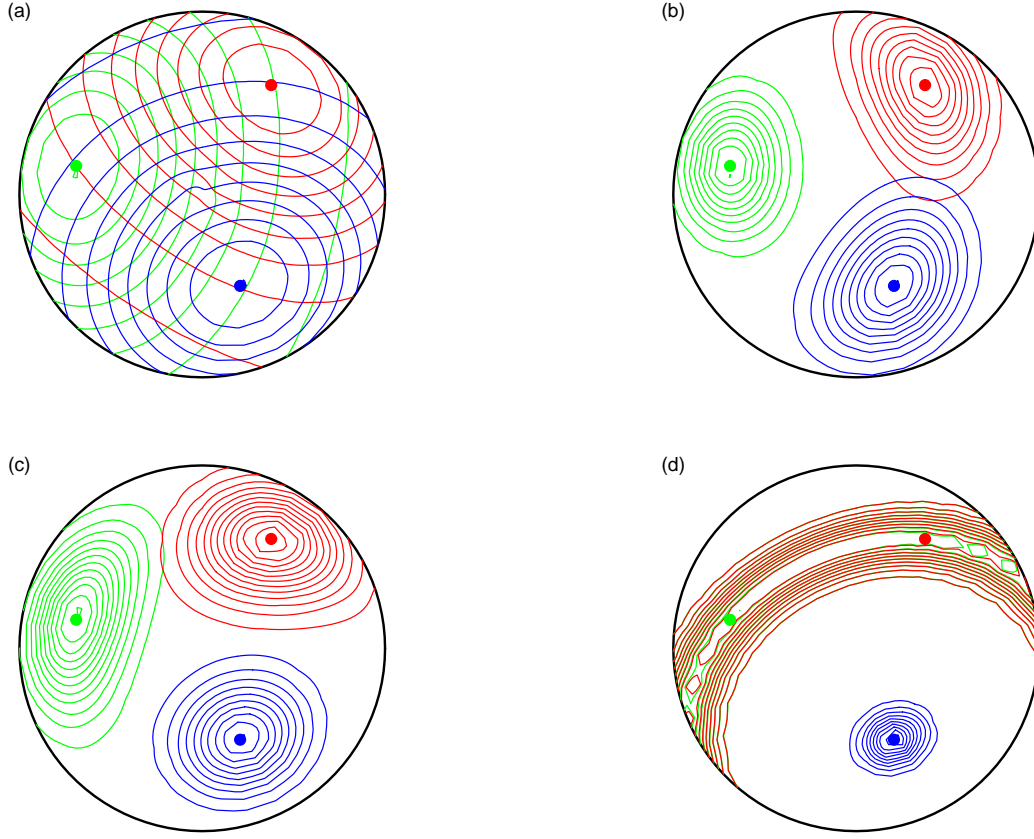


Figure 2.4: Contour plots of the Matrix Fisher density with  $\mathbf{M} = R(\alpha, \beta, \gamma) = R(\frac{7\pi}{8}, \frac{\pi}{4}, \frac{3\pi}{4})$  (marked as filled circles) and  $\mathbf{U}$  and  $\mathbf{D}$  as specified. The three modes  $[\hat{\mathbf{u}} \ \hat{\mathbf{a}} \ \hat{\mathbf{n}}]$  are coloured green, red and blue respectively. Plot (a) has  $\mathbf{D} = \text{diag}(1, 1, 1)$  and no spin, (b) has  $\mathbf{D} = \text{diag}(10, 5, 2)$  and no spin, (c) has  $\mathbf{D} = \text{diag}(10, 5, 2)$  and  $\mathbf{U} = R(\alpha, \beta, \gamma) = R(\frac{2\pi}{3}, \frac{2\pi}{3}, 0)$ , while (d) has  $\mathbf{D} = \text{diag}(0.01, 0.01, 50)$  and no spin.

### Error estimation

While  $\mathbf{D}$  determines the concentration of the distribution of  $R(\Theta)$  around each of the three modes, these values can be difficult to interpret. A more intuitive and interpretable measure of spread is the standard deviation of the angles  $\Theta = (\xi, \delta, \lambda)$  for a given Matrix Fisher distribution. Here we establish an approximate relationship between the concentration matrix  $\mathbf{D}$  and the standard deviations  $\sigma_\xi$  of the strike angle  $\xi$ .

Firstly, we apply a change of variable to represent the Matrix Fisher PDF  $f(\mathbf{X}; \mathbf{F})$  in terms of  $\Theta$ , where  $\mathbf{X} = R(\Theta)$ :

$$\begin{aligned} f(\Theta) &= f(R(\Theta)) \left| \frac{dR(\Theta)}{d\Theta} \right| \\ &\propto f(R(\Theta)) \sin \delta \\ &\propto c(\mathbf{D})^{-1} \exp [\text{tr}(\mathbf{F}^T R(\Theta))] \sin \delta \end{aligned}$$

Now,

$$f(\xi) = \int_0^{\pi/2} \int_0^{2\pi} f(\xi, \delta, \lambda) \sin \delta \, d\lambda \, d\delta$$

and substituting  $\sin \delta d\delta = -d(\cos \delta)$  gives

$$f(\xi) = \int_0^1 \int_0^{2\pi} f(\xi, \cos \delta, \lambda) d\lambda d(\cos \delta)$$

Here we integrate over  $\cos \delta$  since we tabulate  $f(R)$  on a grid in  $\xi, \cos \delta, \lambda$ .

As described in Appendix A.4, the mean of  $\xi$  is

$$\bar{\xi} = \tan^{-1} \left( \frac{\langle \sin \xi \rangle}{\langle \cos \xi \rangle} \right)$$

where in this case

$$\begin{aligned} \langle \sin \xi \rangle &= E(\sin \xi) = \frac{1}{c} \int_0^{2\pi} \sin \xi f(\xi) d\xi \\ &\simeq \frac{1}{c} \frac{2\pi}{n-1} \sum_i^n w_i f(\xi_i) \sin \xi_i \\ \langle \cos \xi \rangle &= E(\cos \xi) = \frac{1}{c} \int_0^{2\pi} \cos \xi f(\xi) d\xi \\ &\simeq \frac{1}{c} \frac{2\pi}{n-1} \sum_i^n w_i f(\xi_i) \cos \xi_i. \end{aligned}$$

In the above we approximate the integrals using the trapezium rule (see Appendix A.7.1), with  $w_1 = w_n = \frac{1}{2}$ ,  $w_j = 1$  otherwise, with  $c = \int f(\xi) d\xi$  a normalisation constant. The variance of  $\xi$  is given by

$$\begin{aligned} \sigma_\xi^2 &= E[(\xi - \bar{\xi})^2] \\ &= \frac{1}{c} \int_0^{2\pi} f(\xi) X d\xi \\ &= \frac{1}{c} \frac{2\pi}{n-1} \sum_i^n w_i f(\xi_i) X_i \end{aligned}$$

where  $X = \min \left( |\xi - \bar{\xi}|^2, (|\xi - \bar{\xi}| + 2\pi)^2, (|\xi - \bar{\xi}| - 2\pi)^2 \right)$  and again we approximate the integral using the trapezium rule.

We can now evaluate  $\sigma_\xi$  for different values of  $\mathbf{D}$  and establish the relationship between these two parameters. We firstly set  $\mathbf{M} = R(\pi, \pi/2, \pi)$ , so that we are evaluating the standard deviation far from the polar singularities of the coordinate system, and also set  $\mathbf{U} = R(0, 0, 0)$ , i.e. no spin. We then set  $\mathbf{D} = \text{diag}(d_1, d_2, k)$ , allow only  $k$  to vary, and evaluate  $\sigma_\xi$  for a range of  $k$  values. Table 2.1 shows values of  $\sigma_\xi$  (in degrees) for four different combinations of  $d_1$  and  $d_2$ .

Figure 2.5 shows there is a roughly linear relationship between  $\log(\sigma_\xi)$  and  $k$  for various values of  $\mathbf{D} = \text{diag}(d_1, d_2, k)$ . The linear relationship is stronger for higher concentration values. The overall relationship can be approximated by regressing  $\log(\sigma_\xi)$  on  $k$  across all chosen values of  $\mathbf{D}$ , and found to be:

$$\sigma_\xi = \exp(2.73738 - 0.02645k)$$

Thus standard deviations of  $\sigma_\xi = 1^\circ, 5^\circ, 10^\circ$  and  $15^\circ$  correspond to  $k$  values of approximately 100, 43, 16 and 1.1.

Table 2.1: Table of the standard deviation of the strike angle  $\sigma_\xi$  (in degrees) for various values of  $\mathbf{D}$ .

$k$	$\sigma_\xi \mathbf{D} = (0.01, 0.01, k)$	$\sigma_\xi \mathbf{D} = (0.01, 50, k)$	$\sigma_\xi \mathbf{D} = (50, 0.01, k)$	$\sigma_\xi \mathbf{D} = (50, 50, k)$
0.0625	57.52	65.49	57.11	3.89
0.125	56.64	64.12	55.99	3.88
0.25	54.87	61.32	53.73	3.87
0.5	51.30	55.63	49.23	3.85
1	44.34	44.52	40.74	3.81
2	32.62	26.75	28.19	3.73
4	19.80	10.06	17.62	3.58
8	12.27	4.03	11.81	3.28
16	8.30	2.98	8.19	2.75
32	5.64	1.99	5.60	1.90
64	2.89	0.91	2.88	0.88
128	0.63	0.19	0.63	0.18
256	0.03	0.01	0.03	0.01
512	0.00	0.00	0.00	0.00

This relationship does not hold in certain cases, however, since  $\max(\sigma_\xi) = \exp(2.73738) = 15.45^\circ$ . Clearly  $\sigma_\xi$  can exceed this value, as shown in Table 2.1, most obviously in cases in which two of the concentration parameters are small ( $< 1$ ). Thus we regress  $\log(\sigma_\xi)$  on  $k$  for  $k \leq 1$  and  $\mathbf{D} \neq (50, 50, k)$ , and obtain the following approximate relationship

$$\sigma_\xi = \exp(4.1196 - 0.3509k)$$

in the case when two or more of the components of  $\mathbf{D}$  are  $< 1$ .

In the case of the Matrix Fisher distribution with scalar concentration parameter, Arnold & Townend (2007) established the following approximate relationship between  $\kappa$  and the standard deviation  $\sigma_\Theta$  of the focal mechanism parameters (in degrees):

$$\sigma_\Theta = \exp(3.9155 - 0.5659 \log \kappa) \quad (2.5)$$

and thus  $\sigma_\Theta$  values of  $1^\circ$ ,  $5^\circ$ ,  $10^\circ$ ,  $15^\circ$ ,  $20^\circ$  and  $30^\circ$ , correspond to  $\kappa$  values of approximately 1000, 60, 17, 8, 5 and 2.5.

## Simulation

To simulate from the Matrix Fisher distribution on  $V_3(\mathbb{R}^3)$  with parameter matrix  $\mathbf{F}$ , we use the method given by Chikuse (2003). The simulation procedure starts with generating a pseudo-random uniform matrix  $\mathbf{X}$  using the following method:

1. Generate 9 independent realisations from the standard normal  $N(0, 1)$  distribution;
2. Arrange these into a  $3 \times 3$  matrix  $\mathbf{L}$ ;
3.  $\mathbf{X} = \mathbf{L}(\mathbf{L}^T \mathbf{L})^{-\frac{1}{2}}$  is a pseudo-random uniform matrix on  $V_3(\mathbb{R}^3)$ .

We then generate a random uniform realisation  $u$  on  $(0, 1)$ . If  $u < \exp[\text{tr}(\mathbf{F}^T \mathbf{X} - \mathbf{D})]$ , where  $\mathbf{D}$  is the diagonal matrix of eigenvalues of  $\mathbf{K}$ , then we accept  $\mathbf{X}$  as a random matrix from the Matrix Fisher distribution with parameter matrix  $\mathbf{F}$ . Otherwise we reject  $\mathbf{X}$  and repeat the procedure starting from Step 1 above.



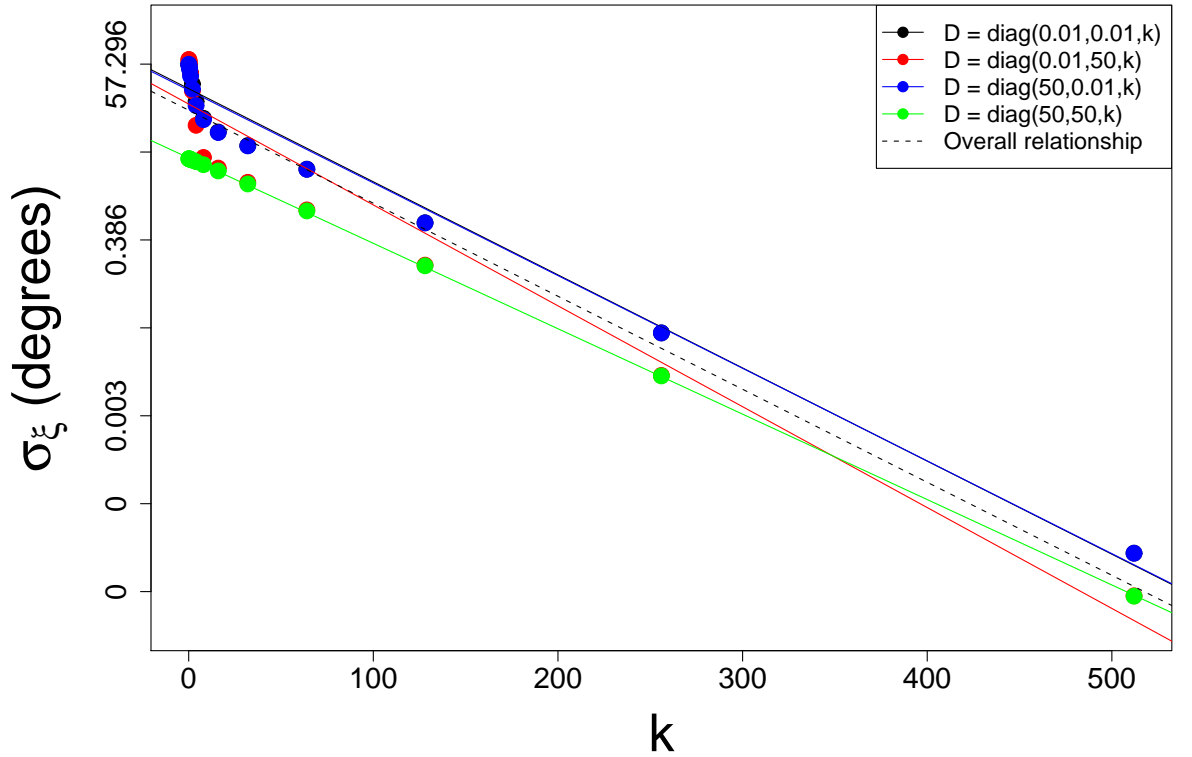


Figure 2.5: Plot of  $\sigma_\xi$  (log scale) versus  $k$  for various values of  $\mathbf{D} = \text{diag}(d_1, d_2, k)$ . Lines of best fit given by regressing  $\log(\sigma_\xi)$  on  $k$  for each  $\mathbf{D}$  are shown. The overall relationship is indicated by the dotted line.

Since the modal matrix  $\mathbf{M}$  is orthogonal and the columns define the modal directions of the distribution, we can always set  $\mathbf{M}$  as the identity, simulate, and then rotate the sample to a desired mean direction specified by three Euler angles. Figure 2.6 shows some simulations with various parameter values.

### Parameter Estimation — Kullback-Leibler Divergence

We use here a Matrix Fisher distribution  $f(R(\Theta)|\mathbf{F})$  to approximate a Bayesian posterior  $P(R(\Theta)|\mathbf{d})$  based on data  $\mathbf{d}$ . We now briefly describe the estimation of the value of  $\mathbf{F}$  that yields the best approximation to  $P(R(\Theta)|\mathbf{d})$ . We use the *Kullback-Leibler divergence*,  $H(\theta)$ , as a measure of the discrepancy between the true distribution  $P(\mathbf{y}|\mathbf{d})$  of some parameter  $\mathbf{y}$ , based on data  $\mathbf{d}$ , and a model distribution  $f(\mathbf{y}|\theta)$  that is defined by a parameter  $\theta$  (see e.g. Gelman et al. 1995). The Kullback-Leibler divergence for a given value of  $\theta$  is

$$\begin{aligned} H(\theta) &= E \left( \log \frac{P(\mathbf{y}|\mathbf{d})}{f(\mathbf{y}|\theta)} \right) \\ &= \int \log \left( \frac{P(\mathbf{y}|\mathbf{d})}{f(\mathbf{y}|\theta)} \right) P(\mathbf{y}|\mathbf{d}) d\mathbf{y} \\ &= \text{const.} - \int P(\mathbf{y}|\mathbf{d}) \log f(\mathbf{y}|\theta) d\mathbf{y} \end{aligned} \quad (2.6)$$

The aim is to find  $\hat{\theta}$  that minimises this divergence. To find  $\hat{\theta}$  we minimise Equation 2.6

$$\hat{\theta} = \underset{\theta}{\text{argmin}} H(\theta) = \underset{\theta}{\text{argmax}} \int P(\mathbf{y}|\mathbf{d}) \log f(\mathbf{y}|\theta) d\mathbf{y}$$

Consider the case in which the model density  $f(\mathbf{y}|\theta)$  is the Matrix Fisher density  $f(\mathbf{X}; \mathbf{F})$ . The density given by Equation 2.4 can be rewritten as

$$f(\mathbf{X}; \mathbf{F}) = c(\mathbf{D})^{-1} \exp [\text{tr}(\mathbf{F}^T \mathbf{X})]$$

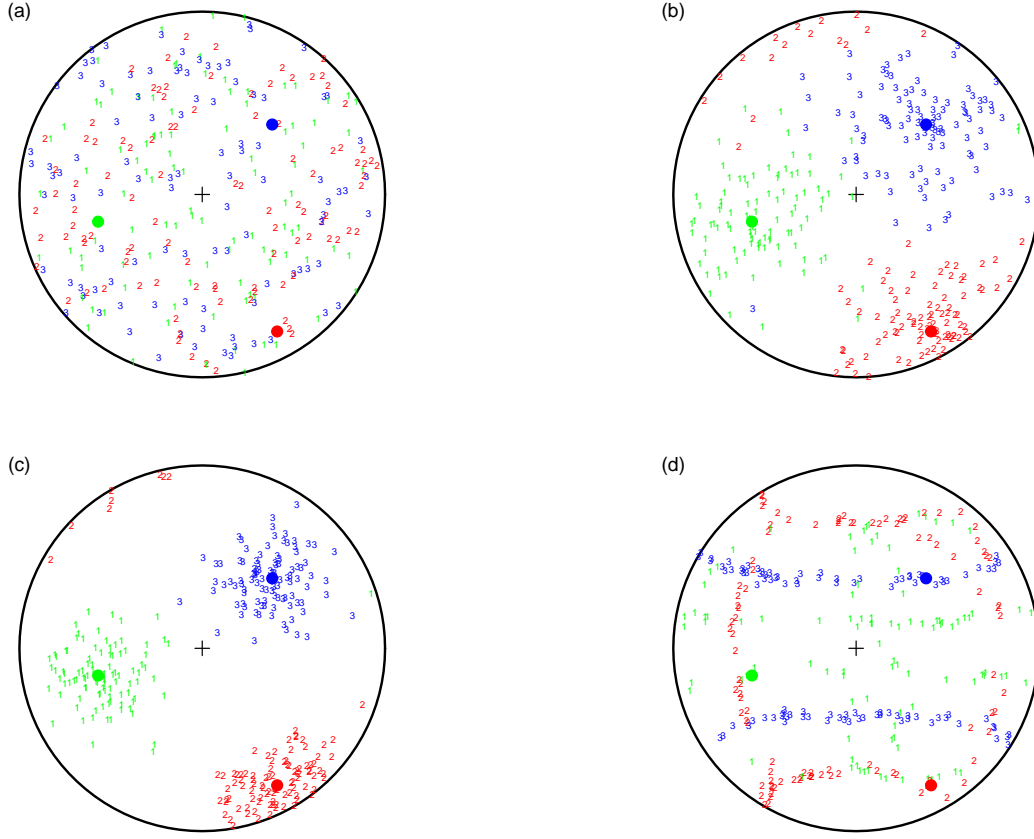


Figure 2.6: Random samples of size 100 drawn from a Matrix Fisher distribution with mode  $\mathbf{M} = R(\alpha, \beta, \gamma) = R(\frac{\pi}{4}, \frac{\pi}{4}, \frac{\pi}{8})$  (marked as filled circles) with no spin and  $\mathbf{D}$  as specified. Each observation is marked (1,2,3) for the  $(x, y, z)$  directions respectively. Values on the upper sphere have been reversed onto the lower sphere. Plot (a) has  $\mathbf{D} = \text{diag}(0, 0, 0)$  (the uniform case), (b) has  $\mathbf{D} = \text{diag}(5, 5, 5)$ , (c) has  $\mathbf{D} = \text{diag}(20, 10, 5)$  and (d) has  $\mathbf{D} = \text{diag}(1000, 0.01, 0.01)$ .

so that

$$\begin{aligned}
 \hat{\mathbf{F}} &= \underset{\mathbf{F}}{\text{argmax}} \int P(\mathbf{X}|\mathbf{d}) \log [c(\mathbf{D})^{-1} \exp (\text{tr}(\mathbf{F}^T \mathbf{X}))] d\mathbf{X} \\
 &= \underset{\mathbf{F}}{\text{argmax}} \left[ \int P(\mathbf{X}|\mathbf{d}) \text{tr} (\mathbf{F}^T \mathbf{X}) d\mathbf{X} - \int P(\mathbf{X}|\mathbf{d}) \log c(\mathbf{D}) d\mathbf{X} \right] \\
 &= \underset{\mathbf{F}}{\text{argmax}} \left[ \text{tr} \left( \mathbf{F}^T \int P(\mathbf{X}|\mathbf{d}) \mathbf{X} d\mathbf{X} \right) - \log c(\mathbf{D}) \int P(\mathbf{X}|\mathbf{d}) d\mathbf{X} \right]
 \end{aligned}$$

If the density  $P(\mathbf{X}|\mathbf{d})$  is unnormalised we divide through by  $\int P(\mathbf{X}|\mathbf{d}) d\mathbf{X}$

$$\begin{aligned}
 \hat{\mathbf{F}} &= \underset{\mathbf{F}}{\text{argmax}} \left[ \frac{\text{tr} (\mathbf{F}^T \int P(\mathbf{X}|\mathbf{d}) \mathbf{X} d\mathbf{X})}{\int P(\mathbf{X}|\mathbf{d}) d\mathbf{X}} - \log c(\mathbf{D}) \frac{\int P(\mathbf{X}|\mathbf{d}) d\mathbf{X}}{\int P(\mathbf{X}|\mathbf{d}) d\mathbf{X}} \right] \\
 &= \underset{\mathbf{F}}{\text{argmax}} \left[ \text{tr} \left( \mathbf{F}^T \frac{\int P(\mathbf{X}|\mathbf{d}) \mathbf{X} d\mathbf{X}}{\int P(\mathbf{X}|\mathbf{d}) d\mathbf{X}} \right) - \log c(\mathbf{D}) \right] \\
 &= \underset{\mathbf{F}}{\text{argmax}} [\text{tr} (\mathbf{F}^T \bar{\mathbf{X}}_w) - \log c(\mathbf{D})]
 \end{aligned} \tag{2.7}$$

where  $\bar{\mathbf{X}}_w$  is the mean weighted by the true density. In this project  $P(\mathbf{X}|\mathbf{d})$  is available on a grid  $\{\mathbf{X}_i\}_{i=1,\dots,k}$ . In that case the weighted mean can be approximated by

$$\bar{\mathbf{X}}_w \simeq \frac{\sum_{i=1}^k w_i P(\mathbf{X}_i|\mathbf{d}) \mathbf{X}_i}{\sum_{i=1}^k w_i P(\mathbf{X}_i|\mathbf{d})}$$

using the trapezium rule (see Appendix A.7.1). Equation 2.7 is equivalent to the log-likelihood of the Matrix Fisher distribution. The maximisation in Equation 2.7 is thus exactly analogous to Maximum Likelihood Estimation of  $\mathbf{F}$  in the case when a random sample of matrices has been drawn from the true density, rather than a tabulated grid of matrix values  $\{\mathbf{X}_i\}_{i=1,\dots,k}$  as in our case.

We now demonstrate how to solve Equation 2.7, in which the following theorem will be useful.

**Theorem 2.3.** *If a matrix  $\mathbf{A}$  commutes with a diagonal matrix with distinct diagonal entries,  $\mathbf{D}$ ,*

$$\mathbf{AD} = \mathbf{DA}$$

*then  $\mathbf{A}$  must also be diagonal.*

*Proof.*

$$\begin{aligned} (\mathbf{AD})_{ij} &= \sum_k \mathbf{A}_{ik} \mathbf{D}_{kj} = \sum_k \mathbf{A}_{ik} \mathbf{D}_{kk} \delta_{kj} = \mathbf{D}_{jj} \mathbf{A}_{ij} \\ (\mathbf{DA})_{ij} &= \sum_k \mathbf{D}_{ik} \mathbf{A}_{kj} = \sum_k \mathbf{D}_{ii} \delta_{ik} \mathbf{A}_{kj} = \mathbf{D}_{ii} \mathbf{A}_{ij} \\ \mathbf{D}_{jj} \mathbf{A}_{ij} &= \mathbf{D}_{ii} \mathbf{A}_{ij} \\ (\mathbf{D}_{jj} - \mathbf{D}_{ii}) \mathbf{A}_{ij} &= 0 \end{aligned}$$

where

$$\delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

Since  $\mathbf{D}_{ii} = \mathbf{D}_{jj}$  only if  $i = j$ , then  $\mathbf{A}_{ij} = 0$  if  $i \neq j$ , so  $\mathbf{A}$  is diagonal. □

Given that  $\mathbf{F}$  can be decomposed into

$$\mathbf{F} = \mathbf{MK} = \mathbf{MUDU}^T$$

we can use Lagrange multipliers to maximise Equation 2.7 subject to the constraints  $\mathbf{M}^T \mathbf{M} = \mathbf{I}$  and  $\mathbf{U}^T \mathbf{U} = \mathbf{I}$ .

Firstly, note that

$$(\mathbf{M}^T \mathbf{M})_{ij} = \sum_{\ell} M_{\ell i} M_{\ell j}$$

and similarly

$$(\mathbf{U}^T \mathbf{U})_{ij} = \sum_{\ell} U_{\ell i} U_{\ell j}$$

and that the identity matrix can be expressed as  $(I)_{ij} = \delta_{ij}$ . Hence the constraints  $\mathbf{M}^T \mathbf{M} = I$  and  $\mathbf{U}^T \mathbf{U} = I$  are equivalent to  $\sum_l M_{li} M_{lj} = \delta_{ij}$  and  $\sum_\ell U_{li} U_{lj} = \delta_{ij}$  for the entry in the  $i^{\text{th}}$  row and  $j^{\text{th}}$  column. So in the case of a  $3 \times 3$  matrix, this means there are nine constraint equations for each of the two constraints  $\mathbf{M}^T \mathbf{M} = I$  and  $\mathbf{U}^T \mathbf{U} = I$ . Each constraint equation adds a separate term given by

$$\lambda_{ij} \left( \delta_{ij} - \sum_\ell M_{li} M_{lj} \right)$$

and

$$\mu_{ij} \left( \delta_{ij} - \sum_\ell U_{li} U_{lj} \right)$$

onto the objective function, where  $\lambda_{ij}$  and  $\mu_{ij}$  are Lagrange multipliers. Hence the objective function to be maximised is

$$\begin{aligned} G(\mathbf{M}, \mathbf{U}, \mathbf{D}) &= \text{tr}(\mathbf{F}^T \bar{\mathbf{X}}) + \sum_{ij} \lambda_{ij} \left( \delta_{ij} - \sum_\ell M_{li} M_{lj} \right) + \sum_{ij} \mu_{ij} \left( \delta_{ij} - \sum_\ell U_{li} U_{lj} \right) - \log c(\mathbf{D}) \\ &= \text{tr}(\mathbf{M}^T \mathbf{A}) + \sum_{ij} \lambda_{ij} \left( \delta_{ij} - \sum_\ell M_{li} M_{lj} \right) + \sum_{ij} \mu_{ij} \left( \delta_{ij} - \sum_\ell U_{li} U_{lj} \right) - \log c(\mathbf{D}) \end{aligned}$$

where  $\mathbf{A} = \bar{\mathbf{X}}_w \mathbf{U} \mathbf{D} \mathbf{U}^T$ . Now we differentiate  $G$  with respect to  $M_{ab}$ :

$$\begin{aligned} G &= \sum_{ij} M_{ij} A_{ij} + \sum_{ij} \lambda_{ij} \left( \delta_{ij} - \sum_\ell M_{li} M_{lj} \right) + \sum_{ij} \mu_{ij} \left( \delta_{ij} - \sum_\ell U_{li} U_{lj} \right) - \log c(\mathbf{D}) \\ \frac{\partial G}{\partial M_{ab}} &= \mathbf{A}_{ab} - \sum_j \lambda_{bj} M_{aj} - \sum_i \lambda_{ib} M_{ai} \\ &= \mathbf{A}_{ab} - (\mathbf{M} \boldsymbol{\lambda}^T)_{ab} - (\mathbf{M} \boldsymbol{\lambda})_{ab} \stackrel{\text{set}}{=} 0 \end{aligned}$$

Hence

$$\mathbf{A} = \mathbf{M}(\boldsymbol{\lambda}^T + \boldsymbol{\lambda}),$$

$\boldsymbol{\lambda}^T + \boldsymbol{\lambda}$  is symmetric and hence  $\mathbf{M}$  is the polar part of  $\mathbf{A}$ . Thus

$$\mathbf{M} = \mathbf{A}(\mathbf{A}^T \mathbf{A})^{-\frac{1}{2}} = \bar{\mathbf{X}}_w \mathbf{K}(\mathbf{K} \bar{\mathbf{X}}_w^T \bar{\mathbf{X}}_w \mathbf{K})^{-\frac{1}{2}}.$$

Now if we put  $\mathbf{B} = \mathbf{M}^T \bar{\mathbf{X}}_w$  then

$$\text{tr}(\mathbf{F}^T \bar{\mathbf{X}}_w) = \text{tr}(\mathbf{U} \mathbf{D} \mathbf{U}^T \mathbf{M}^T \bar{\mathbf{X}}_w) = \text{tr}(\mathbf{U} \mathbf{D} \mathbf{U}^T \mathbf{B}).$$

Differentiating  $G$  with respect to  $U_{ab}$  gives

$$\begin{aligned} G &= \sum_{ijk\ell} U_{ij} D_{jk} U_{\ell k} B_{\ell i} + \sum_{ij} \lambda_{ij} \left( \delta_{ij} - \sum_\ell M_{li} M_{lj} \right) + \sum_{ij} \mu_{ij} \left( \delta_{ij} - \sum_\ell U_{li} U_{lj} \right) - \log c(\mathbf{D}) \\ \frac{\partial G}{\partial U_{ab}} &= \sum_{k\ell} D_{bk} U_{\ell k} B_{\ell a} + \sum_{ij} U_{ij} D_{jb} B_{ai} - \sum_j \mu_{bj} U_{aj} - \sum_i \mu_{ib} U_{ai} \\ &= (\mathbf{B}^T \mathbf{U} \mathbf{D})_{ab} + (\mathbf{B} \mathbf{U} \mathbf{D})_{ab} - (\mathbf{U} \boldsymbol{\mu}^T)_{ab} - (\mathbf{U} \boldsymbol{\mu})_{ab} \stackrel{\text{set}}{=} 0 \end{aligned}$$

and hence

$$\begin{aligned}
(\mathbf{B}^T + \mathbf{B})\mathbf{U}\mathbf{D} &= \mathbf{U}(\mu^T + \mu) \\
(\mathbf{B}^T + \mathbf{B})\mathbf{U}\mathbf{D}\mathbf{U}^T &= \mathbf{U}(\mu^T + \mu)\mathbf{U}^T \\
(\bar{\mathbf{X}}_w^T \mathbf{M} + \mathbf{M}^T \bar{\mathbf{X}}_w)\mathbf{K} &= \mathbf{U}(\mu^T + \mu)\mathbf{U}^T \\
\bar{\mathbf{X}}_w^T \mathbf{M}\mathbf{K} + \mathbf{M}^T \bar{\mathbf{X}}_w \mathbf{K} &= \mathbf{U}\mathbf{H}\mathbf{U}^T
\end{aligned}$$

where  $\mathbf{H} = (\mu^T + \mu)$ . Now recall that

$$\mathbf{M} = \bar{\mathbf{X}}_w^T \mathbf{K} (\mathbf{K} \bar{\mathbf{X}}_w^T \bar{\mathbf{X}}_w \mathbf{K})^{-\frac{1}{2}}$$

which means that

$$\mathbf{M}^T \bar{\mathbf{X}}_w \mathbf{K} = (\mathbf{K} \bar{\mathbf{X}}_w^T \bar{\mathbf{X}}_w \mathbf{K})^{-\frac{1}{2}} \mathbf{K} \bar{\mathbf{X}}_w^T \bar{\mathbf{X}}_w \mathbf{K} = (\mathbf{K} \bar{\mathbf{X}}_w^T \bar{\mathbf{X}}_w \mathbf{K})^{\frac{1}{2}}$$

and so

$$\bar{\mathbf{X}}_w \mathbf{M}\mathbf{K} + (\mathbf{K} \bar{\mathbf{X}}_w^T \bar{\mathbf{X}}_w \mathbf{K})^{\frac{1}{2}} = \mathbf{U}\mathbf{H}\mathbf{U}^T$$

Now the second and third matrices in this expression are symmetric. The first matrix must therefore also be symmetric, i.e.

$$\begin{aligned}
\bar{\mathbf{X}}_w^T \mathbf{M}\mathbf{K} &= \mathbf{K}\mathbf{M}^T \bar{\mathbf{X}}_w \\
\bar{\mathbf{X}}_w^T \bar{\mathbf{X}}_w \mathbf{K} (\mathbf{K} \bar{\mathbf{X}}_w^T \bar{\mathbf{X}}_w \mathbf{K})^{-\frac{1}{2}} \mathbf{K} &= \mathbf{K} (\mathbf{K} \bar{\mathbf{X}}_w^T \bar{\mathbf{X}}_w \mathbf{K})^{-\frac{1}{2}} \mathbf{K} \bar{\mathbf{X}}_w^T \bar{\mathbf{X}}_w \\
\mathbf{K} \bar{\mathbf{X}}_w^T \bar{\mathbf{X}}_w \mathbf{K} (\mathbf{K} \bar{\mathbf{X}}_w^T \bar{\mathbf{X}}_w \mathbf{K})^{-\frac{1}{2}} \mathbf{K} \mathbf{K} &= \mathbf{K} \mathbf{K} (\mathbf{K} \bar{\mathbf{X}}_w^T \bar{\mathbf{X}}_w \mathbf{K})^{-\frac{1}{2}} \mathbf{K} \bar{\mathbf{X}}_w^T \bar{\mathbf{X}}_w \mathbf{K} \\
(\mathbf{K} \bar{\mathbf{X}}_w^T \bar{\mathbf{X}}_w \mathbf{K})^{\frac{1}{2}} \mathbf{K} \mathbf{K} &= \mathbf{K} \mathbf{K} (\mathbf{K} \bar{\mathbf{X}}_w^T \bar{\mathbf{X}}_w \mathbf{K})^{\frac{1}{2}} \\
(\mathbf{K} \bar{\mathbf{X}}_w^T \bar{\mathbf{X}}_w \mathbf{K}) \mathbf{K} \mathbf{K} &= (\mathbf{K} \bar{\mathbf{X}}_w^T \bar{\mathbf{X}}_w \mathbf{K})^{\frac{1}{2}} \mathbf{K} \mathbf{K} (\mathbf{K} \bar{\mathbf{X}}_w^T \bar{\mathbf{X}}_w \mathbf{K})^{\frac{1}{2}}
\end{aligned}$$

Overall, therefore, the right hand side of this expression is symmetric, and hence so is the left hand side, i.e.

$$(\mathbf{K} \bar{\mathbf{X}}_w^T \bar{\mathbf{X}}_w \mathbf{K}) \mathbf{K} \mathbf{K} = [(\mathbf{K} \bar{\mathbf{X}}_w^T \bar{\mathbf{X}}_w \mathbf{K}) \mathbf{K} \mathbf{K}]^T = \mathbf{K} \mathbf{K} (\mathbf{K} \bar{\mathbf{X}}_w^T \bar{\mathbf{X}}_w \mathbf{K}). \quad (2.8)$$

Now  $\bar{\mathbf{X}}_w$  has polar decomposition

$$\bar{\mathbf{X}}_w = \mathbf{R}\mathbf{S}$$

where  $\mathbf{R}$  is the polar part and  $\mathbf{S}$  is the elliptical part. Further,  $\mathbf{S}$  has eigenvalue decomposition

$$\mathbf{S} = \mathbf{V}\mathbf{E}\mathbf{V}^T$$

meaning that

$$\bar{\mathbf{X}}_w^T \bar{\mathbf{X}}_w = \mathbf{V}\mathbf{E}\mathbf{V}^T \mathbf{R}^T \mathbf{R} \mathbf{V}\mathbf{E}\mathbf{V}^T = \mathbf{V}\mathbf{E}^2 \mathbf{V}^T$$

and Equation 2.8 can be written

$$(\mathbf{K}\mathbf{V}\mathbf{E}^2 \mathbf{V}^T \mathbf{K}) \mathbf{K} \mathbf{K} = \mathbf{K} \mathbf{K} (\mathbf{K}\mathbf{V}\mathbf{E}^2 \mathbf{V}^T \mathbf{K})$$

Since  $\mathbf{K}$  is invertible we can cancel  $\mathbf{K}$  twice from both sides of this expression

$$\begin{aligned}
\mathbf{V}\mathbf{E}^2 \mathbf{V}^T \mathbf{K} \mathbf{K} &= \mathbf{K} \mathbf{K} \mathbf{V}\mathbf{E}^2 \mathbf{V}^T \\
\mathbf{V}\mathbf{E}^2 \mathbf{V}^T \mathbf{U} \mathbf{D}^2 \mathbf{U}^T &= \mathbf{U} \mathbf{D}^2 \mathbf{U}^T \mathbf{V}\mathbf{E}^2 \mathbf{V}^T \\
(\mathbf{U}^T \mathbf{V}\mathbf{E}^2 \mathbf{V}^T \mathbf{U}) \mathbf{D}^2 \mathbf{U}^T \mathbf{U} &= \mathbf{U}^T \mathbf{U} \mathbf{D}^2 (\mathbf{U}^T \mathbf{V}\mathbf{E}^2 \mathbf{V}^T \mathbf{U}) \\
(\mathbf{U}^T \mathbf{V}\mathbf{E}^2 \mathbf{V}^T \mathbf{U}) \mathbf{D}^2 &= \mathbf{D}^2 (\mathbf{U}^T \mathbf{V}\mathbf{E}^2 \mathbf{V}^T \mathbf{U})
\end{aligned}$$

Now  $\mathbf{D}^2$  is a diagonal matrix that commutes with  $\mathbf{U}^T \mathbf{V} \mathbf{E}^2 \mathbf{V}^T \mathbf{U}$  and hence by Theorem 2.3  $\mathbf{U}^T \mathbf{V} \mathbf{E}^2 \mathbf{V}^T \mathbf{U}$  is also diagonal. Since  $\mathbf{E}^2$  is diagonal, it follows that  $\mathbf{V}^T \mathbf{U} = \mathbf{I}$  and hence  $\mathbf{U} = \mathbf{V}$ . This means that the eigenvectors of  $\hat{\mathbf{K}}$  are the same as the eigenvectors of  $\mathbf{S}$ .

Now if we collect all these results together we find that

$$\begin{aligned}
\bar{\mathbf{X}}_w &= \mathbf{R} \mathbf{S} = \mathbf{R} \mathbf{V} \mathbf{E} \mathbf{V}^T && \text{(polar decomposition)} \\
\hat{\mathbf{K}} &= \mathbf{V} \hat{\mathbf{D}} \mathbf{V}^T \\
\hat{\mathbf{U}} &= \mathbf{V} \\
\hat{\mathbf{D}} &\text{ must be obtained numerically (see below)} \\
\hat{\mathbf{M}} &= \bar{\mathbf{X}}_w \mathbf{K} (\mathbf{K} \bar{\mathbf{X}}_w^T \bar{\mathbf{X}}_w \mathbf{K})^{-\frac{1}{2}} \\
&= \mathbf{R} \mathbf{V} \mathbf{E} \mathbf{V}^T \mathbf{V} \hat{\mathbf{D}} \mathbf{V}^T (\mathbf{V} \hat{\mathbf{D}} \mathbf{V}^T \mathbf{V} \mathbf{E} \mathbf{V}^T \mathbf{R}^T \mathbf{R} \mathbf{V} \mathbf{E} \mathbf{V}^T \mathbf{V} \hat{\mathbf{D}} \mathbf{V}^T)^{-\frac{1}{2}} \\
&= \mathbf{R} \mathbf{V} \mathbf{E} \mathbf{D} \mathbf{V}^T (\mathbf{V} \mathbf{E}^2 \mathbf{D} \mathbf{V}^T)^{-\frac{1}{2}} \\
&= \mathbf{R} \mathbf{V} \mathbf{E} \mathbf{D} \mathbf{V}^T (\mathbf{V} \mathbf{E} \mathbf{V}^T)^{-1} \\
&= \mathbf{R} \mathbf{V} \mathbf{E} \mathbf{D} \mathbf{V}^T \mathbf{V} \mathbf{E}^{-1} \mathbf{D}^{-1} \mathbf{V}^T \\
&= \mathbf{R} \\
\hat{\mathbf{F}} &= \hat{\mathbf{M}} \hat{\mathbf{K}} = \mathbf{R} \mathbf{V} \hat{\mathbf{D}} \mathbf{V}^T
\end{aligned}$$

So, to find the estimate  $\hat{\mathbf{F}}$  that maximises Equation 2.7, given a grid of matrices  $\{\mathbf{X}_i\}_{i=1,\dots,k}$ , we first compute  $\bar{\mathbf{X}}_w$ , decompose it into  $\mathbf{R}$  and  $\mathbf{S}$ , and then decompose  $\mathbf{S}$  into  $\mathbf{S} = \mathbf{V} \mathbf{E} \mathbf{V}^T$ . We next find the value  $\hat{\mathbf{D}}$  for which Equation 2.7 is maximised based on these values for  $\mathbf{R}$ ,  $\mathbf{S}$  and  $\mathbf{V}$ , and finally form

$$\hat{\mathbf{F}} = \mathbf{R} \mathbf{V} \hat{\mathbf{D}} \mathbf{V}^T$$

We can carry out the maximisation over  $\hat{\mathbf{D}}$  using the `optim()` function in R. This function allows us to specify a lower bound of `diag(0, 0, 0)` for  $\hat{\mathbf{D}}$ , to ensure that the eigenvalues of  $\mathbf{K}$ , and thus the concentrations of the Matrix Fisher distribution, are non-negative.

In the case of the Matrix Fisher distribution with scalar concentration parameter, we simply set  $\mathbf{D} = \kappa \mathbf{I}$  in the `optim()` procedure to obtain the estimate of the scalar concentration parameter  $\kappa$ . The estimate of  $\mathbf{M}$  remains as above.

To demonstrate this method we run the procedure using simulated data. Note that here we do not use a grid of matrices  $\{\mathbf{X}_i\}_{i=1,\dots,k}$ , but a random sample, making this example one of Maximum Likelihood estimation rather than minimisation of the Kullback-Leibler divergence. The two are equivalent, as mentioned above.

We take a sample of size 100 of matrices  $\mathbf{X}_1, \dots, \mathbf{X}_{100}$  from the Matrix Fisher distribution

with arbitrary parameter matrix  $\mathbf{F} = \mathbf{MUDU}^T$  where

$$\mathbf{U} = R(\alpha, \beta, \gamma) = R\left(\pi, \frac{2\pi}{3}, 0\right) = \begin{bmatrix} 0.500 & 0.000 & -0.866 \\ 0.000 & -1.000 & 0.000 \\ -0.866 & 0.000 & -0.500 \end{bmatrix}$$

$$\mathbf{D} = \begin{bmatrix} 8 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

$$\mathbf{M} = R(\alpha, \beta, \gamma) = R\left(\frac{7\pi}{8}, \frac{\pi}{4}, \frac{3\pi}{4}\right) = \begin{bmatrix} 0.191 & 0.733 & -0.653 \\ -0.845 & 0.462 & 0.271 \\ 0.500 & 0.500 & 0.707 \end{bmatrix}$$

implying that,

$$\mathbf{F} = \mathbf{MUDU}^T = \begin{bmatrix} 2.367 & 2.930 & -4.743 \\ -3.659 & 1.848 & 3.953 \\ -0.087 & 2.000 & 3.297 \end{bmatrix}$$

We then calculate  $\hat{\mathbf{F}}$  given the sample mean  $\bar{\mathbf{X}}$ . As the data are sampled from a Matrix Fisher distribution, one would expect  $\mathbf{F}$  and  $\hat{\mathbf{F}}$  to be similar:

$$\hat{\mathbf{U}} = \begin{bmatrix} 0.405 & 0.453 & -0.794 \\ 0.453 & -0.854 & -0.256 \\ -0.794 & -0.256 & -0.551 \end{bmatrix}$$

$$\hat{\mathbf{D}} = \begin{bmatrix} 8.081 & 0 & 0 \\ 0 & 2.863 & 0 \\ 0 & 0 & 2.342 \end{bmatrix}$$

$$\hat{\mathbf{M}} = \begin{bmatrix} 0.182 & 0.728 & -0.661 \\ -0.855 & 0.449 & 0.259 \\ 0.485 & 0.518 & 0.704 \end{bmatrix}$$

and,

$$\hat{\mathbf{F}} = \hat{\mathbf{M}}\hat{\mathbf{U}}\hat{\mathbf{D}}\hat{\mathbf{U}}^T = \begin{bmatrix} 2.498 & 4.282 & -5.731 \\ -3.012 & 0.518 & 2.308 \\ 0.744 & 1.061 & 2.286 \end{bmatrix}$$

The actual and estimated modal ( $\mathbf{M}$ ) matrices are very similar, while the actual and estimated parameter ( $\mathbf{F}$ ), concentration ( $\mathbf{D}$ ) and spin ( $\mathbf{U}$ ) matrices are reasonably similar. In the case of the spin matrix  $\mathbf{U}$ , it is important to remember that the eigenvectors are defined only up to a constant.

We can compare  $\mathbf{F}$  and  $\hat{\mathbf{F}}$  visually by plotting the distributions using each matrix, as seen in Figure 2.7. The similarity of the two contour plots indicates the validity of the estimation method.

### 2.3.3 8-Mode Matrix Fisher distribution

The Matrix Fisher distribution can be generalised to situations in which there are ambiguities regarding the directions defined by the columns of the matrix  $\mathbf{X}$ , i.e. to axial data.

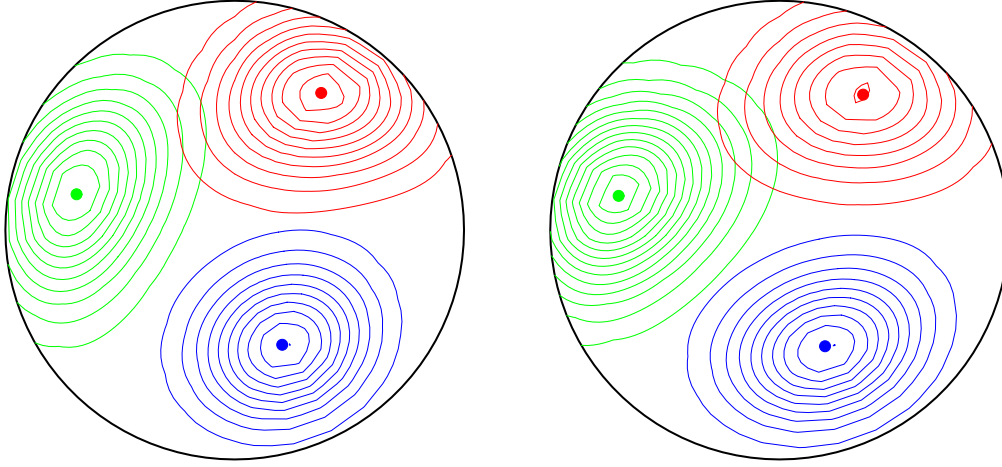


Figure 2.7: Marginal PDF plots of the Matrix Fisher distribution given  $\mathbf{F}$  (left) and the  $\hat{\mathbf{F}}$  (right) using simulated data.

Of particular interest in this project, given the two ambiguities surrounding focal mechanisms mentioned in Section 1.2.3, and one further ambiguity discussed in Section 3.3.1, will be the case in which there are eight equivalent representations of  $\mathbf{X}$ :  $\mathbf{X}$ ,  $\mathbf{XA}_2$ ,  $\mathbf{XC}_2$ ,  $\mathbf{XT}_2$ ,  $\mathbf{XA}_2\mathbf{C}_2$ ,  $\mathbf{XA}_2\mathbf{T}_2$ ,  $\mathbf{XC}_2\mathbf{T}_2$  and  $\mathbf{XA}_2\mathbf{C}_2\mathbf{T}_2$ , where

$$\mathbf{A}_2 = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix} \quad \mathbf{C}_2 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & -1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \quad \mathbf{T}_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

These ambiguities can be interpreted as follows:

- $\mathbf{A}_2$  reverses the first and third columns of  $\mathbf{X}$ . This is due to the axial nature of the fault normal vector  $\hat{\mathbf{n}}$ .
- $\mathbf{C}_2$  swaps the first and third columns of  $\mathbf{X}$ . This is due to the inability to distinguish between the slip vector  $\hat{\mathbf{u}}$  and normal vector  $\hat{\mathbf{n}}$ .
- $\mathbf{T}_2$  reverses the second column of  $\mathbf{X}$ , and reverses the sign of the determinant of  $\mathbf{X}$ . This is due to the lack of dependence of our Bayesian posterior on the orientation of the null vector  $\hat{\mathbf{a}}$  (see Section 3.3.1).

To fit a probability distribution to data of this nature we must generalise the Matrix Fisher distribution to account for the ambiguities. Here we introduce the *8-mode Matrix Fisher*



distribution, which has PDF

$$\begin{aligned}
f(\mathbf{X}; \mathbf{F}) = & \left[ {}_0F_1 \left( \frac{p}{2}; \frac{1}{4} \mathbf{F}^T \mathbf{F} \right) \right]^{-1} \\
& \times \frac{1}{8} \left\{ \exp [\text{tr}(\mathbf{F}^T \mathbf{X})] + \exp [\text{tr}(\mathbf{F}^T \mathbf{X} \mathbf{A}_2)] + \exp [\text{tr}(\mathbf{F}^T \mathbf{X} \mathbf{C}_2)] + \right. \\
& + \exp [\text{tr}(\mathbf{F}^T \mathbf{X} \mathbf{T}_2)] + \exp [\text{tr}(\mathbf{F}^T \mathbf{X} \mathbf{A}_2 \mathbf{C}_2)] + \exp [\text{tr}(\mathbf{F}^T \mathbf{X} \mathbf{A}_2 \mathbf{T}_2)] + \\
& \left. + \exp [\text{tr}(\mathbf{F}^T \mathbf{X} \mathbf{C}_2 \mathbf{T}_2)] + \exp [\text{tr}(\mathbf{F}^T \mathbf{X} \mathbf{A}_2 \mathbf{C}_2 \mathbf{T}_2)] \right\} \quad (2.9)
\end{aligned}$$

where  ${}_0F_1 \left( \frac{p}{2}; \frac{1}{4} \mathbf{F}^T \mathbf{F} \right)$  is the hypergeometric function of matrix argument. This gives equal weight to each of the eight equivalent representations of  $\mathbf{X}$ .

Figure 2.8 shows the shape of the 8-mode distribution under various values of the concentration and spin matrices. The ambiguities cause each of the vectors  $[\hat{\mathbf{u}} \ \hat{\mathbf{a}} \ \hat{\mathbf{n}}]$  to have multiple modes. It can be seen that if there is no spin, the axes  $\hat{\mathbf{u}}$  and  $\hat{\mathbf{n}}$  have identical probability contours (e.g. the blue contours coincide exactly with, and thus overplot, the green contours in plots (a) and (b)), while if spin is present and the concentrations vary (plots (c) and (d)) then the probability contours of the two axes are distinct.

### Simulation

To simulate from the 8-mode Matrix Fisher distribution with parameter matrix  $\mathbf{F}$  we adjust slightly the method of Chikuse (2003) described in Section 2.3.2.

We first generate  $\mathbf{X}$ , a random matrix from the Matrix Fisher distribution with parameter matrix  $\mathbf{F}$ . Then, with uniform probability  $\frac{1}{8}$ , we select randomly one of the 8 transformations  $(\mathbf{I}, \mathbf{A}_2, \mathbf{C}_2, \mathbf{T}_2, \mathbf{A}_2 \mathbf{C}_2, \mathbf{A}_2 \mathbf{T}_2, \mathbf{C}_2 \mathbf{T}_2, \mathbf{A}_2 \mathbf{C}_2 \mathbf{T}_2)$ , and apply the selected transformation to  $\mathbf{X}$ .

Figure 2.9 shows some samples from the 8-mode Matrix Fisher distribution for a variety of parameter matrices. The ambiguities are most evident in plot (c) (high concentration).

### Parameter Estimation — Kullback-Leibler Divergence

The procedure described in Section 2.3.2 must also be adjusted. Here we will adopt an iterative procedure. To find the estimate of  $\mathbf{F}$  that minimises the Kullback-Leibler divergence, given a grid of matrices  $\{\mathbf{X}_i\}_{i=1,\dots,k}$ , calculate the empirical PDF  $P(\mathbf{X}_i|\mathbf{d})$  at each  $\mathbf{X}_i$  and then perform the following steps:

1. Find  $\mathbf{X}_i : f(\mathbf{X}_i; \mathbf{F}) \geq f(\mathbf{X}_j; \mathbf{F}) \ \forall j \neq i$ , and set  $\hat{\mathbf{M}} = \mathbf{X}_i$ . This is the first estimate of  $\mathbf{M}$ , the modal part of  $\mathbf{F}$ ;
2. For every  $\mathbf{X}_i$  in the grid, find which of the eight possible representations of the matrix  $\mathbf{X}_i$  has maximum  $\text{tr}(\hat{\mathbf{M}}^T \mathbf{X})$  — call that representation  $\mathbf{X}_i^*$ ;
3. Calculate the weighted mean using the trapezium rule

$$\bar{\mathbf{X}}_w = \frac{\sum_{i=1}^k w_i P(\mathbf{X}_i|\mathbf{d}) \mathbf{X}_i^*}{\sum_{i=1}^k w_i P(\mathbf{X}_i|\mathbf{d})};$$

4. Use the mean to calculate a new estimate of  $\mathbf{M}$  - the polar part of  $\bar{\mathbf{X}}_w$ ;
5. Now using these values implement the procedure as described at the end of Section 2.3.2 to get  $\hat{\mathbf{F}}$ , an estimate of  $\mathbf{F}$ .

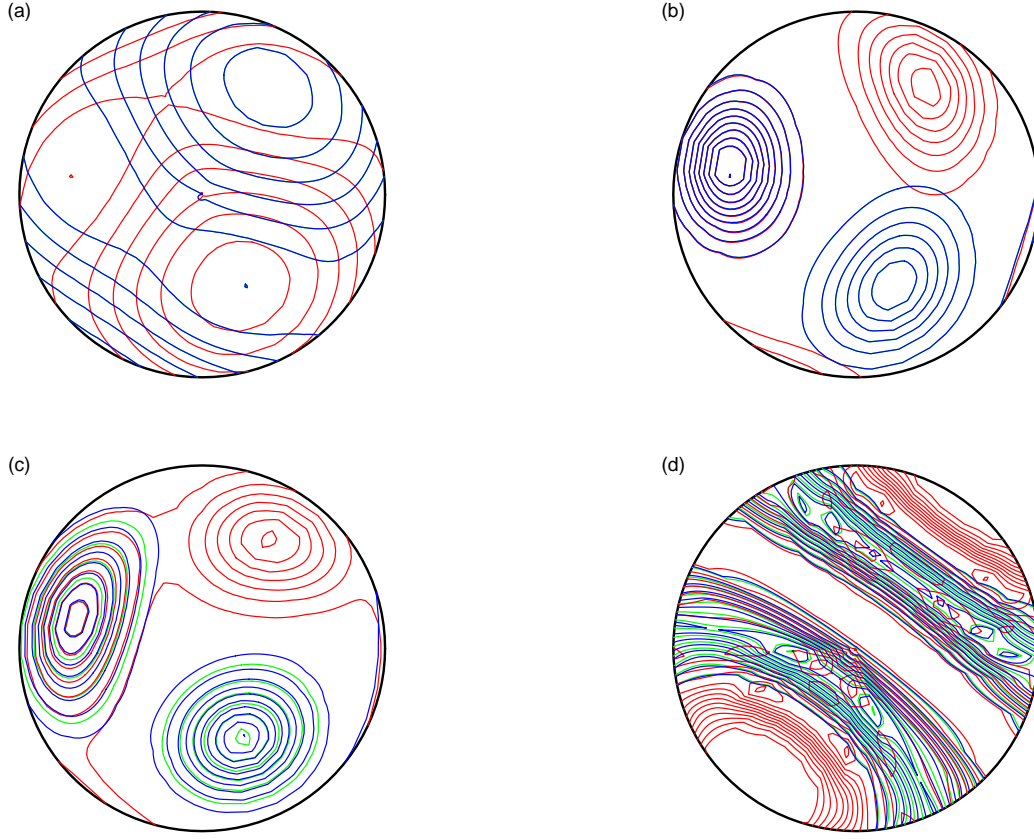


Figure 2.8: Contour plots of the 8-Mode Matrix Fisher density with  $\mathbf{M} = R(\alpha, \beta, \gamma) = R(\frac{7\pi}{8}, \frac{\pi}{4}, \frac{3\pi}{4})$  and  $\mathbf{U}$  and  $\mathbf{D}$  as specified. The three axes  $[\hat{\mathbf{u}} \ \hat{\mathbf{a}} \ \hat{\mathbf{n}}]$  are coloured green, red and blue respectively. Plot (a) has  $\mathbf{D} = \text{diag}(1, 1, 1)$  and no spin, (b) has  $\mathbf{D} = \text{diag}(10, 5, 2)$  and no spin, (c) has  $\mathbf{D} = \text{diag}(10, 5, 2)$  and  $\mathbf{U} = R(\alpha, \beta, \gamma) = R(\frac{2\pi}{3}, \frac{2\pi}{3}, 0)$ , while (d) has  $\mathbf{D} = \text{diag}(0.01, 0.01, 50)$  and  $\mathbf{U} = R(\alpha, \beta, \gamma) = R(\frac{2\pi}{3}, \frac{2\pi}{3}, 0)$ .

From then on, repeat the process from Step 2, except that in Step 2 classify each grid point according to which representation gives maximum  $\text{tr}(\hat{\mathbf{F}}^T \mathbf{X})$ . The process continues until  $\hat{\mathbf{F}}$  converges.

The justification for this procedure is that due to the ambiguities, we cannot just use the mean of the matrices. Instead we take a first guess at the modal matrix of the distribution: the maximum of the posterior PDF (or any maximum, if more than one exist), which we call  $\hat{\mathbf{M}}$ . We then look at all eight matrix representations at each sampled point and choose the one which is closest to the mode  $\hat{\mathbf{M}}$ . This representation,  $\mathbf{X}^*$ , is used to form the mean. Using this mean we get a better estimate of  $\mathbf{M}$ , following which we again check which of the eight matrices is closest to the mode, and then recalculate the average. This process will eventually converge to a single estimate of  $\mathbf{M}$ , and hence of  $\mathbf{F}$ .

To demonstrate this method we ran the procedure using the same simulated data as was used in Section 2.3.2, with one of the eight transformations randomly applied to each. The

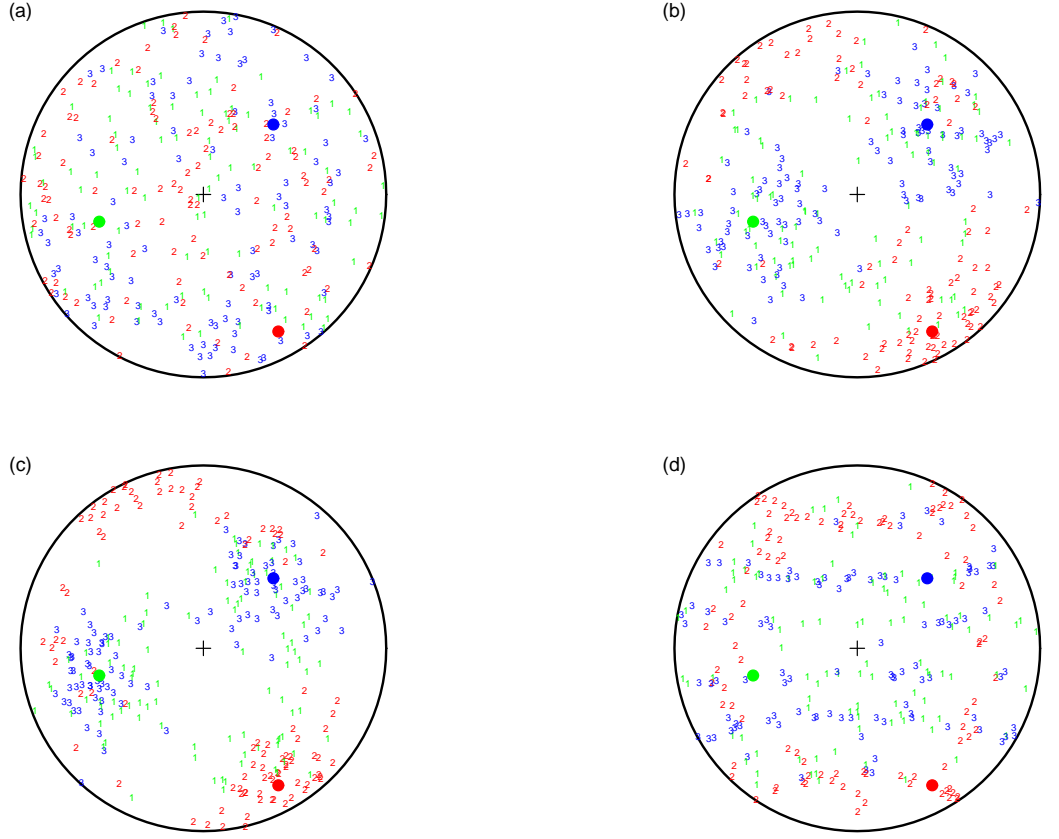


Figure 2.9: Random samples of size 100 drawn from the 8-Mode Matrix Fisher distribution with mode  $\mathbf{M} = R(\alpha, \beta, \gamma) = R(\frac{\pi}{4}, \frac{\pi}{4}, \frac{\pi}{8})$  (marked as filled circles),  $\mathbf{D}$  as specified and no spin. The observations are marked (1,2,3) for the  $(x, y, z)$  axes. Values on the upper sphere have been reversed onto the lower sphere. Plot (a) has  $\mathbf{D} = \text{diag}(0, 0, 0)$  (the uniform case), (b) has  $\mathbf{D} = \text{diag}(5, 5, 5)$ , (c) has  $\mathbf{D} = \text{diag}(10, 10, 10)$  and (d) has  $\mathbf{D} = \text{diag}(1000, 0.01, 0.01)$ .

calculated estimates are

$$\begin{aligned}\hat{\mathbf{U}} &= \begin{bmatrix} 0.473 & -0.723 & 0.503 \\ 0.585 & 0.685 & 0.434 \\ -0.659 & 0.089 & 0.747 \end{bmatrix} \\ \hat{\mathbf{D}} &= \begin{bmatrix} 7.169 & 0 & 0 \\ 0 & 3.391 & 0 \\ 0 & 0 & 2.648 \end{bmatrix} \\ \hat{\mathbf{M}} &= \begin{bmatrix} 0.219 & 0.736 & -0.641 \\ -0.846 & 0.471 & 0.251 \\ 0.486 & 0.487 & 0.725 \end{bmatrix}\end{aligned}$$

hence,

$$\hat{\mathbf{F}} = \hat{\mathbf{M}}\hat{\mathbf{U}}\hat{\mathbf{D}}\hat{\mathbf{U}}^T = \begin{bmatrix} 2.470 & 4.625 & -4.524 \\ -3.374 & 0.965 & 1.593 \\ 1.343 & 1.413 & 1.812 \end{bmatrix}$$

All four matrices are similar to the matrices used to generate the data specified in Section 2.3.2. We again compare  $\mathbf{F}$  and  $\hat{\mathbf{F}}$  visually by plotting the marginal 8-mode Matrix

Fisher distributions (see Figure 2.10). The similarity of the two plots shows  $\hat{\mathbf{F}}$  approximates  $\mathbf{F}$  well, and reinforces the validity of the iteration procedure.

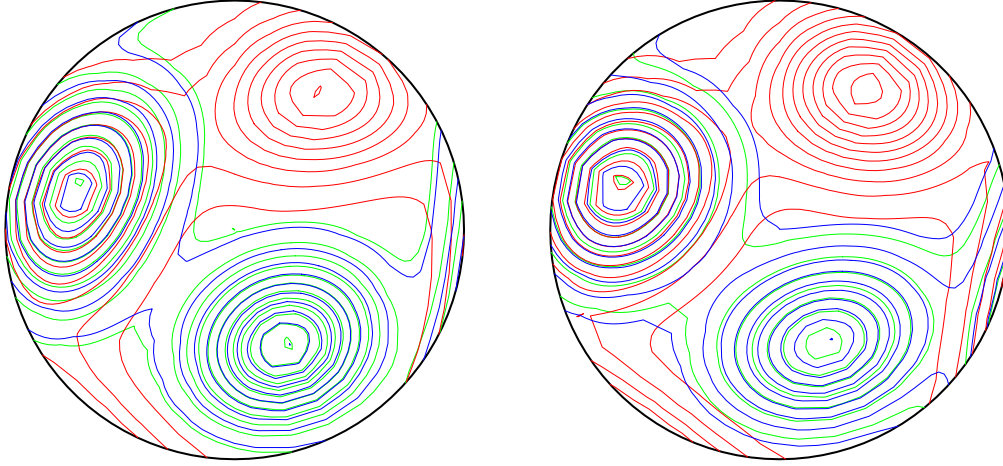


Figure 2.10: Marginal PDF plots of the 8-Mode Matrix Fisher distribution given  $\mathbf{F}$  (left) and the  $\hat{\mathbf{F}}$  (right) using simulated data.

### 2.3.4 Goodness of Fit Testing

Ideally, we would like to perform a test of the goodness of fit that a generalised Matrix Fisher distribution  $f(R(\Theta)|\mathbf{F})$  provides to a Bayesian posterior distribution  $P(R(\Theta)|\mathbf{d})$  based on data  $\mathbf{d}$ . Unfortunately, as will be shown in subsequent chapters, computation of the empirical distribution  $P(R|\mathbf{d})$  is sufficiently computationally intensive that resampling methods such as the bootstrap (in which the earthquake data generation process would be repeated using resampled  $\Theta$ 's from the empirical distribution) are not feasible for statistical testing of the fitted parameter values.

There is no statistical test for goodness of fit in the case where our empirical PDF is evaluated on a grid of points  $\{\mathbf{X}_i\}_{i=1,\dots,k}$  across the sample space, as in the case of the Kullback-Liebler divergence. However, if our parameter estimate had been obtained from a random sample of matrices from the true probability density, the goodness of fit tests on the Stiefel Manifold  $V_3(\mathbb{R}^3)$  described by Jupp (2005) could be applied. There are two test statistics that can be used to test the null hypothesis that the probability density function that generated the observed data is in  $F$ , where  $F$  is a family of probability density functions  $F = \{f(\cdot; \theta) : \theta \in \Theta\}$ . The weighted Rayleigh test statistic is

$$T_{wR} = 3n\text{tr}(\bar{\mathbf{X}}_w^T \bar{\mathbf{X}}_w)$$

where  $\bar{\mathbf{X}}_w$  is a weighted mean

$$\bar{\mathbf{X}}_w = \frac{1}{k} \sum_{i=1}^k \frac{\mathbf{X}_i}{f(\mathbf{X}_i; \hat{\theta})}$$

and the weighted Giné test statistic (based on Giné (1975)'s tests of uniformity) is

$$T_{wG} = \frac{1}{k} \sum_{i=1}^k \sum_{j=1}^k \frac{\left( \frac{1}{2} - \frac{3\pi}{32} \sqrt{\text{tr}(\mathbf{I}_3 - \mathbf{X}_i^T \mathbf{X}_j)} \right)}{f(\mathbf{X}_i; \hat{\theta}) f(\mathbf{X}_j; \hat{\theta})}$$

Significance is evaluated by resampling from the fitted distribution.



# Chapter 3

## Bayesian Methods

### 3.1 Bayes' Theorem

*Bayes' Theorem* is a simple rule in probability theory that forms the basis of the estimation techniques used in this project.

**Theorem 3.1.** [*Bayes' Theorem*] For two random variables  $X$  and  $Y$ :

$$P(X|Y) = \frac{P(Y|X)P(X)}{P(Y)} \quad (3.1)$$

*Proof.* The theorem is proved by noting that  $P(X, Y) = P(Y, X)$ , and

$$\begin{aligned} P(X, Y) &= P(X|Y)P(Y) \\ P(Y, X) &= P(Y|X)P(X) \end{aligned}$$

by the definition of conditional probability. Therefore,

$$\begin{aligned} P(X|Y)P(Y) &= P(Y|X)P(X) \\ \implies P(X|Y) &= \frac{P(Y|X)P(X)}{P(Y)}. \end{aligned}$$

□

Sivia (1996) explains the importance of the theorem for data analysis. Often we observe the result of some event (our data), and we want to establish the underlying cause of this outcome (the parameters of interest). This is not always an easy task. The reverse, working out the probability that some event occurs given we know the cause, is much easier. For example, suppose we flip a coin 10 times and obtain 6 heads. We then determine the probability that this is a fair coin, given the observations. Intuitively this is not simple, but if we were to determine the probability of a fair coin producing 6 heads in 10 flips, this probability is simply determined by the binomial distribution.

If we replace  $X$  and  $Y$  in Equation 3.1 by parameters and data, then:

$$P(\text{parameters}|\text{data}) \propto P(\text{data}|\text{parameters})P(\text{parameters})$$

Here, the probability of the parameters, given the data, is related to the probability that the data occurred given the parameters, the latter of which we are more likely to be able to calculate (Sivia 1996).  $P(\text{parameters})$  is known as the *prior* probability — our prior knowledge of the parameters. This prior knowledge is influenced by the *likelihood function*  $P(\text{data}|\text{parameters})$ . The likelihood is a function of the parameters; the data remain fixed. This outputs the *posterior* probability  $P(\text{parameters}|\text{data})$ , our knowledge of the parameters

given the data (Sivia 1996). Note that the denominator in Equation 3.1 has been dropped, because it does not depend on the parameters, and is simply a normalisation constant. Note that this does require an enumeration of the parameter space, and the specification of a set of prior probabilities for all possible hypotheses.

## 3.2 Examples

In this section we will consider two examples of situations in which a Bayesian approach can be applied. The first example, the lighthouse problem, is a canonical problem that provides a relatively straightforward introduction to a real-world situation in which Bayesian techniques can be used. The second example, the earthquake hypocentre location problem, is particularly relevant to this project, as it underpins the algorithm we will use to locate earthquakes. These locations are necessary inputs to the main problem solved in this project; the determination of focal mechanism parameters  $\Theta$ .

### 3.2.1 Lighthouse Problem

We now consider an example of a situation in which a Bayesian approach can be applied. The problem is defined as follows: “A lighthouse is somewhere off a piece of straight coast-line at a position  $a$  along the shore and a distance  $b$  out at sea. It emits a series of short highly collimated flashes at random intervals and hence at random azimuths. These pulses are intercepted on the coast by photo-detectors that record only the fact that a flash has occurred, but not the angle from which it came.  $N$  flashes have so far been recorded at positions  $\{x_k\}$ . Where is the lighthouse?” (Sivia 1996). Hereafter we let curly brackets denote a set.

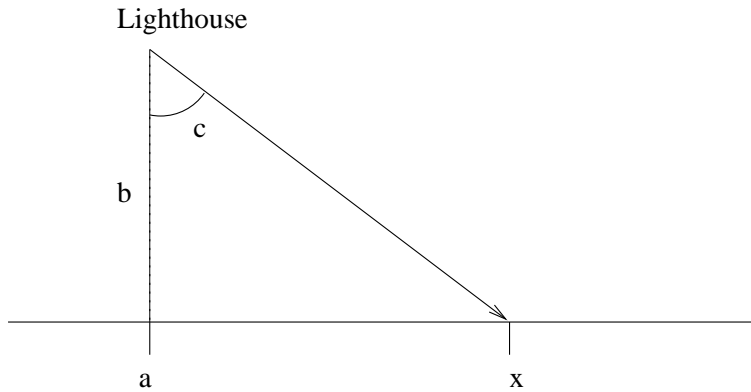


Figure 3.1: Visual representation of the lighthouse problem

Given that we know nothing about the lighthouse location, it is reasonable to assume a uniform prior on the azimuth  $c$  of the  $k^{th}$  observation:

$$P(c|a, b) = \frac{1}{\pi} \quad (3.2)$$

as  $(-\frac{\pi}{2} \leq c \leq \frac{\pi}{2})$ . Trigonometry tells us that  $b \tan(c) = x_k - a$ . We can use the change of variable technique (see Section A.5) to rewrite the PDF above as:

$$P(x_k|a, b) = \frac{b}{\pi[b^2 + (x - a)^2]}$$

So the probability of the  $k^{th}$  flash being recorded at  $x_k$ , given that we know where the lighthouse is, follows a Cauchy distribution. We will now assume in this example that  $b$  is known,



and we wish to find the posterior probability of  $a$ :

$$P(a|\{x_k\}, b) \propto P(\{x_k\}|a, b)P(a|b)$$

We know nothing about  $a$ , so we will give it a uniform prior, which reflects our ignorance (i.e. every possible distance along the shore is equally probable):

$$P(a|b) = P(a) = \begin{cases} A & (a_{min} \leq a \leq a_{max}) \\ 0 & \text{otherwise} \end{cases} \quad (3.3)$$

where  $A$  is equal to  $\frac{1}{a_{max}-a_{min}}$ . Now the recording of each signal is independent, so the likelihood function is given by:

$$P(\{x_k\}|a, b) = \prod_{k=1}^N P(x_k|a, b) \quad (3.4)$$

Now we know the prior (Equation 3.3) and the likelihood function (Equation 3.4), which we can put into Bayes' Theorem to obtain an expression for the posterior PDF:

$$P(a|\{x_k\}, b) \propto A \times \prod_{k=1}^N \frac{b}{\pi[b^2 + (x_k - a)^2]}$$

Table 3.1 illustrates how the various PDFs relate to those defined in Section 3.1.

Table 3.1: How PDFs in the lighthouse example relate to the PDFs in Section 3.1.

	Prior	Likelihood	Posterior
In general	$P(\text{parameters})$	$P(\text{data} \text{parameters})$	$P(\text{parameters} \text{data})$
Lighthouse example	$P(a)$	$P(\{x_k\} a, b)$	$P(a x_k, b)$

To get the best estimate of  $a$ , we need to maximise the posterior PDF. It becomes easier to deal with the log of the posterior PDF — the maximum will remain the same.

$$\begin{aligned} \log[P(a|\{x_k\}, b)] &= \log A + \sum_{k=1}^N [\log(b) - \log \pi - \log(b^2 + (x_k - a)^2)] \\ &= \text{constant} - \sum_{k=1}^N \log(b^2 + (x_k - a)^2) \end{aligned}$$

To maximise this, we differentiate with respect to  $a$  and set it equal to zero:

$$\frac{d \log[P(a|x_k, b)]}{da} = 2 \sum_{k=1}^N \frac{x_k - a}{b^2 + (x_k - a)^2} \stackrel{\text{set}}{=} 0$$

This equation cannot be easily rearranged to express  $a$  in terms of  $x_k$  and  $b$ , so we look at the problem numerically. We calculate the posterior PDF for many different values of  $a$ ; the largest PDF value will correspond to the best estimate of  $a$ . We can perform this procedure easily using R, generating random azimuths using Equation 3.2, and converting these into positions  $\{x_k\}$ . We also use fixed values of  $b = 1$ , and  $a = 2$ , to generate the data. We can then plot the posterior PDF against the lighthouse position,  $a$ , to find the best estimate of  $a$  (see Figure 3.2).

The plots of the posterior PDF begin to narrow as the number of flashes increases, centering on  $a = 2$ . This is the value of the estimate of  $a$ , and, as mentioned, was the value of  $a$  used to generate the data.

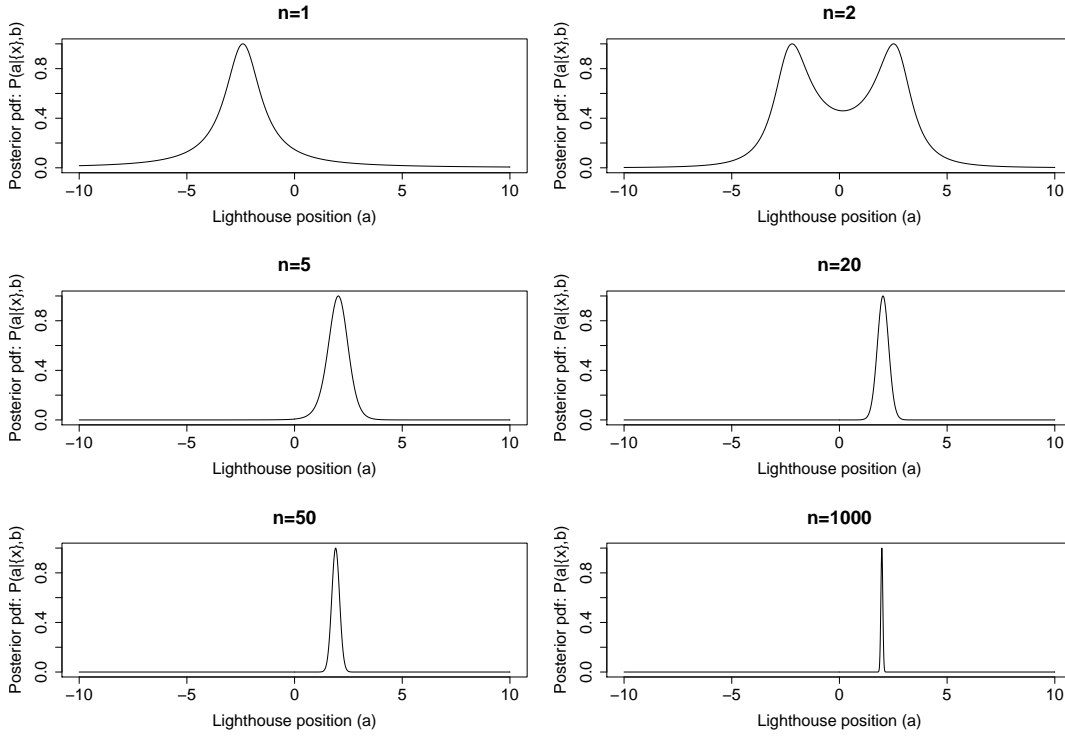


Figure 3.2: Posterior PDF of the position,  $a$ , of the lighthouse given  $b = 1$ . The number of flashes observed is given by  $n$ .

### 3.2.2 Earthquake Hypocentre Location

A second example of a situation in which a Bayesian approach can be adopted is the problem of locating an earthquake's hypocentre. This location algorithm is based on Tarantola & Valette (1982)'s approach. Locating an earthquake is an example of a Bayesian problem in that we wish to use data (seismometer arrival times) to calculate parameter values that describe the system (the coordinates of the hypocentre) (Tarantola 2005).

Tarantola & Valette (1982) apply the inversion approach to the hypocentre location problem. The unknown parameters are the hypocentre coordinates,  $\mathbf{x} = (x, y, z)$ , and the origin time,  $\mathbf{T}$ , while the known data are the seismic wave (P-, or both P- and S-) arrival times,  $\{t_i^{obs}\}$ , at stations. We assume we also know the locations of the  $n$  stations,  $s_i$ , and the velocity model,  $v$ . P-wave arrival times are the main wave type used in hypocentre location, while S-wave arrival times can be used to supplement the P-wave data and better constrain the hypocentre. In that case two velocity models are used, but the formulation below is unaffected.

We seek an expression for the posterior PDF of the hypocentre location  $P(\mathbf{x}|\{t_i\})$ . This example is slightly more complicated than the lighthouse example, as we have two unknown parameters;  $\mathbf{x}$  and  $\mathbf{T}$ .

Tarantola & Valette (1982) consider two sources of error — those errors caused by the assumed velocity model, and those caused by uncertainty in picking the exact arrival times at stations. Let  $\{t_i^{cal}\} = \{t_i^{cal}(\mathbf{x}, \mathbf{T})\}$  be the theoretical (calculated) arrival time at station  $i$  from a hypocentre location  $\mathbf{x}$  given a velocity model. We assume that  $\{t_i^{cal}\}$  is normally distributed with mean  $\{t_i^m\}$  and covariance matrix  $\mathbf{C}_T$ , where  $\{t_i^m\}$  represents the theoretical arrival time at station  $i$  from a hypocentre location  $\mathbf{x}$  given a perfectly implemented velocity model. We also assume the data  $\{t_i^{obs}\}$  are normally distributed with mean  $\{t_i^{cal}\}$  and covariance matrix  $\mathbf{C}_t$  (Tarantola & Valette 1982).

From the above it follows that the likelihood of the data,  $\{t_i^{obs}\}$ , given the parameters, is:

$$P(\{t_i^{obs}\}|\mathbf{x}, \mathbf{T}) = \exp \left\{ -\frac{1}{2}[\mathbf{t}^{obs} - \mathbf{t}^{cal}]^T (\mathbf{C}_t + \mathbf{C}_T)^{-1} [\mathbf{t}^{obs} - \mathbf{t}^{cal}] \right\} \quad (3.5)$$

This encompasses both the errors caused by the velocity model and the observational errors. So, the posterior PDF of the unknown parameters  $\mathbf{x}$  and  $\mathbf{T}$  is given by:

$$P(\mathbf{x}, \mathbf{T}|\{t_i^{obs}\}) \propto P(\mathbf{x}, \mathbf{T})P(\{t_i^{obs}\}|\mathbf{x}, \mathbf{T}) \quad (3.6)$$

$$\propto P(\mathbf{x}, \mathbf{T}) \exp \left\{ -\frac{1}{2}[\mathbf{t}^{obs} - \mathbf{t}^{cal}]^T (\mathbf{C}_t + \mathbf{C}_T)^{-1} [\mathbf{t}^{obs} - \mathbf{t}^{cal}] \right\} \quad (3.7)$$

where  $P(\mathbf{x}, \mathbf{T})$  describes our prior information about the parameters. Tarantola & Valette (1982) then assume a uniform prior on  $\mathbf{T}$ , since we generally have no prior information about the origin time. Hence,

$$\begin{aligned} P(\mathbf{x}, \mathbf{T}) &= P(\mathbf{x}).P(\mathbf{T}) && \text{(due to independence)} \\ &\propto P(\mathbf{x}) && \text{(due to uniformity of } P(\mathbf{T}) \text{)} \end{aligned} \quad (3.8)$$

We now define the theoretical travel time  $\{h_i^{cal}(\mathbf{x})\}$  between a hypocentre location  $\mathbf{x}$  and station  $i$ , which is simply the difference in time between the calculated arrival time and the origin time  $\mathbf{T}$ :

$$\mathbf{h}^{cal}(\mathbf{x}) = \mathbf{t}^{cal} - \mathbf{T} \quad (3.9)$$

Since we are more interested in the location of the earthquake than the origin time, we wish to obtain the posterior PDF of the hypocentre coordinates alone. We do this by integrating over the range of the origin time  $\mathbf{T}$  in Equation 3.7:

$$\begin{aligned} P(\mathbf{x}|\{t_i^{obs}\}) &= \int P(\mathbf{x}, \mathbf{T}|\{t_i^{obs}\})d\mathbf{T} \\ &= P(\mathbf{x}, \mathbf{T}) \int \exp \left\{ -\frac{1}{2}[\mathbf{t}^{obs} - \mathbf{t}^{cal}]^T (\mathbf{C}_t + \mathbf{C}_T)^{-1} [\mathbf{t}^{obs} - \mathbf{t}^{cal}] \right\} d\mathbf{T} \quad \text{(by 3.7)} \\ &\propto P(\mathbf{x}) \int \exp \left\{ -\frac{1}{2}[\mathbf{t}^{obs} - \mathbf{t}^{cal}]^T (\mathbf{C}_t + \mathbf{C}_T)^{-1} [\mathbf{t}^{obs} - \mathbf{t}^{cal}] \right\} d\mathbf{T} \quad \text{(by 3.8)} \\ &\propto \int \exp \left\{ -\frac{1}{2}[\mathbf{t}^{obs} - \mathbf{h}^{cal} - \mathbf{T}]^T \cdot \mathbf{P} \cdot [\mathbf{t}^{obs} - \mathbf{h}^{cal} - \mathbf{T}] \right\} d\mathbf{T} \quad \text{(by 3.9)} \end{aligned}$$

where  $\mathbf{P} = (\mathbf{C}_t + \mathbf{C}_T)^{-1}$ ,

$$\begin{aligned} &\propto \int \exp \left\{ -\frac{1}{2} \sum_{ij} [t_i^{obs} - h_i^{cal} - \mathbf{T}] \cdot P_{ij} \cdot [t_j^{obs} - h_j^{cal} - \mathbf{T}] \right\} d\mathbf{T} \\ &\propto \int \exp \left\{ -\frac{1}{2} \left[ \sum_{ij} P_{ij} \mathbf{T}^2 - 2\mathbf{T} \sum_{ij} P_{ij} (t_j^{obs} - h_j^{cal}) + \right. \right. \\ &\quad \left. \left. + \sum_{ij} (t_j^{obs} - h_j^{cal}) \cdot P_{ij} \cdot (t_i^{obs} - h_i^{cal}) \right] \right\} d\mathbf{T} \end{aligned}$$

Now let

$$\begin{aligned} a &= \sum_{ij} P_{ij} \\ b &= \sum_{ij} P_{ij} \cdot (t_j^{obs} - h_j^{cal}) \\ c &= \sum_{ij} (t_j^{obs} - h_j^{cal}) \cdot P_{ij} \cdot (t_i^{obs} - h_i^{cal}). \end{aligned}$$

Substituting these into the above gives:

$$\begin{aligned} P(\mathbf{x}|\{t_i^{obs}\}) &\propto \int \exp \left\{ -\frac{1}{2} (a\mathbf{T}^2 - 2\mathbf{T}b + c) \right\} d\mathbf{T} \\ &\propto \int \exp \left\{ -\frac{1}{2} \left( a\mathbf{T}^2 - 2\mathbf{T}b + \frac{b^2}{a} + c - \frac{b^2}{a} \right) \right\} d\mathbf{T} \\ &\propto \int \exp \left\{ -\frac{1}{2} \left[ a \left( \mathbf{T} - \frac{b}{a} \right)^2 + \left( c - \frac{b^2}{a} \right) \right] \right\} d\mathbf{T} \\ &\propto \exp \left\{ -\frac{1}{2} \left( c - \frac{b^2}{a} \right) \right\} \int \exp \left\{ -\frac{a}{2} \left( \mathbf{T} - \frac{b}{a} \right)^2 \right\} d\mathbf{T} \\ &\quad \text{Now using } \int \exp(-sx^2)dx = \sqrt{\frac{\pi}{s}} \text{ gives} \\ &\propto \exp \left\{ -\frac{1}{2} \left( c - \frac{b^2}{a} \right) \right\} \sqrt{\frac{2\pi}{a}} \end{aligned}$$

We now substitute  $a$ ,  $b$  and  $c$  back in:

$$\begin{aligned} P(\mathbf{x}|\{t_i^{obs}\}) &\propto \sqrt{\frac{2\pi}{\sum_{ij} P_{ij}}} \exp \left\{ -\frac{1}{2} \left[ (t_i^{obs} - h_i^{cal}) \cdot P_{ij} \cdot (t_j^{obs} - h_j^{cal}) - \frac{(\sum_{ij} P_{ij} (t_j^{obs} - h_j^{cal}))^2}{\sum_{ij} P_{ij}} \right] \right\} \\ &\propto \sqrt{\frac{2\pi}{\sum_{ij} P_{ij}}} \exp \left\{ -\frac{1}{2} \sum_{ij} \left[ t_i^{obs} - h_i^{cal} - \frac{\sum_{kl} P_{kl} (t_l^{obs} - h_l^{cal})}{\sum_{kl} P_{kl}} \right] \right. \\ &\quad \left. \cdot P_{ij} \cdot \left[ t_j^{obs} - h_j^{cal} - \frac{\sum_{kl} P_{kl} (t_l^{obs} - h_l^{cal})}{\sum_{kl} P_{kl}} \right] \right\} \\ &\propto \sqrt{\frac{2\pi}{\sum_{ij} P_{ij}}} \exp \left\{ -\frac{1}{2} \sum_{ij} \left( [\tilde{\mathbf{t}}^{obs} - \tilde{\mathbf{h}}^{cal}(\mathbf{x})]^T \cdot \mathbf{P} \cdot [\tilde{\mathbf{t}}^{obs} - \tilde{\mathbf{h}}^{cal}(\mathbf{x})] \right) \right\} \quad (3.10) \end{aligned}$$

where,  $\tilde{\mathbf{t}}^{obs}$  is the observed arrival time minus the weighted mean of observed arrival times

$$\tilde{t}_i^{obs} = t_i^{obs} - \frac{\sum_{kl} P_{kl} \cdot t_l^{obs}}{\sum_{kl} P_{kl}}$$

and  $\tilde{h}_i^{cal}(\mathbf{x})$  is the computed travel time between  $\mathbf{x}$  and station  $i$ , minus the weighted mean of computed travel times

$$\tilde{h}_i^{cal} = h_i^{cal} - \frac{\sum_{kl} P_{kl} \cdot h_l^{obs}}{\sum_{kl} P_{kl}}$$

(Tarantola & Valette 1982).

To summarise, Equation 3.10 gives the posterior PDF for the spatial location of an earthquake hypocentre, given the arrival time data  $\{t_i\}$ . Note that this requires estimation of  $\mathbf{P}$  and a valid velocity model for the region of interest.

### 3.3 Application to Focal Mechanisms

We now turn our attention to the main aim of this thesis: the estimation of focal mechanism parameters from a set of P-wave first motion data measured by an array of seismometers.

#### 3.3.1 Velocity Model Known

In this section we develop a probabilistic relationship between the focal mechanism parameters  $\Theta = (\xi, \delta, \lambda)$  and the available data using Bayesian methods, taking into account the relevant uncertainties (see Section 1.2.4). First we assume that the velocity model is perfectly known.

##### Seismological Model

Given a focal mechanism represented by  $\Theta = (\xi, \delta, \lambda)$ , we can compute the theoretical P-wave amplitude  $A_i^*$  at the  $i^{th}$  station using Equation 4.89 of Aki & Richards (2002):

$$A_i^* = 2(\hat{\mathbf{p}}_i \cdot \hat{\mathbf{n}})(\hat{\mathbf{p}}_i \cdot \hat{\mathbf{u}}) \quad (3.11)$$

where  $\hat{\mathbf{u}}$  is the slip vector,  $\hat{\mathbf{n}}$  is the fault normal, and they are both defined in terms of the focal mechanism parameters (Aki & Richards 2002):

$$\hat{\mathbf{u}}(\Theta) = (\sin \xi \cos \delta \sin \lambda + \cos \xi \cos \lambda, -\cos \xi \cos \delta \sin \lambda + \sin \xi \cos \lambda, -\sin \delta \sin \lambda) \quad (3.12)$$

$$\hat{\mathbf{n}}(\Theta) = (-\sin \xi \sin \delta, \cos \xi \sin \delta, -\cos \delta) \quad (3.13)$$

Here  $\hat{\mathbf{p}}_i$  is the unit vector from the hypocentre,  $\mathbf{x}$ , to the point on the focal sphere corresponding to station  $i$ . The vector  $\hat{\mathbf{p}}_i$  has azimuth  $\phi_i$  and take-off angle  $\theta_i$ :

$$\hat{\mathbf{p}}_i(\phi_i, \theta_i) = (\sin \theta_i \cos \phi_i, \sin \theta_i \sin \phi_i, \cos \theta_i) \quad (3.14)$$

In estimating a focal mechanism, we are more interested in the polarity of a P-wave first motion than its amplitude. The approach of Brillinger et al. (1980) is adopted here. Let  $Y$  define P-wave first motions as follows:

$$Y = \begin{cases} +1 & \text{if the first motion is recorded as positive (a compression)} \\ -1 & \text{if the first motion is recorded as negative (a dilatation)} \end{cases}$$

Then we can treat observed polarities at the  $i^{th}$  station as Bernoulli random variables:

$$\begin{aligned} P(Y_i = +1) &= \pi_i \\ P(Y_i = -1) &= 1 - \pi_i \end{aligned}$$

or alternatively

$$P(Y_i = y_i) = \pi_i^{\frac{1}{2}(1+y_i)} (1 - \pi_i)^{\frac{1}{2}(1-y_i)} \quad (y_i = -1, 1)$$

To compute  $\pi_i$ , we assume that the observed amplitude  $A_i$  is normally distributed:

$$\begin{aligned} A_i | A_i^*, c &\sim N(A_i^*, \sigma_a^2) \\ A_i | A_i^*, -c &\sim N(-A_i^*, \sigma_a^2) \end{aligned}$$

where  $c$  indicates a correctly wired station, and  $-c$  indicates cross-wiring (meaning that one can expect to observe the negative of the true amplitude). Note that we assume  $\sigma_a$  to

be constant across stations, in the absence of previous empirical studies. The conditional probabilities of observing a positive amplitude, given  $c$  and  $-c$  respectively, are:

$$\begin{aligned}
P(A_i > 0 | A_i^*, c) &= 1 - \Phi\left(\frac{0 - A_i^*}{\sigma_a}\right) \\
&= 1 - \Phi\left(\frac{-A_i^*}{\sigma_a}\right) \\
&= \Phi\left(\frac{A_i^*}{\sigma_a}\right) \\
P(A_i > 0 | A_i^*, -c) &= 1 - \Phi\left(\frac{0 - (-A_i^*)}{\sigma_a}\right) \\
&= 1 - \Phi\left(\frac{A_i^*}{\sigma_a}\right)
\end{aligned}$$

where  $\Phi$  is the Normal cumulative distribution function. Now let  $\pi_p$  be the probability of a correctly wired station. Then, to work out  $\pi_i$ , the probability of a positive observed first motion, we note that  $\pi_i$  is equal to the probability of a positive observed amplitude  $A_i$ :

$$\begin{aligned}
\pi_i &= P(A_i > 0 | A_i^*) \\
&= P(A_i > 0 | A_i^*, c)P(c) + P(A_i > 0 | A_i^*, -c)P(-c) \\
&= \Phi\left(\frac{A_i^*}{\sigma_a}\right)\pi_p + \left(1 - \Phi\left(\frac{A_i^*}{\sigma_a}\right)\right)(1 - \pi_p) \\
&= \Phi\left(\frac{A_i^*}{\sigma_a}\right)\pi_p + 1 - \Phi\left(\frac{A_i^*}{\sigma_a}\right) - \pi_p + \pi_p\Phi\left(\frac{A_i^*}{\sigma_a}\right) \\
&= 1 - \pi_p + (\pi_p - 1 + \pi_p)\Phi\left(\frac{A_i^*}{\sigma_a}\right) \\
&= 1 - \pi_p + (2\pi_p - 1)\Phi\left(\frac{A_i^*}{\sigma_a}\right) \\
&= \pi'_p + (1 - 2\pi'_p)\Phi\left(\frac{A_i^*}{\sigma_a}\right)
\end{aligned} \tag{3.15}$$

where  $\pi'_p = 1 - \pi_p$  is the probability of an incorrectly wired station. Substituting Equation 3.11 into the above gives:

$$\pi_i = \pi'_p + (1 - 2\pi'_p)\Phi\left(\frac{2(\hat{\mathbf{p}}_i \cdot \hat{\mathbf{n}})(\hat{\mathbf{p}}_i \cdot \hat{\mathbf{u}})}{\sigma_a}\right) \tag{3.16}$$

With precise data,  $\pi_p$  is large ( $\pi'_p$  small) and  $\sigma_a$  small. This model has the property that the larger the magnitude of  $A_i^*$ , the greater the probability of the P-wave first motion having been observed correctly (Brillinger et al. 1980). To summarise, the conditional probability of an observed polarity given the true amplitude and the relevant errors, is:

$$P(Y_i | A_i^*, \sigma_a, \pi_p) = \pi_i^{\frac{1}{2}(1+y_i)} (1 - \pi_i)^{\frac{1}{2}(1-y_i)} \quad (y_i = -1, 1) \tag{3.17}$$

where  $\pi_i$  is defined in Equation 3.16.

## Probabilistic Constraints

From here on, let curly brackets  $\{\}$  denote a set of values over the seismometer array ( $i = 1, \dots, n$ ). We are interested in attaining a probabilistic relationship between the following:

- **The data:** arrival times  $\{t_i\}$  and observed P-wave polarities  $\{Y_i\}$ . Let

$$\mathbf{d} = (\{t_i\}, \{Y_i\})$$

denote the data as a whole.

- **The unknown parameters of interest:** focal mechanism parameters  $\Theta = (\xi, \delta, \lambda)$ .
- **The nuisance parameters (unknown parameters that enter the analysis but are not of interest):** the earthquake hypocentre  $\mathbf{x}$ , the probability of a correctly wired station  $\pi_p$ , the amplitude noise  $\sigma_a$ , and the theoretical amplitude  $A_i^*$ . Let

$$\psi = (\psi_1, \mathbf{x}) = (\pi_p, \sigma_a, A_i^*, \mathbf{x})$$

denote the nuisance parameters, where  $\psi_1 = (\pi_p, \sigma_a, A_i^*)$ .

- **The known parameters:** station locations  $\{s_i\}$ , the P-wave arrival time errors  $\{\sigma_{t_i}\}$ , the velocity model  $v$  and the covariance matrix  $\mathbf{C}_T = \sigma_T^2 \cdot \exp \left\{ -\frac{1}{2} \frac{D_{ij}^2}{\Delta^2} \right\}$  describing the theoretical errors in the hypocentre location model (see Sections 3.2.2 and 4.2.1). Let

$$\omega = (\{s_i\}, \{\sigma_{t_i}\}, v, \mathbf{C}_T)$$

denote the known parameters.

Our goal is to attain an expression for the posterior probability density of the focal mechanism parameters, in terms of the data and the known parameters:

$$P(\Theta | \mathbf{d}, \omega)$$

We split  $\mathbf{d}$  into its components and apply Bayes Rule as follows:

$$P(\Theta | \{t_i\}, \{Y_i\}, \omega) \propto P(\Theta) P(\{Y_i\} | \{t_i\}, \Theta, \omega) \quad (3.18)$$

where  $P(\Theta)$  is the focal mechanism prior of our choice, and  $P(\{Y_i\} | \{t_i\}, \Theta, \omega)$  is the likelihood function of the data. Note that the prior could be denoted  $P(\Theta | \omega)$ , where  $\omega$  is all the background information that we use to formulate the prior. In practice, however, we will later adopt a non-informative parameterless prior, and so denote the prior  $P(\Theta)$ . In the following let MR denote the use of the marginalisation rule and PR the product rule. We will now work with the likelihood function to obtain a solution for the posterior probability. The likelihood is

$$\begin{aligned} P(\{Y_i\} | \{t_i\}, \Theta, \omega) & \stackrel{\text{MR}}{=} \int P(\{Y_i\}, \psi | \{t_i\}, \Theta, \omega) d\psi \\ & \stackrel{\text{PR}}{=} \int P(\{Y_i\} | \psi, \{t_i\}, \Theta, \omega) P(\psi | \{t_i\}, \Theta, \omega) d\psi \\ & = \int \left[ \prod_{i=1}^n P(Y_i | \psi, t_i, \Theta, \omega) \right] P(\psi | \{t_i\}, \Theta, \omega) d\psi \end{aligned}$$

now split  $\psi$  into  $\mathbf{x}$  and  $\psi_1$  and apply the product rule

$$\stackrel{\text{PR}}{=} \iint \left[ \prod_{i=1}^n P(Y_i | \mathbf{x}, \psi_1, t_i, \Theta, \omega) \right] P(\mathbf{x} | \psi_1, \{t_i\}, \Theta, \omega) P(\psi_1 | \{t_i\}, \Theta, \omega) d\psi_1 d\mathbf{x}$$

now  $Y_i \perp (\mathbf{x}, t_i, \Theta, \omega) | \psi_1$  and  $\mathbf{x} \perp \psi_1, \Theta$

$$= \iint \left[ \prod_{i=1}^n P(Y_i | \psi_1) \right] P(\mathbf{x} | \{t_i\}, \omega) P(\psi_1 | \{t_i\}, \Theta, \omega) d\psi_1 d\mathbf{x}$$

now split  $\psi_1$  into its components and apply the product rule

$$\begin{aligned}
&= \iiint \left[ \prod_{i=1}^n \int P(Y_i|A_i^*, \sigma_a, \pi_p) P(A_i^*|\sigma_a, \pi_p, \{t_i\}, \Theta, \omega) dA_i^* \right] P(\mathbf{x}|\{t_i\}, \omega) \\
&\quad \times P(\sigma_a) P(\pi_p) d\mathbf{x} d\sigma_a d\pi_p \\
&= \iiint \left[ \prod_{i=1}^n \int P(Y_i|A_i^*, \sigma_a, \pi_p) \delta(A_i^* - 2(\hat{\mathbf{p}}_i \cdot \hat{\mathbf{n}})(\hat{\mathbf{p}}_i \cdot \hat{\mathbf{u}})) dA_i^* \right] P(\mathbf{x}|\{t_i\}, \omega) \\
&\quad \times P(\sigma_a) P(\pi_p) d\mathbf{x} d\sigma_a d\pi_p \\
&= \iiint \left[ \prod_{i=1}^n P(Y_i|A_i^* = a_i, \sigma_a, \pi_p) \right] P(\mathbf{x}|\{t_i\}, \omega) P(\sigma_a) P(\pi_p) d\mathbf{x} d\sigma_a d\pi_p
\end{aligned}$$

where  $a_i = 2(\hat{\mathbf{p}}_i \cdot \hat{\mathbf{n}})(\hat{\mathbf{p}}_i \cdot \hat{\mathbf{u}})$ , and  $\delta(\cdot)$  is the Dirac delta function (see Appendix A.9). Approximating this integral we obtain

$$\simeq \iint \sum_{j=1}^m \left[ \prod_{i=1}^n P(Y_i|A_{ij}^* = a_{ij}, \sigma_a, \pi_p) \right] P(\sigma_a) P(\pi_p) d\sigma_a d\pi_p \quad (3.19)$$

where  $a_{ij} = 2(\hat{\mathbf{p}}_{ij} \cdot \hat{\mathbf{n}})(\hat{\mathbf{p}}_{ij} \cdot \hat{\mathbf{u}})$ , and  $m$  is the number of hypocentre locations sampled. In Equation 3.19 we have used Monte Carlo integration; the integral  $\int f(\mathbf{x})P(\mathbf{x}|\{t_i\}, \omega)d\mathbf{x}$  is approximated by evaluating the integrand at a random sample of hypocentre locations (see Appendix A.7.2 for details).

Note that for seismometer  $i$ ,  $\hat{\mathbf{p}}_i$  becomes  $\hat{\mathbf{p}}_{ij}$ , as it is now evaluated at a sample of hypocentre locations  $\mathbf{x}_j$ ,  $j = 1 \dots m$ . Thus,

$$\hat{\mathbf{p}}_{ij}(\phi_{ij}, \theta_{ij}) = (\sin \theta_{ij} \cos \phi_{ij}, \sin \theta_{ij} \sin \phi_{ij}, \cos \theta_{ij}) \quad (3.20)$$

and hence the amplitude  $A_i^*$ , a function of  $\hat{\mathbf{p}}$ , is evaluated at each hypocentre location and is now denoted  $A_{ij}^*$ .

Substituting Equation 3.17 into Equation 3.19 and then into Equation 3.18 we have an expression for the posterior PDF of the focal mechanism that we can evaluate, given by:

$$P(\Theta|\mathbf{d}, \omega) \propto P(\Theta) \iint \sum_{j=1}^m \left[ \prod_{i=1}^n \pi_{ij}^{\frac{1}{2}(1+y_i)} (1 - \pi_{ij})^{\frac{1}{2}(1-y_i)} \right] P(\sigma_a) P(\pi_p) d\sigma_a d\pi_p \quad (3.21)$$

where  $\pi_{ij}$  is given by

$$\pi_{ij} = \pi'_p + (1 - 2\pi'_p) \Phi \left( \frac{2(\hat{\mathbf{p}}_{ij} \cdot \hat{\mathbf{n}})(\hat{\mathbf{p}}_{ij} \cdot \hat{\mathbf{u}})}{\sigma_a} \right) \quad (3.22)$$

$P(\sigma_a)$  and  $P(\pi_p)$  are priors of our choice for the error terms  $\sigma_a$  and  $\pi_p$ , and  $P(\Theta)$  is the focal mechanism prior. From hereon we shall refer to Equation 3.21 as the velocity model known posterior PDF, or *VMK posterior PDF*.

We use R to evaluate the VMK posterior PDF of the focal mechanism parameters, by numerical evaluation of Equation 3.21. For each earthquake we cover the entire parameter space of  $\Theta = (\xi, \delta, \lambda)$  by calculating the VMK posterior PDF at every point on a grid. We use 21 evenly spaced values over each of  $\xi$ ,  $\cos \delta$  and  $\lambda$ , resulting in 9261 points. We use  $\cos \delta$  rather than  $\delta$  to correct for the fact that if we took an even spacing over  $\delta$ , we would have



higher node density near  $\delta = 0^\circ$  than near  $\delta = 90^\circ$ .

We can then obtain an optimal solution for  $\Theta$  based on the VMK posterior PDF. We find  $\arg\max_{\Theta} P(\Theta|\mathbf{d}, \omega)$ , the  $\Theta$  value on our grid for which  $P(\Theta|\mathbf{d}, \omega)$  is a maximum, and then conduct a local numerical optimisation of Equation 3.21 to obtain a final  $\Theta$  for which  $P(\Theta|\mathbf{d}, \omega)$  is maximised. We refer to this value as the *maximum a posteriori estimate* (or MAP estimate) of  $\Theta$ .

### A Further Ambiguity

The posterior PDF given by Equation 3.21 is dependent on the amplitude  $A_{ij}$ , which is given by the equation

$$A_{ij}^* = 2(\hat{\mathbf{p}}_{ij} \cdot \hat{\mathbf{n}})(\hat{\mathbf{p}}_{ij} \cdot \hat{\mathbf{u}})$$

and hence the only dependence on the focal mechanism parameters  $\Theta$  in the posterior PDF is through the vectors  $\hat{\mathbf{n}}$  and  $\hat{\mathbf{u}}$ . There is no dependence on the orientation of the null vector  $\hat{\mathbf{a}}$ , which is used in the construction of the rotation matrix  $\mathbf{X} = R(\Theta) = [\hat{\mathbf{u}} \ \hat{\mathbf{a}} \ \hat{\mathbf{n}}]$ , the distribution of which we are interested in. Thus as far as the posterior PDF is concerned,  $[\hat{\mathbf{u}} \ \hat{\mathbf{a}} \ \hat{\mathbf{n}}]$  is equivalent to

$$[\hat{\mathbf{u}} \ -\hat{\mathbf{a}} \ +\hat{\mathbf{n}}] = [\hat{\mathbf{u}} \ \hat{\mathbf{a}} \ \hat{\mathbf{n}}]\mathbf{T}_2 \quad \text{where } \mathbf{T}_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

This, combined with the two ambiguities discussed in Section 1.2.3, means that there are 8 equivalent representations of  $R(\Theta)$ :  $R(\Theta)$ ,  $R(\Theta)\mathbf{A}_2$ ,  $R(\Theta)\mathbf{C}_2$ ,  $R(\Theta)\mathbf{T}_2$ ,  $R(\Theta)\mathbf{A}_2\mathbf{C}_2$ ,  $R(\Theta)\mathbf{A}_2\mathbf{T}_2$ ,  $R(\Theta)\mathbf{C}_2\mathbf{T}_2$  and  $R(\Theta)\mathbf{A}_2\mathbf{C}_2\mathbf{T}_2$ . Each will have the same posterior PDF value. Note that the transformation by  $\mathbf{T}_2$  allows  $\det \mathbf{X} = \pm 1$ .

### 3.3.2 Velocity Model Unknown

In this section we adapt the probabilistic relationship between the focal mechanism parameters  $\Theta = (\xi, \delta, \lambda)$  and the available data established in the previous section. Here we assume that the velocity model is imperfectly known. Hardebeck & Shearer (2002) found that a change in velocity model had more impact on focal mechanism estimation than a change in hypocentre location.

#### Probabilistic Constraints

The difference from Section 3.3.1 is that here the velocity model  $v$  is unknown, and thus becomes a nuisance parameter. The nuisance parameters are now

$$\psi = (\psi_1, v, \mathbf{x}) = (\pi_p, \sigma_a, A_i^*, v, \mathbf{x})$$

where  $\psi_1 = (\pi_p, \sigma_a, A_i^*)$ . Again we apply Bayes Rule

$$P(\Theta|\{t_i\}, \{Y_i\}, \omega) \propto P(\Theta)P(\{Y_i\}|\{t_i\}, \Theta, \omega)$$

and now work with the likelihood function to obtain a solution for the posterior probability.

$$\begin{aligned}
P(\{Y_i\}|\{t_i\}, \Theta, \omega) &\stackrel{\text{MR}}{=} \int P(\{Y_i\}, \psi|\{t_i\}, \Theta, \omega) d\psi \\
&\stackrel{\text{PR}}{=} \int P(\{Y_i\}|\psi, \{t_i\}, \Theta, \omega) P(\psi|\{t_i\}, \Theta, \omega) d\psi \\
&= \int \left[ \prod_{i=1}^n P(Y_i|\psi, t_i, \Theta, \omega) \right] P(\psi|\{t_i\}, \Theta, \omega) d\psi
\end{aligned}$$

now split  $\psi$  into  $\mathbf{x}$ ,  $v$  and  $\psi_1$  and apply the product rule

$$\begin{aligned}
&\stackrel{\text{PR}}{=} \iiint \left[ \prod_{i=1}^n P(Y_i|\mathbf{x}, v, \psi_1, t_i, \Theta, \omega) \right] P(\mathbf{x}|\psi_1, v, \{t_i\}, \Theta, \omega) P(\psi_1|v, \{t_i\}, \Theta, \omega) \\
&\quad \times P(v|\{t_i\}, \Theta, \omega) d\psi_1 d\mathbf{x} dv
\end{aligned}$$

now  $Y_i \perp\!\!\!\perp (v, \mathbf{x}, t_i, \Theta, \omega) | \psi_1$ ,  $v \perp\!\!\!\perp (\{t_i\}, \Theta, \omega)$  and  $\mathbf{x} \perp\!\!\!\perp \psi_1, \Theta$

$$= \iiint \left[ \prod_{i=1}^n P(Y_i|\psi_1) \right] P(\mathbf{x}|\{t_i\}, \omega) P(\psi_1|v, \{t_i\}, \Theta, \omega) P(v) d\psi_1 d\mathbf{x} dv$$

now split  $\psi_1$  into its components and apply the product rule

$$\begin{aligned}
&= \iiint \left[ \prod_{i=1}^n \int P(Y_i|A_i^*, \sigma_a, \pi_p) P(A_i^*|\sigma_a, \pi_p, v, \{t_i\}, \Theta, \omega) dA_i^* \right] P(\mathbf{x}|v, \{t_i\}, \omega) \\
&\quad \times P(\sigma_a) P(\pi_p) P(v) d\mathbf{x} d\sigma_a d\pi_p dv \\
&= \iiint \left[ \prod_{i=1}^n \int P(Y_i|A_i^*, \sigma_a, \pi_p) \delta(A_i^* - 2(\hat{\mathbf{p}}_i \cdot \hat{\mathbf{n}})(\hat{\mathbf{p}}_i \cdot \hat{\mathbf{u}})) dA_i^* \right] P(\mathbf{x}|v, \{t_i\}, \omega) \\
&\quad \times P(\sigma_a) P(\pi_p) P(v) d\mathbf{x} d\sigma_a d\pi_p dv \\
&= \iiint \left[ \prod_{i=1}^n P(Y_i|A_i^* = a_i, \sigma_a, \pi_p) \right] P(\mathbf{x}|v, \{t_i\}, \omega) P(\sigma_a) P(\pi_p) P(v) d\mathbf{x} d\sigma_a d\pi_p dv
\end{aligned}$$

where  $a_i = 2(\hat{\mathbf{p}}_i \cdot \hat{\mathbf{n}})(\hat{\mathbf{p}}_i \cdot \hat{\mathbf{u}})$ .

$$\simeq \iiint \sum_{k=1}^p \left[ \prod_{i=1}^n P(Y_{ik}|A_{ik}^* = a_{ik}, \sigma_a, \pi_p) \right] P(\mathbf{x}_k|v_k, \{t_i\}, \omega) P(\sigma_a) P(\pi_p) d\mathbf{x} d\sigma_a d\pi_p \quad (3.23)$$

$$\simeq \iint \sum_{k=1}^p \sum_{j=1}^m \left[ \prod_{i=1}^n P(Y_{ijk}|A_{ijk}^* = a_{ijk}, \sigma_a, \pi_p) \right] P(\sigma_a) P(\pi_p) d\sigma_a d\pi_p \quad (3.24)$$

where  $a_{ijk} = 2(\hat{\mathbf{p}}_{ijk} \cdot \hat{\mathbf{n}})(\hat{\mathbf{p}}_{ijk} \cdot \hat{\mathbf{u}})$ ,  $m$  is the number of hypocentre locations sampled, and  $p$  is the number of velocity models sampled.

In Equation 3.23 Monte Carlo integration is applied to approximate the integral of  $P(v)$  by evaluating the integrand at a random sample of velocity models taken from the prior  $P(v)$ . Thus  $\hat{\mathbf{p}}_i$  becomes  $\hat{\mathbf{p}}_{ik}$ , and  $P(\mathbf{x}|v, \{t_i\}, \omega)$  becomes  $P(\mathbf{x}_k|v_k, \{t_i\}, \omega)$  as for each sampled velocity model the hypocentre location PDF varies.

In Equation 3.24 Monte Carlo integration is once again applied to approximate the integral of  $P(\mathbf{x}_k|v_k, \{t_i\}, \omega)$  by evaluating the integrand at a random sample of hypocentre locations. Here  $\hat{\mathbf{p}}_{ik}$  becomes  $\hat{\mathbf{p}}_{ijk}$ , as it is now evaluated for  $p$  different samples of  $m$  hypocentre locations  $\mathbf{x}_{jk}$ ,  $j = 1 \dots m$ ,  $k = 1 \dots p$ , where  $p$  and  $m$  are as described above.

Substituting Equation 3.17 into Equation 3.24 leaves an expression for the posterior PDF of the focal mechanism that we can evaluate, given by:

$$P(\Theta|\mathbf{d}, \omega) \propto P(\Theta) \int \int \sum_{k=1}^p \sum_{j=1}^m \left[ \prod_{i=1}^n \pi_{ijk}^{\frac{1}{2}(1+y_i)} (1 - \pi_{ijk})^{\frac{1}{2}(1-y_i)} \right] P(\sigma_a) P(\pi_p) d\sigma_a d\pi_p \quad (3.25)$$

Here  $\pi_{ijk}$  is given by

$$\pi_{ijk} = \pi'_p + (1 - 2\pi'_p) \Phi \left( \frac{2(\hat{\mathbf{p}}_{ijk} \cdot \hat{\mathbf{n}})(\hat{\mathbf{p}}_{ijk} \cdot \hat{\mathbf{u}})}{\sigma_a} \right)$$

and again  $P(\sigma_a)$  and  $P(\pi_p)$  are priors of our choice for the error terms  $\sigma_a$  and  $\pi_p$ , and  $P(\Theta)$  is the focal mechanism prior.

To distinguish this case from the previously established VMK posterior PDF, we shall from hereon refer to Equation 3.25 as the velocity model unknown posterior PDF, or *VMU posterior PDF*. Note that the VMK (Equation 3.21) and VMU (Equation 3.25) posterior PDFs are similar. In the VMU case we effectively sum over  $p$  different VMK posterior PDFs, weighted by their prior probabilities  $P(v)$ .

### 3.3.3 Probability Density of P- and T-axes

We can convert a PDF over focal mechanism parameters  $\Theta$  to one over the P- or T-axes (see Appendix A.8.1)  $\hat{\mathbf{v}}^P$  and  $\hat{\mathbf{v}}^T$ . For a given T-axis,  $\hat{\mathbf{v}}^T \parallel (\mathbf{u} + \mathbf{n})$ . The values of  $\hat{\mathbf{u}}$  and  $\hat{\mathbf{n}}$  that correspond to  $\hat{\mathbf{v}}^T$  are not unique, as  $\hat{\mathbf{u}} + \hat{\mathbf{n}} = (\hat{\mathbf{u}} + \Delta) + (\hat{\mathbf{n}} - \Delta)$ . In fact, for given  $\hat{\mathbf{v}}^T$ ,  $\hat{\mathbf{u}}$  and  $\hat{\mathbf{n}}$  can swing around by up to  $360^\circ$ , as shown in Figure 3.3.

We now define a coordinate system  $\mathbb{T}$ , expressed in geographic coordinates. The vector  $\hat{\mathbf{v}}^T$  is in the  $z$ -direction. Let  $\mathbf{m}$  be an arbitrary vector not parallel to  $\hat{\mathbf{v}}^T$ . Then  $\hat{\mathbf{b}}$ , a unit vector in the direction of  $\mathbf{m} \times \hat{\mathbf{v}}^T$ , defines the  $x$ -direction, and  $\hat{\mathbf{c}} = \hat{\mathbf{v}}^T \times \hat{\mathbf{b}}$  defines the  $y$ -direction of the coordinate system. Thus the coordinate system  $\mathbb{T}$ , expressed in geographical coordinates, has coordinate axis unit vectors  $R_{\mathbb{GT}} = [\hat{\mathbf{b}} \ \hat{\mathbf{c}} \ \hat{\mathbf{v}}^T]$ .

Let  $\zeta$  be the angle from the  $\hat{\mathbf{b}}$  axis to the projection of  $\hat{\mathbf{u}}$  onto the  $xy$  plane. Note that  $\hat{\mathbf{u}}$  and  $\hat{\mathbf{n}}$  are at  $45^\circ$  to the  $xy$  plane. Then

$$\hat{\mathbf{u}} = \frac{1}{\sqrt{2}} \begin{bmatrix} \cos \zeta \\ \sin \zeta \\ 1 \end{bmatrix} \quad \hat{\mathbf{n}} = \frac{1}{\sqrt{2}} \begin{bmatrix} -\cos \zeta \\ -\sin \zeta \\ 1 \end{bmatrix} \quad \hat{\mathbf{a}} = \hat{\mathbf{n}} \times \hat{\mathbf{u}}$$

and  $R(\zeta)_{\mathbb{T}} = [\hat{\mathbf{u}} \ \hat{\mathbf{a}} \ \hat{\mathbf{n}}]$ . We calculate  $P(\hat{\mathbf{v}}^T)$  as follows

$$\begin{aligned}
P(\hat{\mathbf{v}}^T) &= \int_{R_{\mathbb{G}}: \hat{\mathbf{v}}^T \parallel (\hat{\mathbf{u}} + \hat{\mathbf{n}})} P(R_{\mathbb{G}}) dR_{\mathbb{G}} \\
&= \int_0^{2\pi} P(R(\zeta)_{\mathbb{G}}) d\zeta \\
&= \int_0^{2\pi} P(R_{\mathbb{G}\mathbb{T}} R(\zeta)_{\mathbb{T}}) d\zeta \\
&= \int_0^{2\pi} P[\Theta(R_{\mathbb{G}\mathbb{T}} R(\zeta)_{\mathbb{T}})] d\zeta \\
&\simeq \frac{2\pi}{n-1} \sum_{i=1}^n w_i P[\Theta(R_{\mathbb{G}\mathbb{T}} R(\zeta_i)_{\mathbb{T}})]
\end{aligned}$$

where we approximate the integral over  $\zeta$  using the trapezium rule over a grid of  $n$   $\zeta$  values evenly spaced from 0 to  $2\pi$ . Also,  $\Theta(R_{\mathbb{G}\mathbb{T}} R(\zeta_i)_{\mathbb{T}})$  denotes the angles  $\Theta$  that correspond to the matrix  $R_{\mathbb{G}\mathbb{T}} R(\zeta_i)_{\mathbb{T}}$  (see Appendix A.1.1). We obtain  $P[\Theta(R_{\mathbb{G}\mathbb{T}} R(\zeta_i)_{\mathbb{T}})]$  by linear interpolation from the grid-tabulated posterior PDF  $P(\Theta|\mathbf{d}, \omega)$ .

The PDF over the P-axis is calculated similarly. We remain in coordinate system  $\mathbb{T}$ , and now think of the T-/z-axis as the P-axis. Since  $\hat{\mathbf{v}}^P \parallel (\hat{\mathbf{u}} - \hat{\mathbf{n}})$ , then  $R(\zeta)_{\mathbb{T}} = [-\hat{\mathbf{n}} \ \hat{\mathbf{a}} \ \hat{\mathbf{u}}]$ , and the calculation remains otherwise the same.

We can calculate the MAP estimate of the P- and T-axes directly from our MAP estimate of  $\Theta$ , using the equations  $\hat{\mathbf{v}}^P = \frac{1}{2}(\hat{\mathbf{u}} - \hat{\mathbf{n}})$  and  $\hat{\mathbf{v}}^T = \frac{1}{2}(\hat{\mathbf{u}} + \hat{\mathbf{n}})$ . In general the MAP P- and T-axes will sit close to the maxima of the 2-dimensional (over spherical coordinates  $\theta, \phi$ ) PDFs  $P(\hat{\mathbf{v}}^P)$  and  $P(\hat{\mathbf{v}}^T)$  respectively. However, since the MAP estimates are obtained from the 3-dimensional PDF  $P(\Theta|\mathbf{d}, \omega)$ , there may be slight differences.

Having now developed the theoretical and numerical approaches to the evaluation of the posterior PDFs and appropriate summaries, we now turn to some practical matters of their implementation.

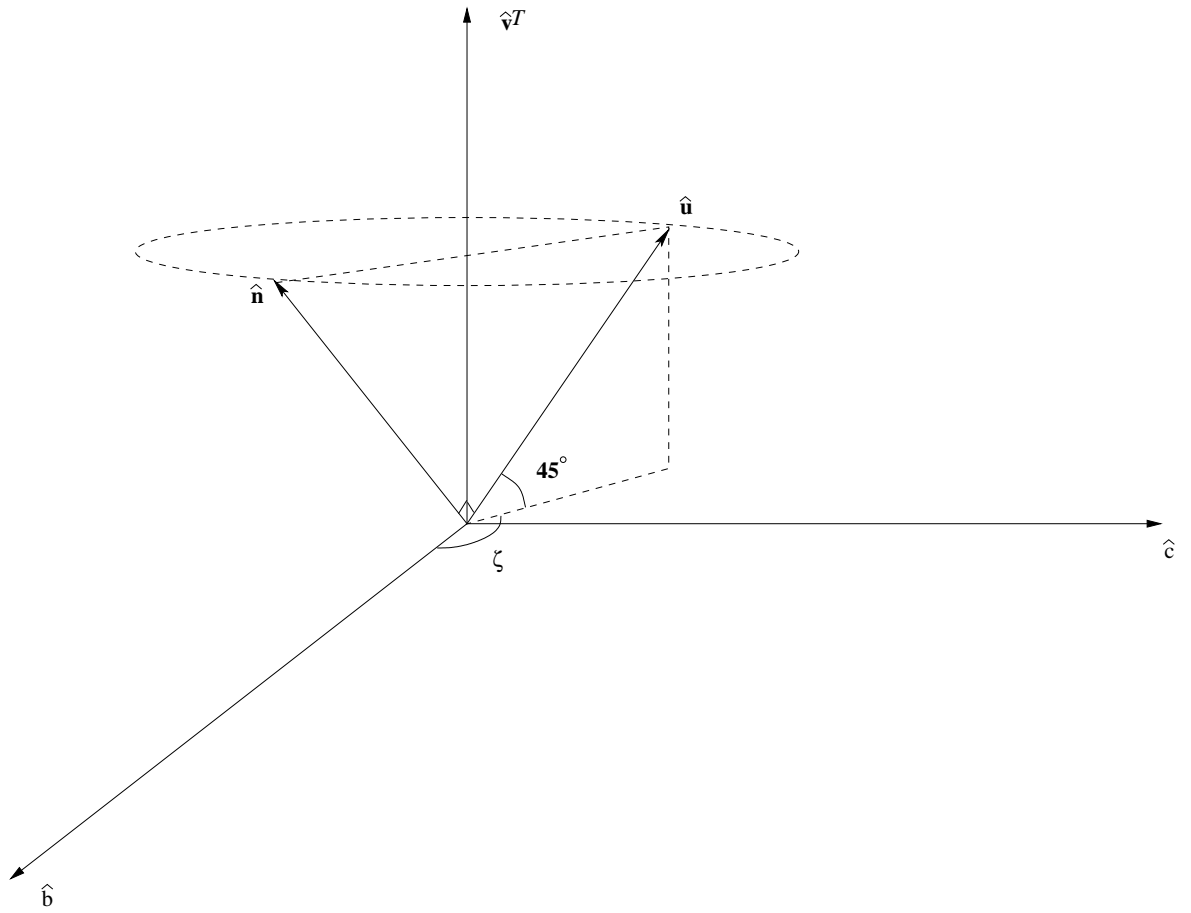


Figure 3.3: Figure showing how vectors  $\hat{u}$  and  $\hat{n}$  correspond to the T-axis. For a given T-axis,  $\hat{u}$  and  $\hat{n}$  must stay locked at right angles, but can swing around as defined by an angle  $\zeta$  from the b-axis in the  $\mathbb{T}$  coordinate system.



# Chapter 4

## Computing

This is a computationally intensive project in which several different software packages are utilised. This chapter outlines this software and describes how each package is used.

### 4.1 R

We have chosen to implement the models in this project using the computer package R (R Development Core Team 2006). R is easy to use, and flexible due to its extensive range of user-submitted packages and powerful graphical capabilities.

### 4.2 NonLinLoc

Evaluation of Equations 3.21 and 3.25 requires a sample of hypocentre locations with their corresponding take-off angles for each station. Non-Linear Location, or NonLinLoc, is a software package used “for velocity model construction, travel-time calculation and probabilistic, non-linear, global-search earthquake location in 3D structures, and for visualisation of 3D volume data and location results” (Lomax 2007). NonLinLoc will be used in this project to give estimates of the take-off parameters for a sample of possible  $\{\mathbf{x}_j\}$  hypocentre positions in a Cartesian coordinate system, and their posterior probabilities  $P(\{\mathbf{x}_j\}|\{t_i\})$  defined by Equation 3.10.

#### 4.2.1 Running NonLinLoc

##### Input data

NonLinLoc requires input phase data in one of a number of specific formats. In this project, we use phase data obtained from <http://www.geonet.org.nz>, and convert it to NonLinLoc Phase file format. GeoNet uses quality codes to describe the uncertainty of each arrival time pick — from 0 for the clearest picks, to 4 for the noisiest picks (Clarke 2007). As the NonLinLoc Phase file format requires arrival time errors, we convert these quality codes into values for  $\sigma_{t_i}$  as shown in Table 4.1. The noisiest picks are given an error of 9999, which gives zero weight to that particular phase. Station location data obtained from GeoNet is also formatted to fit the format required by NonLinLoc.

##### Control File

NonLinLoc includes a highly customisable control file, with a variety of user-specified options for running the earthquake location program. Below is a list of noteworthy commands. The syntax involves a keyword followed by one or more parameters.

Table 4.1: GeoNet qualities and arrival time errors  $\sigma_{t_i}$

GeoNet quality	$\sigma_{t_i}$ (sec)
0	0.1
1	0.5
2	1
3	2
4	9999

- **TRANS** — Set to **SIMPLE**, NonLinLoc will transform geographic coordinates to Cartesian coordinates. The **SIMPLE** mode also means NonLinLoc will assume a flat earth, rectangular, left-handed,  $(x, y, z)$  coordinate system (positive  $x$  = East, positive  $y$  = North, positive  $z$  = down). The parameters define the geographic coordinates that will be taken to be the origin of the Cartesian grid (Lomax 2007).
- **VGGRID** — Specifies the dimensions of the velocity model grid.
- **GTFILES** — Specifies input (velocity model) and output (time and take-off angle data) files' names and locations for use with the program **Grid2Time** (see Section 4.2.2). Also specifies wave type (P or S).
- **GTMODE** — Specifies whether the velocity model is 2D or 3D, and whether take-off angles are to be stored.
- **GTSRCE** — Specifies station names and locations.
- **LOCFILES** — Specifies the input (earthquake phase data, and time data from **Grid2Time**) and output (location data) files' names and locations for use with the program **NLLoc**.
- **LOCSEARCH** — Defines the search method - either a Grid-Search, a stochastic Metropolis-Gibbs sampling approach, or the oct-tree importance sampling algorithm.
- **LOCGRID** — Defines the size, origin and distance between nodes of the grid.
- **LOCMETH** — Specifies the location method - we set it to the inversion method of Tarantola & Valette (1982) described in Section 3.2.2, whether to use an S-wave velocity model in the hypocentre location routine, and allows various data quality controls to be implemented.
- **LOCDELAY** — Specifies station corrections. Only used in the velocity model unknown case (Section 5.2).
- **LOCGAU** — Specifies the theoretical covariance matrix  $\mathbf{C}_T$ , of the form  $\mathbf{C}_T = \sigma_T \exp\left(-\frac{1D_{ij}^2}{2\Delta}\right)$  where  $D_{ij}$  is the distance between stations  $i$  and  $j$ ,  $\sigma_T$  the theoretical arrival time error, and  $\Delta$  is the correlation length that controls covariance between stations. Here we use NonLinLoc default values of  $\sigma_T = 0.2$  and  $\Delta = 1$ .

For a more in depth overview of all the commands in the control file, refer to the NonLinLoc website (Lomax 2007).

## 4.2.2 Programs

The NonLinLoc package includes a range of different programs. The three programs we are interested in are **Vel2Grid**, **Grid2Time** and **NLLoc**.



Vel2Grid converts velocity model specifications into a 3D grid file in binary format, for use with Grid2Time and NLLoc.

Grid2Time calculates the travel-times between a station and all nodes of an  $x, y, z$  spatial grid — the velocity model grid — using the Eikonal finite-difference scheme of Podvin & Lecomte (1991). Grid2Time can also calculate the take-off angles ( $\phi, \theta$ ) for each point in the grid for each station (i.e. for each node,  $x, y, z$ , on the grid, the take-off angles for a ray leaving a theoretical earthquake at that point to a station,  $s_i$ , are calculated). This is done by analysing the gradients of the travel-times along each axis ( $x, y$  and  $z$ ) at the point, to create a vector gradient of travel-times. The direction opposite to this vector determines the azimuth and take-off angle (Lomax 2007).

A quality factor is also determined for each set of take-off angles. Essentially, the quality of the take-off angle determination will be low if there may be two rays that arrive at the station almost simultaneously, making it difficult to determine which ray's take-off angles to report (Lomax 2007).

Grid2Time is run with the relevant velocity model as the input file. This outputs two files for each station: the travel-time grid file, and the angles grid file. P-wave ( $V_p$ ) or both P- and S-wave ( $V_s$ ) velocity models can be used as inputs.

Following this, the NLLoc program is run. NLLoc uses Tarantola & Valette (1982)'s Bayesian method of calculating the posterior PDF of the hypocentre location outlined in Section 3.2.2. We then specify one of three techniques available in NLLoc to search the posterior PDF for the maximum likelihood hypocentre location: a systematic Grid-Search, a Metropolis-Gibbs algorithm, or an Oct-tree importance sampling algorithm. We use the Oct-tree search in this project. The advantages of this method are that it is much faster than the grid search method, more global and complete than the Metropolis sampling algorithm, and only requires us to specify the initial grid size and the number of samples to be taken (Lomax 2007).

The Oct-tree sampling method is started by defining a coarse grid on which to search. The probability at the centre of each grid cell is determined from the posterior PDF, Equation 3.10, and is multiplied by the volume of the cell to give the probability that the hypocentre is anywhere inside that cell. The probabilities are ordered in a list  $L_p$ . The algorithm then runs as follows:

- Identify the cell with the largest probability from  $L_p$ . It is then divided into 8 new smaller cells;
- The probability is calculated for these 8 cells;
- The list  $L_p$  is updated by inserting these new probabilities.

These steps are then repeated until either the maximum specified number of nodes to process has been reached, or the smallest specified node size is reached (Lomax 2007). Samples are then drawn from the Oct-tree structure to give a sample from  $P(\mathbf{x}|\{t_i\}, \omega)$ .

The NonLinLoc input files are the travel-time grid files generated by Grid2Time, and the earthquake phase data provided by GeoNet. This outputs, among other results, estimates of the PDF for a sample  $\{\mathbf{x}_j\}$  of possible  $(x, y, z)$  hypocentre locations in an output file known as a scatter file. Unfortunately, the scatter file does not include take-off angle data for each of these hypocentre locations. In the following section, we show how to interpolate this information from the angles grid file that was output by the Grid2Time program.

### 4.2.3 Obtaining Take-off Parameters from NonLinLoc Output

The azimuth, take-off angle and a quality number for each station for a theoretical earthquake occurring at each node on the velocity model grid are generated by the program `Grid2Time`, and are stored in a binary file, i.e. for each point on a spatial grid the azimuth and take-off angle are tabulated. The sample of hypocentre locations  $\{\mathbf{x}_j\}$  output by `NLLoc` do not coincide in general with the points of this grid, so we must interpolate to obtain an azimuth and take-off angle for each sampled hypocentre, in order to define the take-off vector (Equation 3.20).

Only some stations in the data have P-wave polarities ( $Y_i = \pm 1$ , see Section 3.3.1) recorded. While we use all stations for the earthquake location method carried out by `NonLinLoc`, we only interpolate the take-off parameters for stations that have polarity data. As P-wave polarity data  $Y_i$  is required in the formulation of our posterior PDF, we need not calculate take-off parameters for stations without polarities. Additionally, in this project, S-wave information is only used for hypocentre location purposes, and therefore we need not obtain S-wave take-off parameters.

Azimuth is stored as a 16 bit integer, 0 to 3600, in tenths of degrees, measured clockwise from north. Take-off angle is stored as a 12 bit integer, 0 (down) to 1800 (up), also in tenths of degrees. Quality number is a 4 bit integer, indicating 0 (low quality) to 10 (high quality). We wish to interpolate linearly the azimuth, take-off angle and quality number from the nodes of the velocity model grid to the sampled hypocentre locations  $\{\mathbf{x}_j\}$  given in the scatter file.

Initially, to make the binary files for each station easier to read into R, they are converted to text files using a C program written by Richard Arnold, `rbuf2txt`. We then use R to carry out the linear interpolation, weighting each azimuth and dip by its quality number, as described below.

The azimuth and take-off angle at a hypocentre location are calculated by a weighted average of the azimuth and take-off angle values at the 8 nodes of the velocity model grid that surround that location. To calculate the weight that each of the 8 points has, three values,  $u$ ,  $v$  and  $w$ , are calculated:

$$u = \frac{x - x_0}{x_1 - x_0} \quad v = \frac{y - y_0}{y_1 - y_0} \quad w = \frac{z - z_0}{z_1 - z_0}$$

Where  $(x, y, z)$  are the coordinates of the hypocentre location,  $(x_0, y_0, z_0)$  are the coordinates of the node of the velocity grid closest to the origin of the entire grid, and  $(x_1, y_1, z_1)$  are the coordinates of the node of the velocity grid furthest from the origin. Essentially,  $u$  gives the proportion of the distance traveled between adjacent  $x$  nodes of the velocity model grid when the hypocentre location is encountered, and similarly for  $v$  and  $w$ . This is displayed in figure 4.1.

Since interpolation is a form of weighted average, and we are dealing with angles, we adopt the averaging approach outlined in Appendix A.4. Thus,

$$\phi_{ij} = \tan^{-1} \left( \frac{\langle \sin \phi \rangle_{ij}}{\langle \cos \phi \rangle_{ij}} \right)$$

gives the relevant interpolated azimuth value for station  $i$  at hypocentre location  $j$ , where

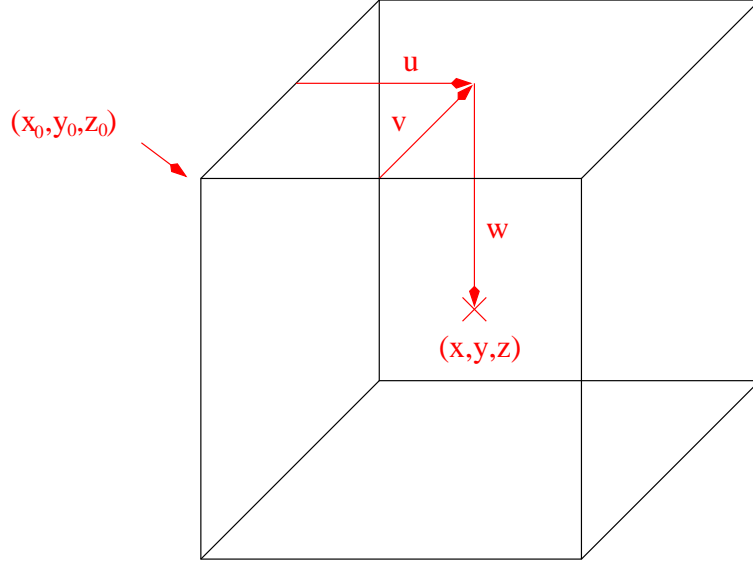


Figure 4.1: Three-dimensional linear interpolation:  $(x, y, z)$  marks the hypocentre location we wish to interpolate to, while  $u$ ,  $v$  and  $w$  give the proportion of the distance traveled between adjacent nodes of the velocity model grid where the hypocentre location is encountered.

$\langle \sin \phi \rangle_{ij}$  and  $\langle \cos \phi \rangle_{ij}$  are given by

$$\langle \sin \phi \rangle_{ij} = \frac{\sum_{k=1}^8 weight_{ik} \sin(\phi_{ik})}{\sum_{k=1}^8 weight_{ik}}$$

$$\langle \cos \phi \rangle_{ij} = \frac{\sum_{k=1}^8 weight_{ik} \cos(\phi_{ik})}{\sum_{k=1}^8 weight_{ik}}.$$

where  $\phi_{ik}$  is the azimuth for station  $i$  at the  $k^{th}$  of the surrounding 8 nodes of the velocity model grid surrounding location  $\mathbf{x}_j$ , and  $weight$  is given by:

$$\begin{aligned} \sum_{k=1}^8 weight_{ik} = & (1-u)(1-v)(1-w)qual_{i1} + (1-u)(1-v)w \cdot qual_{i2} \\ & + (1-u)v(1-w)qual_{i3} + u(1-v)(1-w)qual_{i4} \\ & + (1-u)v \cdot w \cdot qual_{i5} + u(1-v)w \cdot qual_{i6} \\ & + u \cdot v(1-w)qual_{i7} + u \cdot v \cdot w \cdot qual_{i8} \end{aligned} \quad (4.1)$$

Further,

$$\theta_{ij} = \cos^{-1} \left( \frac{\langle \cos \theta \rangle_{ij}}{\sqrt{\langle \sin \theta \cos \phi \rangle_{ij}^2 + \langle \sin \theta \sin \phi \rangle_{ij}^2 + \langle \cos \theta \rangle_{ij}^2}} \right)$$

gives the interpolated take-off angle for station  $i$  at hypocentre location  $j$ , where

$$\langle \cos \theta \rangle_{ij} = \frac{\sum_{k=1}^8 weight_{ik} \cos(\theta_{ik})}{\sum_{k=1}^8 weight_{ik}}$$

$$\langle \sin \theta \cos \phi \rangle_{ij} = \frac{\sum_{k=1}^8 weight_{ik} \sin(\theta_{ik}) \cos(\phi_{ik})}{\sum_{k=1}^8 weight_{ik}}$$

$$\langle \sin \theta \sin \phi \rangle_{ij} = \frac{\sum_{k=1}^8 weight_{ik} \sin(\theta_{ik}) \sin(\phi_{ik})}{\sum_{k=1}^8 weight_{ik}}$$

where  $\phi_{ik}$  and  $\theta_{ik}$  are the azimuth and take-off angle respectively for station  $i$  at the  $k^{th}$  of the surrounding 8 nodes of the velocity model grid, and *weight* is given in Equation 4.1.

Once the interpolation is carried out we have the polarity  $Y_i$  for station  $i$ , azimuth  $\phi_{ij}$  and take-off angle  $\theta_{ij}$  for station  $i$  and sampled hypocentre location  $\mathbf{x}_j$ , along with the hypocentre location PDF  $P(\mathbf{x}_j|\{t_i\})$ . This leaves us with all the information needed to construct the Bayesian posterior PDFs (Equations 3.21 and 3.25), as shown in Table 4.2.

Table 4.2: Sources of each element of Equation 3.21 and 3.25

Parameter/Value	Description	Source
$\{Y_i\}$	Polarities at stations	Data
$\{s_i\}$	Location of stations	Data
$\{\mathbf{x}_j\}$	Sample of hypocentres	NonLinLoc
$P(\mathbf{x}_j \{t_i\}, \omega)$	Hypocentre PDF	NonLinLoc
$\hat{\mathbf{p}}_{ij} / \hat{\mathbf{p}}_{ijk}$	Ray take-off vector	NonLinLoc
$\sigma_a$	Amplitude noise	User defined
$\pi'_p$	Probability of cross wiring	User defined

### 4.3 Velest

Velest is a program that uses a *coupled hypocentre-velocity model* method to determine *minimum 1-dimensional velocity models*. Velest is used in this project to obtain a random sample of velocity models from a given prior  $P(v)$ , as is required in Equation 3.25.

The coupled hypocentre-velocity model method is similar to Tarantola & Valette (1982)'s Bayesian method of hypocentre location described in Section 3.2.2 and implemented in NonLinLoc. The difference is that here it is assumed the velocity model is unknown to begin with, and a solution for the velocity model is solved simultaneously with hypocentre locations (Kissling 1988).

An indication of the quality of a velocity model can be given by the difference  $t_i^{res}$  between the observed and theoretical arrival times at station  $i$

$$t_i^{res} = t_i^{obs} - t_i^{cal} \quad (4.2)$$

where the theoretical arrival times  $t_i^{cal} = t_i^{cal}(\hat{\mathbf{x}}, \hat{\mathbf{T}}, \hat{v}, y_i)$ , depend on an estimated hypocentre location  $\hat{\mathbf{x}}$ , estimated origin time  $\hat{\mathbf{T}}$ , the implemented velocity model  $\hat{v}$  and station locations  $y_i$ . The observed arrival times  $t_i^{obs} = t_i^{obs}(\mathbf{x}_0, \mathbf{T}_0, v_0, y_i)$  depend on the true hypocentre location  $\mathbf{x}_0$ , true origin time  $\mathbf{T}_0$ , the true velocity model  $v_0$  and station locations  $y_i$  (Kissling 1988).

Velest takes an initial input velocity model and hypocentre locations and uses this to calculate arrival times  $t_i^{cal}$ . The program then adjusts hypocentral and velocity model parameters. To do this, a relationship between  $t_i^{res}$  and the required adjustments is established. A first order Taylor series expansion of  $t_i^{obs}$  about the estimated parameters ( $\mathbf{x}_0 = \hat{\mathbf{x}}, \mathbf{T}_0 = \hat{\mathbf{T}}, v_0 =$

$\hat{v}, y_i = y_i)$  gives

$$\begin{aligned} t_i^{obs}(\mathbf{x}_0, \mathbf{T}_0, v_0, y_i) &= t_i^{obs}(\hat{\mathbf{x}}, \hat{\mathbf{T}}, \hat{v}, y_i) + \frac{\partial t_i}{\partial \mathbf{x}}(\mathbf{x}_0 - \hat{\mathbf{x}}) + \frac{\partial t_i}{\partial \mathbf{T}}(\mathbf{T}_0 - \hat{\mathbf{T}}) + \frac{\partial t_i}{\partial v}(v_0 - \hat{v}) \\ &\quad + \frac{\partial t_i}{\partial y_i}(y_i - y_i) \\ &= t_i^{obs}(\hat{\mathbf{x}}, \hat{\mathbf{T}}, \hat{v}, y_i) + \frac{\partial t_i}{\partial \mathbf{x}}(\mathbf{x}_0 - \hat{\mathbf{x}}) + \frac{\partial t_i}{\partial \mathbf{T}}(\mathbf{T}_0 - \hat{\mathbf{T}}) + \frac{\partial t_i}{\partial v}(v_0 - \hat{v}) \end{aligned}$$

Substituting this into Equation 4.2 gives

$$\begin{aligned} t_i^{res} &= t_i^{obs}(\hat{\mathbf{x}}, \hat{\mathbf{T}}, \hat{v}, y_i) + \frac{\partial t_i}{\partial \mathbf{x}}(\mathbf{x}_0 - \hat{\mathbf{x}}) + \frac{\partial t_i}{\partial \mathbf{T}}(\mathbf{T}_0 - \hat{\mathbf{T}}) + \frac{\partial t_i}{\partial v}(v_0 - \hat{v}) - t_i^{cal}(\hat{\mathbf{x}}, \hat{\mathbf{T}}, \hat{v}, y_i) \\ &= \frac{\partial t_i}{\partial \mathbf{x}}(\mathbf{x}_0 - \hat{\mathbf{x}}) + \frac{\partial t_i}{\partial \mathbf{T}}(\mathbf{T}_0 - \hat{\mathbf{T}}) + \frac{\partial t_i}{\partial v}(v_0 - \hat{v}) \quad (\text{since } t_i^{obs}(\hat{\mathbf{x}}, \hat{\mathbf{T}}, \hat{v}, y_i) = t_i^{cal}(\hat{\mathbf{x}}, \hat{\mathbf{T}}, \hat{v}, y_i)) \\ &= \frac{\partial t_i}{\partial \mathbf{x}}\Delta\mathbf{x} + \frac{\partial t_i}{\partial \mathbf{T}}\Delta\mathbf{T} + \frac{\partial t_i}{\partial v}\Delta v \end{aligned} \tag{4.3}$$

where  $\Delta\mathbf{x}$  is the required adjustment in estimated hypocentre location,  $\Delta\mathbf{T}$  is the adjustment in estimated origin time, and  $\Delta v$  is the adjustment in the velocity model (Kissling 1988, Kissling, Ellsworth, Eberhart-Phillips & Kradolfer 1994).

The minimum 1D velocity model is the velocity model with minimum root mean square (RMS) misfit of  $\{t_i^{res}\}$ , where RMS is defined as

$$\text{RMS}(t_i^{res}) = \sqrt{\frac{1}{n} \sum_i (t_i^{res})^2}$$

This minimum 1D velocity model is given by solving Equation 4.3. This equation is non-linear, and hence is solved numerically by Velest (Kissling 1988).

The iterative procedure of Velest is as follows:

1. Solve the coupled hypocentre-velocity model problem for the estimated hypocentres and velocity model. This gives adjusted hypocentres and an adjusted velocity model;
2. Recalculate  $t_i^{cal}$  based on these adjustments;
3. Check the RMS of the new velocity model — if it is lower, retain it. Return to 1.

Due to the non-linearity of the coupled hypocentre-velocity method, multiple local RMS minima may occur over the velocity model solution space. Different input models may find different local minima. A priori it is unknown where these minima occur so a number of Velest runs are conducted using a variety of different input velocity models, resulting in a set of minimum 1D velocity models (Kissling 1995, Clarke 2007). Further specific details on how Velest was run in this project are given in Section 5.2.2.

## 4.4 Grid Computing

As mentioned in Section 3.3.2, the VMU posterior PDF (Equation 3.25) is calculated using Monte Carlo integration; summing over  $p$  different VMK posterior PDFs. Calculation of this is particularly computationally intensive given the large sample of velocity models used in this project ( $p \simeq 1000$ ). Running the required programs and models on one machine would take several days. To reduce this computation time we make use of the School of Mathematics, Statistics and Computer Science's computational grid, which comprises approximately

170 NetBSD (Unix) workstations.

The grid is particularly useful for multiple runs of the same program with differing parameter values, as is the case here. We break the job down into sets of 10 velocity models, and have each computer on the grid evaluate the VMK posterior PDF for its set of 10 models. We then retrieve each VMK posterior PDF from the grid and sum over all  $\sim 1000$  to obtain the VMU posterior PDF. This reduces the computation time from days to hours.

# Chapter 5

## Applications

In this chapter we discuss the application of our Bayesian methods of focal mechanism estimation to earthquake phase data from New Zealand. We consider two cases; the velocity model known case, with data from the Raukumara Peninsula, and the velocity model unknown case, with data from Kawerau.

### 5.1 Velocity model known — Raukumara Peninsula

We use data from the Raukumara Peninsula, New Zealand, to illustrate the case in which we presume that the uncertainties in the hypocentre location are caused solely by P-wave arrival timing errors, and that the velocity model is error-free and known. The Raukumara Peninsula is particularly suitable to study this objective, as the 3-dimensional  $V_p$  (P-wave velocity) and  $V_s$  (S-wave velocity) structures of the area have been determined by Reyners et al. (1999).

#### 5.1.1 Tectonic Setting

New Zealand lies at the boundary of the Pacific and Australian tectonic plates. To the east of the North Island, the Pacific plate subducts beneath the overlying Australian plate. The two plates converge at approximately 45 mm/yr in the region of interest; the Raukumara Peninsula, on the East Cape of the North Island of New Zealand. The plate interface occurs at a depth of approximately 15 km beneath the east of the Raukumara Peninsula (Reyners et al. 1999).

The Raukumara Peninsula (see Figure 5.1) lies 300 km southwest of the Tonga-Kermadec and Hikurangi subduction zone junction. At this junction, crust to the north experiences subduction along the Kermadec Trench, while to the south the subduction is influenced by the Hikurangi Plateau (Reyners & McGinty 1999).

#### 5.1.2 Velocity Model

The velocity model we use here is based on a 3D velocity model obtained by Reyners et al. (1999), who deployed 36 seismographs over the Raukumara Peninsula between July and December 1994. This data enabled Reyners et al. (1999) to determine the  $V_p$  and  $V_s$  structure of the region.

The velocity model spans an irregular grid rotated  $36^\circ$  east relative to north (see Figure 5.1). It has 13  $x$ -nodes over a distance of 130 km, 11  $y$ -nodes over 200 km, and 9  $z$ -nodes to a depth of 100 km. These nodes are at non-constant distances, while the format for use with NonLinLoc requires constant grid spacing. Hence, we carry out a linear interpolation of the velocities in Reyners et al. (1999)’s model, to obtain velocities at a constant 1km grid

spacing over the entire volume. We use both the  $V_p$  and  $V_s$  velocity models for the earthquake location routine carried out by NonLinLoc. Although we do not use S-wave data in the formulation of our posterior PDFs, the  $V_s$  velocity model is used to better constrain the hypocentre location.

The  $V_p$  velocity structure is shown in Figure 5.2. Velocity generally increases with depth. It can be seen that at shallow depths ( $< 20$  km)  $V_p$  is lower in the southeast, and higher in the northwest of the model. The subducted plate dips towards the northwest, meaning the plate interface is shallower in the southeast, resulting in higher  $V_p$  in the southeast at depths  $> 20$  km (i.e. in the mantle just below the interface) (Reyners et al. 1999). The  $V_s$  velocity structure is shown in Figure 5.3.

### 5.1.3 Data

We use GeoNet phase data in the region for the period from 1 January 1990 to 30 September 2005. Polarity data were provided by Reyners & McGinty (1999) who re-analysed earthquakes that occurred during the July to December 1994 period, and read many more first motions than were obtained by routine processing of the GeoNet data. These first motions were matched to the equivalent earthquakes in the GeoNet phase data catalogue. Station location data were provided by Martin Reyners (for the locations of the portable seismographs used in the Reyners & McGinty (1999) and Reyners et al. (1999) studies) and GeoNet (for the locations of the permanent stations in the area).

It was decided, on the basis of discussion with Victoria University of Wellington and GNS Science staff, that for an event to be worth analysing, a minimum of 10 stations with P-wave polarity observations were required. Any fewer than 10 polarities would mean that the focal mechanism of the event would be poorly constrained. There were 193 earthquakes with  $\geq 10$  polarities.

Reyners & McGinty (1999) calculated 117 focal mechanism solutions for Raukumara earthquakes. Those solutions provide an opportunity to compare our MAP estimates to an established method of focal mechanism estimation. Thus we only select here events for which a solution was obtained by Reyners & McGinty (1999).

Using these criteria, 87 earthquakes were selected for analysis. All 87 events occurred between July and December 1994, the period for which Reyners & McGinty (1999) re-analysed polarity readings. Appendix B summarises the 87 Raukumara earthquakes. Hypocentre location and origin time are as calculated by NonLinLoc, while magnitudes are taken from the GeoNet catalogue.

### 5.1.4 Posterior PDF Particulars

Recall that the posterior probability for the velocity model known case is given by Equation 3.21:

$$P(\Theta|\mathbf{d}, \omega) \propto P(\Theta) \iint \sum_{j=1}^m \left[ \prod_{i=1}^n \pi_{ij}^{\frac{1}{2}(1+y_i)} (1 - \pi_{ij})^{\frac{1}{2}(1-y_i)} \right] P(\sigma_a) P(\pi_p) d\sigma_a d\pi_p$$

We calculate this posterior PDF for all 87 events, under the following conditions:

- a uniform prior on  $\Theta$ :  $P(\Theta) \propto \sin \theta \Leftrightarrow P(R(\Theta)) \propto 1$



- $P(\sigma_a) = \delta(\sigma_a - \sigma_{a_0})$ , and  $P(\pi_p) = \delta(\pi_p - \pi_{p_0})$  where  $\delta$  denotes the Dirac delta function and  $\sigma_{a_0}$  and  $\pi_{p_0}$  are values for  $\sigma_a$  and  $\pi_p$ , assumed to be known. Due to the properties of the delta function (see Appendix A.9), the posterior PDF becomes

$$P(\Theta|\mathbf{d}, \omega) \propto P(\Theta) \sum_{j=1}^m \left[ \prod_{i=1}^n \pi_{ij}^{\frac{1}{2}(1+y_i)} (1 - \pi_{ij})^{\frac{1}{2}(1-y_i)} \right]$$

where  $\pi_{ij}$  is given by

$$\pi_{ij} = \pi'_{p_0} + (1 - 2\pi'_{p_0}) \Phi \left( \frac{2(\hat{\mathbf{p}}_{ij} \cdot \hat{\mathbf{n}})(\hat{\mathbf{p}}_{ij} \cdot \hat{\mathbf{u}})}{\sigma_{a_0}} \right)$$

This approach is equivalent to taking fixed values for  $\sigma_a$  and  $\pi_p$ . For this to be valid we require appropriate values for these parameters.

Hardebeck & Shearer (2002) found that around 20% of ambiguously determined polarities were inconsistent. Thus we take a (conservative) value of  $\pi'_{p_0} = 0.2$ .

For the amplitude noise  $\sigma_a$  we take a value of  $\sigma_{a_0} = \frac{1}{6}$ , based on values in Zollo & Bernard (1991) and Brillinger et al. (1980).

### 5.1.5 Results

Here we present results for nine of the Raukumara events, selected to show three poorly constrained solutions, three intermediately constrained solutions and three well constrained solutions. The determinant of the concentration matrix  $\hat{\mathbf{D}}$  is used as a measure of constraint. For each event, identified by its unique CUSP id, we present:

- Hypocentre summary information according to NonLinLoc.
- Stereonet showing P-wave first motion polarities, along with the beachball corresponding to the MAP estimate for  $\Theta$  based on our Bayesian posterior PDF, and the beachball corresponding to the solution found by Reyners & McGinty (1999).
- Posterior PDF of P- and T-axes.
- The estimates of the parameters of both the full and scalar concentration 8-Mode Matrix Fisher distributions that provide the best fit to our posterior PDF.
- Marginal plots of the posterior PDF  $P(\Theta|\mathbf{d}, \omega)$ , and both the full and scalar concentration 8-Mode Matrix Fisher distributions.

Table 5.1 contains a summary of this information for all nine events. A common method of comparing two focal mechanism solutions is the *angular difference*,  $a$ , between the two rotation matrices that define the solutions (see e.g. Hardebeck & Shearer 2002, Arnold & Townend 2007, Kagan 2007). Section A.1.3 describes the calculation of the angular difference. In Table 5.1  $\sigma_\Theta$  is given by Equation 2.5.

Table 5.1: Summary table of results for the selected Raukumara events.

	CUSPID	MAP $\hat{\Theta}$	$\hat{\kappa}$	$\sigma_{\Theta}$	Reyners & McGinty $\hat{\Theta}$	$a$
1.	646630	(22.62°, 71.94°, 250.96°)	2.66	28.86°	(353.43°, 82.50°, 202.32°)	51.40°
2.	672060	(229.32°, 68.49°, 296.00°)	2.55	29.55°	(209.60°, 67.27°, 277.35°)	21.40°
3.	668273	(343.51°, 74.54°, 233.53°)	2.75	28.29°	(23.23°, 60.28°, 232.59°)	42.45°
4.	640980	(24.46°, 80.72°, 72.64°)	3.52	24.60°	(191.71°, 103.94°, 305.19°)	20.23°
5.	636036	(124.13°, 72.01°, 12.88°)	3.63	24.17°	(308.00°, 90.00°, 337.20°)	20.59°
6.	635767	(253.18°, 29.80°, 141.65°)	3.85	23.39°	(38.04°, 127.58°, 254.31°)	32.77°
7.	669233	(128.39°, 59.81°, 19.72°)	6.13	17.98°	(202.83°, 105.35°, 217.61°)	6.95°
8.	665895	(165.52°, 81.30°, 240.41°)	6.55	17.32°	(178.24°, 82.20°, 240.75°)	12.71°
9.	675146	(154.11°, 72.14°, 134.26°)	6.50	17.40°	(348.02°, 96.42°, 218.02°)	18.38°

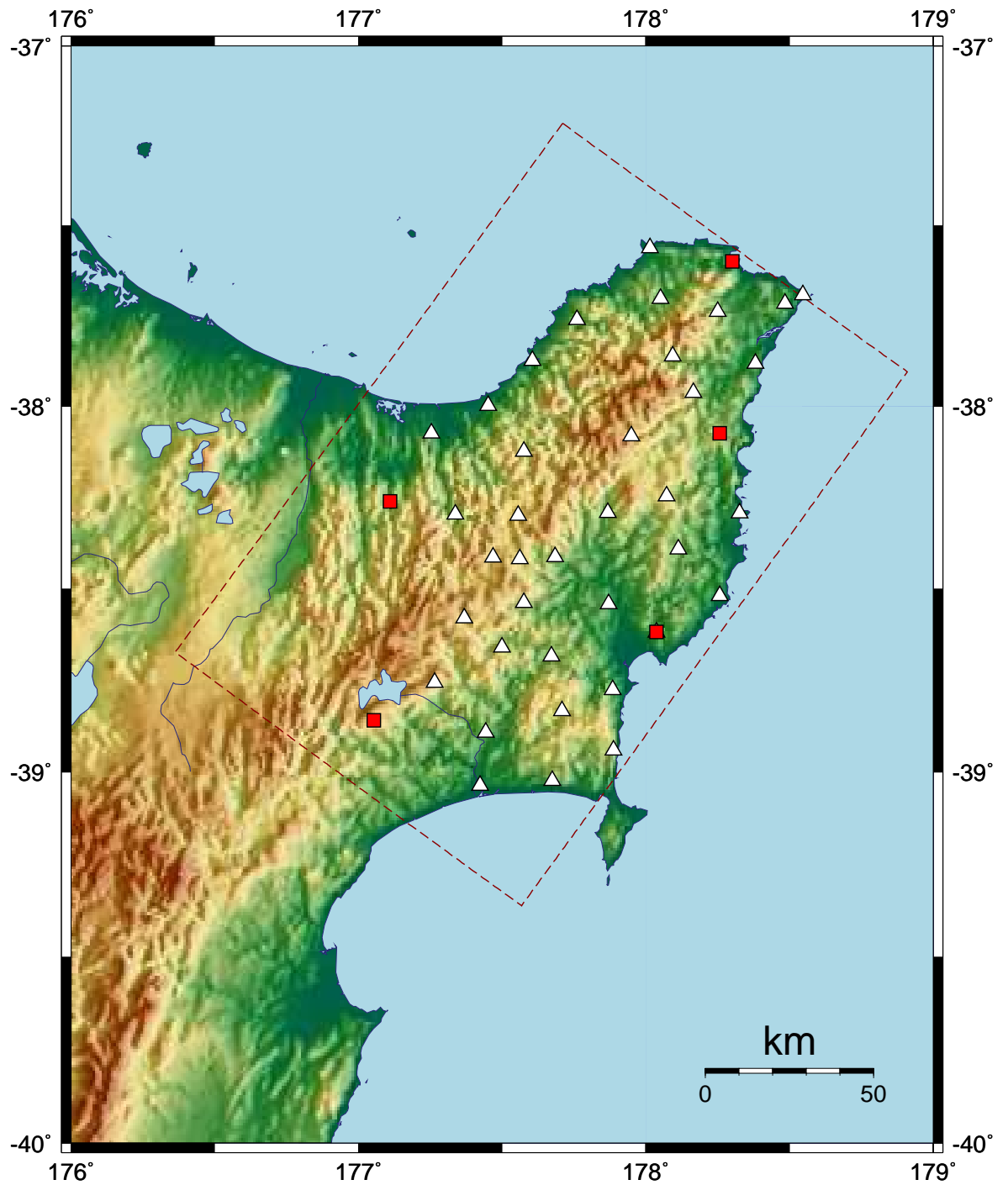


Figure 5.1: Map of the Raukumara Peninsula. Symbols show the boundary of the Reyners et al. (1999) velocity model (dark red line), temporary seismometers deployed by Reyners et al. (1999) (white triangles), and permanent seismometers within the velocity model bounds as at time of the Reyners et al. (1999) study (red squares).

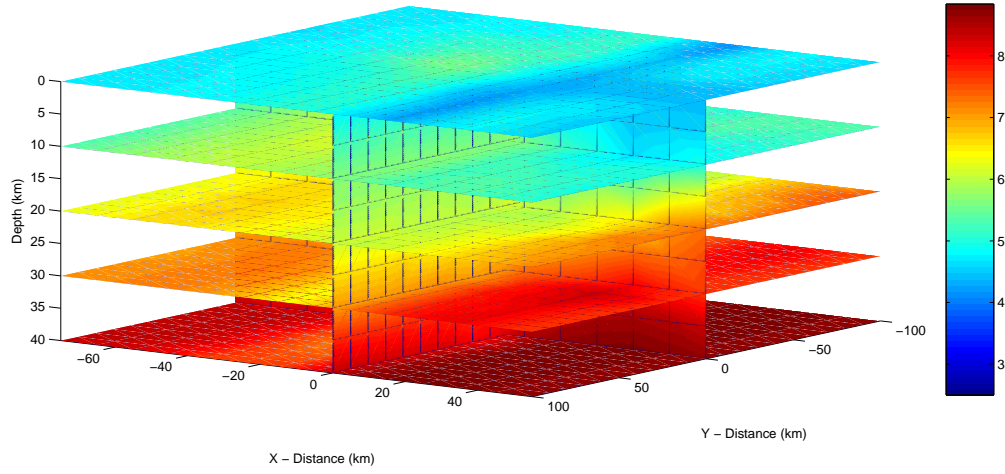


Figure 5.2: Cross section of the  $V_p$  velocity model for the Raukumara Peninsula by Reyners et al. (1999). The model is rotated  $36^\circ$  clockwise of north – this perspective is from the south. The  $x$ -axis is positive to the southeast, the  $y$ -axis positive to the southwest. The colours denote velocity in  $\text{km s}^{-1}$  according to the colour bar.

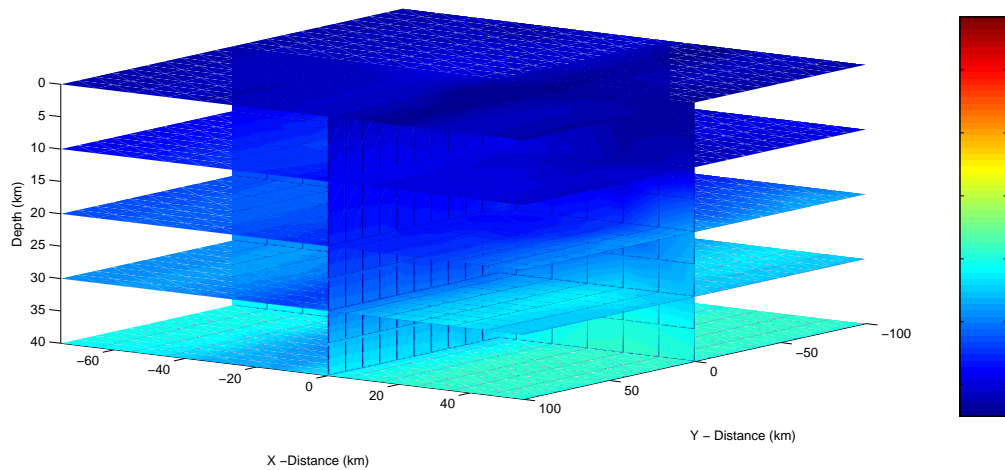


Figure 5.3: Cross section of the  $V_s$  velocity model for the Raukumara Peninsula by Reyners et al. (1999). The model is rotated  $36^\circ$  clockwise of north – this perspective is from the south. The  $x$ -axis is positive to the southeast, the  $y$ -axis positive to the southwest. The colours denote velocity in  $\text{km s}^{-1}$  according to the colour bar.

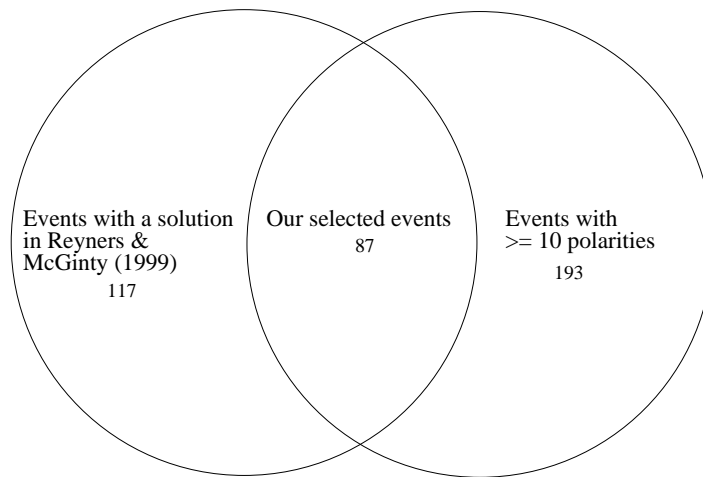


Figure 5.4: Diagram showing our event selection criteria. For an event to be selected, it must have  $\geq 10$  polarity readings, and must have a corresponding solution by Reyners & McGinty (1999).

## 1. EVENT 646630

Date	Time	Latitude	Longitude	Depth	Magnitude
01/09/1994	09:19:42.86	-38.467	178.098	25.879	2.8

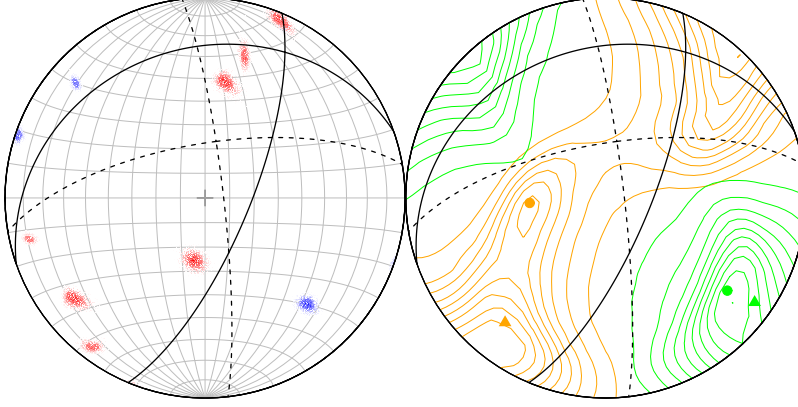


Figure 5.5: Stereonet for event 646630 (left). The solid dark line indicates the MAP focal mechanism solution  $[\Theta = (22.62^\circ, 71.94^\circ, 250.96^\circ)]$  while the dotted line indicates the solution given by Reyners & McGinty (1999). Angular difference between the two solutions =  $51.403^\circ$ . Blue points are compressions, red points are dilatations. PT contour plot for event 646630 (right). Orange denotes the P-axis, green the T-axis. The circle denotes the MAP estimate, while the triangle denotes the solution given by Reyners & McGinty (1999).

### Parameter Estimates

8-mode Matrix Fisher Distribution:

$$\hat{\mathbf{F}} = \begin{bmatrix} -7.748 & 4.199 & -1.669 \\ 2.443 & -0.853 & 4.232 \\ 1.352 & -0.603 & 0.629 \end{bmatrix}, \quad \hat{\mathbf{M}} = \begin{bmatrix} -0.787 & 0.614 & 0.058 \\ 0.061 & -0.016 & 0.998 \\ 0.613 & 0.789 & -0.025 \end{bmatrix},$$

$$\hat{\mathbf{U}} = \begin{bmatrix} 0.837 & -0.241 & -0.241 \\ -0.433 & 0.257 & -0.864 \\ 0.334 & 0.936 & 0.111 \end{bmatrix}, \quad \hat{\mathbf{D}} = \begin{bmatrix} 9.786 & 0 & 0 \\ 0 & 3.445 & 0 \\ 0 & 0 & 0.074 \end{bmatrix}$$

8-mode Matrix Fisher Distribution with scalar concentration:

$$\hat{\mathbf{M}} = \begin{bmatrix} -0.974 & 0.208 & 0.089 \\ 0.075 & -0.074 & 0.994 \\ 0.213 & 0.975 & -0.025 \end{bmatrix}, \quad \hat{\kappa} = 2.657$$

## 1. EVENT 646630

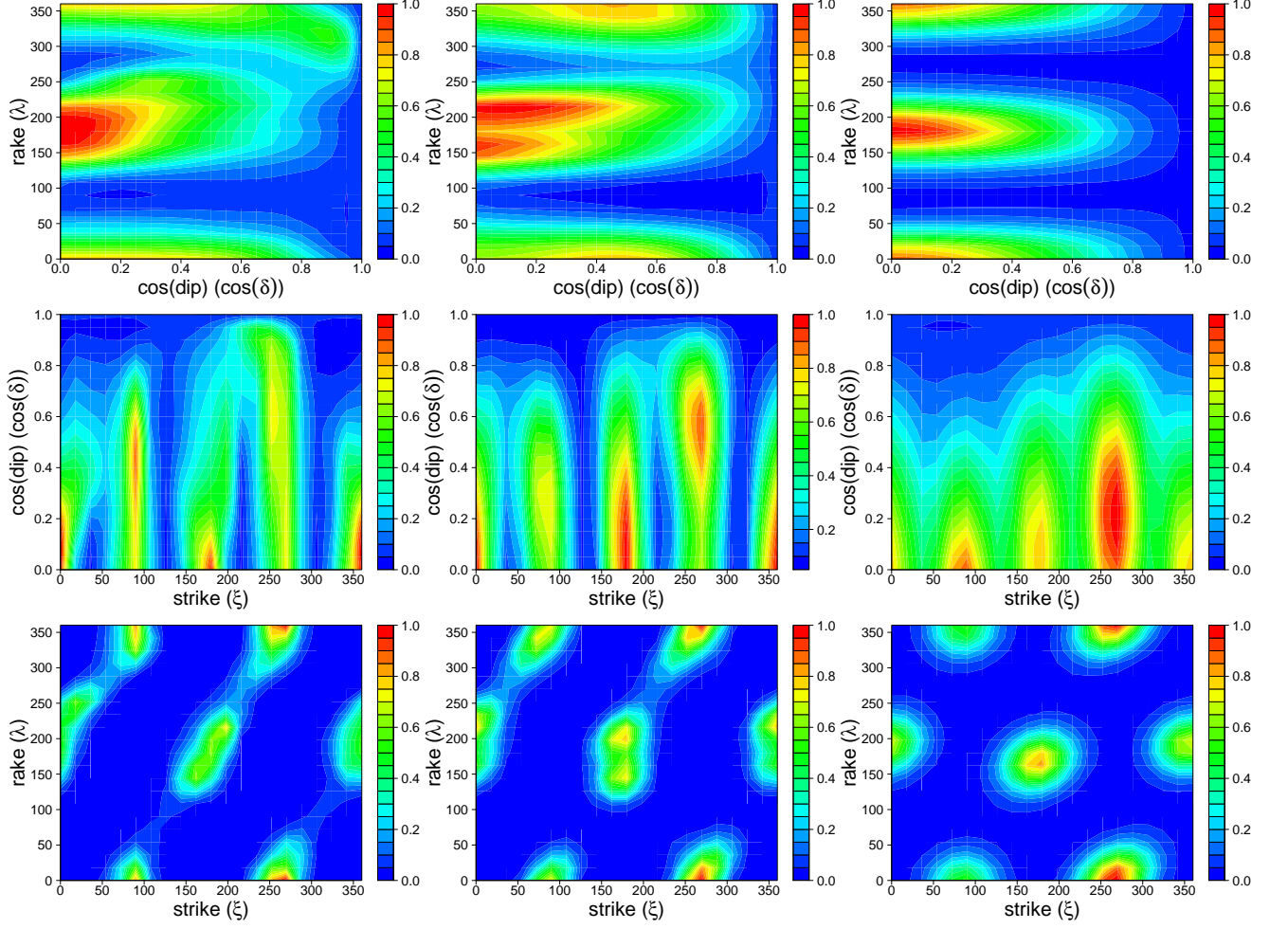


Figure 5.6: Marginal PDF plots for Bayesian posterior PDF  $P(\Theta|\mathbf{d}, \omega)$  (left), fitted 8-mode Matrix Fisher distribution  $P(\Theta|\hat{\mathbf{F}})$  (centre) and fitted 8-mode Matrix Fisher distribution with scalar concentration parameter  $P(\Theta|\hat{\mathbf{M}}, \hat{\kappa})$  (right) for event 646630.

## 2. EVENT 672060

Date	Time	Latitude	Longitude	Depth	Magnitude
16/10/1994	09:43:48.61	-37.834	178.378	10.449	2.9

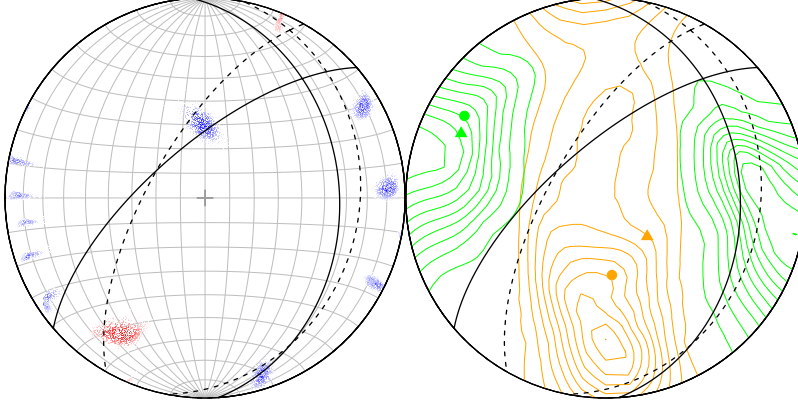


Figure 5.7: Stereonet for event 672060 (left). The solid dark line indicates the MAP focal mechanism solution  $[\Theta = (229.32^\circ, 68.49^\circ, 296.00^\circ)]$  while the dotted line indicates the solution given by Reyners & McGinty (1999). Angular difference between the two solutions =  $21.403^\circ$ . Blue points are compressions, red points are dilatations. PT contour plot for event 672060 (right). Orange denotes the P-axis, green the T-axis. The circle denotes the MAP estimate, while the triangle denotes the solution given by Reyners & McGinty (1999).

### Parameter Estimates

8-mode Matrix Fisher Distribution:

$$\hat{\mathbf{F}} = \begin{bmatrix} -2.902 & -2.207 & 0.858 \\ -9.267 & -5.909 & -6.192 \\ 1.499 & 0.807 & -0.870 \end{bmatrix}, \quad \hat{\mathbf{M}} = \begin{bmatrix} -0.149 & -0.780 & 0.607 \\ -0.677 & -0.367 & -0.638 \\ 0.721 & -0.506 & -0.474 \end{bmatrix},$$

$$\hat{\mathbf{U}} = \begin{bmatrix} -0.754 & -0.339 & -0.339 \\ -0.486 & -0.290 & -0.825 \\ -0.443 & 0.895 & -0.054 \end{bmatrix}, \quad \hat{\mathbf{D}} = \begin{bmatrix} 12.971 & 0 & 0 \\ 0 & 2.916 & 0 \\ 0 & 0 & 0.266 \end{bmatrix}$$

8-mode Matrix Fisher Distribution with scalar concentration:

$$\hat{\mathbf{M}} = \begin{bmatrix} -0.359 & -0.760 & 0.542 \\ -0.754 & -0.106 & -0.649 \\ 0.550 & -0.641 & -0.474 \end{bmatrix}, \quad \hat{\kappa} = 2.548$$



## 2. EVENT 672060

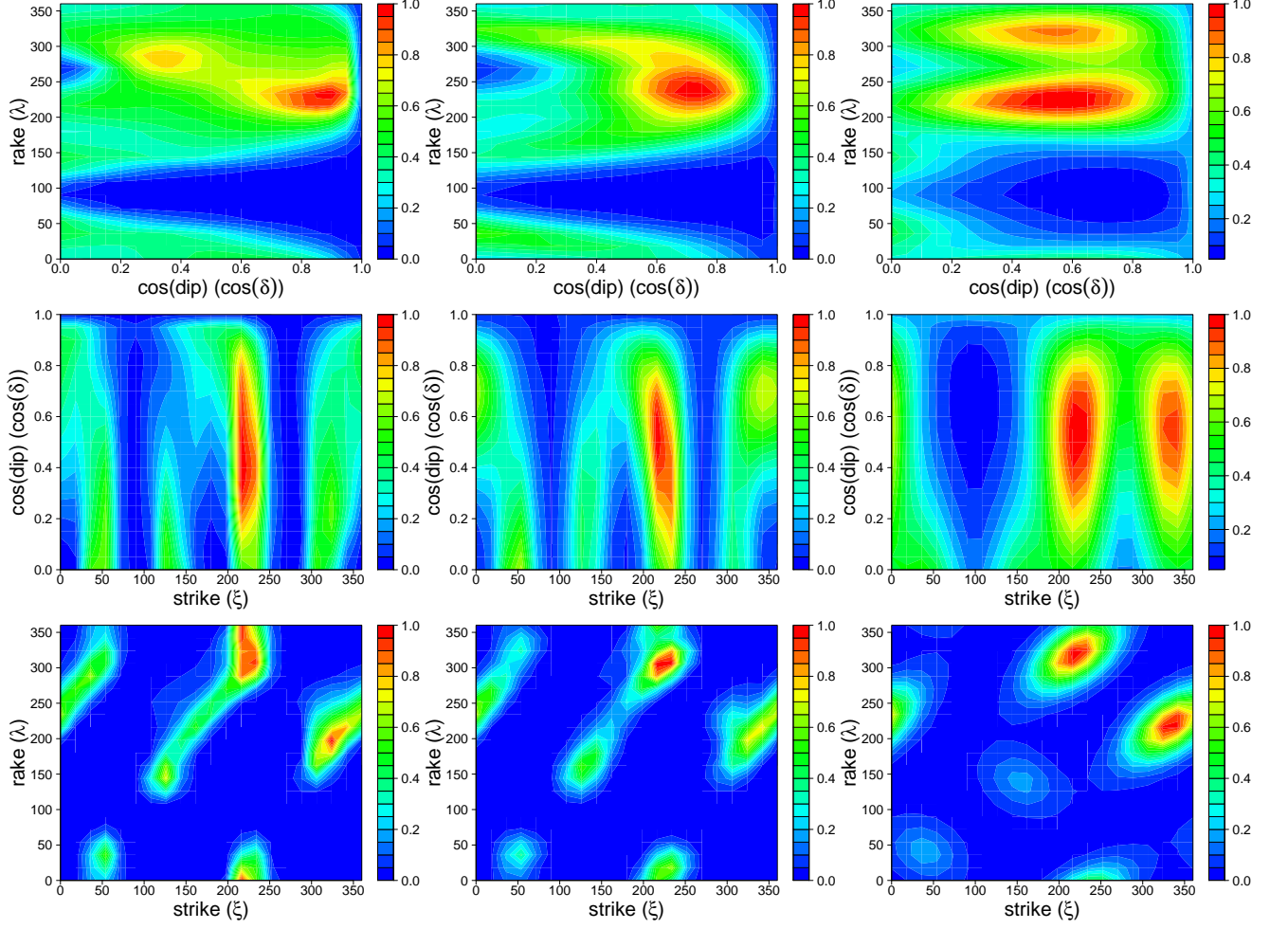


Figure 5.8: Marginal PDF plots for Bayesian posterior PDF  $P(\Theta|\mathbf{d}, \omega)$  (left), fitted 8-mode Matrix Fisher distribution  $P(\Theta|\hat{\mathbf{F}})$  (centre) and fitted 8-mode Matrix Fisher distribution with scalar concentration parameter  $P(\Theta|\hat{\mathbf{M}}, \hat{\kappa})$  (right) for event 672060.

### 3. EVENT 668273

Date	Time	Latitude	Longitude	Depth	Magnitude
02/10/1994	22:38:48.96	-37.866	178.084	30.957	2.9

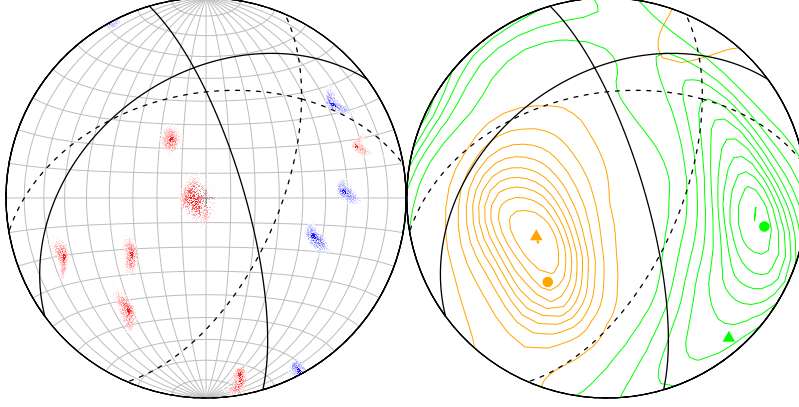


Figure 5.9: Stereonet for event 668273 (left). The solid dark line indicates the MAP focal mechanism solution  $[\Theta = (343.51^\circ, 74.54^\circ, 233.53^\circ)]$  while the dotted line indicates the solution given by Reyners & McGinty (1999). Angular difference between the two solutions =  $42.45^\circ$ . Blue points are compressions, red points are dilatations. PT contour plot for event 668273 (right). Orange denotes the P-axis, green the T-axis. The circle denotes the MAP estimate, while the triangle denotes the solution given by Reyners & McGinty (1999).

#### Parameter Estimates

8-mode Matrix Fisher Distribution:

$$\hat{\mathbf{F}} = \begin{bmatrix} -1.809 & -0.436 & 0.370 \\ 1.483 & 1.419 & 3.373 \\ 7.223 & 3.799 & -1.185 \end{bmatrix}, \quad \hat{\mathbf{M}} = \begin{bmatrix} -0.646 & 0.761 & -0.056 \\ 0.180 & 0.224 & 0.958 \\ 0.741 & 0.609 & -0.282 \end{bmatrix},$$

$$\hat{\mathbf{U}} = \begin{bmatrix} 0.882 & -0.025 & -0.025 \\ 0.467 & 0.173 & -0.867 \\ -0.060 & 0.985 & 0.164 \end{bmatrix}, \quad \hat{\mathbf{D}} = \begin{bmatrix} 8.600 & 0 & 0 \\ 0 & 3.612 & 0 \\ 0 & 0 & 0.422 \end{bmatrix}$$

8-mode Matrix Fisher Distribution with scalar concentration:

$$\hat{\mathbf{M}} = \begin{bmatrix} -0.432 & 0.902 & 0.017 \\ 0.220 & 0.087 & 0.972 \\ 0.875 & 0.424 & -0.282 \end{bmatrix}, \quad \hat{\kappa} = 2.752$$

### 3. EVENT 668273

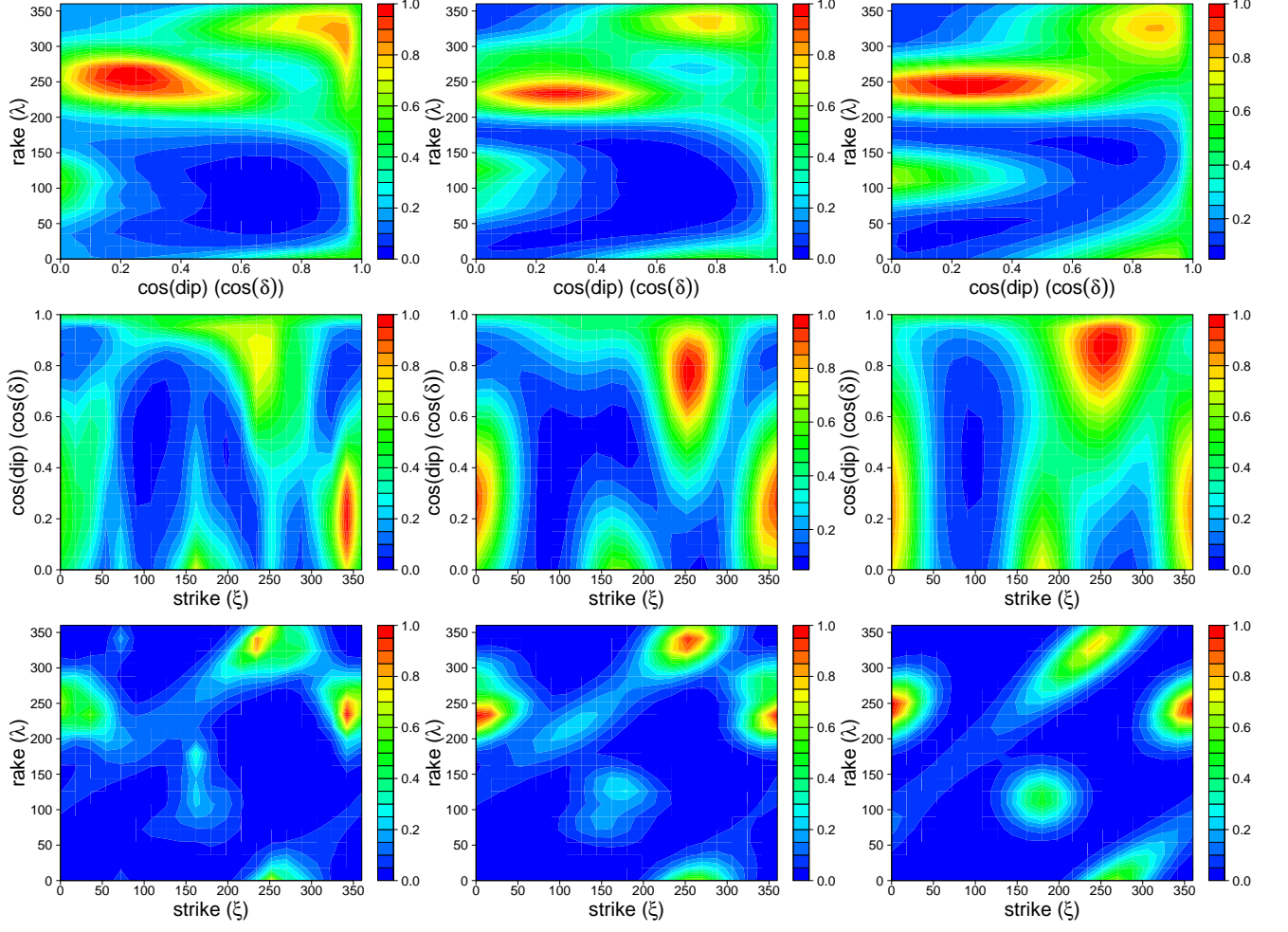


Figure 5.10: Marginal PDF plots for Bayesian posterior PDF  $P(\Theta|\mathbf{d}, \omega)$  (left), fitted 8-mode Matrix Fisher distribution  $P(\Theta|\hat{\mathbf{F}})$  (centre) and fitted 8-mode Matrix Fisher distribution with scalar concentration parameter  $P(\Theta|\hat{\mathbf{M}}, \hat{\kappa})$  (right) for event 668273.

#### 4. EVENT 640980

Date	Time	Latitude	Longitude	Depth	Magnitude
21/08/1994	13:36:52.95	-38.198	178.14	18.018	2.8

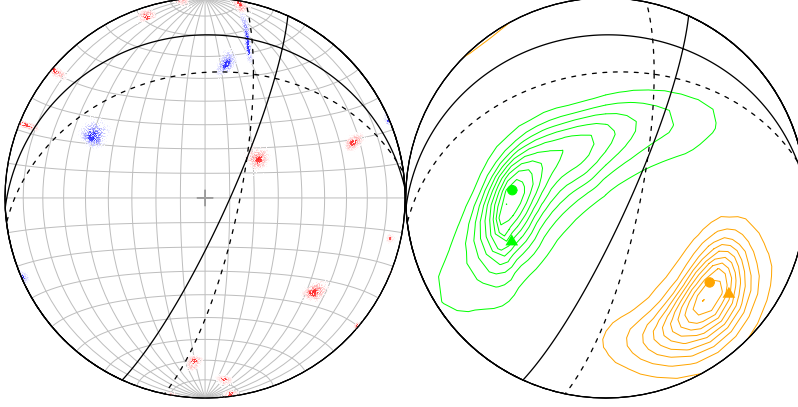


Figure 5.11: Stereonet for event 640980 (left). The solid dark line indicates the MAP focal mechanism solution  $[\Theta = (24.46^\circ, 80.72^\circ, 72.64^\circ)]$  while the dotted line indicates the solution given by Reyners & McGinty (1999). Angular difference between the two solutions =  $20.229^\circ$ . Blue points are compressions, red points are dilatations. PT contour plot for event 640980 (right). Orange denotes the P-axis, green the T-axis. The circle denotes the MAP estimate, while the triangle denotes the solution given by Reyners & McGinty (1999).

#### Parameter Estimates

8-mode Matrix Fisher Distribution:

$$\hat{\mathbf{F}} = \begin{bmatrix} 8.762 & -5.091 & -8.509 \\ -4.196 & 1.625 & 5.049 \\ -6.893 & 0.833 & 1.831 \end{bmatrix}, \quad \hat{\mathbf{M}} = \begin{bmatrix} 0.379 & -0.794 & -0.475 \\ -0.122 & -0.552 & 0.825 \\ -0.917 & -0.254 & -0.306 \end{bmatrix},$$

$$\hat{\mathbf{U}} = \begin{bmatrix} -0.725 & 0.689 & 0.689 \\ 0.322 & 0.351 & 0.879 \\ 0.609 & 0.634 & -0.476 \end{bmatrix}, \quad \hat{\mathbf{D}} = \begin{bmatrix} 16.072 & 0 & 0 \\ 0 & 3.599 & 0 \\ 0 & 0 & 1.067 \end{bmatrix}$$

8-mode Matrix Fisher Distribution with scalar concentration:

$$\hat{\mathbf{M}} = \begin{bmatrix} 0.389 & -0.737 & -0.553 \\ -0.112 & -0.633 & 0.766 \\ -0.914 & -0.235 & -0.306 \end{bmatrix}, \quad \hat{\kappa} = 3.524$$

#### 4. EVENT 640980

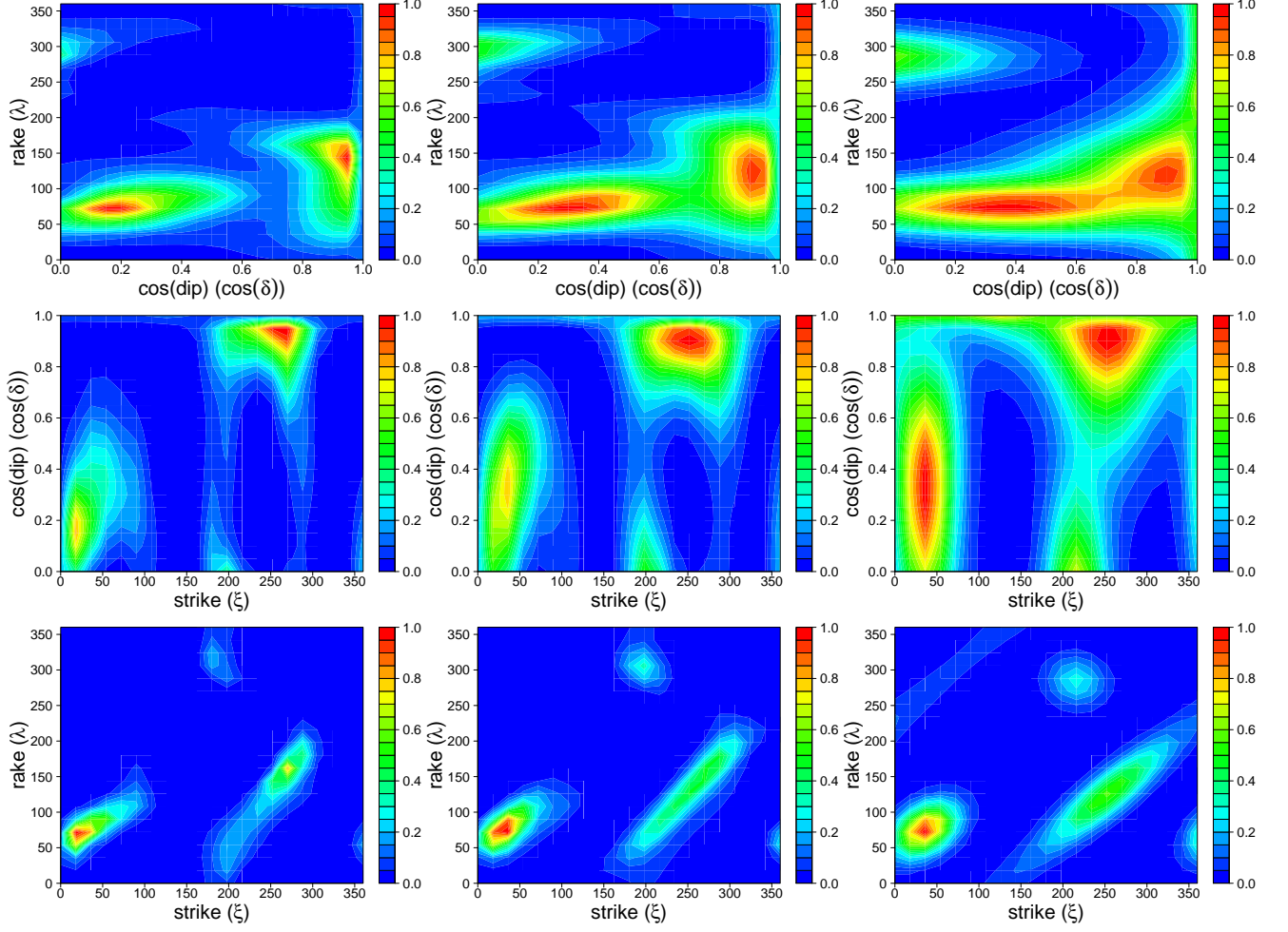


Figure 5.12: Marginal PDF plots for Bayesian posterior PDF  $P(\Theta|\mathbf{d}, \omega)$  (left), fitted 8-mode Matrix Fisher distribution  $P(\Theta|\hat{\mathbf{F}})$  (centre) and fitted 8-mode Matrix Fisher distribution with scalar concentration parameter  $P(\Theta|\hat{\mathbf{M}}, \hat{\kappa})$  (right) for event 640980.

## 5. EVENT 636036

Date	Time	Latitude	Longitude	Depth	Magnitude
03/08/1994	15:47:23.60	-38.518	177.848	33.398	2.8

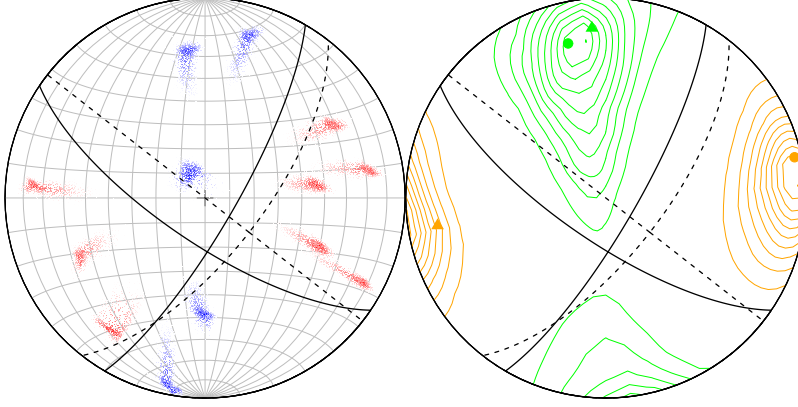


Figure 5.13: Stereonet for event 636036 (left). The solid dark line indicates the MAP focal mechanism solution  $[\Theta = (124.13^\circ, 72.01^\circ, 12.88^\circ)]$  while the dotted line indicates the solution given by Reyners & McGinty (1999). Angular difference between the two solutions =  $20.591^\circ$ . Blue points are compressions, red points are dilatations. PT contour plot for event 636036 (right). Orange denotes the P-axis, green the T-axis. The circle denotes the MAP estimate, while the triangle denotes the solution given by Reyners & McGinty (1999).

### Parameter Estimates

8-mode Matrix Fisher Distribution:

$$\hat{\mathbf{F}} = \begin{bmatrix} -4.127 & 2.514 & -1.483 \\ 10.471 & -2.686 & -6.978 \\ 1.732 & -1.610 & -1.888 \end{bmatrix}, \quad \hat{\mathbf{M}} = \begin{bmatrix} -0.543 & 0.495 & -0.679 \\ 0.796 & 0.045 & -0.604 \\ -0.268 & -0.868 & -0.419 \end{bmatrix},$$

$$\hat{\mathbf{U}} = \begin{bmatrix} 0.833 & 0.353 & 0.353 \\ -0.258 & -0.432 & -0.864 \\ -0.489 & 0.830 & -0.269 \end{bmatrix}, \quad \hat{\mathbf{D}} = \begin{bmatrix} 13.553 & 0 & 0 \\ 0 & 3.896 & 0 \\ 0 & 0 & 1.190 \end{bmatrix}$$

8-mode Matrix Fisher Distribution with scalar concentration:

$$\hat{\mathbf{M}} = \begin{bmatrix} -0.574 & 0.449 & -0.685 \\ 0.791 & 0.087 & -0.606 \\ -0.213 & -0.889 & -0.419 \end{bmatrix}, \quad \hat{\kappa} = 3.635$$

## 5. EVENT 636036

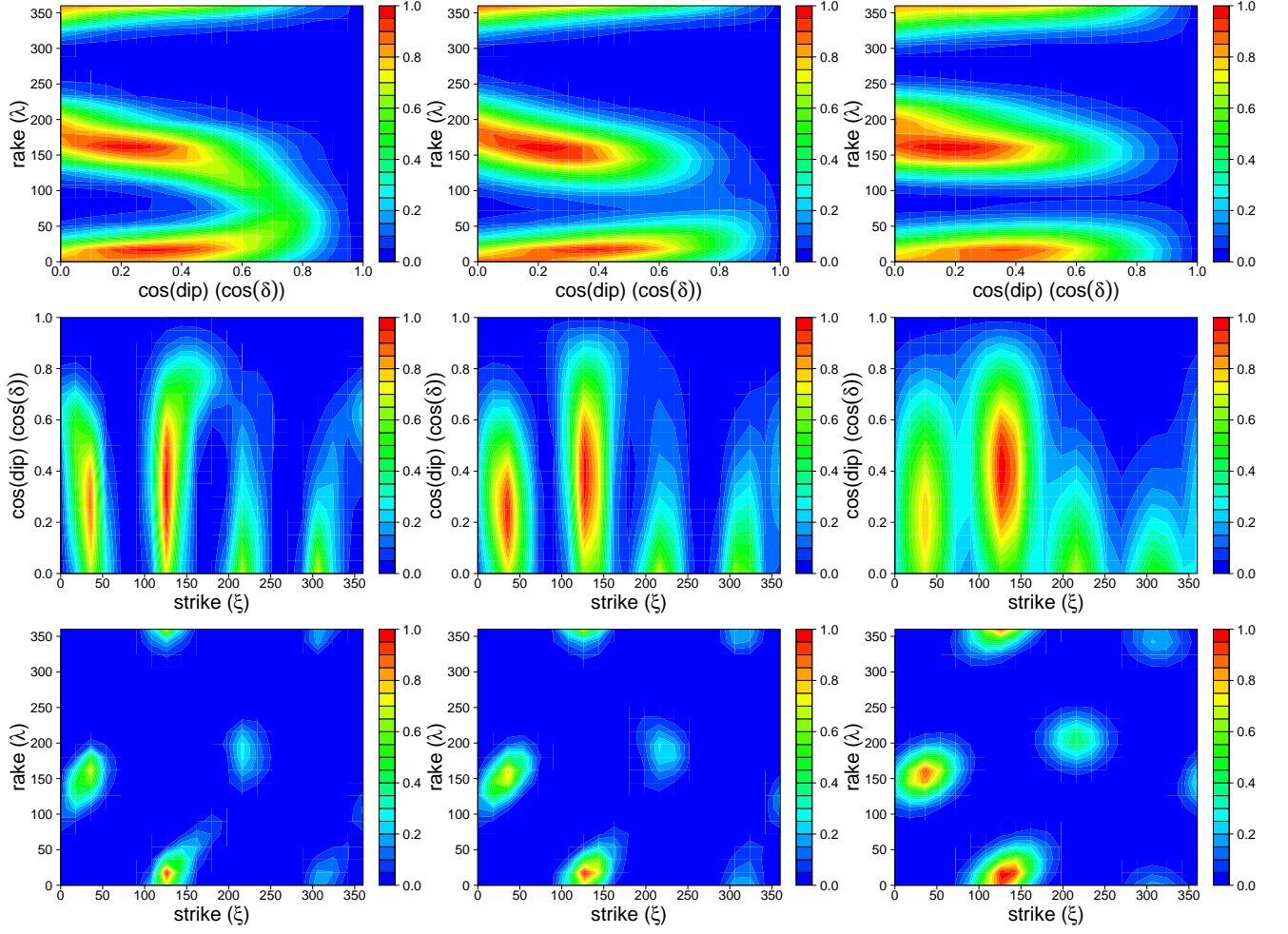


Figure 5.14: Marginal PDF plots for Bayesian posterior PDF  $P(\Theta|\mathbf{d}, \omega)$  (left), fitted 8-mode Matrix Fisher distribution  $P(\Theta|\hat{\mathbf{F}})$  (centre) and fitted 8-mode Matrix Fisher distribution with scalar concentration parameter  $P(\Theta|\hat{\mathbf{M}}, \hat{\kappa})$  (right) for event 636036.



## 6. EVENT 635767

Date	Time	Latitude	Longitude	Depth	Magnitude
31/07/1994	05:58:40.01	-38.361	177.872	23.73	3.0

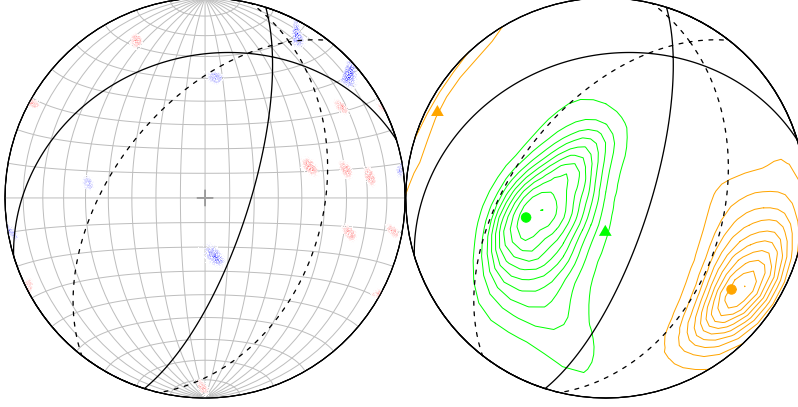


Figure 5.15: Stereonet for event 635767 (left). The solid dark line indicates the MAP focal mechanism solution  $[\Theta = (253.18^\circ, 29.80^\circ, 141.65^\circ)]$  while the dotted line indicates the solution given by Reyners & McGinty (1999). Angular difference between the two solutions =  $32.772^\circ$ . Blue points are compressions, red points are dilatations. PT contour plot for event 635767 (right). Orange denotes the P-axis, green the T-axis. The circle denotes the MAP estimate, while the triangle denotes the solution given by Reyners & McGinty (1999).

### Parameter Estimates

8-mode Matrix Fisher Distribution:

$$\hat{\mathbf{F}} = \begin{bmatrix} -0.765 & 0.470 & 2.140 \\ 7.809 & 4.612 & -6.472 \\ 0.995 & 2.659 & -6.776 \end{bmatrix}, \quad \hat{\mathbf{M}} = \begin{bmatrix} -0.231 & 0.843 & 0.486 \\ 0.877 & 0.397 & -0.271 \\ -0.421 & 0.364 & -0.831 \end{bmatrix},$$

$$\hat{\mathbf{U}} = \begin{bmatrix} -0.565 & -0.762 & -0.762 \\ -0.405 & -0.078 & -0.911 \\ 0.719 & -0.643 & -0.265 \end{bmatrix}, \quad \hat{\mathbf{D}} = \begin{bmatrix} 12.848 & 0 & 0 \\ 0 & 4.100 & 0 \\ 0 & 0 & 1.276 \end{bmatrix}$$

8-mode Matrix Fisher Distribution with scalar concentration:

$$\hat{\mathbf{M}} = \begin{bmatrix} -0.183 & 0.876 & 0.446 \\ 0.897 & 0.335 & -0.289 \\ -0.403 & 0.347 & -0.831 \end{bmatrix}, \quad \hat{\kappa} = 3.852$$



## 6. EVENT 635767

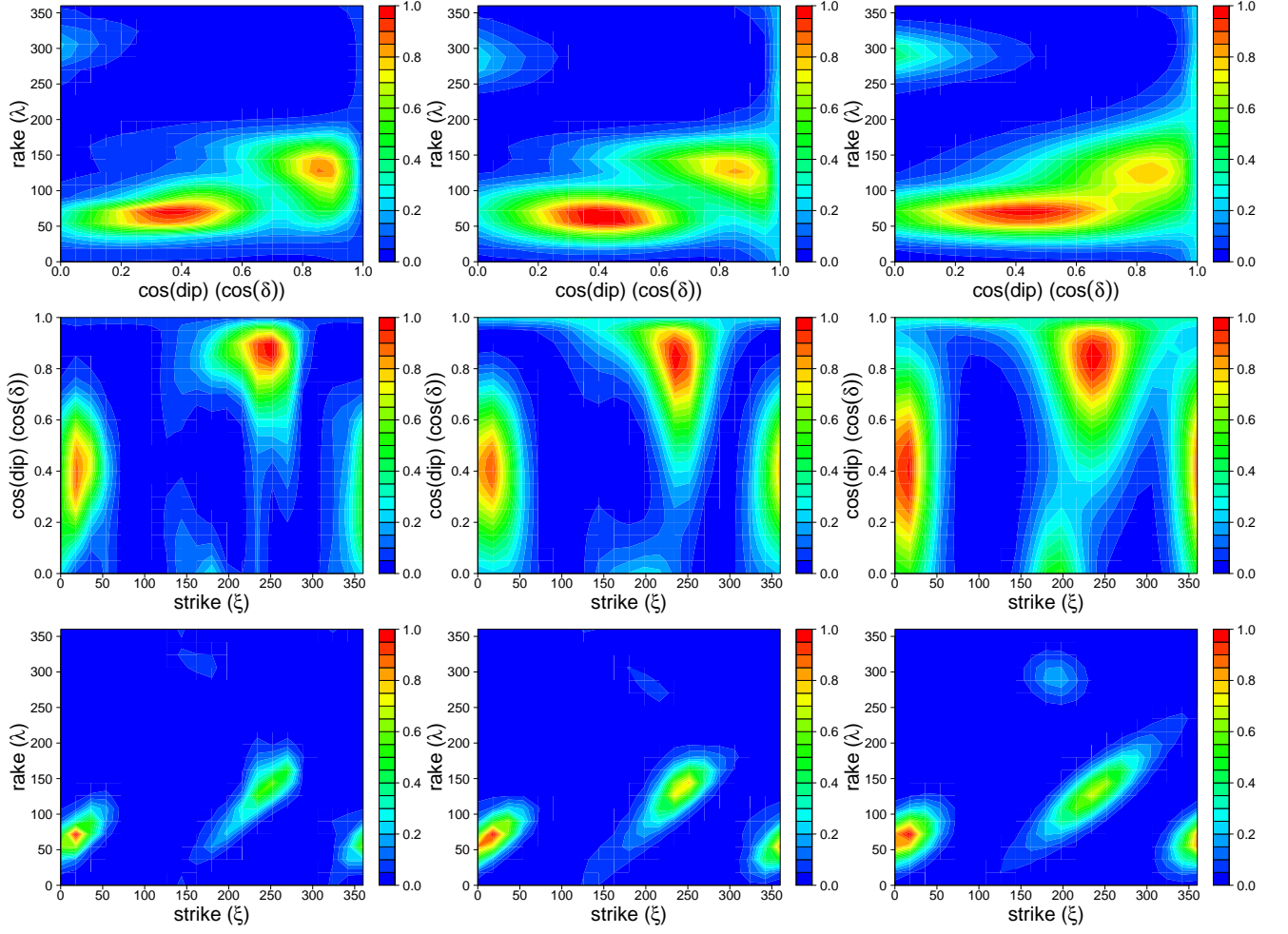


Figure 5.16: Marginal PDF plots for Bayesian posterior PDF  $P(\Theta|\mathbf{d}, \omega)$  (left), fitted 8-mode Matrix Fisher distribution  $P(\Theta|\hat{\mathbf{F}})$  (centre) and fitted 8-mode Matrix Fisher distribution with scalar concentration parameter  $P(\Theta|\hat{\mathbf{M}}, \hat{\kappa})$  (right) for event 635767.

## 7. EVENT 669233

Date	Time	Latitude	Longitude	Depth	Magnitude
03/10/1994	20:51:10.99	-38.544	177.814	26.123	3.0

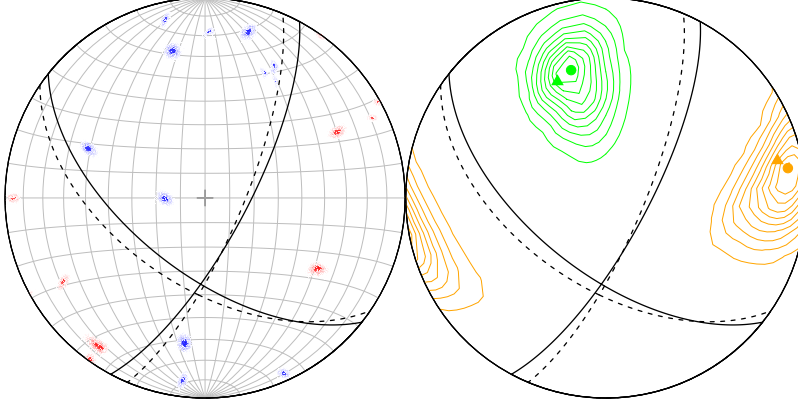


Figure 5.17: Stereonet for event 669233 (left). The solid dark line indicates the MAP focal mechanism solution  $[\Theta = (128.39^\circ, 59.81^\circ, 19.72^\circ)]$  while the dotted line indicates the solution given by Reyners & McGinty (1999). Angular difference between the two solutions =  $6.946^\circ$ . Blue points are compressions, red points are dilatations. PT contour plot for event 669233 (right). Orange denotes the P-axis, green the T-axis. The circle denotes the MAP estimate, while the triangle denotes the solution given by Reyners & McGinty (1999).

### Parameter Estimates

8-mode Matrix Fisher Distribution:

$$\hat{\mathbf{F}} = \begin{bmatrix} -6.597 & 5.783 & -13.371 \\ 8.347 & 3.209 & -7.075 \\ -4.443 & 0.706 & -4.984 \end{bmatrix}, \quad \hat{\mathbf{M}} = \begin{bmatrix} -0.466 & 0.655 & -0.595 \\ 0.829 & 0.087 & -0.553 \\ -0.310 & -0.751 & -0.583 \end{bmatrix},$$

$$\hat{\mathbf{U}} = \begin{bmatrix} 0.284 & 0.958 & 0.958 \\ -0.359 & 0.154 & 0.920 \\ 0.889 & -0.243 & 0.388 \end{bmatrix}, \quad \hat{\mathbf{D}} = \begin{bmatrix} 17.661 & 0 & 0 \\ 0 & 10.842 & 0 \\ 0 & 0 & 1.177 \end{bmatrix}$$

8-mode Matrix Fisher Distribution with scalar concentration:

$$\hat{\mathbf{M}} = \begin{bmatrix} -0.461 & 0.615 & -0.640 \\ 0.829 & 0.042 & -0.557 \\ -0.316 & -0.787 & -0.583 \end{bmatrix}, \quad \hat{\kappa} = 6.130$$

## 7. EVENT 669233

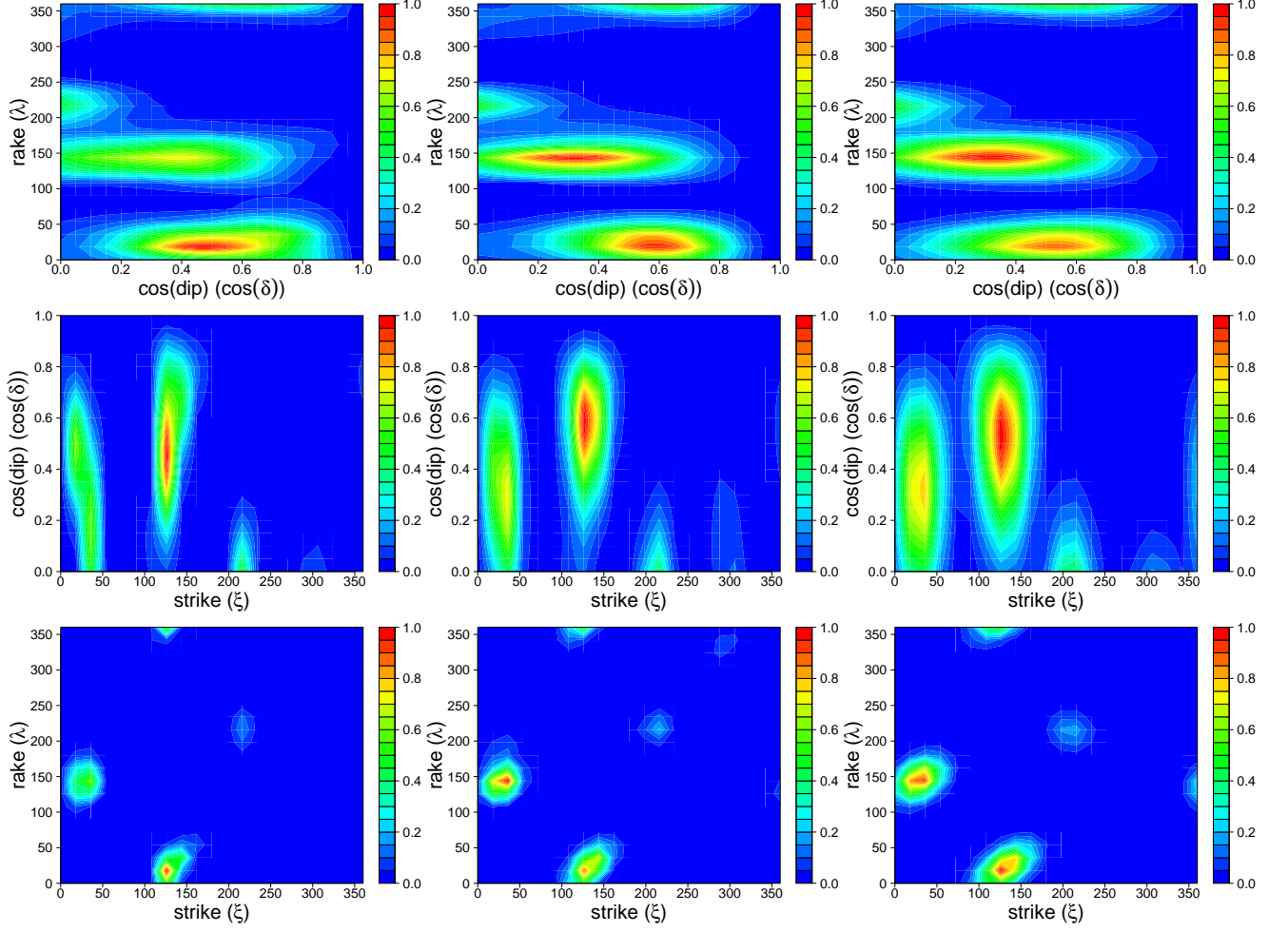


Figure 5.18: Marginal PDF plots for Bayesian posterior PDF  $P(\Theta|\mathbf{d}, \omega)$  (left), fitted 8-mode Matrix Fisher distribution  $P(\Theta|\hat{\mathbf{F}})$  (centre) and fitted 8-mode Matrix Fisher distribution with scalar concentration parameter  $P(\Theta|\hat{\mathbf{M}}, \hat{\kappa})$  (right) for event 669233.

## 8. EVENT 665895

Date	Time	Latitude	Longitude	Depth	Magnitude
09/10/1994	11:34:02.22	-38.549	178.062	20.898	3.0

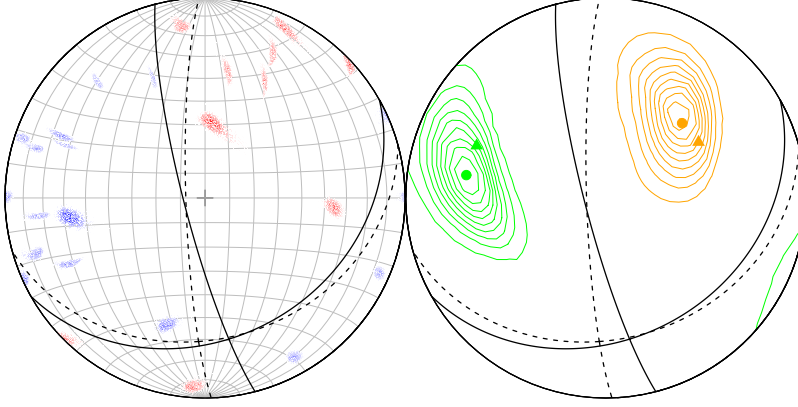


Figure 5.19: Stereonet for event 665895 (left). The solid dark line indicates the MAP focal mechanism solution  $[\Theta = (165.52^\circ, 81.30^\circ, 240.41^\circ)]$  while the dotted line indicates the solution given by Reyners & McGinty (1999). Angular difference between the two solutions =  $12.709^\circ$ . Blue points are compressions, red points are dilatations. PT contour plot for event 665895 (right). Orange denotes the P-axis, green the T-axis. The circle denotes the MAP estimate, while the triangle denotes the solution given by Reyners & McGinty (1999).

### Parameter Estimates

8-mode Matrix Fisher Distribution:

$$\hat{\mathbf{F}} = \begin{bmatrix} 11.004 & -6.181 & -6.035 \\ -3.193 & -0.898 & -9.477 \\ 9.127 & -2.334 & -1.959 \end{bmatrix}, \quad \hat{\mathbf{M}} = \begin{bmatrix} 0.486 & -0.835 & -0.258 \\ -0.306 & 0.114 & -0.945 \\ 0.819 & 0.538 & -0.200 \end{bmatrix},$$

$$\hat{\mathbf{U}} = \begin{bmatrix} 0.822 & 0.474 & 0.474 \\ -0.387 & 0.056 & -0.921 \\ -0.418 & 0.879 & 0.229 \end{bmatrix}, \quad \hat{\mathbf{D}} = \begin{bmatrix} 16.815 & 0 & 0 \\ 0 & 10.206 & 0 \\ 0 & 0 & 1.485 \end{bmatrix}$$

8-mode Matrix Fisher Distribution with scalar concentration:

$$\hat{\mathbf{M}} = \begin{bmatrix} 0.512 & -0.814 & -0.275 \\ -0.310 & 0.123 & -0.943 \\ 0.801 & 0.568 & -0.200 \end{bmatrix}, \quad \hat{\kappa} = 6.551$$

## 8. EVENT 665895

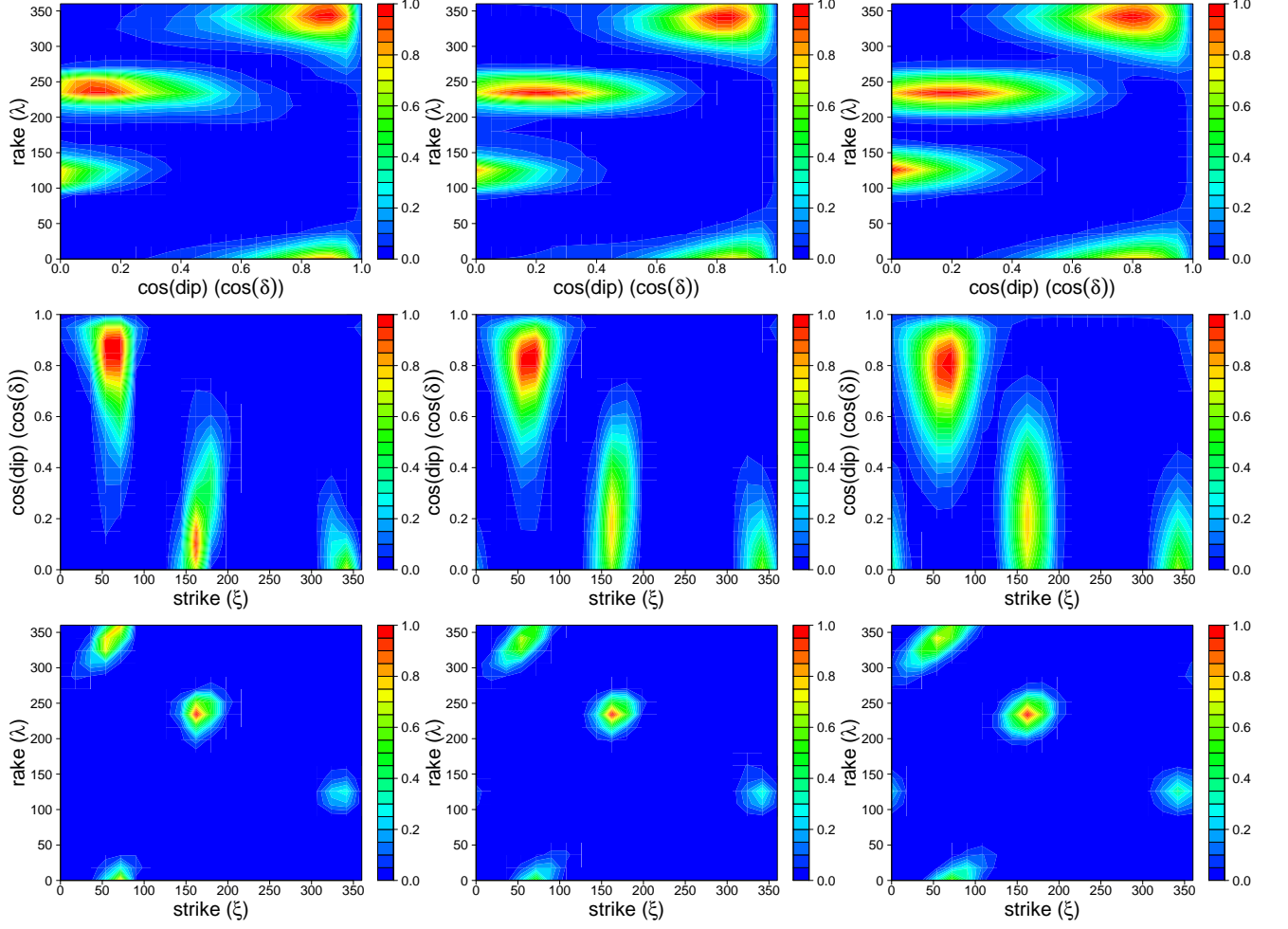


Figure 5.20: Marginal PDF plots for Bayesian posterior PDF  $P(\Theta|\mathbf{d}, \omega)$  (left), fitted 8-mode Matrix Fisher distribution  $P(\Theta|\hat{\mathbf{F}})$  (centre) and fitted 8-mode Matrix Fisher distribution with scalar concentration parameter  $P(\Theta|\hat{\mathbf{M}}, \hat{\kappa})$  (right) for event 665895.

## 9. EVENT 675146

Date	Time	Latitude	Longitude	Depth	Magnitude
24/10/1994	01:18:43.58	-38.539	178.098	25.488	2.9

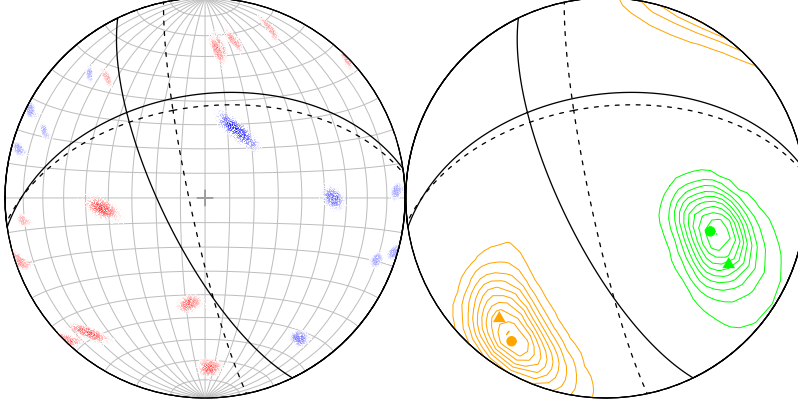


Figure 5.21: Stereonet for event 675146 (left). The solid dark line indicates the MAP focal mechanism solution  $[\Theta = (154.11^\circ, 72.14^\circ, 134.26^\circ)]$  while the dotted line indicates the solution given by Reyners & McGinty (1999). Angular difference between the two solutions =  $18.377^\circ$ . Blue points are compressions, red points are dilatations. PT contour plot for event 675146 (right). Orange denotes the P-axis, green the T-axis. The circle denotes the MAP estimate, while the triangle denotes the solution given by Reyners & McGinty (1999).

### Parameter Estimates

8-mode Matrix Fisher Distribution:

$$\hat{\mathbf{F}} = \begin{bmatrix} 15.673 & 5.546 & -1.024 \\ -6.887 & -5.436 & -11.816 \\ -7.719 & -1.964 & -1.827 \end{bmatrix}, \quad \hat{\mathbf{M}} = \begin{bmatrix} 0.741 & 0.554 & -0.381 \\ -0.137 & -0.430 & -0.892 \\ -0.658 & 0.713 & -0.243 \end{bmatrix},$$

$$\hat{\mathbf{U}} = \begin{bmatrix} 0.873 & -0.372 & -0.372 \\ 0.375 & 0.099 & -0.922 \\ 0.312 & 0.923 & 0.226 \end{bmatrix}, \quad \hat{\mathbf{D}} = \begin{bmatrix} 20.998 & 0 & 0 \\ 0 & 10.892 & 0 \\ 0 & 0 & 1.124 \end{bmatrix}$$

8-mode Matrix Fisher Distribution with scalar concentration:

$$\hat{\mathbf{M}} = \begin{bmatrix} 0.775 & 0.521 & -0.357 \\ -0.162 & -0.382 & -0.910 \\ -0.610 & 0.763 & -0.243 \end{bmatrix}, \quad \hat{\kappa} = 6.500$$

## 9. EVENT 675146

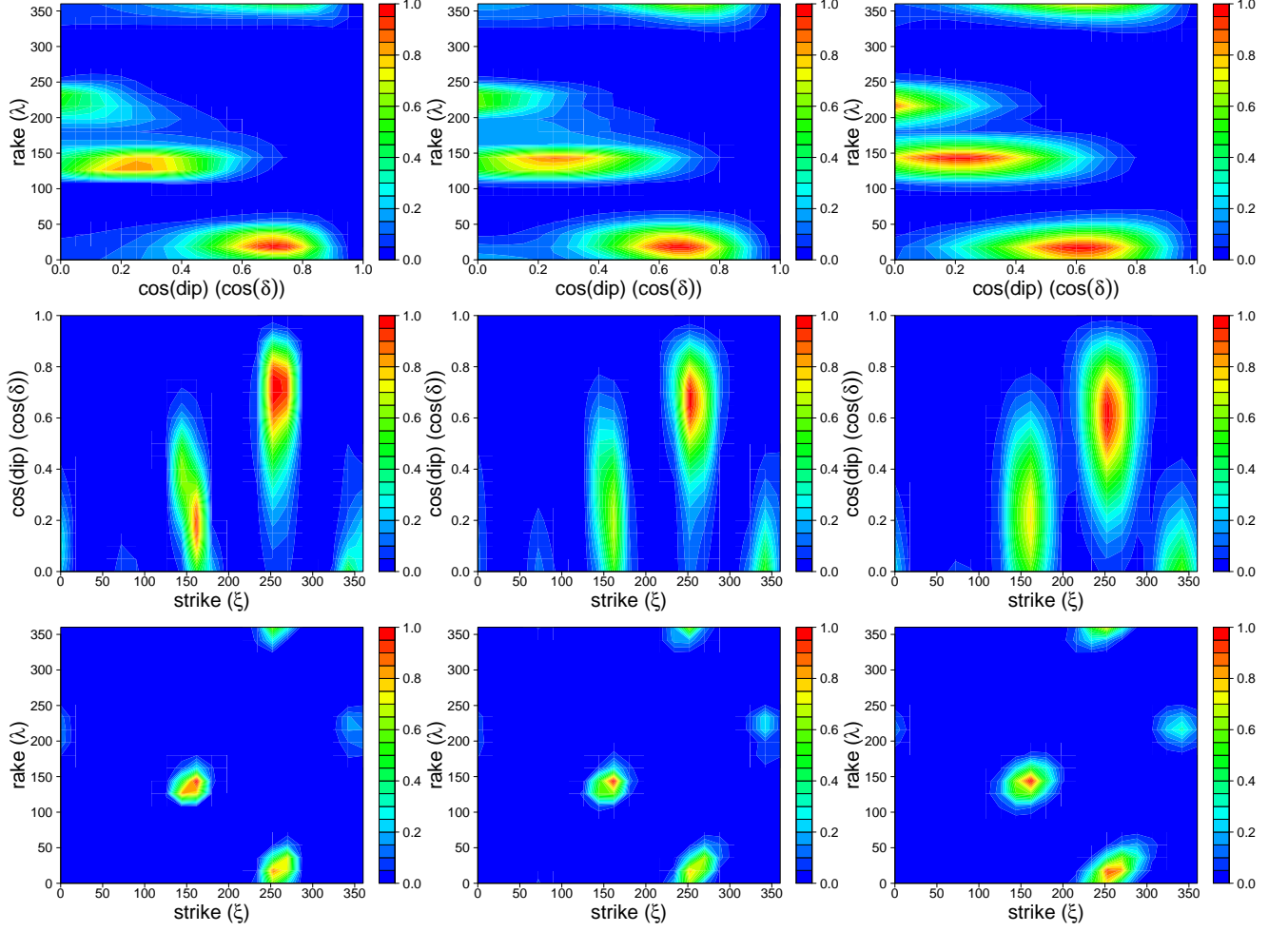


Figure 5.22: Marginal PDF plots for Bayesian posterior PDF  $P(\Theta|\mathbf{d}, \omega)$  (left), fitted 8-mode Matrix Fisher distribution  $P(\Theta|\hat{\mathbf{F}})$  (centre) and fitted 8-mode Matrix Fisher distribution with scalar concentration parameter  $P(\Theta|\hat{\mathbf{M}}, \hat{\kappa})$  (right) for event 675146.

The first event (CUSPID 646630, our first poorly constrained event) shows a large disparity between the MAP solution of  $\Theta$  and that of Reyners & McGinty (1999, RM99), indicated by the angular difference of  $51.403^\circ$  and the fairly large disparity between the sets of nodal planes on the stereonet (Figure 5.5, p70). The nodal planes should separate completely the compressions (blue) and dilatations (red), as the MAP solution does. The probability contours of the P-axis appear bimodal, with a mode near both the MAP estimate and the RM99 estimate of  $\hat{\mathbf{v}}^P$ . The 8-mode Matrix Fisher distribution approximates fairly well the Bayesian posterior PDF (Figure 5.6, p71), while the Matrix Fisher distribution with scalar concentration parameter is a slightly poorer approximation of the empirical distribution, e.g. it cannot reproduce the bimodal maxima in the distribution of the rake.

The second event (CUSPID 672060) shows reasonable agreement between the MAP and RM99 solutions. The MAP estimate completely separates the compressions and dilatations (see Figure 5.7, p72) whereas the RM99 solution does not. The 8-mode Matrix Fisher approximation is again close to the empirical distribution (Figure 5.8, p73). The third event (CUSPID 668273) again has a large disparity between the MAP estimate and that of RM99, and a misfit polarity is evident on the stereonet, near  $(\phi, \theta) = (\pi/2, \pi/2)$  (see Figure 5.9, p74). The location of this dilatational point (red) amongst a group of compressions (blue) indicates it may have been an incorrectly read polarity.

In general for these three poorly constrained events, we have seen large disparity between the MAP estimates and the RM99 estimates of  $\Theta$ . Further, the P- and T-axis contours are poorly constrained and can cross the nodal planes. The 8-mode Matrix Fisher distribution provides a reasonable approximation to the empirical distribution, and the Matrix Fisher distribution with scalar concentration parameter provides a slightly poorer approximation. The low values of  $\hat{\kappa}$  (2.66, 2.55, and 2.75, for Events 1, 2 and 3 respectively), give an overall indication of the poor constraint on the posterior PDFs.

The first intermediately constrained event (CUSPID 640980) shows reasonable agreement between the MAP and RM99 solutions of  $\Theta$ . There is one polarity reading that crosses a nodal plane boundary (see Figure 5.11, p76). The P- and T-axis contour plots appear better constrained than the previous three events, as demonstrated by the narrower spread around the modes. The 8-mode PDF provides a good fit to the empirical distribution (Figure 5.12, p77). The second (CUSPID 636036) intermediately constrained event exhibits a similarly constrained posterior PDF, and here the scalar concentration PDF appears to be almost as good an approximation as the full Matrix Fisher (Figure 5.14, p79). The third (CUSPID 635767) intermediately constrained event shows less agreement between the MAP solution of  $\Theta$  and that of RM99 (see Figure 5.15, p80), but again the P- and T-axis plots are fairly well constrained, and the Matrix Fisher approximations show good agreement with the Bayesian posterior PDF (Figure 5.16, p81).

The first well constrained solution (CUSPID 669233) shows very good agreement between the MAP and RM99 solutions of  $\Theta$ , with an angular difference of just  $6.946^\circ$  (see Figure 5.17, p82). Both Matrix Fisher approximations are very close to the Bayesian posterior PDF (Figure 5.18, p5.18). The same is true of both the second (CUSPID 665895) and third (CUSPID 675146) well constrained events. In the third of these events there appears to be one misfit polarity (see the upper left quadrant of Figure 5.21, p86). The well constrained events are characterised by high  $\hat{\kappa}$  (6.13, 6.55 and 6.50) compared to the poorer constrained events. The P- and T-axis contours of all three events are tightly constrained. In general the well constrained events have a higher number of polarity readings and better focal sphere coverage than the poorly and intermediately constrained events.



The better constrained events agree closely the established focal mechanism solutions of RM99. The Matrix Fisher approximation tends to match well the Bayesian posterior PDF, with the match appearing better for well constrained events. The full 8-mode Matrix Fisher distribution with parameter matrix  $\hat{\mathbf{F}}$  generally provides a better fit than the scalar concentration version, at the cost of longer computation time and increased complexity, although the difference in quality of fit is small for the better constrained events.

Figure 5.23 shows a map of all 87 MAP focal mechanism estimates obtained in this study. The map corresponds closely to Figure 5.24, the solutions obtained in the RM99 study, except for some small discrepancies which we discuss below.

Figure 5.25 shows a histogram of angular differences between our solutions and those of RM99. For 75% of events, the solutions are within  $27.3^\circ$  of each other, indicating that solutions obtained by our method are generally similar to those obtained by RM99. The angular differences between the two sets may be partly explained by the different focal mechanism estimation methodology used and the hypocentral uncertainties considered here, but may also be partly explained by differences in hypocentre location resulting from our interpolation of the Reyners et al. (1999) velocity model to a constant grid spacing. Figure 5.26 shows a plot of angular difference between solutions versus the distance between the epicentres as located by RM99 and NonLinLoc in this study. There is no obvious relationship evident.

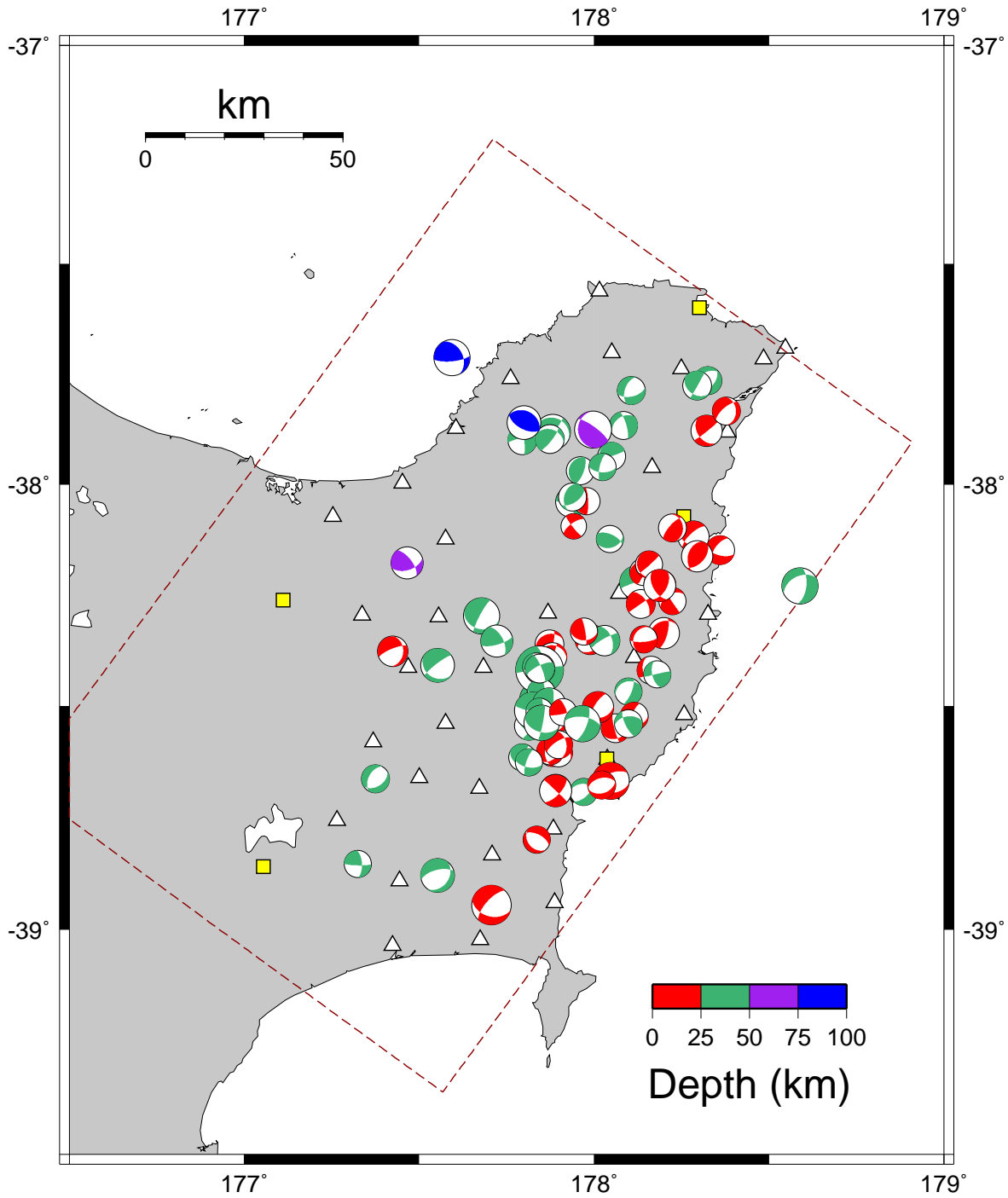


Figure 5.23: Map of Raukumara showing MAP focal mechanism estimates obtained using our method. Hypocentre locations are as calculated by NonLinLoc. Beachballs are scaled relative to their magnitudes. The dark red line indicates the boundary of the velocity model.

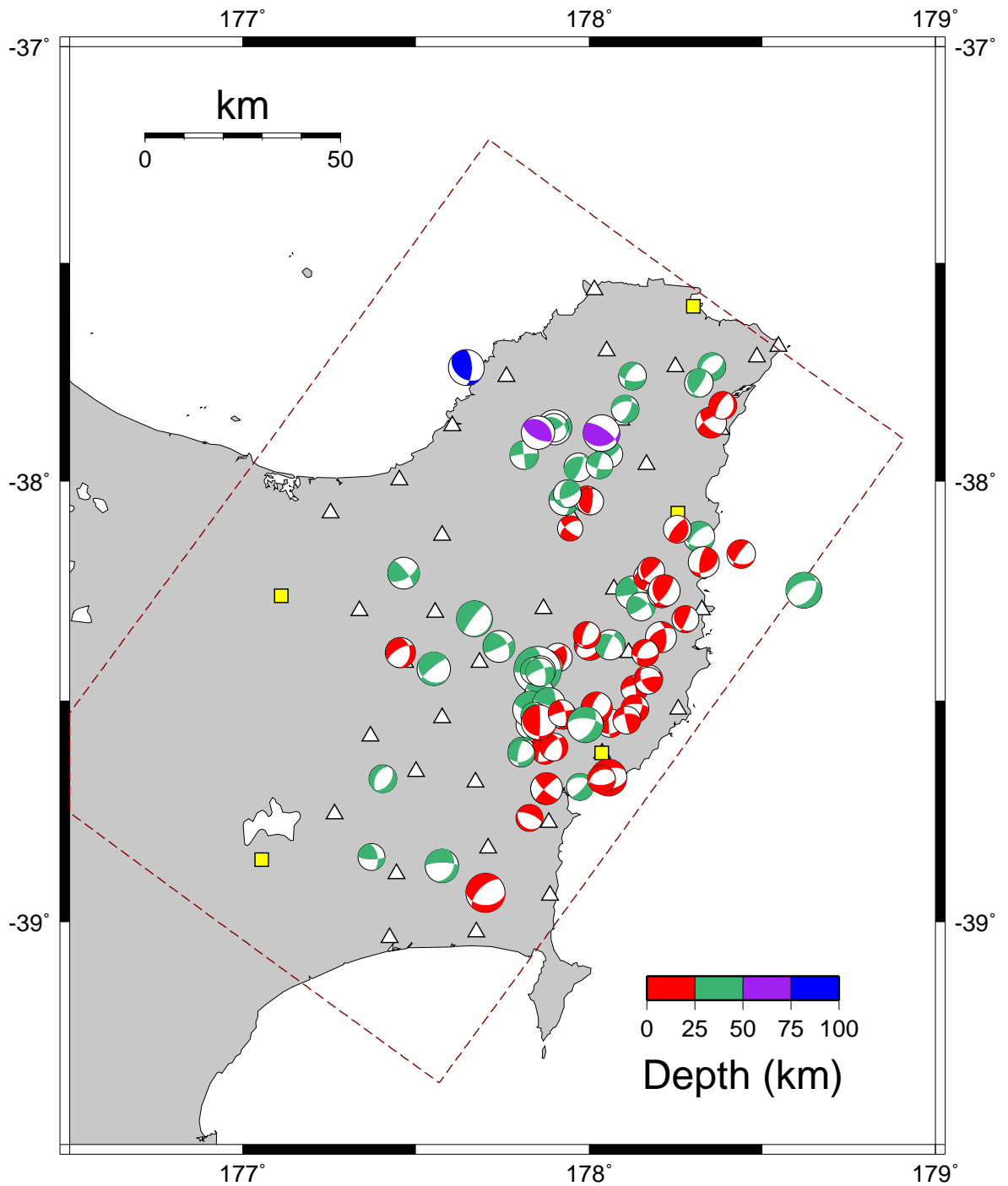


Figure 5.24: Map of Raukumara showing focal mechanisms obtained by Reyners & McGinty (1999). Hypocentre locations also from Reyners & McGinty (1999). Beachballs are scaled relative to their magnitudes. The dark red line indicates the boundary of the velocity model.

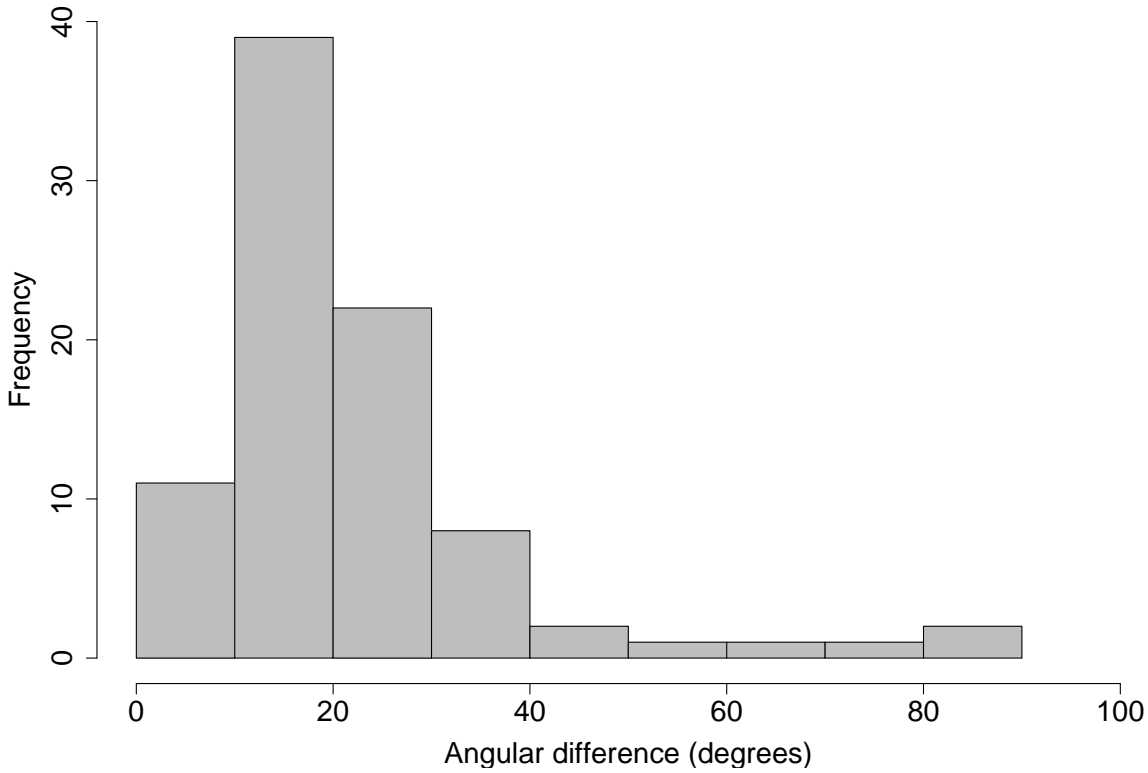


Figure 5.25: Histogram of  $n = 87$  angular differences between our MAP focal mechanism estimates and the focal mechanisms of Reyners & McGinty (1999).

## 5.2 Velocity model unknown — Kawerau

In this Section the Bayesian model is applied in the situation in which neither the hypocentre location of the earthquake nor the velocity model is known, with applications to data from the Kawerau region of New Zealand.

### 5.2.1 Tectonic Setting

The Kawerau geothermal field is situated in the eastern Taupo Volcanic Zone (TVZ), in the North Island of New Zealand. The field, a roughly circular area of 19-35 km<sup>2</sup>, is in the most seismically active part of the TVZ, where many shallow earthquakes occur. The age of the field has been estimated at around 200,000 years, and its energy output estimated at 100 MW (Bibby et al. 1995). The Kawerau geothermal field lies to the east of the Taupo fault belt, and to the west of the North Island dextral fault belt (Clarke 2007). As with the Raukumara Peninsula, seismicity here is caused by the Pacific plate subducting underneath the Australian plate beneath the region. Previous focal mechanism studies in the TVZ have found mechanisms that are predominantly normal or normal with a strike-slip component, with large variation in strike (Hurst et al. 2002).

### 5.2.2 Velocity Models

The velocity models used in this section were obtained by Clarke (2007) using Velest (Kissling 1995) (see Section 4.3) and phase data from GeoNet (see Section 5.2.3).

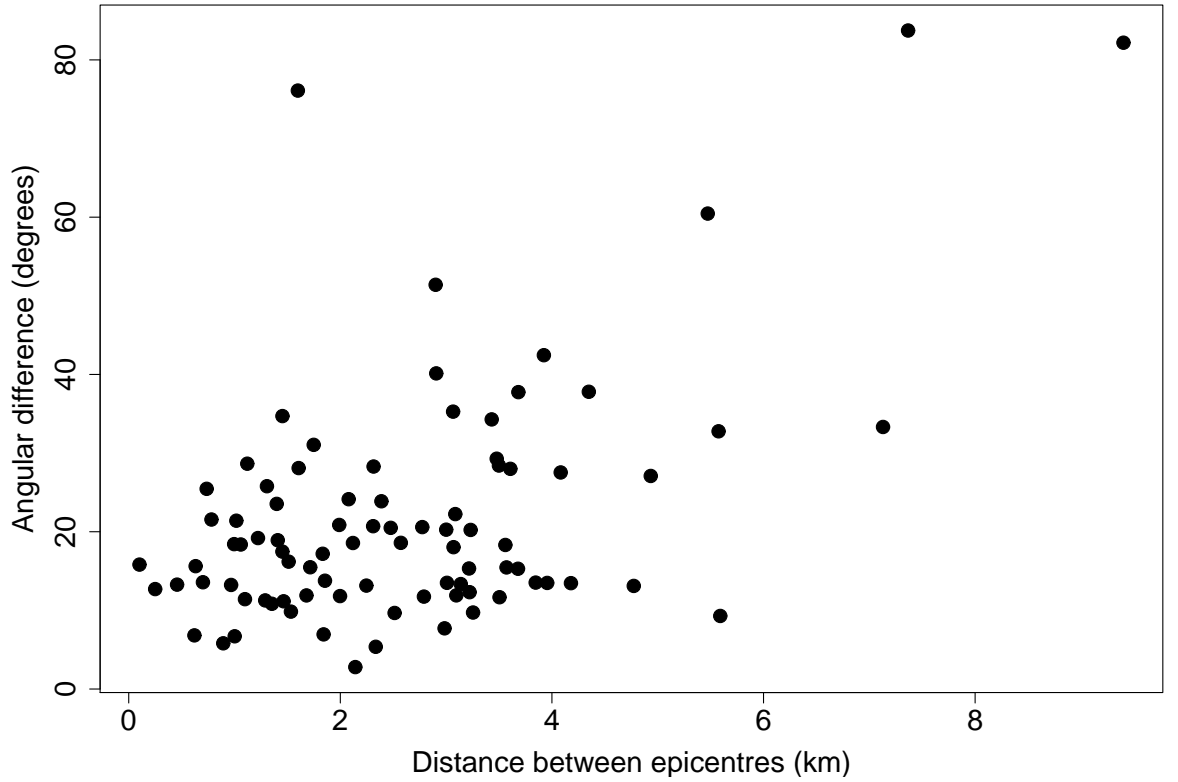


Figure 5.26: Plot of angular difference versus distance between epicentres for our MAP focal mechanism estimates and the focal mechanisms of Reyners & McGinty (1999).

We drew a random sample of velocity models using a method described by Clarke (2007), which we take as our prior  $P(v)$  for the velocity model. Twelve layer boundaries were selected at  $-3, 0, 2, 4, 6, 8, 10, 15, 20, 25, 30$  and  $40$  km depth. Velest does not adjust the position of the boundaries in its procedure. The procedure for assigning initial velocities was:

- Pick  $u_1 \sim \text{Uniform}(1, 6)$  - the velocity in km/s for the top ( $-3$ km) layer.
- Pick  $u_2 \sim \text{Uniform}(6, 10)$  - the velocity in km/s for the bottom ( $40$ km) layer.
- Pick  $u_3 \sim \text{Uniform}(u_1, u_2)$  - an intermediate velocity in km/s for the  $10$ km layer.
- Pick a sample of size 5 from  $\text{Uniform}(u_1, u_3)$ , order from lowest to highest, and assign as velocities in km/s to the  $0, 2, 4, 6$  and  $8$ km layers respectively.
- Pick a sample of size 4 from  $\text{Uniform}(u_3, u_2)$ , order from lowest to highest, and assign as velocities in km/s to the  $15, 20, 25$  and  $30$ km layers respectively.

Selection of the intermediate velocity third means the models will have different gradients in the upper and lower parts. This ensures a wide range of input models are selected (Clarke 2007).

Clarke (2007) generated approximately 1000 P-wave velocity models in this manner, and used these as a basis for joint P- and S-wave velocity model inversions using Velest. Initial P-wave velocities for this joint procedure were randomly selected within 1 standard deviation of the mean in each layer from the P-wave only models. Initial S-wave velocities were chosen randomly, in a similar manner as the initial random P-wave models, except that  $u_1 \sim \text{Uniform}(0, 3)$ ,  $u_2 \sim \text{Uniform}(3, 9)$ , and  $V_p \geq V_s$  in every layer. For our hypocentre

location routine we use the models output from the joint P- and S-wave inversion. These 1000 models are shown in Figure 5.28.

Veltest also outputs station corrections for each model, which are values of  $t_i^{res}$  for a given velocity model and station  $i$ , averaged over all events (Kissling 1995). These station corrections adjust for the true 3D variation in velocity that a 1D model cannot account for.

We convert the 1D velocity models to 3D for use with NonLinLoc, using the Vel2Grid program. This requires us to select the bounds for the model, as a 1D velocity model inherently has none. As stations must be inside the model bounds for Grid2Time to calculate their take-off angles, we use the tightest constraints on latitude and longitude such that all stations for which polarities are recorded are encompassed by the model. This gives us latitude bounds of  $-37.73^\circ$  to  $-38.92^\circ$ , and longitude bounds of  $176.26^\circ$  to  $177.12^\circ$ , while we take a depth of 50 km as a lower boundary.

### 5.2.3 Data

As the velocity models we will use are based on the research of Clarke (2007), we use here the same phase data that was used in that study to construct the velocity models. Using GeoNet phase data, Clarke (2007) selected earthquakes in the Kawerau region, with hypocentre latitudes between  $-38^\circ$  and  $-38.2^\circ$ , longitudes between  $175.55^\circ$  and  $176.85^\circ$ , and depths shallower than 20 km. From this set of 1875 earthquakes, the most reliable events were selected. The selection criteria were:

- The event must have a minimum of 8 P-wave phases and 3 S-wave phases to ensure it is able to be located reliably.
- Seismic stations receiving waves from the event must have a maximum azimuthal gap of less than  $180^\circ$  to guard against epicentral bias.
- The nearest station to the event must be maximum of 10km away from the epicentre (as determined by the GeoNet hypocentre location) to ensure depth is determined accurately.

Meanwhile, we apply a further criterion that an event must have seven or more polarity readings to ensure adequate coverage of the focal sphere. This number is slightly relaxed from the value of 10 used in Section 5.1, as the GeoNet data contains relatively few polarity readings. This selection criterion narrows our set of earthquakes to seven.

### 5.2.4 Posterior PDF Particulars

We assign equal prior weight to each of the  $\sim 1000$  velocity models, and these together constitute our prior for the velocity model  $P(v)$ . Further work can and should be put into establishing a better motivated prior for the velocity model, however for the purposes of this project we have simply used the results of Clarke (2007) to provide  $P(v)$ . Thus our posterior PDF becomes

$$P(\Theta|\mathbf{d}, \omega) \propto P(\Theta) \sum_{k=1}^p \sum_{j=1}^m \left[ \prod_{i=1}^n \pi_{ijk}^{\frac{1}{2}(1+y_i)} (1 - \pi_{ijk})^{\frac{1}{2}(1-y_i)} \right]$$

where

$$\pi_{ijk} = \pi'_{p_0} + (1 - 2\pi'_{p_0}) \Phi \left( \frac{2(\hat{\mathbf{p}}_{ijk} \cdot \hat{\mathbf{n}})(\hat{\mathbf{p}}_{ijk} \cdot \hat{\mathbf{u}})}{\sigma_{a_0}} \right)$$

and we adopt here the same set values for  $\pi'_p$  and  $\sigma_a$  that were used for the velocity model known case:  $\pi'_{p_0} = 0.2$  and  $\sigma_{a_0} = \frac{1}{6}$ .

## 5.2.5 Results

We present here results in the same format as in Section 5.1.5, for all seven events that meet our selection criteria. For each event we present first the velocity model known results, followed by the results in the velocity model unknown case. In the velocity model known case, we use here the mean velocity model (see Table 5.2) given by the mean velocity in each layer over the set of 1000 models. The results of this will be used as a basis to which we can compare the effect of adding uncertainty over the velocity structure into our probability model.

Table 5.2: Mean P- and S-wave velocity models with corresponding standard deviations for Kawerau.

Depth (km)	P-wave velocity $V_p$ (km/s)	Standard deviation (km/s)	S-wave velocity $V_s$ (km/s)	Standard deviation (km/s)
-3	4.254	0.275	2.481	0.297
0	4.485	0.372	2.620	0.247
2	4.973	0.153	2.756	0.232
4	5.335	0.147	3.124	0.131
6	5.808	0.074	3.396	0.126
8	5.918	0.068	3.530	0.111
10	6.073	0.040	3.627	0.074
15	6.113	0.055	3.744	0.107
20	6.228	0.123	3.828	0.121
25	6.649	0.253	3.956	0.151
30	7.389	0.228	4.097	0.176
40	7.980	0.422	4.477	0.350

We do not have previously published focal mechanism solutions for our selected events, although there have been previous focal mechanism studies in the TVZ (see e.g. Hurst et al. 2002), to which we may compare the fault types of our solutions. In addition we use HASH by Hardebeck & Shearer (2002) as a means of comparing solutions from an established focal mechanism estimation method to our MAP solutions for the selected events. Table 5.3 contains a summary of the results for our seven selected events: the estimates of  $\Theta$  for the VMK and VMU cases, the HASH estimate of  $\Theta$ , and the angular differences between the VMU MAP estimates (our maximal model) and the VMK and HASH estimates.

In the velocity model unknown case we slightly alter the stereonets, as we have sampled hypocentre locations from  $\sim 1000$  runs of NonLinLoc. This means that duplicate hypocentre locations, and therefore points on the focal sphere, can occur. Instead of simply overplotting, we take a grid of points over spherical coordinates  $(\phi, \theta)$ , and count the number of points in each cell. This gives a probability of a first motion for each cell, from which we can plot the contours of the first motions.

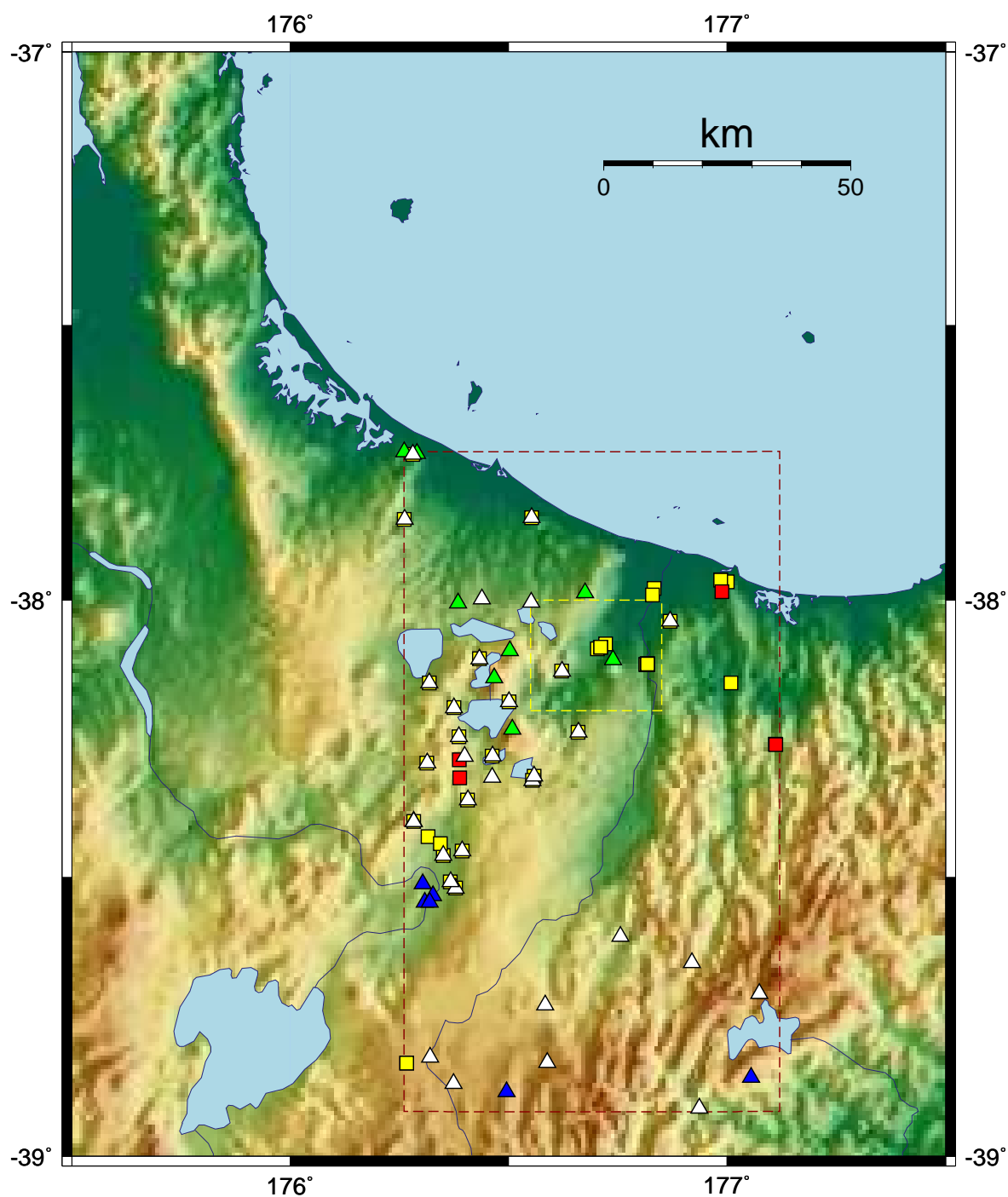


Figure 5.27: Map of the Kawerau area. Lines show the boundary of Clarke (2007)'s selected earthquakes (yellow line) and the boundary imposed on the velocity models in this study (dark red line). Symbols show the national seismograph network (red squares), strong motion network (yellow squares), Rotorua network (green triangles), temporary stations (white triangles), and other networks (blue triangles).



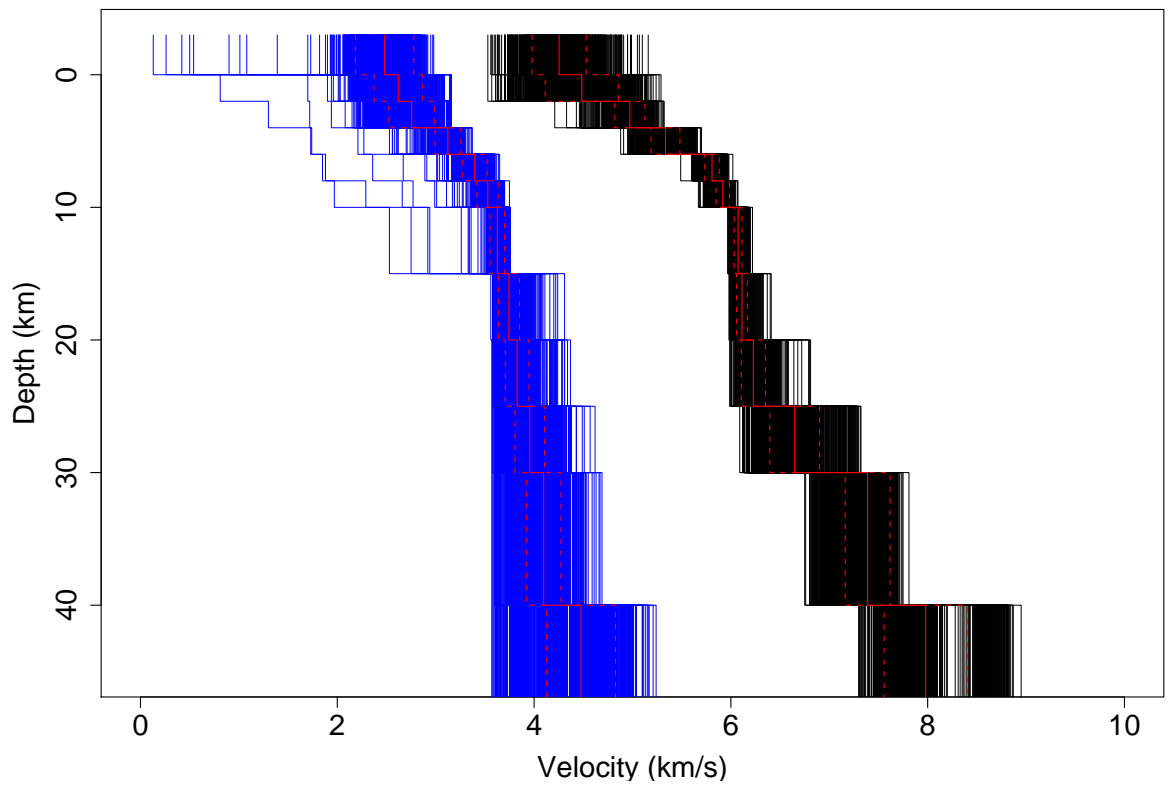


Figure 5.28: Plot of 1000 P- (black) and S-wave (blue) velocity models for Kawerau, as obtained by (Clarke 2007) using Velest. The solid red line marks the mean velocity in each layer, while the dashed lines mark one standard deviation from the mean.

Table 5.3: Summary table of results for the selected Kawerau events.

	Velocity Model Unknown			Velocity Model Known			HASH	Angular difference $a$	
CUSPID	MAP $\hat{\Theta}$	$\hat{\kappa}$	$\sigma_{\Theta}$	MAP $\hat{\Theta}$	$\hat{\kappa}$	$\sigma_{\Theta}$	$\hat{\Theta}$	HASH $\leftrightarrow$ VMU	VMK $\leftrightarrow$ VMU
<b>1.</b> 731019	(127.20°, 46.83°, 357.70°)	2.54	29.62°	(130.88°, 46.28°, 1.28°)	2.54	29.57°	(205°, 57°, −161°)	39.16°	2.92°
<b>2.</b> 745516	(219.13°, 70.46°, 213.00°)	2.58	29.35°	(219.60°, 71.43°, 214.61°)	2.71	28.52°	(212°, 83°, −172°)	27.34°	1.81°
<b>3.</b> 788921	(250.38°, 95.81°, 170.73°)	2.30	31.34°	(71.99°, 77.54°, 200.30°)	2.69	28.65°	(78°, 67°, −167°)	18.78°	12.75°
<b>4.</b> 802105	(225.16°, 62.10°, 154.55°)	2.30	31.32°	(327.83°, 67.96°, 31.52°)	3.14	26.25°	(48°, 82°, −155°)	36.01°	1.12°
<b>5.</b> 802106	(224.40°, 65.06°, 181.43°)	2.32	31.18°	(132.93°, 87.61°, 334.19°)	2.59	29.28°	(39°, 78°, 172°)	37.98°	1.63°
<b>6.</b> 1697233	(233.93°, 29.02°, 264.41°)	2.10	32.95°	(60.65°, 62.05°, 273.25°)	2.75	28.29°	(62°, 20°, −106°)	45.75°	0.96°
<b>7.</b> 1728730	(273.08°, 32.38°, 312.36°)	2.00	33.86°	(278.70°, 35.65°, 317.51°)	2.36	30.89°	(31°, 84°, 145°)	99.01°	4.55°

## 1. EVENT 731019 – Velocity model known

Date	Time	Latitude	Longitude	Depth	Magnitude
06/02/1995	11:50:25.69	-38.101	176.689	-3.000	2.9

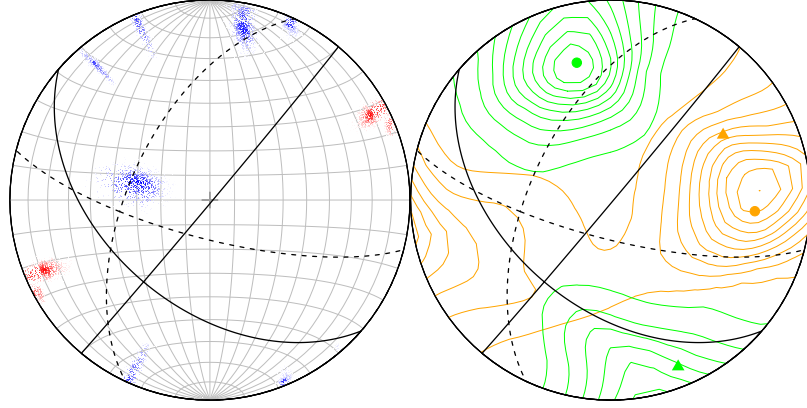


Figure 5.29: Stereonet for event 731019 (left). The solid dark line indicates the MAP focal mechanism solution  $[\Theta = (130.88^\circ, 46.28^\circ, 1.28^\circ)]$  while the dotted line indicates the solution given by HASH. Angular difference between the two solutions =  $41.852^\circ$ . Blue points are compressions, red points are dilatations. PT contour plot for event 731019 (right). Orange denotes the P-axis, green the T-axis. The circle denotes the MAP estimate, while the triangle denotes the solution given by HASH.

### Parameter Estimates

8-mode Matrix Fisher Distribution:

$$\begin{aligned} \hat{\mathbf{F}} &= \begin{bmatrix} -2.458 & 4.127 & -6.982 \\ 2.044 & 2.162 & -3.319 \\ -0.235 & 0.632 & -1.818 \end{bmatrix}, & \hat{\mathbf{M}} &= \begin{bmatrix} -0.542 & 0.557 & -0.630 \\ 0.840 & 0.390 & -0.378 \\ 0.035 & -0.734 & -0.679 \end{bmatrix}, \\ \hat{\mathbf{U}} &= \begin{bmatrix} 0.158 & 0.986 & 0.986 \\ -0.501 & 0.126 & 0.856 \\ 0.851 & -0.109 & 0.514 \end{bmatrix}, & \hat{\mathbf{D}} &= \begin{bmatrix} 9.326 & 0 & 0 \\ 0 & 2.886 & 0 \\ 0 & 0 & 0.390 \end{bmatrix} \end{aligned}$$

8-mode Matrix Fisher Distribution with scalar concentration:

$$\hat{\mathbf{M}} = \begin{bmatrix} -0.552 & 0.340 & -0.761 \\ 0.832 & 0.285 & -0.476 \\ 0.055 & -0.896 & -0.679 \end{bmatrix}, \hat{\kappa} = 2.545$$

## 1. EVENT 731019 – Velocity model known

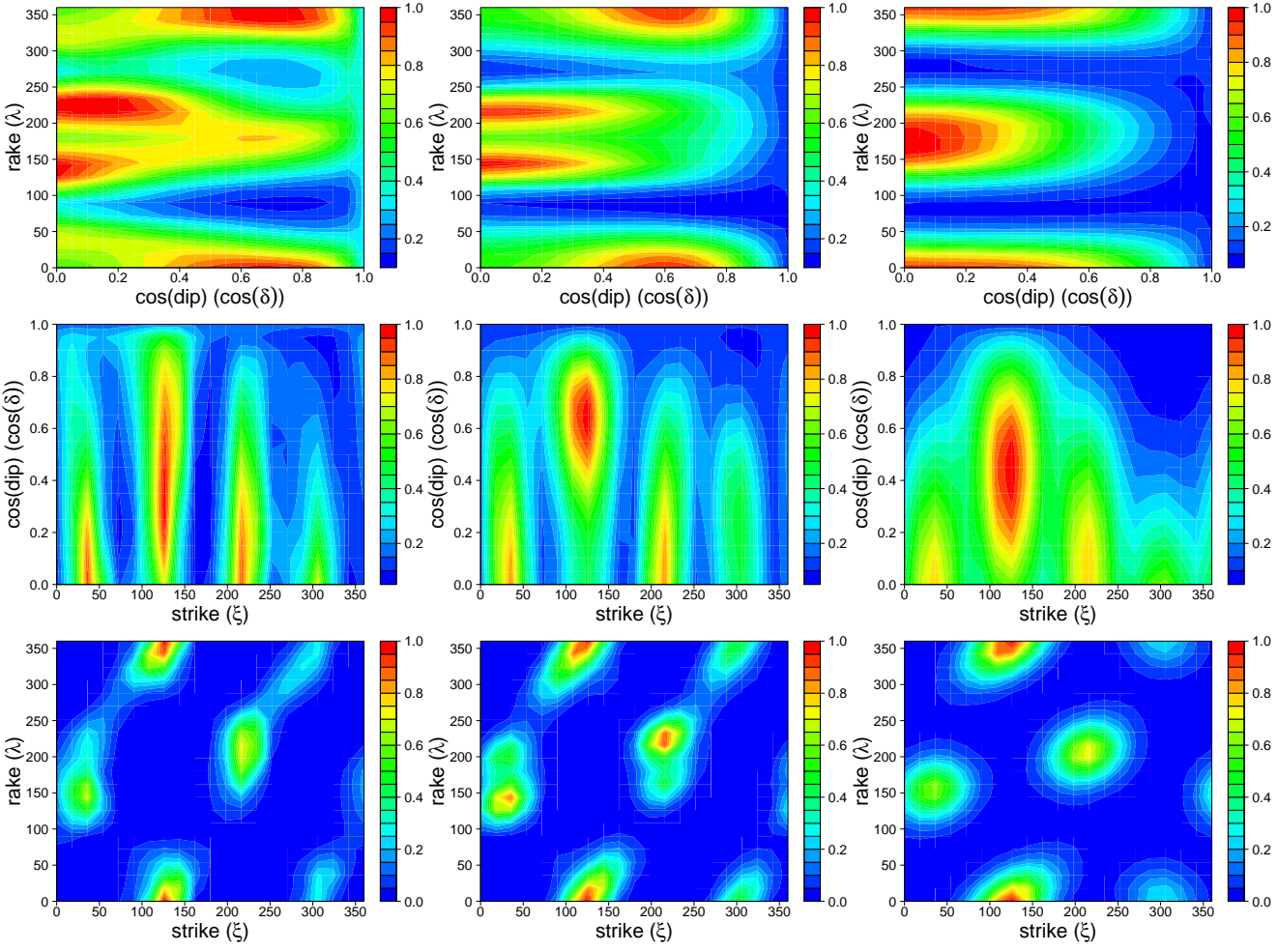


Figure 5.30: Marginal PDF plots for Bayesian posterior PDF  $P(\Theta|\mathbf{d}, \omega)$  (left), fitted 8-mode Matrix Fisher distribution  $P(\Theta|\hat{\mathbf{F}})$  (centre) and fitted 8-mode Matrix Fisher distribution with scalar concentration parameter  $P(\Theta|\hat{\mathbf{M}}, \hat{\kappa})$  (right) for event 731019.

## 1. EVENT 731019 – Velocity model unknown

Date	Time	Latitude	Longitude	Depth	Magnitude
06/02/1995	11:50:25.69	-38.101	176.689	0.000	2.9

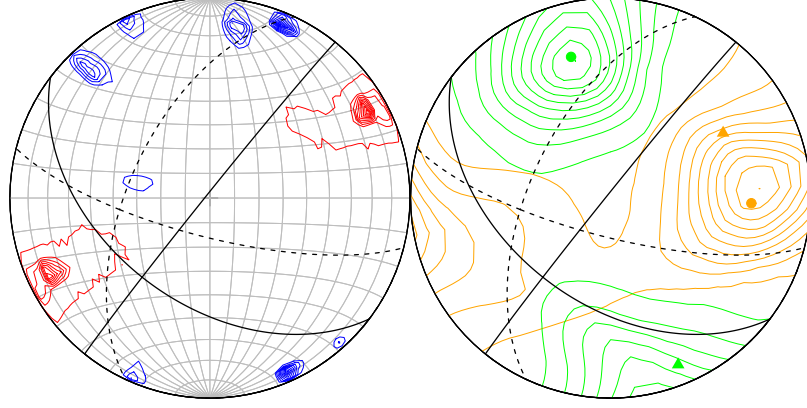


Figure 5.31: Stereonet for event 731019 (left). The solid dark line indicates the MAP focal mechanism solution  $[\Theta = (127.20^\circ, 46.83^\circ, 357.70^\circ)]$  while the dotted line indicates the solution given by HASH. Angular difference between the two solutions =  $39.156^\circ$ . Blue points are compressions, red points are dilatations. PT contour plot for event 731019 (right). Orange denotes the P-axis, green the T-axis. The circle denotes the MAP estimate, while the triangle denotes the solution given by HASH.

### Parameter Estimates

8-mode Matrix Fisher Distribution:

$$\hat{\mathbf{F}} = \begin{bmatrix} -2.363 & 4.081 & -6.857 \\ 2.075 & 2.188 & -3.334 \\ -0.186 & 0.646 & -1.820 \end{bmatrix}, \quad \hat{\mathbf{M}} = \begin{bmatrix} -0.543 & 0.557 & -0.629 \\ 0.839 & 0.399 & -0.371 \\ 0.045 & -0.729 & -0.683 \end{bmatrix},$$

$$\hat{\mathbf{U}} = \begin{bmatrix} 0.145 & 0.988 & 0.988 \\ -0.505 & 0.119 & 0.855 \\ 0.851 & -0.098 & 0.516 \end{bmatrix}, \quad \hat{\mathbf{D}} = \begin{bmatrix} 9.205 & 0 & 0 \\ 0 & 2.889 & 0 \\ 0 & 0 & 0.387 \end{bmatrix}$$

8-mode Matrix Fisher Distribution with scalar concentration:

$$\hat{\mathbf{M}} = \begin{bmatrix} -0.552 & 0.340 & -0.761 \\ 0.831 & 0.295 & -0.471 \\ 0.064 & -0.893 & -0.683 \end{bmatrix}, \quad \hat{\kappa} = 2.538$$

## 1. EVENT 731019 – Velocity model unknown

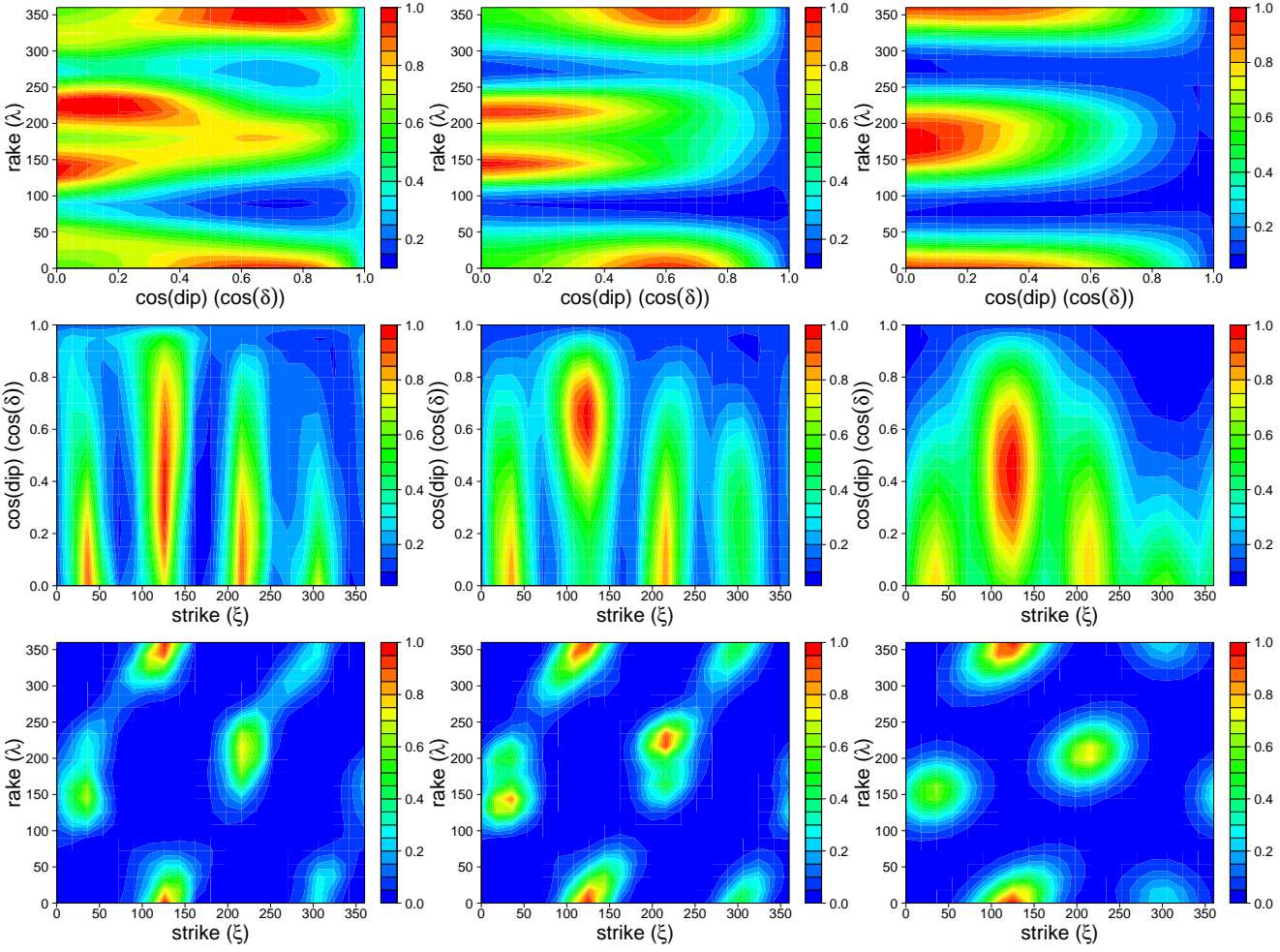


Figure 5.32: Marginal PDF plots for Bayesian posterior PDF  $P(\Theta|\mathbf{d}, \omega)$  (left), fitted 8-mode Matrix Fisher distribution  $P(\Theta|\hat{\mathbf{F}})$  (centre) and fitted 8-mode Matrix Fisher distribution with scalar concentration parameter  $P(\Theta|\hat{\mathbf{M}}, \hat{\kappa})$  (right) for event 731019.

## 2. EVENT 745516 – Velocity model known

Date	Time	Latitude	Longitude	Depth	Magnitude
18/02/1995	03:48:27.91	-38.102	176.707	2.000	2.5

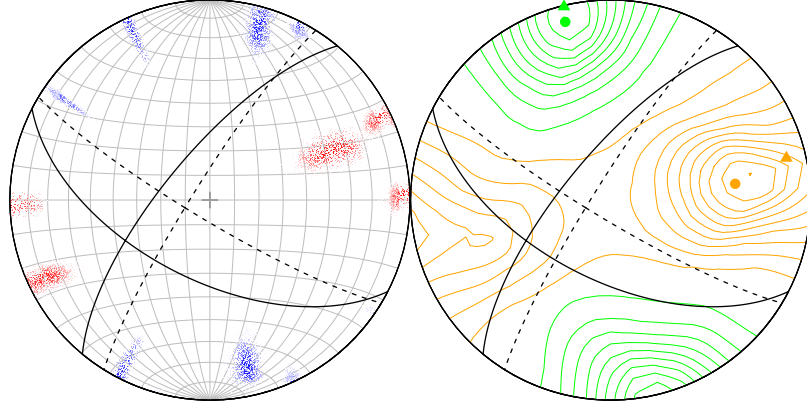


Figure 5.33: Stereonet for event 745516 (left). The solid dark line indicates the MAP focal mechanism solution  $[\Theta = (219.60^\circ, 71.43^\circ, 214.61^\circ)]$  while the dotted line indicates the solution given by HASH. Angular difference between the two solutions =  $28.398^\circ$ . Blue points are compressions, red points are dilatations. PT contour plot for event 745516 (right). Orange denotes the P-axis, green the T-axis. The circle denotes the MAP estimate, while the triangle denotes the solution given by HASH.

### Parameter Estimates

8-mode Matrix Fisher Distribution:

$$\begin{aligned} \hat{\mathbf{F}} &= \begin{bmatrix} 7.350 & -3.352 & 2.966 \\ 4.250 & -3.585 & -3.211 \\ 0.954 & -0.405 & -0.309 \end{bmatrix}, & \hat{\mathbf{M}} &= \begin{bmatrix} 0.748 & -0.298 & 0.593 \\ 0.346 & -0.588 & -0.731 \\ 0.567 & 0.752 & -0.336 \end{bmatrix}, \\ \hat{\mathbf{U}} &= \begin{bmatrix} 0.871 & 0.068 & 0.068 \\ -0.486 & 0.276 & -0.829 \\ 0.078 & 0.959 & 0.273 \end{bmatrix}, & \hat{\mathbf{D}} &= \begin{bmatrix} 9.806 & 0 & 0 \\ 0 & 4.500 & 0 \\ 0 & 0 & 0.215 \end{bmatrix} \end{aligned}$$

8-mode Matrix Fisher Distribution with scalar concentration:

$$\hat{\mathbf{M}} = \begin{bmatrix} 0.824 & -0.117 & 0.555 \\ 0.507 & -0.284 & -0.814 \\ 0.253 & 0.952 & -0.336 \end{bmatrix}, \hat{\kappa} = 2.713$$

## 2. EVENT 745516 – Velocity model known

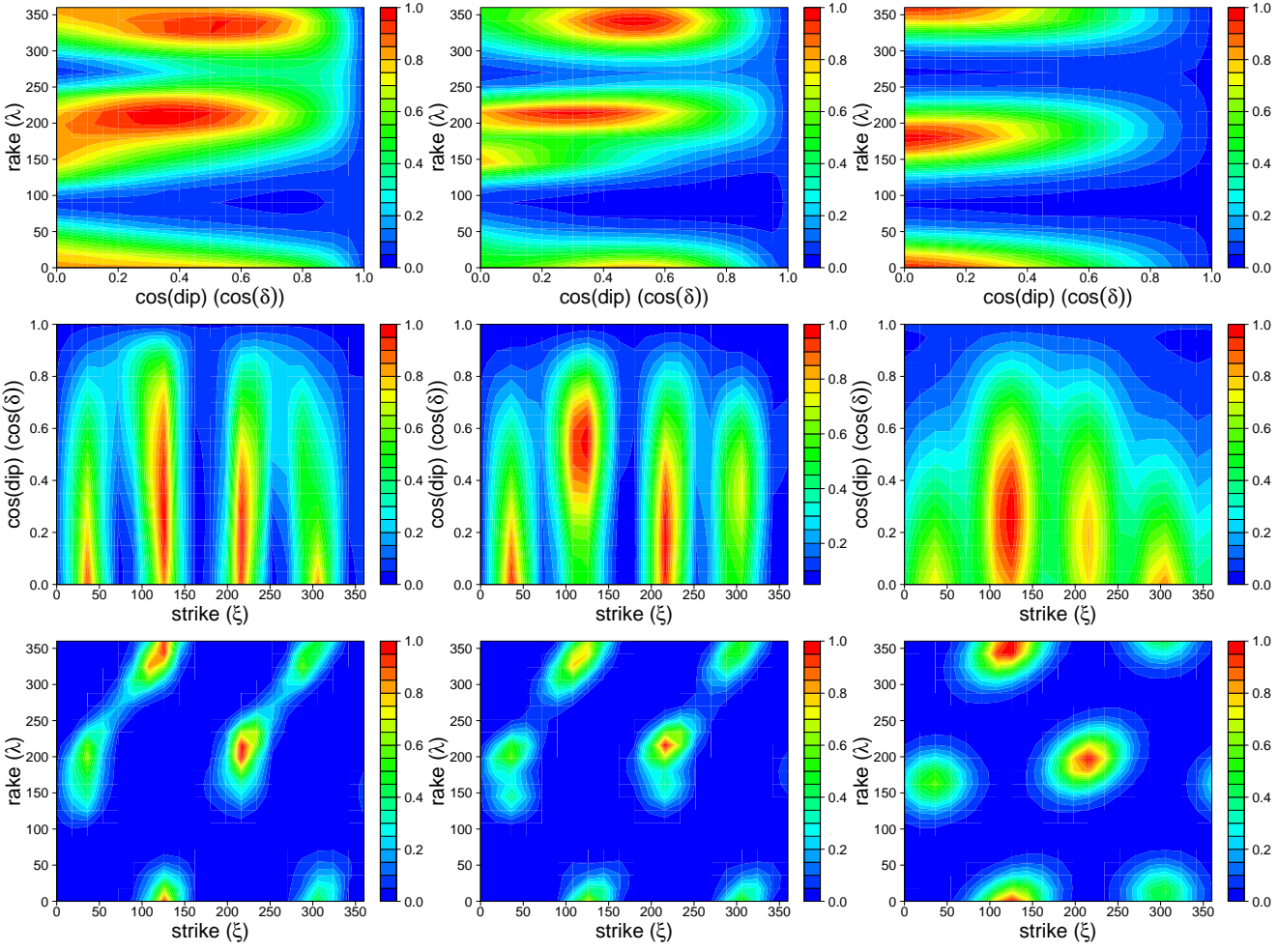


Figure 5.34: Marginal PDF plots for Bayesian posterior PDF  $P(\Theta|\mathbf{d}, \omega)$  (left), fitted 8-mode Matrix Fisher distribution  $P(\Theta|\hat{\mathbf{F}})$  (centre) and fitted 8-mode Matrix Fisher distribution with scalar concentration parameter  $P(\Theta|\hat{\mathbf{M}}, \hat{\kappa})$  (right) for event 745516.



## 2. EVENT 745516 – Velocity model unknown

Date	Time	Latitude	Longitude	Depth	Magnitude
18/02/1995	03:48:27.91	-38.102	176.707	4.000	2.5

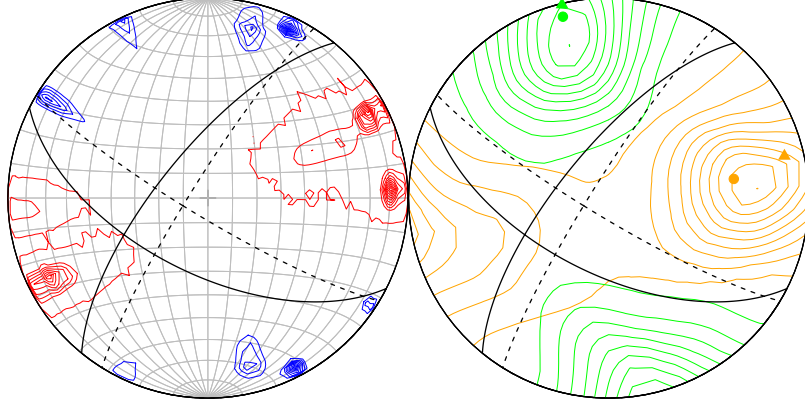


Figure 5.35: Stereonet for event 745516 (left). The solid dark line indicates the MAP focal mechanism solution  $[\Theta = (219.13^\circ, 70.46^\circ, 213.00^\circ)]$  while the dotted line indicates the solution given by HASH. Angular difference between the two solutions =  $27.341^\circ$ . Blue points are compressions, red points are dilatations. PT contour plot for event 745516 (right). Orange denotes the P-axis, green the T-axis. The circle denotes the MAP estimate, while the triangle denotes the solution given by HASH.

### Parameter Estimates

8-mode Matrix Fisher Distribution:

$$\begin{aligned} \hat{\mathbf{F}} &= \begin{bmatrix} 6.162 & -3.302 & 1.840 \\ 4.067 & -3.188 & -2.859 \\ 1.299 & -0.392 & -0.033 \end{bmatrix}, & \hat{\mathbf{M}} &= \begin{bmatrix} 0.686 & -0.425 & 0.590 \\ 0.365 & -0.501 & -0.785 \\ 0.629 & 0.754 & -0.189 \end{bmatrix}, \\ \hat{\mathbf{U}} &= \begin{bmatrix} 0.853 & 0.133 & 0.133 \\ -0.521 & 0.164 & -0.838 \\ -0.029 & 0.978 & 0.209 \end{bmatrix}, & \hat{\mathbf{D}} &= \begin{bmatrix} 8.766 & 0 & 0 \\ 0 & 3.468 & 0 \\ 0 & 0 & 0.337 \end{bmatrix} \end{aligned}$$

8-mode Matrix Fisher Distribution with scalar concentration:

$$\hat{\mathbf{M}} = \begin{bmatrix} 0.792 & -0.254 & 0.555 \\ 0.486 & -0.290 & -0.825 \\ 0.371 & 0.923 & -0.189 \end{bmatrix}, \hat{\kappa} = 2.580$$

## 2. EVENT 745516 – Velocity model unknown

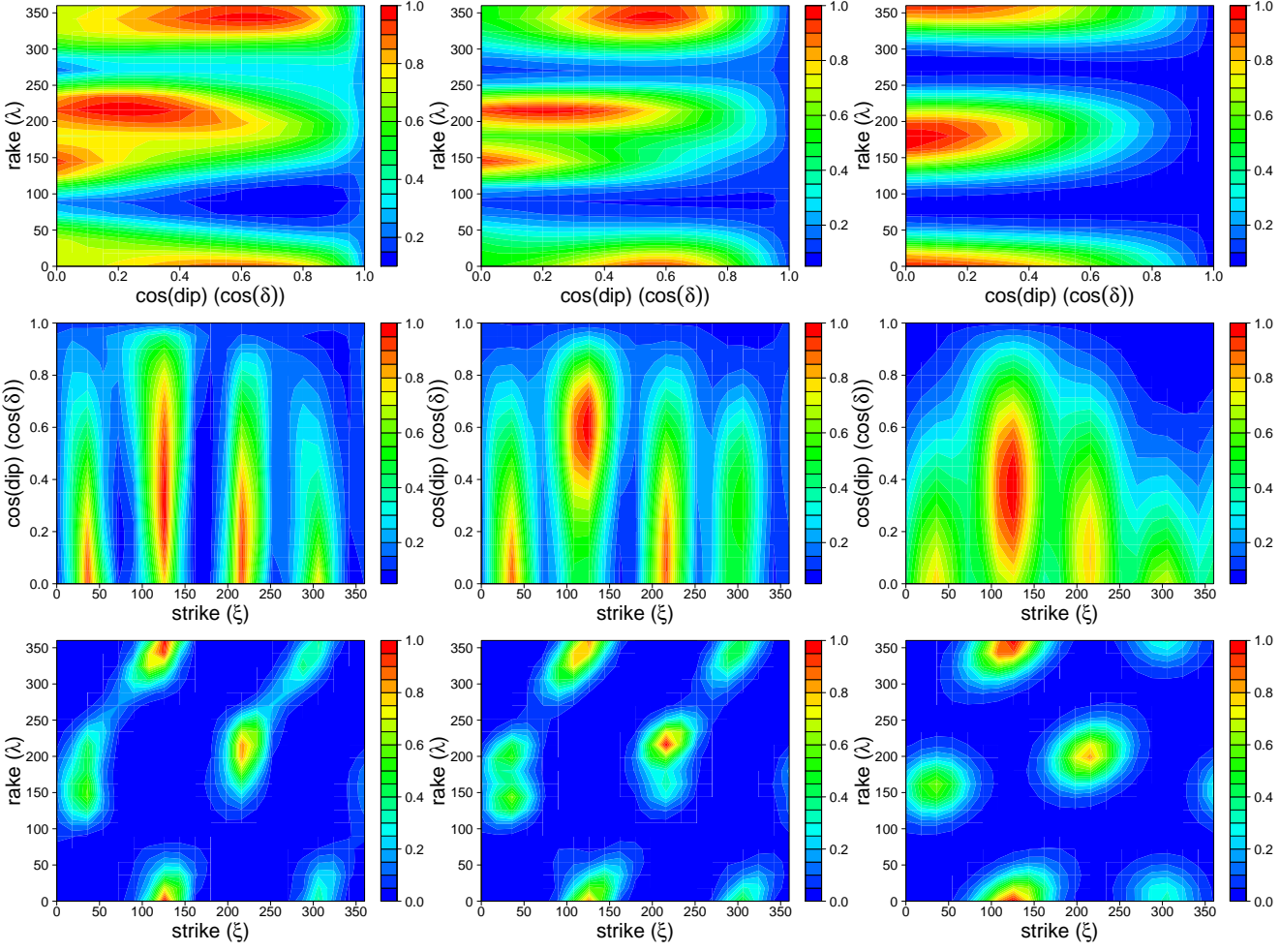


Figure 5.36: Marginal PDF plots for Bayesian posterior PDF  $P(\Theta|\mathbf{d}, \omega)$  (left), fitted 8-mode Matrix Fisher distribution  $P(\Theta|\hat{\mathbf{F}})$  (centre) and fitted 8-mode Matrix Fisher distribution with scalar concentration parameter  $P(\Theta|\hat{\mathbf{M}}, \hat{\kappa})$  (right) for event 745516.

### 3. EVENT 788921 – Velocity model known

Date	Time	Latitude	Longitude	Depth	Magnitude
01/04/1995	11:28:04.73	-38.178	176.599	6.000	2.4

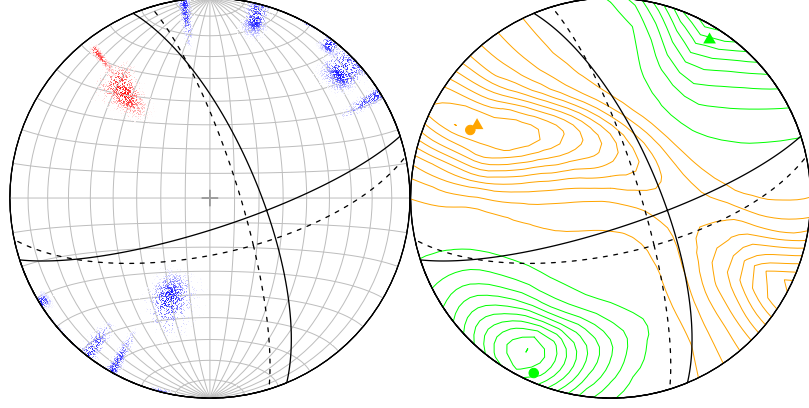


Figure 5.37: Stereonet for event 788921 (left). The solid dark line indicates the MAP focal mechanism solution  $[\Theta = (71.99^\circ, 77.54^\circ, 200.30^\circ)]$  while the dotted line indicates the solution given by HASH. Angular difference between the two solutions =  $15.068^\circ$ . Blue points are compressions, red points are dilatations. PT contour plot for event 788921 (right). Orange denotes the P-axis, green the T-axis. The circle denotes the MAP estimate, while the triangle denotes the solution given by HASH.

#### Parameter Estimates

8-mode Matrix Fisher Distribution:

$$\hat{\mathbf{F}} = \begin{bmatrix} -3.783 & -5.117 & -10.223 \\ -3.307 & 0.314 & -0.500 \\ 0.996 & -0.716 & -1.052 \end{bmatrix}, \quad \hat{\mathbf{M}} = \begin{bmatrix} -0.276 & -0.532 & -0.800 \\ -0.825 & 0.558 & -0.086 \\ 0.492 & 0.637 & -0.594 \end{bmatrix},$$

$$\hat{\mathbf{U}} = \begin{bmatrix} -0.334 & 0.931 & 0.931 \\ -0.419 & -0.289 & -0.861 \\ -0.845 & -0.225 & 0.486 \end{bmatrix}, \quad \hat{\mathbf{D}} = \begin{bmatrix} 12.150 & 0 & 0 \\ 0 & 3.358 & 0 \\ 0 & 0 & 0.047 \end{bmatrix}$$

8-mode Matrix Fisher Distribution with scalar concentration:

$$\hat{\mathbf{M}} = \begin{bmatrix} -0.237 & -0.144 & -0.961 \\ -0.893 & 0.421 & 0.157 \\ 0.382 & 0.896 & -0.594 \end{bmatrix}, \quad \hat{\kappa} = 2.691$$

### 3. EVENT 788921 – Velocity model known

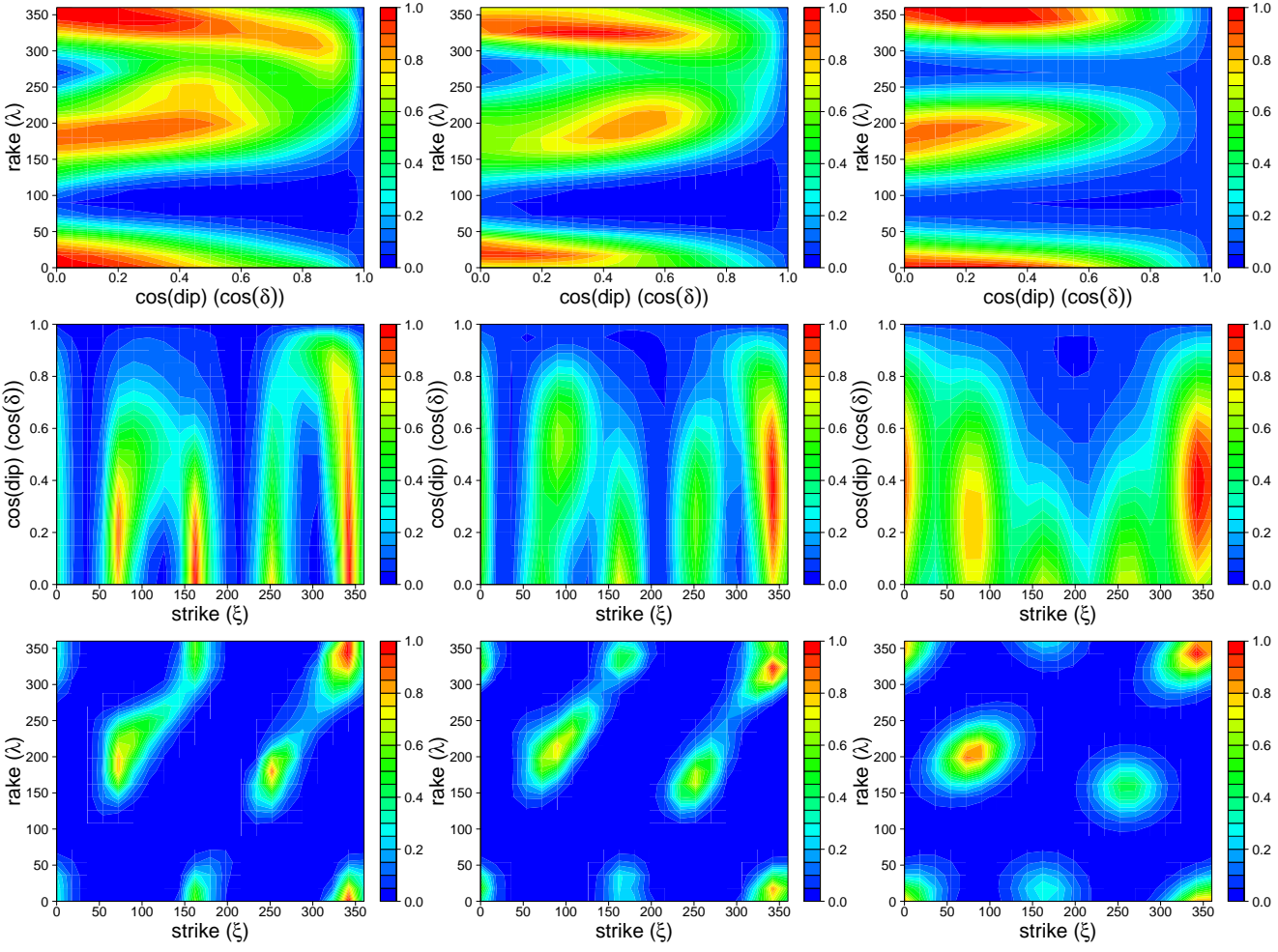


Figure 5.38: Marginal PDF plots for Bayesian posterior PDF  $P(\Theta|\mathbf{d}, \omega)$  (left), fitted 8-mode Matrix Fisher distribution  $P(\Theta|\hat{\mathbf{F}})$  (centre) and fitted 8-mode Matrix Fisher distribution with scalar concentration parameter  $P(\Theta|\hat{\mathbf{M}}, \hat{\kappa})$  (right) for event 788921.

### 3. EVENT 788921 – Velocity model unknown

Date	Time	Latitude	Longitude	Depth	Magnitude
01/04/1995	11:28:04.73	-38.178	176.599	8.000	2.4

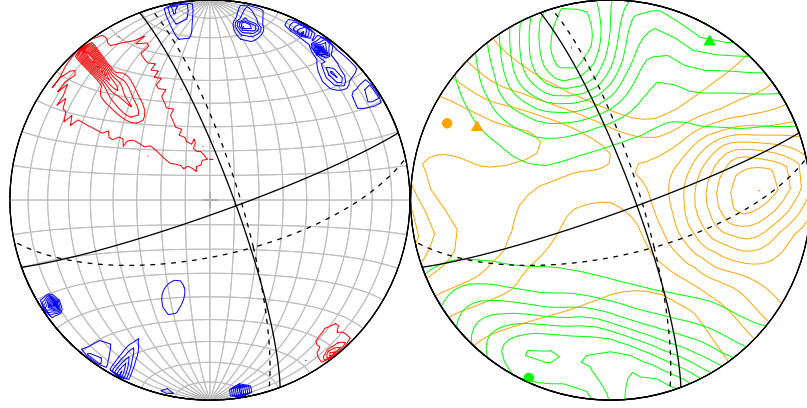


Figure 5.39: Stereonet for event 788921 (left). The solid dark line indicates the MAP focal mechanism solution  $[\Theta = (250.38^\circ, 95.81^\circ, 170.73^\circ)]$  while the dotted line indicates the solution given by HASH. Angular difference between the two solutions =  $18.778^\circ$ . Blue points are compressions, red points are dilatations. PT contour plot for event 788921 (right). Orange denotes the P-axis, green the T-axis. The circle denotes the MAP estimate, while the triangle denotes the solution given by HASH.

#### Parameter Estimates

8-mode Matrix Fisher Distribution:

$$\hat{\mathbf{F}} = \begin{bmatrix} 2.098 & -2.388 & 4.271 \\ 1.091 & 1.425 & -2.277 \\ 0.604 & 2.279 & -2.130 \end{bmatrix}, \quad \hat{\mathbf{M}} = \begin{bmatrix} 0.673 & -0.211 & 0.709 \\ 0.693 & -0.156 & -0.704 \\ 0.259 & 0.965 & 0.041 \end{bmatrix},$$

$$\hat{\mathbf{U}} = \begin{bmatrix} 0.154 & 0.960 & 0.960 \\ -0.546 & 0.280 & 0.789 \\ 0.823 & 0.006 & 0.568 \end{bmatrix}, \quad \hat{\mathbf{D}} = \begin{bmatrix} 6.410 & 0 & 0 \\ 0 & 2.320 & 0 \\ 0 & 0 & 0.619 \end{bmatrix}$$

8-mode Matrix Fisher Distribution with scalar concentration:

$$\hat{\mathbf{M}} = \begin{bmatrix} 0.678 & -0.118 & 0.726 \\ 0.715 & -0.124 & -0.688 \\ 0.171 & 0.985 & 0.041 \end{bmatrix}, \quad \hat{\kappa} = 2.297$$

### 3. EVENT 788921 – Velocity model unknown

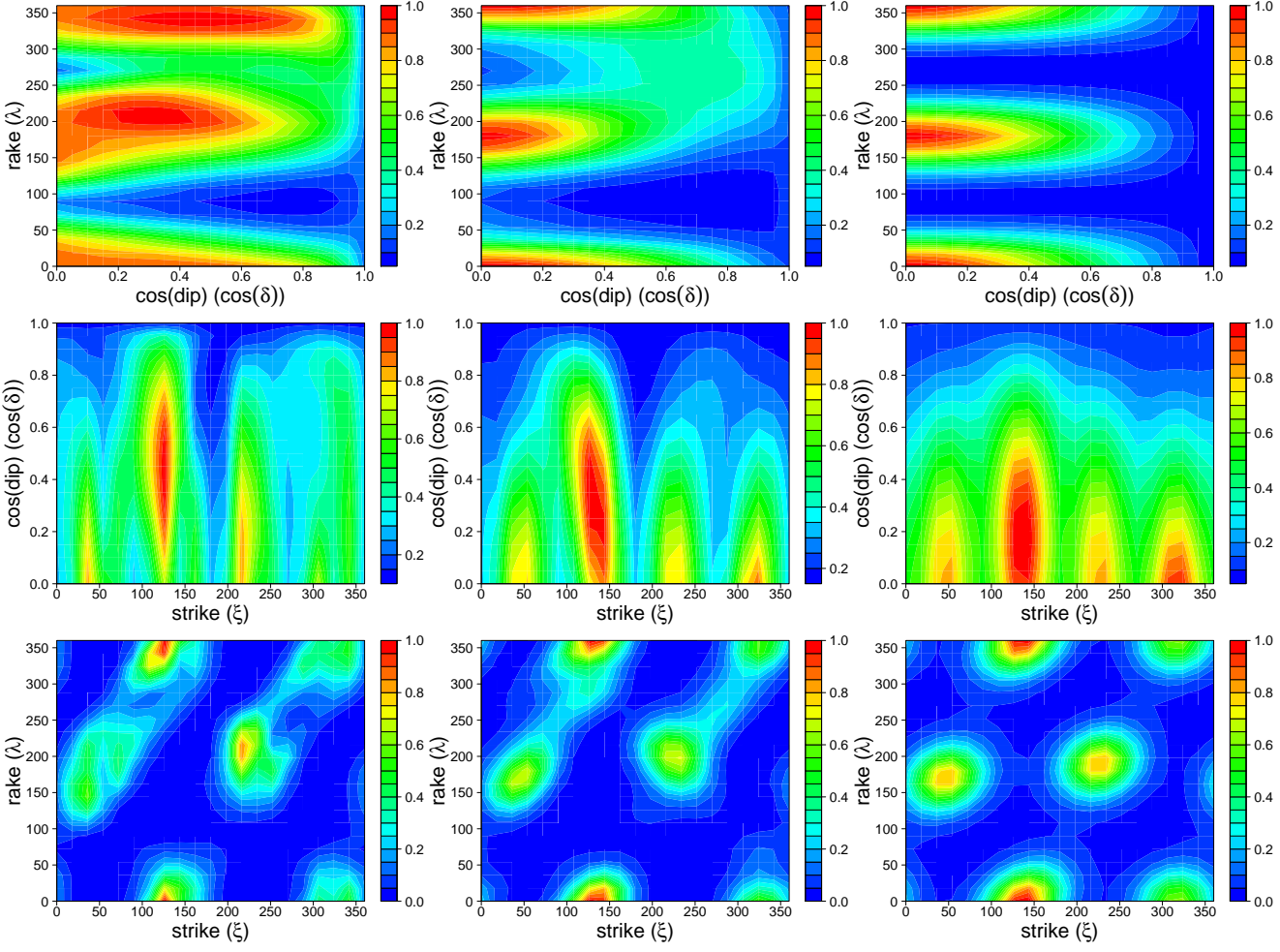


Figure 5.40: Marginal PDF plots for Bayesian posterior PDF  $P(\Theta|\mathbf{d}, \omega)$  (left), fitted 8-mode Matrix Fisher distribution  $P(\Theta|\hat{\mathbf{F}})$  (centre) and fitted 8-mode Matrix Fisher distribution with scalar concentration parameter  $P(\Theta|\hat{\mathbf{M}}, \hat{\kappa})$  (right) for event 788921.

#### 4. EVENT 802105 – Velocity model known

Date	Time	Latitude	Longitude	Depth	Magnitude
03/05/1995	16:10:22.26	-38.174	176.645	10.000	2.5

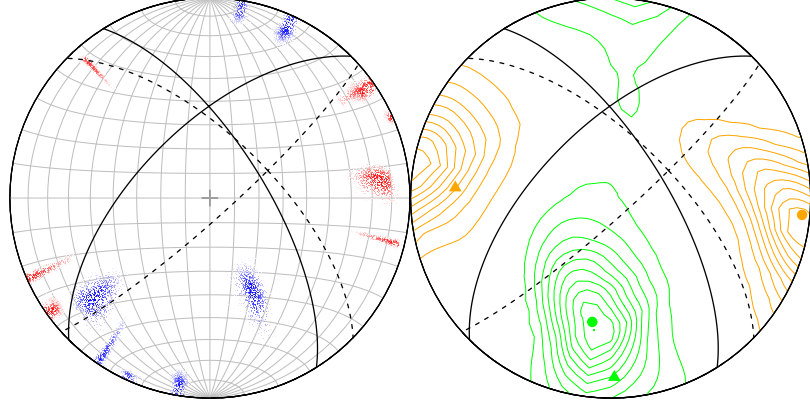


Figure 5.41: Stereonet for event 802105 (left). The solid dark line indicates the MAP focal mechanism solution [ $\Theta = (327.83^\circ, 67.96^\circ, 31.52^\circ)$ ] while the dotted line indicates the solution given by HASH. Angular difference between the two solutions =  $37.105^\circ$ . Blue points are compressions, red points are dilatations. PT contour plot for event 802105 (right). Orange denotes the P-axis, green the T-axis. The circle denotes the MAP estimate, while the triangle denotes the solution given by HASH.

#### Parameter Estimates

8-mode Matrix Fisher Distribution:

$$\hat{\mathbf{F}} = \begin{bmatrix} 2.452 & -3.531 & 5.088 \\ -3.806 & -3.843 & 7.715 \\ -1.575 & -0.132 & -0.743 \end{bmatrix}, \quad \hat{\mathbf{M}} = \begin{bmatrix} 0.645 & -0.667 & 0.373 \\ -0.620 & -0.173 & 0.765 \\ -0.446 & -0.725 & -0.525 \end{bmatrix},$$

$$\hat{\mathbf{U}} = \begin{bmatrix} -0.178 & 0.978 & 0.978 \\ -0.476 & -0.186 & 0.860 \\ 0.861 & 0.100 & 0.499 \end{bmatrix}, \quad \hat{\mathbf{D}} = \begin{bmatrix} 10.746 & 0 & 0 \\ 0 & 4.495 & 0 \\ 0 & 0 & 0.707 \end{bmatrix}$$

8-mode Matrix Fisher Distribution with scalar concentration:

$$\hat{\mathbf{M}} = \begin{bmatrix} 0.642 & -0.609 & 0.466 \\ -0.620 & -0.055 & 0.782 \\ -0.450 & -0.792 & -0.525 \end{bmatrix}, \quad \hat{\kappa} = 3.141$$



#### 4. EVENT 802105 – Velocity model known

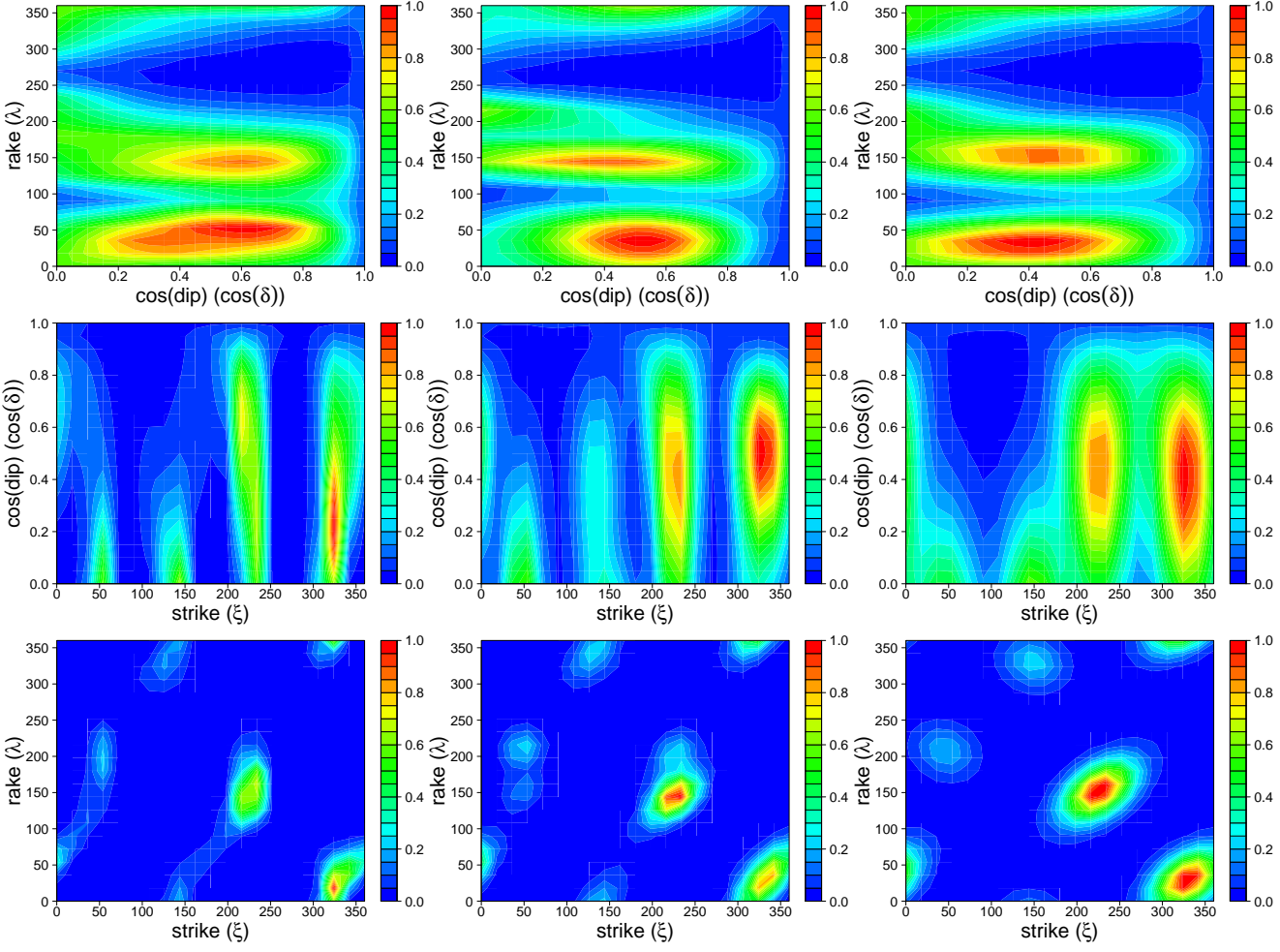


Figure 5.42: Marginal PDF plots for Bayesian posterior PDF  $P(\Theta|\mathbf{d}, \omega)$  (left), fitted 8-mode Matrix Fisher distribution  $P(\Theta|\hat{\mathbf{F}})$  (centre) and fitted 8-mode Matrix Fisher distribution with scalar concentration parameter  $P(\Theta|\hat{\mathbf{M}}, \hat{\kappa})$  (right) for event 802105.



#### 4. EVENT 802105 – Velocity model unknown

Date	Time	Latitude	Longitude	Depth	Magnitude
03/05/1995	16:10:22.26	-38.174	176.645	15.000	2.5

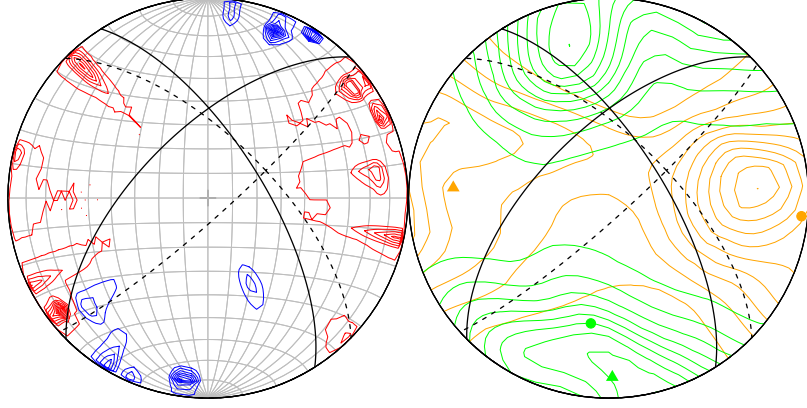


Figure 5.43: Stereonet for event 802105 (left). The solid dark line indicates the MAP focal mechanism solution  $[\Theta = (225.16^\circ, 62.10^\circ, 154.55^\circ)]$  while the dotted line indicates the solution given by HASH. Angular difference between the two solutions =  $36.007^\circ$ . Blue points are compressions, red points are dilatations. PT contour plot for event 802105 (right). Orange denotes the P-axis, green the T-axis. The circle denotes the MAP estimate, while the triangle denotes the solution given by HASH.

#### Parameter Estimates

8-mode Matrix Fisher Distribution:

$$\hat{\mathbf{F}} = \begin{bmatrix} 2.346 & -3.011 & 5.193 \\ 1.022 & 1.588 & -2.882 \\ 0.207 & 1.596 & -1.742 \end{bmatrix}, \quad \hat{\mathbf{M}} = \begin{bmatrix} 0.659 & -0.273 & 0.701 \\ 0.718 & -0.048 & -0.694 \\ 0.223 & 0.961 & 0.164 \end{bmatrix},$$

$$\hat{\mathbf{U}} = \begin{bmatrix} 0.209 & 0.977 & 0.977 \\ -0.505 & 0.152 & 0.850 \\ 0.837 & -0.153 & 0.525 \end{bmatrix}, \quad \hat{\mathbf{D}} = \begin{bmatrix} 7.375 & 0 & 0 \\ 0 & 2.100 & 0 \\ 0 & 0 & 0.485 \end{bmatrix}$$

8-mode Matrix Fisher Distribution with scalar concentration:

$$\hat{\mathbf{M}} = \begin{bmatrix} 0.677 & -0.078 & 0.731 \\ 0.721 & -0.123 & -0.681 \\ 0.143 & 0.989 & 0.164 \end{bmatrix}, \quad \hat{\kappa} = 2.299$$

#### 4. EVENT 802105 – Velocity model unknown

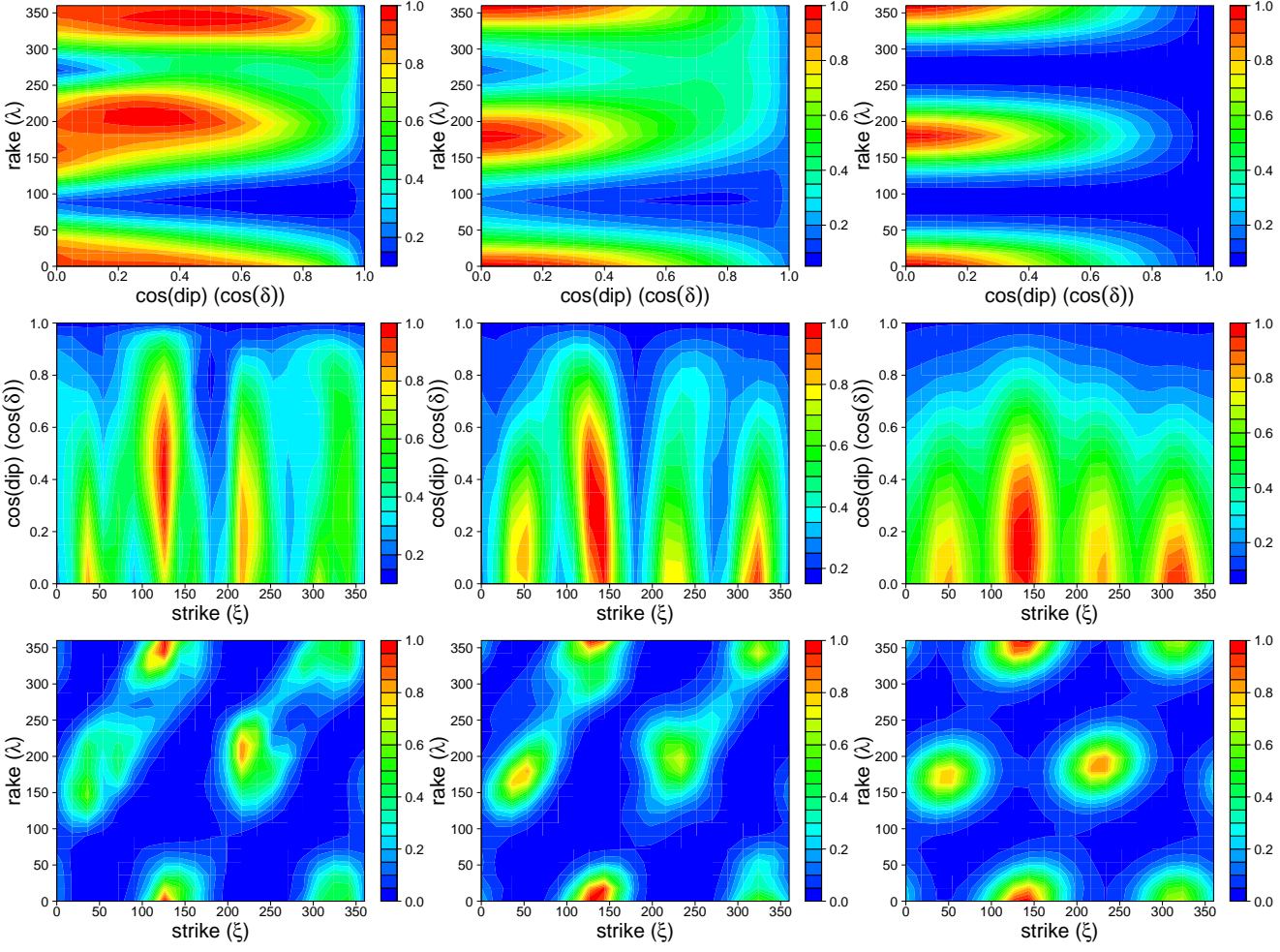


Figure 5.44: Marginal PDF plots for Bayesian posterior PDF  $P(\Theta|\mathbf{d}, \omega)$  (left), fitted 8-mode Matrix Fisher distribution  $P(\Theta|\hat{\mathbf{F}})$  (centre) and fitted 8-mode Matrix Fisher distribution with scalar concentration parameter  $P(\Theta|\hat{\mathbf{M}}, \hat{\kappa})$  (right) for event 802105.

## 5. EVENT 802106 – Velocity model known

Date	Time	Latitude	Longitude	Depth	Magnitude
03/05/1995	16:13:44.42	-38.174	176.645	20.000	2.4

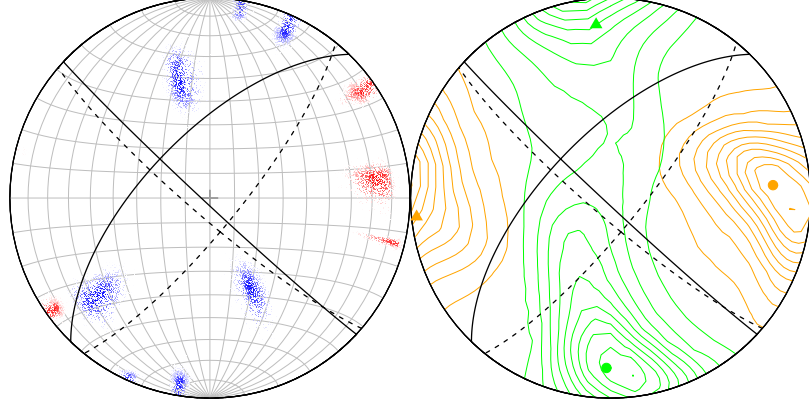


Figure 5.45: Stereonet for event 802106 (left). The solid dark line indicates the MAP focal mechanism solution  $[\Theta = (132.93^\circ, 87.61^\circ, 334.19^\circ)]$  while the dotted line indicates the solution given by HASH. Angular difference between the two solutions =  $38.563^\circ$ . Blue points are compressions, red points are dilatations. PT contour plot for event 802106 (right). Orange denotes the P-axis, green the T-axis. The circle denotes the MAP estimate, while the triangle denotes the solution given by HASH.

### Parameter Estimates

8-mode Matrix Fisher Distribution:

$$\begin{aligned} \hat{\mathbf{F}} &= \begin{bmatrix} -0.507 & 3.145 & -5.190 \\ 4.740 & 4.235 & -8.225 \\ 1.220 & 0.485 & -0.775 \end{bmatrix}, & \hat{\mathbf{M}} &= \begin{bmatrix} -0.624 & 0.596 & -0.505 \\ 0.678 & 0.090 & -0.730 \\ 0.390 & 0.798 & 0.461 \end{bmatrix}, \\ \hat{\mathbf{U}} &= \begin{bmatrix} -0.344 & 0.933 & 0.933 \\ -0.445 & -0.257 & 0.858 \\ 0.827 & 0.250 & 0.504 \end{bmatrix}, & \hat{\mathbf{D}} &= \begin{bmatrix} 11.767 & 0 & 0 \\ 0 & 2.994 & 0 \\ 0 & 0 & 0.155 \end{bmatrix} \end{aligned}$$

8-mode Matrix Fisher Distribution with scalar concentration:

$$\hat{\mathbf{M}} = \begin{bmatrix} -0.633 & -0.339 & -0.696 \\ 0.684 & 0.176 & -0.708 \\ 0.362 & -0.924 & 0.461 \end{bmatrix}, \hat{\kappa} = 2.591$$

## 5. EVENT 802106 – Velocity model known

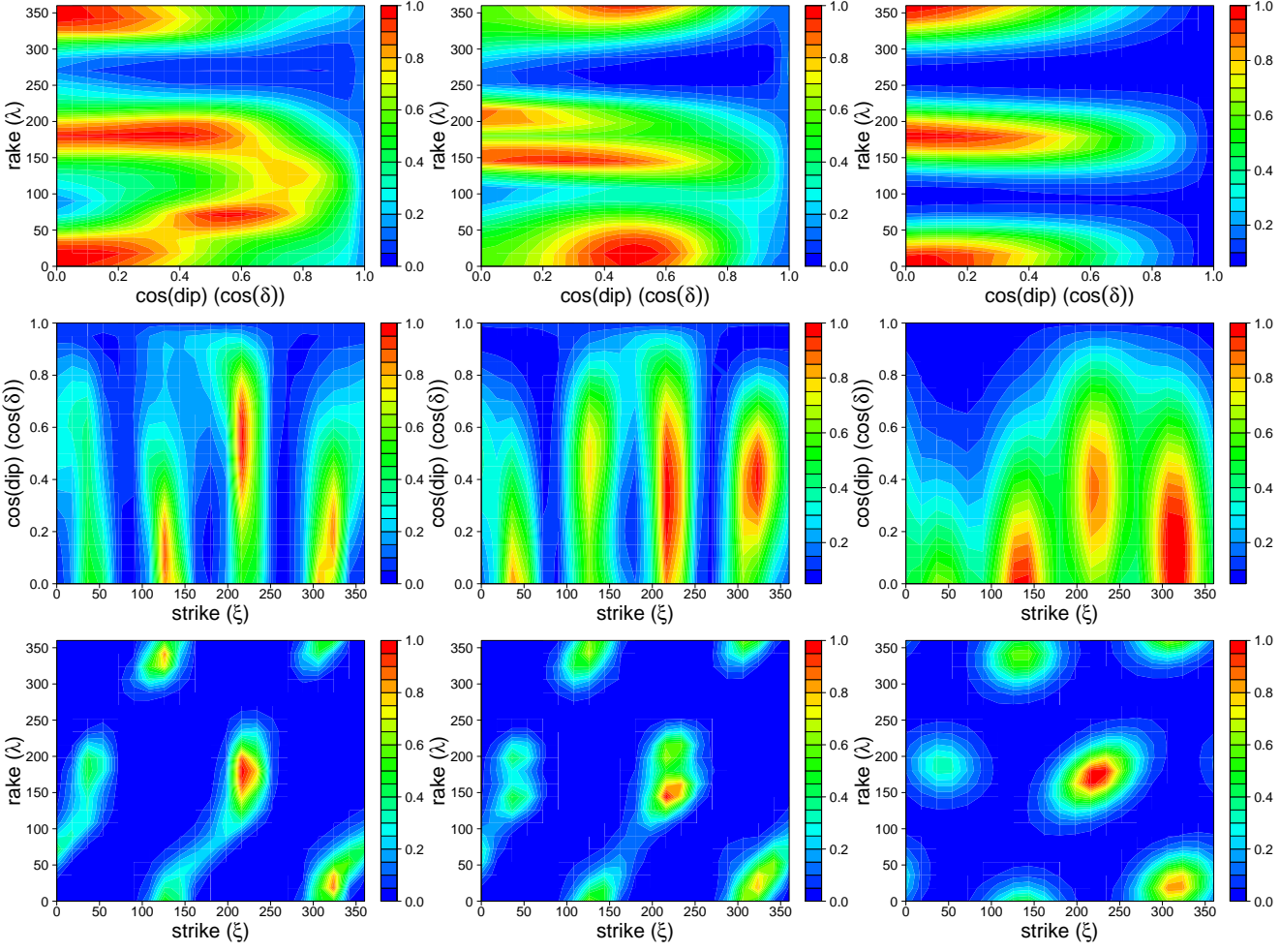


Figure 5.46: Marginal PDF plots for Bayesian posterior PDF  $P(\Theta|\mathbf{d}, \omega)$  (left), fitted 8-mode Matrix Fisher distribution  $P(\Theta|\hat{\mathbf{F}})$  (centre) and fitted 8-mode Matrix Fisher distribution with scalar concentration parameter  $P(\Theta|\hat{\mathbf{M}}, \hat{\kappa})$  (right) for event 802106.

## 5. EVENT 802106 – Velocity model unknown

Date	Time	Latitude	Longitude	Depth	Magnitude
03/05/1995	16:13:44.42	-38.174	176.645	25.000	2.4

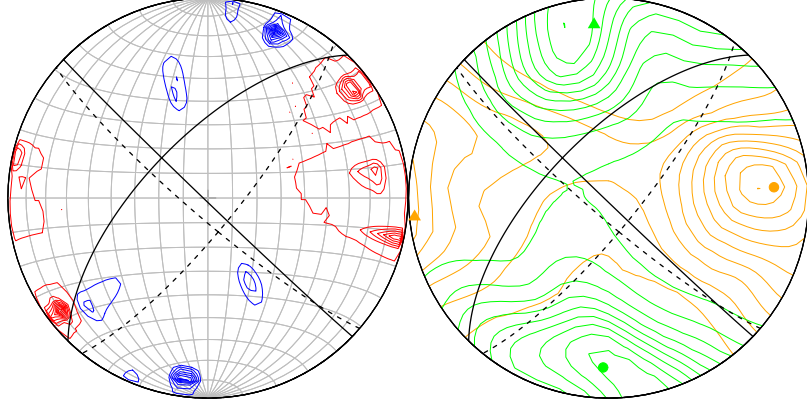


Figure 5.47: Stereonet for event 802106 (left). The solid dark line indicates the MAP focal mechanism solution  $[\Theta = (224.40^\circ, 65.06^\circ, 181.43^\circ)]$  while the dotted line indicates the solution given by HASH. Angular difference between the two solutions =  $37.975^\circ$ . Blue points are compressions, red points are dilatations. PT contour plot for event 802106 (right). Orange denotes the P-axis, green the T-axis. The circle denotes the MAP estimate, while the triangle denotes the solution given by HASH.

### Parameter Estimates

8-mode Matrix Fisher Distribution:

$$\hat{\mathbf{F}} = \begin{bmatrix} 2.423 & -3.058 & 5.097 \\ 1.157 & 1.525 & -3.044 \\ 0.094 & 1.395 & -1.624 \end{bmatrix}, \quad \hat{\mathbf{M}} = \begin{bmatrix} 0.667 & -0.287 & 0.688 \\ 0.711 & -0.032 & -0.703 \\ 0.224 & 0.957 & 0.183 \end{bmatrix},$$

$$\hat{\mathbf{U}} = \begin{bmatrix} 0.216 & 0.975 & 0.975 \\ -0.501 & 0.060 & 0.863 \\ 0.838 & -0.216 & 0.502 \end{bmatrix}, \quad \hat{\mathbf{D}} = \begin{bmatrix} 7.320 & 0 & 0 \\ 0 & 2.227 & 0 \\ 0 & 0 & 0.424 \end{bmatrix}$$

8-mode Matrix Fisher Distribution with scalar concentration:

$$\hat{\mathbf{M}} = \begin{bmatrix} 0.686 & -0.055 & 0.726 \\ 0.715 & -0.135 & -0.686 \\ 0.135 & 0.989 & 0.183 \end{bmatrix}, \quad \hat{\kappa} = 2.318$$

## 5. EVENT 802106 – Velocity model unknown

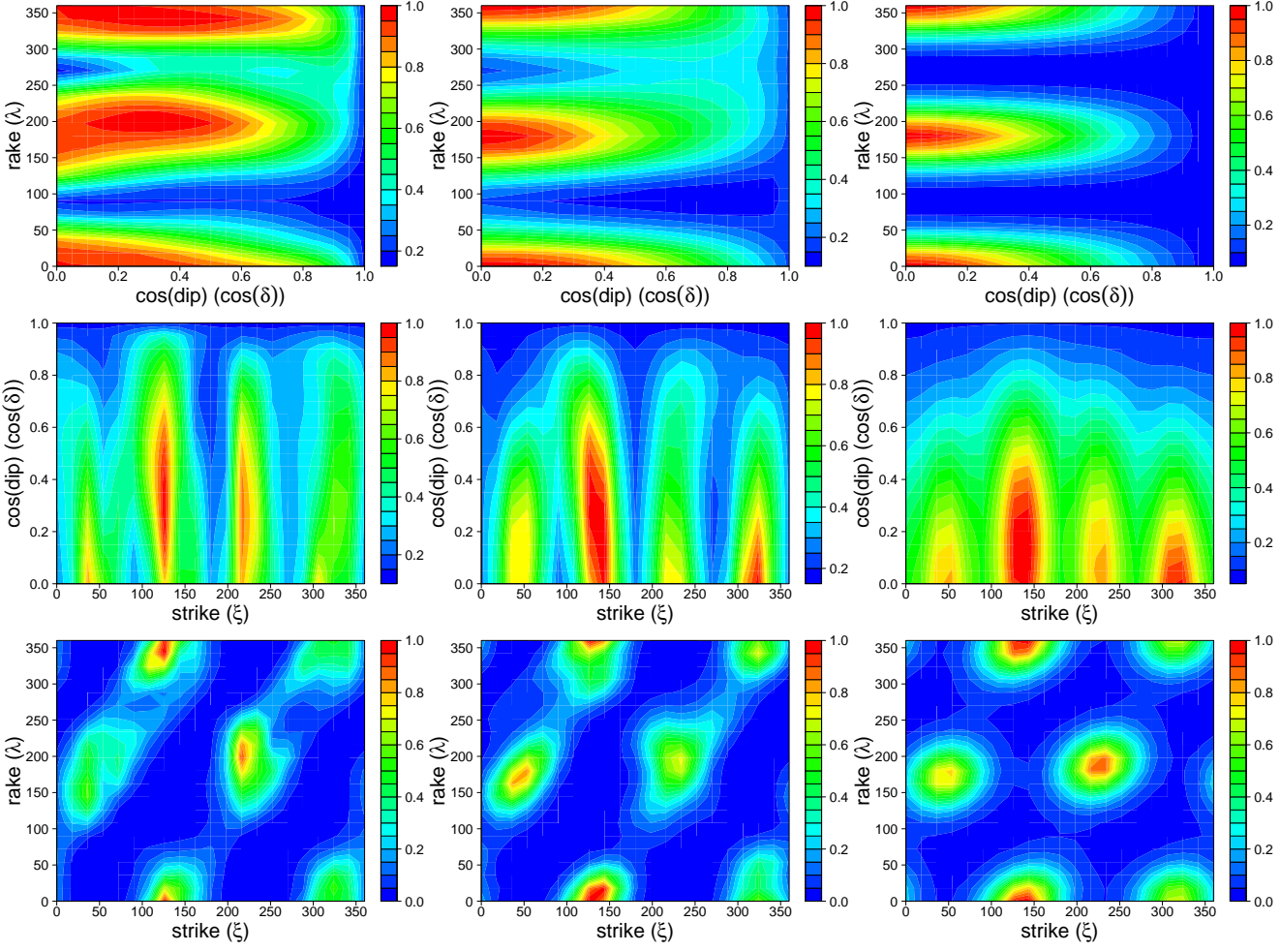


Figure 5.48: Marginal PDF plots for Bayesian posterior PDF  $P(\Theta|\mathbf{d}, \omega)$  (left), fitted 8-mode Matrix Fisher distribution  $P(\Theta|\hat{\mathbf{F}})$  (centre) and fitted 8-mode Matrix Fisher distribution with scalar concentration parameter  $P(\Theta|\hat{\mathbf{M}}, \hat{\kappa})$  (right) for event 802106.

## 6. EVENT 1697233 – Velocity model known

Date	Time	Latitude	Longitude	Depth	Magnitude
27/02/2001	11:51:15.60	-38.124	176.701	30.000	2.6

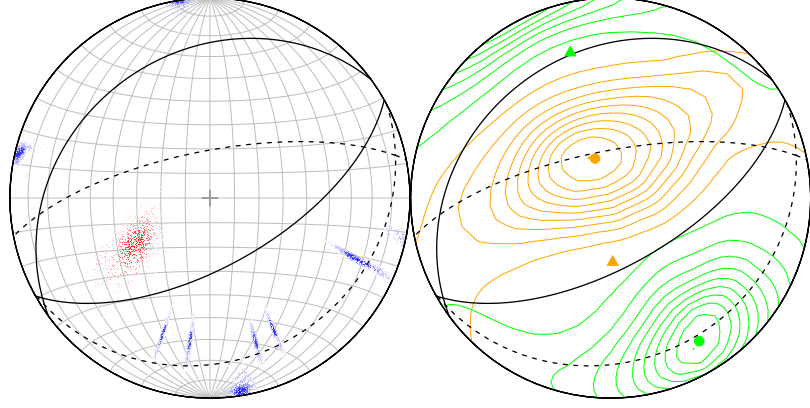


Figure 5.49: Stereonet for event 1697233 (left). The solid dark line indicates the MAP focal mechanism solution  $[\Theta = (60.65^\circ, 62.05^\circ, 273.25^\circ)]$  while the dotted line indicates the solution given by HASH. Angular difference between the two solutions =  $46.511^\circ$ . Blue points are compressions, red points are dilatations. PT contour plot for event 1697233 (right). Orange denotes the P-axis, green the T-axis. The circle denotes the MAP estimate, while the triangle denotes the solution given by HASH.

### Parameter Estimates

8-mode Matrix Fisher Distribution:

$$\begin{aligned} \hat{\mathbf{F}} &= \begin{bmatrix} -4.707 & 3.198 & -3.708 \\ 0.267 & 0.486 & 1.380 \\ 5.479 & -3.001 & -0.394 \end{bmatrix}, & \hat{\mathbf{M}} &= \begin{bmatrix} -0.292 & 0.562 & -0.774 \\ 0.388 & 0.809 & 0.441 \\ 0.874 & -0.172 & -0.454 \end{bmatrix}, \\ \hat{\mathbf{U}} &= \begin{bmatrix} 0.817 & -0.271 & -0.271 \\ -0.497 & 0.113 & -0.860 \\ 0.291 & 0.956 & -0.042 \end{bmatrix}, & \hat{\mathbf{D}} &= \begin{bmatrix} 8.769 & 0 & 0 \\ 0 & 3.190 & 0 \\ 0 & 0 & 0.671 \end{bmatrix} \end{aligned}$$

8-mode Matrix Fisher Distribution with scalar concentration:

$$\hat{\mathbf{M}} = \begin{bmatrix} -0.378 & 0.448 & -0.810 \\ 0.246 & 0.892 & 0.379 \\ 0.892 & -0.056 & -0.454 \end{bmatrix}, \hat{\kappa} = 2.752$$



## 6. EVENT 1697233 – Velocity model known

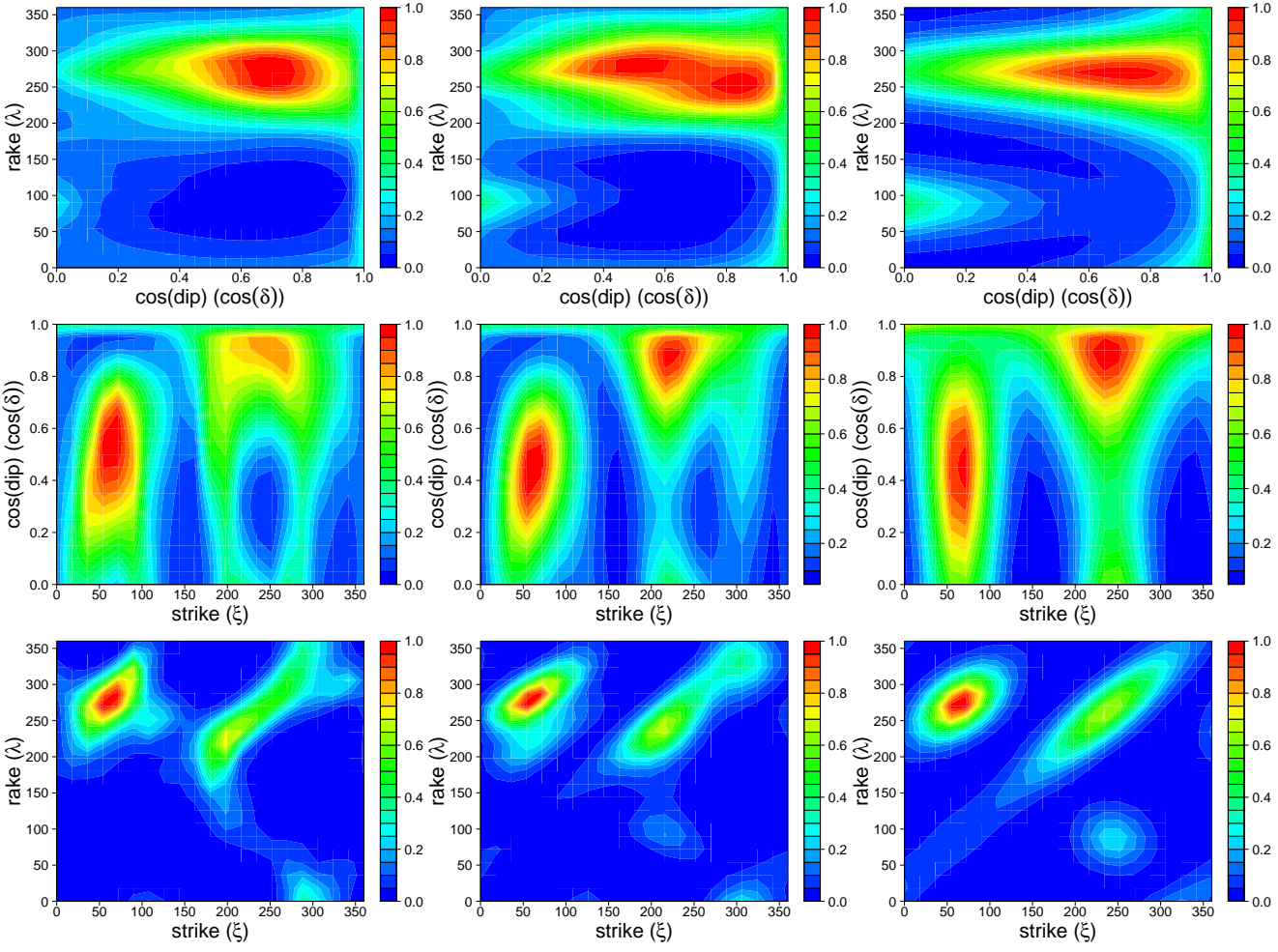


Figure 5.50: Marginal PDF plots for Bayesian posterior PDF  $P(\Theta|\mathbf{d}, \omega)$  (left), fitted 8-mode Matrix Fisher distribution  $P(\Theta|\hat{\mathbf{F}})$  (centre) and fitted 8-mode Matrix Fisher distribution with scalar concentration parameter  $P(\Theta|\hat{\mathbf{M}}, \hat{\kappa})$  (right) for event 1697233.



## 6. EVENT 1697233 – Velocity model unknown

Date	Time	Latitude	Longitude	Depth	Magnitude
27/02/2001	11:51:15.60	-38.124	176.701	40.000	2.6

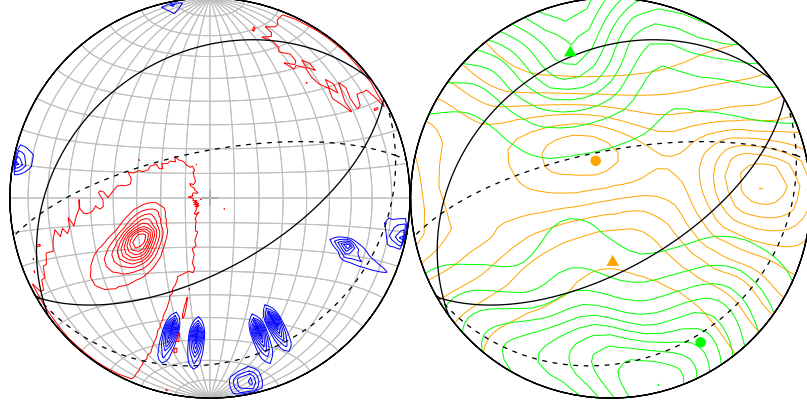


Figure 5.51: Stereonet for event 1697233 (left). The solid dark line indicates the MAP focal mechanism solution  $[\Theta = (233.93^\circ, 29.02^\circ, 264.41^\circ)]$  while the dotted line indicates the solution given by HASH. Angular difference between the two solutions =  $45.754^\circ$ . Blue points are compressions, red points are dilatations. PT contour plot for event 1697233 (right). Orange denotes the P-axis, green the T-axis. The circle denotes the MAP estimate, while the triangle denotes the solution given by HASH.

### Parameter Estimates

8-mode Matrix Fisher Distribution:

$$\hat{\mathbf{F}} = \begin{bmatrix} 5.432 & -3.776 & 2.941 \\ 2.514 & -1.757 & -0.806 \\ 0.040 & 0.223 & -0.549 \end{bmatrix}, \quad \hat{\mathbf{M}} = \begin{bmatrix} 0.705 & -0.355 & 0.614 \\ 0.393 & -0.525 & -0.755 \\ 0.591 & 0.773 & -0.230 \end{bmatrix},$$

$$\hat{\mathbf{U}} = \begin{bmatrix} 0.776 & -0.281 & -0.281 \\ -0.541 & 0.166 & -0.825 \\ 0.326 & 0.945 & -0.023 \end{bmatrix}, \quad \hat{\mathbf{D}} = \begin{bmatrix} 7.684 & 0 & 0 \\ 0 & 1.932 & 0 \\ 0 & 0 & 0.201 \end{bmatrix}$$

8-mode Matrix Fisher Distribution with scalar concentration:

$$\hat{\mathbf{M}} = \begin{bmatrix} 0.752 & -0.027 & 0.659 \\ 0.645 & -0.179 & -0.743 \\ 0.138 & 0.983 & -0.230 \end{bmatrix}, \quad \hat{\kappa} = 2.102$$

## 6. EVENT 1697233 – Velocity model unknown

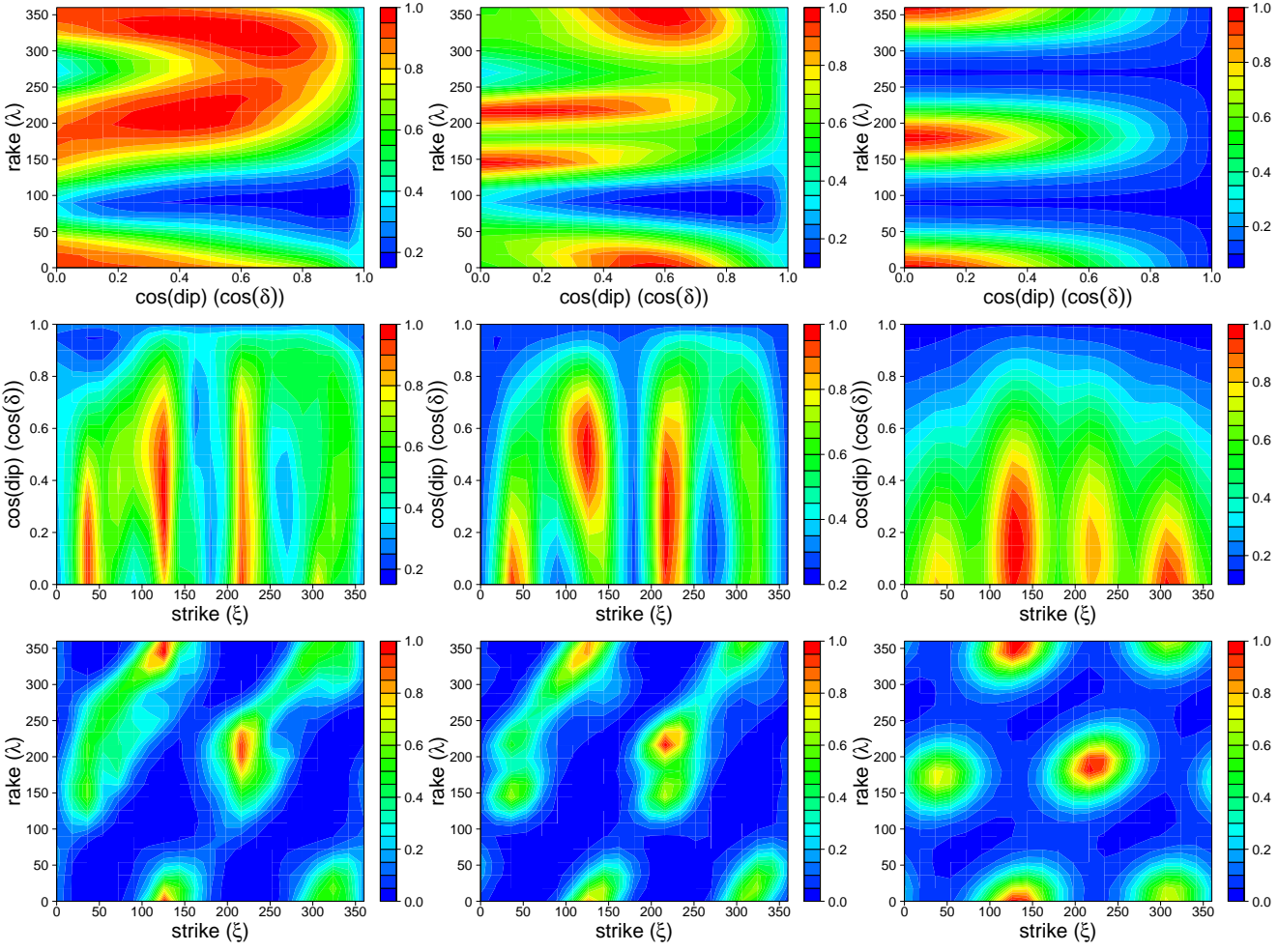


Figure 5.52: Marginal PDF plots for Bayesian posterior PDF  $P(\Theta|\mathbf{d}, \omega)$  (left), fitted 8-mode Matrix Fisher distribution  $P(\Theta|\hat{\mathbf{F}})$  (centre) and fitted 8-mode Matrix Fisher distribution with scalar concentration parameter  $P(\Theta|\hat{\mathbf{M}}, \hat{\kappa})$  (right) for event 1697233.

## 7. EVENT 1728730 – Velocity model known

Date	Time	Latitude	Longitude	Depth	Magnitude
01/05/2001	09:53:26.68	-38.152	176.695	-3.000	2.3

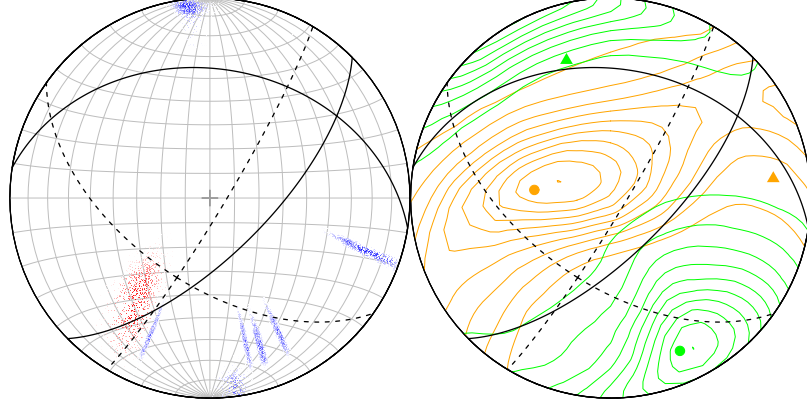


Figure 5.53: Stereonet for event 1728730 (left). The solid dark line indicates the MAP focal mechanism solution  $[\Theta = (278.70^\circ, 35.65^\circ, 317.51^\circ)]$  while the dotted line indicates the solution given by HASH. Angular difference between the two solutions =  $95.427^\circ$ . Blue points are compressions, red points are dilatations. PT contour plot for event 1728730 (right). Orange denotes the P-axis, green the T-axis. The circle denotes the MAP estimate, while the triangle denotes the solution given by HASH.

### Parameter Estimates

8-mode Matrix Fisher Distribution:

$$\hat{\mathbf{F}} = \begin{bmatrix} 2.963 & -3.600 & 5.780 \\ -1.582 & -0.375 & 0.022 \\ -0.215 & 2.923 & -4.952 \end{bmatrix}, \quad \hat{\mathbf{M}} = \begin{bmatrix} 0.650 & -0.680 & 0.338 \\ -0.627 & -0.732 & -0.267 \\ 0.429 & -0.039 & -0.902 \end{bmatrix},$$

$$\hat{\mathbf{U}} = \begin{bmatrix} 0.273 & 0.960 & 0.960 \\ -0.500 & 0.193 & 0.844 \\ 0.822 & -0.202 & 0.533 \end{bmatrix}, \quad \hat{\mathbf{D}} = \begin{bmatrix} 9.244 & 0 & 0 \\ 0 & 2.315 & 0 \\ 0 & 0 & 0.296 \end{bmatrix}$$

8-mode Matrix Fisher Distribution with scalar concentration:

$$\hat{\mathbf{M}} = \begin{bmatrix} 0.699 & -0.507 & 0.504 \\ -0.564 & -0.824 & -0.047 \\ 0.439 & -0.251 & -0.902 \end{bmatrix}, \quad \hat{\kappa} = 2.356$$

## 7. EVENT 1728730 – Velocity model known

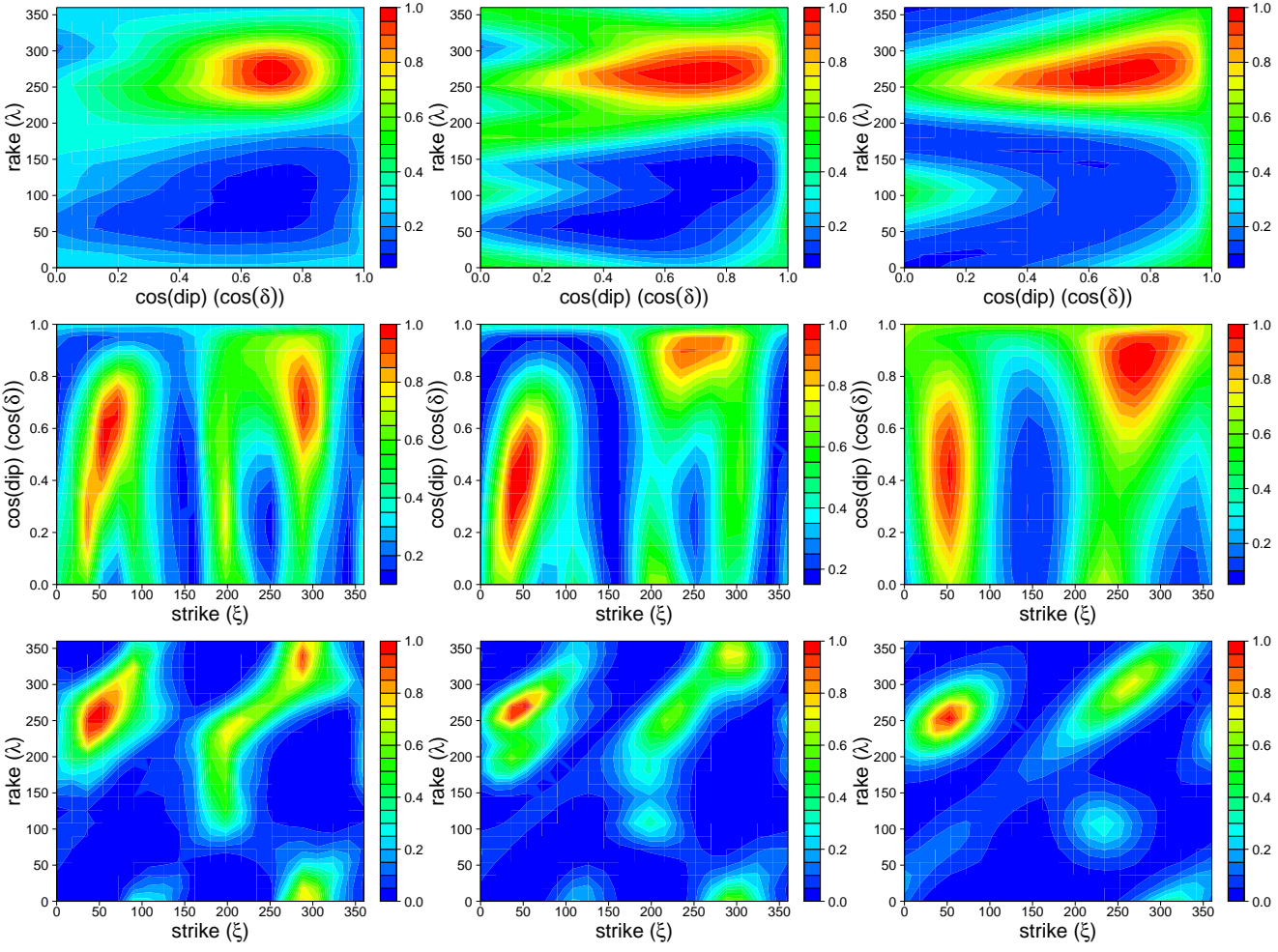


Figure 5.54: Marginal PDF plots for Bayesian posterior PDF  $P(\Theta|\mathbf{d}, \omega)$  (left), fitted 8-mode Matrix Fisher distribution  $P(\Theta|\hat{\mathbf{F}})$  (centre) and fitted 8-mode Matrix Fisher distribution with scalar concentration parameter  $P(\Theta|\hat{\mathbf{M}}, \hat{\kappa})$  (right) for event 1728730.

## 7. EVENT 1728730 – Velocity model unknown

Date	Time	Latitude	Longitude	Depth	Magnitude
01/05/2001	09:53:26.68	-38.152	176.695	0.000	2.3

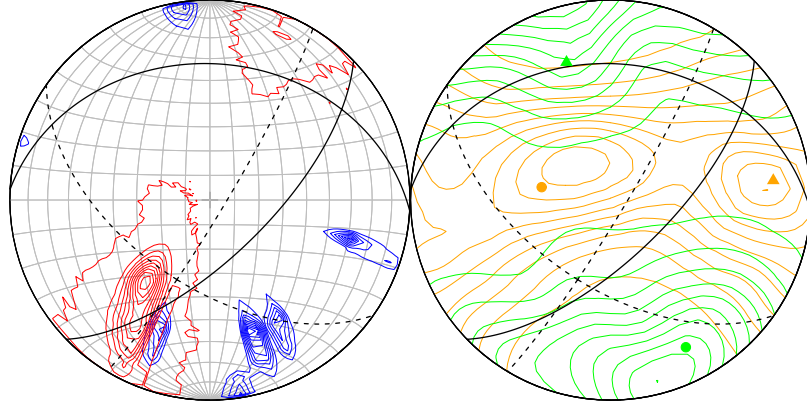


Figure 5.55: Stereonet for event 1728730 (left). The solid dark line indicates the MAP focal mechanism solution  $[\Theta = (273.08^\circ, 32.38^\circ, 312.36^\circ)]$  while the dotted line indicates the solution given by HASH. Angular difference between the two solutions =  $99.014^\circ$ . Blue points are compressions, red points are dilatations. PT contour plot for event 1728730 (right). Orange denotes the P-axis, green the T-axis. The circle denotes the MAP estimate, while the triangle denotes the solution given by HASH.

### Parameter Estimates

8-mode Matrix Fisher Distribution:

$$\begin{aligned} \hat{\mathbf{F}} &= \begin{bmatrix} 4.807 & -3.504 & 2.717 \\ 2.076 & -1.757 & -1.118 \\ -1.086 & 0.920 & -0.404 \end{bmatrix}, & \hat{\mathbf{M}} &= \begin{bmatrix} 0.751 & -0.294 & 0.592 \\ 0.494 & -0.344 & -0.798 \\ 0.438 & 0.892 & -0.113 \end{bmatrix}, \\ \hat{\mathbf{U}} &= \begin{bmatrix} 0.759 & -0.202 & -0.202 \\ -0.568 & 0.262 & -0.780 \\ 0.320 & 0.944 & 0.084 \end{bmatrix}, & \hat{\mathbf{D}} &= \begin{bmatrix} 7.028 & 0 & 0 \\ 0 & 2.050 & 0 \\ 0 & 0 & 0.081 \end{bmatrix} \end{aligned}$$

8-mode Matrix Fisher Distribution with scalar concentration:

$$\hat{\mathbf{M}} = \begin{bmatrix} 0.832 & -0.054 & 0.552 \\ 0.458 & -0.494 & -0.739 \\ 0.313 & 0.868 & -0.113 \end{bmatrix}, \hat{\kappa} = 2.004$$

## 7. EVENT 1728730 – Velocity model unknown

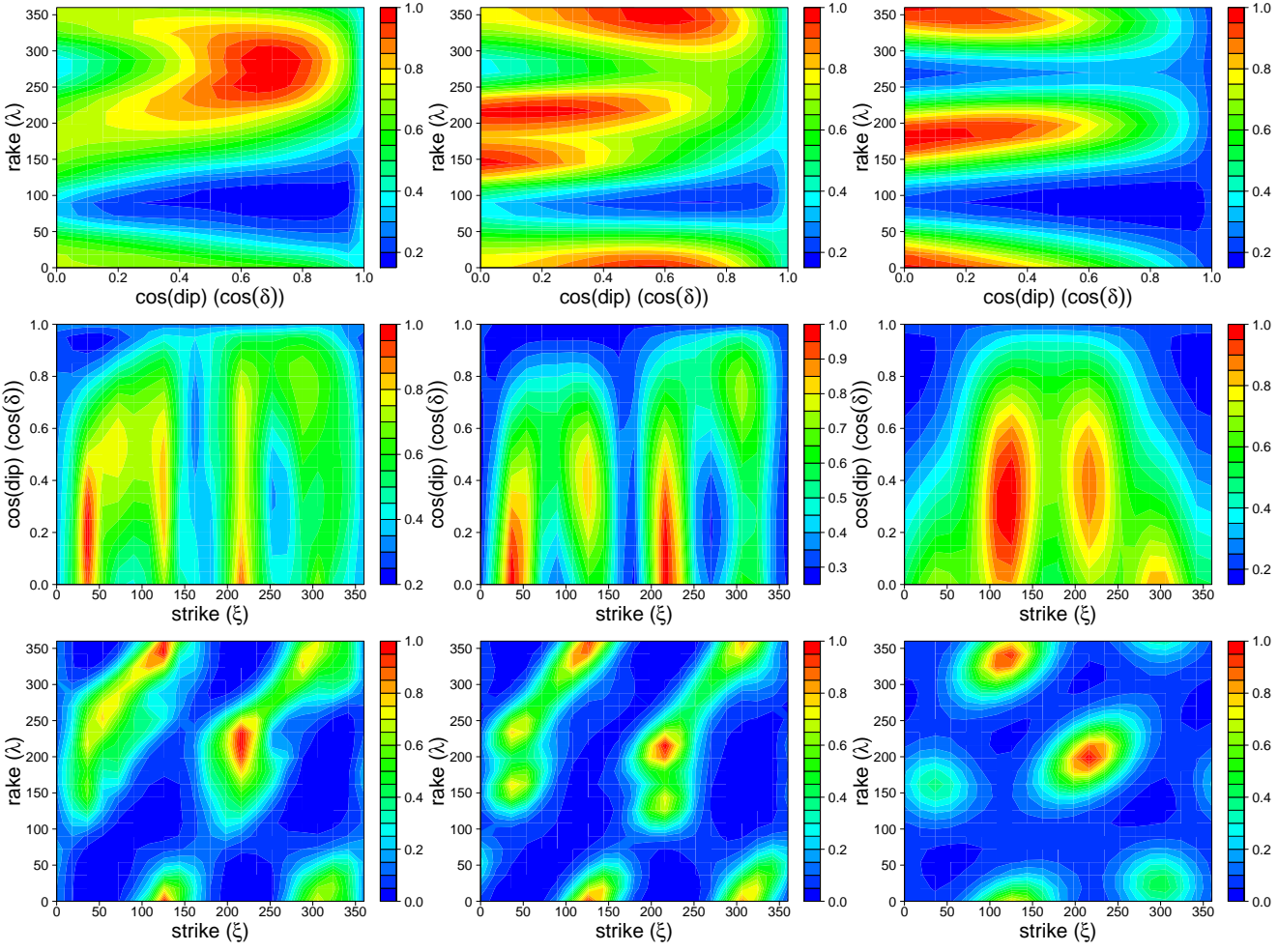


Figure 5.56: Marginal PDF plots for Bayesian posterior PDF  $P(\Theta|\mathbf{d}, \omega)$  (left), fitted 8-mode Matrix Fisher distribution  $P(\Theta|\hat{\mathbf{F}})$  (centre) and fitted 8-mode Matrix Fisher distribution with scalar concentration parameter  $P(\Theta|\hat{\mathbf{M}}, \hat{\kappa})$  (right) for event 1728730.

For the first event (CUSPID 731019) there is very little difference between the VMK and VMU results. In both cases the HASH solution is somewhat distant ( $30^\circ - 40^\circ$ ) from our MAP solution, but our MAP solution again fits the first motions well (see Figures 5.29, p99 and 5.31, p101). There is an angular difference of just  $2.92^\circ$  degrees between the VMK and VMU MAP solutions. The posterior and Matrix Fisher PDFs are similar for both cases (Figures 5.30, p100 and 5.32, p102), with the concentration parameters in the VMU case being slightly lower — presumably caused by the added uncertainty in velocity model. In both cases the full Matrix Fisher distribution provides a better fit than the scalar concentration version.

The second event (CUSPID 745516) again exhibits very little difference between the VMK and VMU results (pp 103-106). The HASH estimates of  $\Theta$ ,  $\hat{v}^P$  and  $\hat{v}^T$  are reasonably close to the corresponding MAP solutions, and again there is only a tiny difference ( $1.81^\circ$ ) between the MAP estimates of  $\Theta$  in the VMK and VMU cases (see Figures 5.33 p103 and 5.35 p105). The concentration parameters of the posterior PDF in the VMU case are again lower. The Matrix Fisher fit is good for both cases, but appears slightly better for the VMK case (Figure 5.34, p104).

In the third event (CUSPID 788921) we see some differences between the VMK and VMU results. The MAP solutions are similar in the two cases (see Figures 5.37 p107 and 5.39 p109), however the parameter estimates, P- and T-axis contours, and posterior PDFs are all strikingly different. Notably the modes of the posterior PDF shift slightly from the VMK to VMU case (Figures 5.38 p108 and 5.40 p110), and the concentration parameters are reduced. The fourth and fifth events (CUSPIDs 802105 and 802106) also show differences between their respective VMK and VMU results, characterised by a shift in the modes and reduced constraint on the posterior PDFs, however the MAP estimates remain similar. For all three of these events the Matrix Fisher PDFs fit reasonably well, however this fit appears better for the VMK cases. As the fourth and fifth events occurred in the same location just three minutes (in time) apart, we would expect them to have the same focal mechanism. The posterior PDFs and fitted parameters for the VMK case for both events are reasonably similar (see Figures 5.42, p112 and 5.46, p116), while the VMU posterior PDFs (Figures 5.44, p114 and 5.48, p118) and parameter estimates are almost identical.

The sixth event (CUSPID 1697233) exhibits a rather large disparity between the VMK and VMU posterior PDFs (Figures 5.50 p120 and 5.52 p122), most notably in the  $\lambda$  versus  $\cos \delta$  marginal plot. Here, adding the uncertainty in velocity model has resulted in a change from a single mode in the VMK case to a bimodal plot in the VMU case, with the modes being either side of the original single mode. In the VMU case the Matrix Fisher fit is not particularly good. The seventh event (CUSPID 1728730) exhibits a poorer constrained posterior PDF in the VMU case compared to the VMK case, evident in the plots (Figures 5.54 p124 and 5.56 p126) and the lower concentration parameters. The HASH estimate of  $\Theta$  is very different from our MAP estimate in both cases, and the P- and T-axis contours are poorly constrained, reflecting the poor focal sphere coverage of the first motions in this event (see Figures 5.53 p123 and 5.55 p125).

In general the MAP estimate of  $\Theta$  did not change much between our VMK and VMU cases. Adding uncertainty in velocity model has resulted in posterior PDFs of  $\Theta$  that tend to have broader maxima which constrain  $\Theta$  less tightly. Some events showed very little difference between the VMK and VMU PDFs, while others showed large differences. We can only conjecture over what causes the posterior PDF of an event to change dramatically when we add uncertainty in velocity model into the formulation. Compare Figures 5.50 and 5.52. In the posterior PDF (particularly the  $\lambda$  vs  $\cos \delta$  marginal plot) the single modes in Figure 5.50



are split in two in Figure 5.52. This could indicate a posterior that is bimodal in velocity models  $v$ : with models where the velocity remains low for most depths being as favoured as models where the velocity is uniformly high, but models with strong gradients in velocity not being favoured by the data. Future work into a better motivated prior for  $v$  would better inform us of how the velocity model uncertainty affects the posterior PDFs of  $\Theta$ .

The Matrix Fisher distribution, though fitting reasonably well in the VMU case, provides a better fit in the VMK case. This corresponds to what we observed in Section 5.1.5, in that the fit was better for the more well constrained events. Again, in the events seen here, the full distribution is a more accurate fit than the simplified scalar concentration version.

It should be noted that the data quality for Kawerau is poor. There are relatively few P-wave polarity data compared to the Raukumara data, and due to the shallow depth of the events, only the outer region of the focal sphere is well covered. This has resulted in poorly constrained posterior PDFs of  $\Theta$ , as demonstrated by the plots and the fact that all the estimates of  $\kappa$  lie in the range  $\sim (2, 3.1)$ , similar or more poorly constrained than the most poorly constrained of our chosen Raukumara events. A deployment of seismometers in the region could resolve these data issues, but this is beyond the scope of this project. Note that the HASH focal mechanism solutions are all of poor quality, as denoted by a quality code (“D” for each of the seven mechanisms), and the RMS difference of the acceptable nodal planes from the preferred solutions given in the results (all  $> 39^\circ$ ).

A map of focal mechanism solutions obtained using the average velocity model is shown in Figure 5.57. Our estimated focal mechanisms are predominantly normal (i.e. the dilatational quadrant is in the centre of the focal sphere), with some having a strike-slip component. However, the most south-westerly located events appear to have strike-slip mechanisms. These results are consistent with the earlier focal mechanism study in the TVZ by Hurst et al. (2002), which found mechanisms that were predominantly normal, or normal with a strike-slip component.



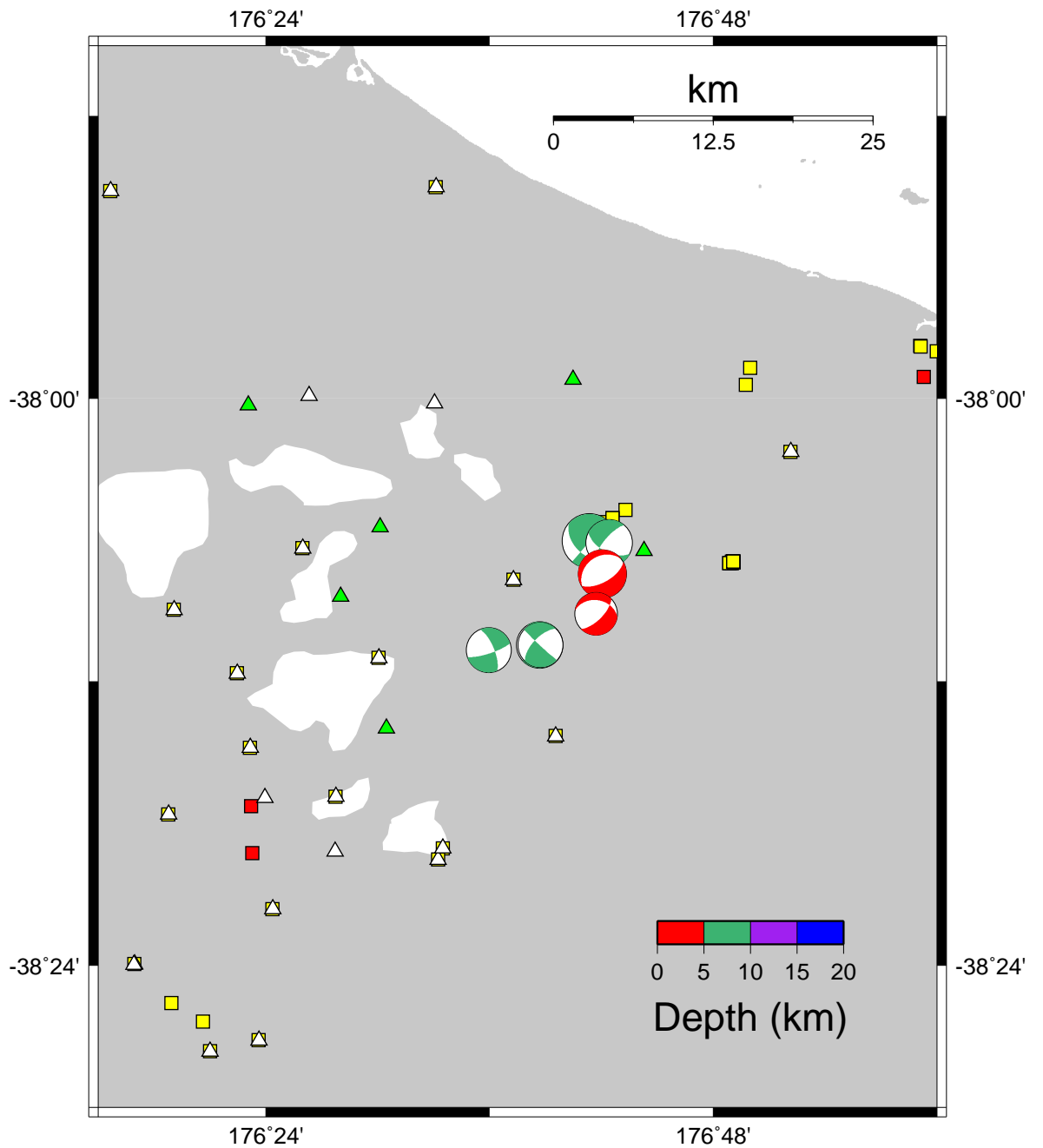


Figure 5.57: Map of Kawerau showing MAP focal mechanism estimates obtained using our method. Hypocentre locations are as calculated by NonLinLoc. Beachballs are scaled relative to their magnitudes.



# Chapter 6

## Conclusions

In this project we have introduced a new probabilistic (Bayesian) method of focal mechanism estimation that directly accounts for uncertainty in hypocentre location, seismic velocity structure, and P-wave polarities. We have examined the case when the velocity model is assumed to be precisely known (VMK), with application to data from the Raukumara Peninsula, and the case when the velocity model is imperfectly known (VMU), with application to data from Kawerau. Introducing uncertainty over the velocity structure had the effect of reducing the concentration of the resulting posterior PDFs of the focal mechanism parameters ( $\Theta$ ).

Our MAP estimates of  $\Theta$  have been shown to accurately divide the compressional and dilatational first motions. Given reasonable data quality, the MAP estimates of focal mechanism parameters that result from our method have also been shown to be consistent with established methods of focal mechanism estimation, with the advantage of providing a full posterior distribution of  $\Theta$  values.

We have explored the use of two generalised Matrix Fisher distributions — the 8-mode Matrix Fisher distribution, and the 8-mode Matrix Fisher distribution with scalar concentration parameter — for approximating the posterior PDF of the focal mechanism parameters. The full 8-mode Matrix Fisher distribution provides a superior fit to the empirical distributions, although, interestingly, for well-constrained events the scalar concentration distribution also fits well. This appears to justify the approach taken by Arnold & Townend (2007), who assumed that focal mechanism errors follow a Matrix Fisher distribution with scalar concentration parameter, in their work on estimating tectonic stress. It would be interesting to know if the additional information of the full Matrix Fisher distribution has a significant impact on the outputs of the stress estimation procedure, or whether a scalar representation is sufficient.

The advantages of our method over previously published methods are, therefore, that it (1) models the data generation process and incorporates observational errors, particularly those arising from imperfectly known earthquake locations; (2) allows exploration of the entire parameter space; (3) leads to natural point estimates of focal mechanism parameters; and (4) that the resulting posterior PDF can be well approximated by generalised Matrix Fisher distributions.

This work suggests a number of possible avenues for further research. As mentioned in Section 5.2.4, the approach taken here would benefit from further work into establishing a well motivated prior for the velocity model in the VMU case, which would better inform us of how the velocity model uncertainty affects the posterior PDFs of  $\Theta$ . Future work could also investigate the formulation and use of more informative priors for the focal mechanism parameters  $\Theta$ , as we have only considered the state of total ignorance (a uniform prior) here.

Additionally, we have not examined the effect of varying the amplitude noise  $\sigma_a$  and polarity error  $\pi'_p$  parameters, and have not focused on robustly estimating these parameters. Sensitivity analysis could be conducted, in order to assess how the posterior PDF changes as the parameters are varied. Establishing a well motivated prior for these parameters, rather than using fixed values as in this project, could also be investigated. Furthermore, future work could address the effect of varying other parameters that we have assumed known in this project —  $\sigma_{t_i}$ , which we obtained from GeoNet quality codes as described in Table 4.1, and  $C_T$ , for which we have used the NonLinLoc default values.

The Kawerau case study described in Section 5.2 demonstrated the effect of poor data quality, specifically that the low number of P-wave polarity readings read in routine CUSP processing resulted in poorly constrained posterior PDFs of  $\Theta$ . In future, P-wave data could perhaps be supplemented with S-wave polarisation data or amplitude ratios, which would help constrain the solutions when P-wave polarities are scarce. An adequate means of incorporating the S-wave data and its inherent uncertainties into the formulation of the posterior PDF would need to be established, but this would not affect the underlying Bayesian framework of our method.

To conclude, we have developed here a robust new method of focal mechanism estimation by directly accounting for the relevant uncertainties. Robust focal mechanism estimates are important tools in assessing the tectonic characteristics of a region, and are inputs to the problem of estimating tectonic stress — changes in which may place constraints on the processes involved in earthquake occurrence and volcanism. Thus the method developed here can be seen as addressing one component of the wider problem of earthquake source characterisation and tectonic interpretation.

# Glossary

The glossary contains a list of symbols used in this project and their meanings, followed by a list of terms used in this project and their definitions.

## Symbols

$\delta$	dip angle
$\hat{\mathbf{a}}$	null vector
$\hat{\mathbf{n}}$	fault normal vector
$\hat{\mathbf{p}}_i$	unit vector from the hypocentre, $\mathbf{x}$ , to the point on the focal sphere corresponding to station $i$
$\hat{\mathbf{u}}$	slip vector
$\hat{\mathbf{v}}^P$	unit vector in the direction of the P-axis
$\hat{\mathbf{v}}^T$	unit vector in the direction of the T-axis
$\lambda$	rake angle
$\mathbf{D}$	concentration matrix of a Matrix Fisher distribution
$\mathbf{d}$	the data
$\mathbf{F}$	parameter matrix of a Matrix Fisher distribution
$\mathbf{M}$	modal matrix of a Matrix Fisher distribution
$\mathbf{U}$	spin matrix of a Matrix Fisher distribution
$\mathbf{x}$	hypocentre location
$\omega$	known parameters
$\phi$	azimuth
$\pi'_p$	probability of an incorrect polarity
$\pi_i$	probability of a positive observed first motion at station $i$
$\psi$	nuisance parameters
$\sigma_a$	the standard deviation of the amplitude of the first motion
$\sigma_{t_i}$	P-wave arrival time error at station $i$
$\mathbf{C}_T$	covariance matrix of $\{t_i^{cal}\}$
$\mathbf{C}_t$	covariance matrix of $\{t_i^{obs}\}$

$\Theta$	the set of focal mechanism parameters $(\xi, \delta, \lambda)$
$\theta$	take-off angle
$v$	velocity model
$\xi$	strike angle
$A_i$	amplitude of the first motion at station $i$
$A_i^*$	theoretical P-wave amplitude at station $i$
$H(\cdot)$	The Kullback-Leibler divergence
$h_i^{cal}(\mathbf{x})$	calculated travel time between a hypocentre location $\mathbf{x}$ and station $i$
$R(\Theta)$	rotation matrix with columns $[\hat{\mathbf{u}} \ \hat{\mathbf{a}} \ \hat{\mathbf{n}}]$ , used to define the focal mechanism
$s_i$	location of seismic station $i$
$t_i^{cal}$	calculated seismic wave arrival time at station $i$ given the implemented velocity model
$t_i^{obs}$	observed seismic wave arrival time at station $i$
$t_i^{res}$	the difference between the observed and theoretical arrival times at station $i$
$Y_i$	P-wave polarity at station $i$
<b>T</b>	earthquake origin time

## Terms

<i>8-Mode Matrix Fisher distribution</i>	a generalisation of the Matrix Fisher distribution to situations in which there is a two-fold ambiguity in the direction defined by each column of the orthogonal matrix random variable
<i>angular difference</i>	minimum rotation about any axis needed to make two rotation matrices coincide
<i>auxiliary plane</i>	one of two nodal planes of a focal mechanism, and perpendicular to the fault plane
<i>axial data</i>	directional data in which the positive and negative directions are equivalent
<i>azimuth</i>	angle measured clockwise from north
<i>Bingham distribution</i>	probability distribution for axial data in 3D space, see Mardia & Jupp (2000)
<i>circular data</i>	directional data in 2D space
<i>compression</i>	an upwards first motion

<i>coupled hypocentre-velocity model</i>	solving simultaneously for the hypocentre location and velocity model
<i>dilatation</i>	a downwards first motion
<i>dip angle</i>	the angle that the fault slants downwards from the horizontal to the right looking along the strike direction
<i>directional data</i>	observations that are directions, or unit vectors, in space
<i>epicentre</i>	the point on the earth's surface directly above the hypocentre of an earthquake
<i>equivalent body forces</i>	a model of the faulting process; the forces that would yield the observed seismic wave radiation pattern
<i>fault normal vector</i>	vector normal to the fault plane
<i>fault plane</i>	planar surface on which an earthquake occurs
<i>first motion</i>	the direction of motion, or polarity, of the first P-wave arrival at a seismic station
<i>focal mechanism</i>	geometrical representation of fault slip during an earthquake
<i>focal sphere</i>	imaginary sphere of negligible radius surrounding the earthquake source
<i>force couple</i>	two forces acting together
<i>hypocentre</i>	the location of an earthquake
<i>Kullback-Leibler divergence</i>	a measure of discrepancy between two probability distributions
<i>likelihood</i>	function of the parameters of interest given the observed data
<i>Matrix Fisher distribution</i>	probability distribution for matrices on the Stiefel Manifold, see Downs (1972)
<i>minimum 1D velocity model</i>	velocity model with minimum root mean square (RMS) misfit of $\{t_i^{res}\}$
<i>moment tensor</i>	a quantity that depends on source strength and fault orientation
<i>nodal planes</i>	the fault plane and the auxiliary plane
<i>NonLinLoc</i>	a software package used “for velocity model construction, travel-time calculation and probabilistic, non-linear, global-search earthquake location”
<i>null vector</i>	vector defined by $\hat{\mathbf{n}} \times \hat{\mathbf{u}}$ , i.e. perpendicular to the slip and normal vectors
<i>orientation data</i>	observations made up of $n$ directions describing a $p$ dimensional object
<i>orthogonal group</i>	A Stiefel Manifold where $n = p$ ; the group of $p \times p$ orthogonal matrices

<i>P-wave</i>	longitudinal seismic wave that travels fastest from the earthquake source
<i>posterior</i>	probability density function describing our knowledge of the parameters given the data
<i>prior</i>	probability density function describing our prior knowledge of the parameters
<i>rake angle</i>	the direction of motion of the upper side of the fault with respect to the lower side of the fault, measured in the fault plane anti-clockwise from the direction of the strike
<i>S-wave polarisation</i>	splitting the S-wave into two perpendicular components, SV and SH
<i>S-wave</i>	transverse seismic wave that travels more slowly than the P-wave from the earthquake source
<i>seismic moment</i>	a measure of the magnitude of an earthquake
<i>seismic noise</i>	fluctuations in the seismic wave signal caused by external factors such as human activity
<i>seismometer</i>	instruments that measure and record ground motions
<i>slip vector</i>	vector in the direction the fault slipped during the earthquake
<i>special orthogonal group</i>	The group of $p \times p$ orthogonal matrices with determinant 1
<i>spherical data</i>	directional data in 3D space
<i>stereonet</i>	a 2D projection of the lower hemisphere of the focal sphere
<i>Stiefel manifold</i>	the set of all $p \times n$ matrices that describe the orientation of an object in $p$ -dimensions, defined by $n$ directions, and for which $\mathbf{X}^T \mathbf{X} = \mathbf{I}_n$
<i>strike angle</i>	the angle measured clockwise from north to the strike direction
<i>strike direction</i>	the direction of a horizontal line in the fault plane
<i>take-off angle</i>	angle measured from the downward vertical to the point on the focal sphere where a P-wave left the earthquake source
<i>velocity model</i>	a simplified representation of the seismic velocity structure of the earth
<i>VMK posterior PDF</i>	velocity model known posterior PDF of the focal mechanism parameters
<i>VMU posterior PDF</i>	velocity model unknown posterior PDF of the focal mechanism parameters
<i>von Mises distribution</i>	probability distribution for circular data, see Mardia & Jupp (2000)



<i>von Mises-Fisher distribution</i>	probability distribution for spherical data, see Mardia & Jupp (2000)
Grid2Time	program that calculates travel-times and take-off angles between a station and all nodes of an $x, y, z$ spatial grid
NLLoc	program that implements Tarantola & Valette (1982)'s Bayesian method of hypocentre location
Vel2Grid	program that converts velocity model specifications into a 3D grid file in binary format for use with Grid2Time and NLLoc
Velest	a program that determines minimum 1-dimensional velocity models



# References

- Abadir, K. M. & Magnus, J. R. (2005), *Matrix Algebra*, Econometric Exercises, Cambridge University Press, Cambridge.
- Aki, K. & Richards, P. G. (2002), *Quantitative Seismology*, 2nd edn, University Science Books, Sausalito.
- Anton, H. (2000), *Elementary linear algebra*, 8th edn, Wiley, New York.
- Arfken, G. (1985), *Mathematical methods for physicists*, 3rd edn, Academic Press Inc., San Diego.
- Arnold, R. & Townend, J. (2007), 'A Bayesian Approach to Estimating Tectonic Stress from Seismological Data', *Geophys. J. Int.* **170**, 1336 – 1356.
- Bengtsson, H. (2007), *R.matlab: Read and write of MAT files together with R-to-Matlab connectivity*. R package version 1.1.3, <http://www.braju.com/R/>.
- Best, D. & Fisher, N. (1979), 'Efficient Simulation of the von Mises Distribution', *Appl. Statist.* **28**, 152–157.
- Bibby, H. M., Caldwell, T. G., Davey, F. J. & Webb, T. H. (1995), 'Geophysical evidence on the structure of the Taupo Volcanic Zone and its hydrothermal circulation', *Journal of Volcanology and Geothermal Research* **68**, 29–58.
- Bouley, S. (2005), Bayesian estimation of earthquake focal mechanisms using data from the Raukumara Peninsula, New Zealand. Report on Victoria University of Wellington internship.
- Brillinger, D. R., Udias, A. & Bolt, B. A. (1980), 'A Probability Model for Regional Focal Mechanism Solutions', *Bulletin of the Seismological Society of America* **70**, 149–170.
- Chikuse, Y. (2003), *Statistics on Special Manifolds*, Lecture Notes in Statistics, Springer.
- Clarke, D. (2007), Velocity Modelling and Earthquake Relocation in the Rotorua and Kawerau Geothermal Areas, Taupo Volcanic Zone, New Zealand, Master's thesis, Victoria University of Wellington, New Zealand.
- Cox, A. & Hart, R. (1986), *Plate Tectonics - How It Works*, Blackwell Scientific, Oxford.
- De Natale, G., Ferraro, A. & Virieux, J. (1991), 'A probability method for local earthquake focal mechanisms', *Geophysical Research Letters* **18**, 613–616.
- Downs, T. (1972), 'Orientation Statistics', *Biometrika* **59**(3), 665–676.
- Fowler, C. (1990), *The solid earth: an introduction to global geophysics*, Cambridge University Press, Cambridge.
- Gelman, A., Carlin, J. B., Stern, H. S. & Rubin, D. B. (1995), *Bayesian data analysis*, Chapman & Hall/CRC, Florida.

- Giné, M. E. (1975), 'Invariant tests for uniformity on compact riemannian manifolds based pm sobolev norms', *Annals of Statistics* **3**(6), 1243–1266.
- Hardebeck, J. L. & Shearer, P. M. (2002), 'A New Method for Determining First-Motion Focal Mechanisms', *Bulletin of the Seismological Society of America* **92**(6), 2264–2276.
- Hardebeck, J. L. & Shearer, P. M. (2003), 'Using S/P Amplitude Ratios to Constrain the Focal Mechanisms of Small Earthquakes', *Bulletin of the Seismological Society of America* **93**(6), 2434–2444.
- Hogg, R. & Tanis, E. (2001), *Probability and Statistical Inference*, 6th edn, Prentice Hall, New Jersey.
- Hurst, A. W., Bibby, H. M. & Robinson, R. R. (2002), 'Earthquake focal mechanisms in the central Taupo Volcanic Zone and their relation to faulting and deformation', *New Zealand Journal of Geology & Geophysics* **45**, 527–536.
- James, A. T. (1964), 'Distributions of Matrix Variates and Latent Roots Derived from Normal Samples', *The Annals of Mathematical Statistics* **35**(2), 475–501.
- Jupp, P. E. (2005), 'Sobolev tests of goodness of fit of distributions on compact Riemannian manifolds', *The Annals of Statistics* **33**(6), 2957–2966.
- Kagan, Y. Y. (2005), 'Double-couple earthquake focal mechanisms: random rotation and display', *Geophys. J. Int.* **163**, 1065–1072.
- Kagan, Y. Y. (2007), 'Simplified algorithms for calculating double-couple rotation', *Geophys. J. Int.* **171**, 411–418.
- Khatri, C. & Mardia, K. V. (1977), 'The von Mises-Fisher matrix distribution in orientation statistics', *Journal of the Royal Statistical Society. Series B (Methodological)* **39**(1), 95–106.
- Kissling, E. (1988), 'Geotomography with local earthquake data', *Reviews of Geophysics* **26**(4), 659–698.
- Kissling, E. (1995), *Velost user's guide*, Institute of Geophysics, ETH Zuerich.
- Kissling, E., Ellsworth, W. L., Eberhart-Phillips, D. & Kradolfer, U. (1994), 'Initial reference models in local earthquake tomography', *Journal of Geophysical Research* **99**(B10), 19635–19646.
- Koev, P. & Edelman, A. (2006), 'The efficient evaluation of the hypergeometric function of a matrix argument', *Mathematics of Computation* **75**(254), 833–846.
- Kolecki, J. C. (2002), An Introduction to Tensors for Students of Physics and Engineering, Technical Report NASA/TM-2002-211716, National Aeronautics and Space Administration.
- Lomax, A. (2007), 'The NonLinLoc Software Guide', <http://alomax.free.fr/nlloc/>.
- Mardia, K. V. & Jupp, P. E. (2000), *Directional Statistics*, Wiley, Chichester.
- McLaren, M. L. (1976), 'Algorithm AS94: Coefficients of the Zonal Polynomials', *Applied Statistics* **25**(1), 82–87.
- Muirhead, R. J. (1982), *Aspects of multivariate statistical theory*, Wiley, New York.

- Nakamura, M. (2002), ‘Determination of focal mechanism solution using initial motion polarity of P and S waves’, *Physics of the Earth and Planetary Interiors* **130**, 17–29.
- Podvin, P. & Lecomte, I. (1991), ‘Finite difference computation of traveltimes in very contrasted velocity models: a massively parallel approach and its associated tools’, *Geophys. J. Int.* **105**, 271–284.
- R Development Core Team (2006), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, <http://www.R-project.org>.
- Rabinowitz, N. & Hofstetter, A. (1992), ‘A rapid algorithm for estimating fault plane solution using polarity-amplitude data: application of a non-linear programming approach’, *Physics of the Earth and Planetary Interiors* **239-254**.
- Rau, R., Wu, F. & Shin, T. (1996), ‘Regional Network Focal Mechanism Determination Using 3D Velocity Model and SH/P Amplitude Ratios’, *Bulletin of the Seismological Society of America* **1270-1283**.
- Reasenbergs, P. A. & Oppenheimer, D. (1985), ‘FPFIT, FPLOT, and FPPAGE: Fortran computer programs for calculating and displaying earthquake fault-plane solutions’, *U.S. Geol. Surv. Open-File Rep.* **85-739**.
- Reyners, M., Eberhart-Phillips, D. & Stuart, G. (1999), ‘A three-dimensional image of shallow subduction: crustal structure of the Raukumara Peninsula, New Zealand’, *Geophysics Journal International* **137**, 873 – 890.
- Reyners, M. & McGinty, P. (1999), ‘Shallow subduction tectonics in the Raukumara Peninsula, New Zealand, as illuminated by earthquake focal mechanisms’, *Journal of Geophysical Research* **104**, 3025–3034.
- Reyners, M., Robinson, R. & McGinty, P. (1997), ‘Plate coupling in the northern South Island and southernmost North Island, New Zealand, as illuminated by earthquake focal mechanisms’, *Journal of Geophysical Research* **102(B7)**, 15197 – 15210.
- Robert, C. & Casella, G. (1999), *Monte Carlo Statistical Methods*, 1st edn, Springer-Verlag, New York.
- Robinson, R. & Webb, T. (1996), “‘AMPRAT’ and ‘MECHTOOL’: Programs for determining focal mechanisms of local earthquakes”, *Institute of Geological & Nuclear Sciences science reports* **Sci. Rep. 96/7**.
- Sivia, D. (1996), *Data Analysis: A Bayesian Tutorial*, 1st edn, Oxford University Press, Oxford, UK.
- Snoke, J. A. (2003), *International Handbook of Earthquake and Engineering Seismology*, Academic Press, chapter FOCMEC: Focal MECHANism Determinations, pp. 29–30.
- Stein, S. & Wysession, M. (2003), *An Introduction to Seismology, Earthquakes, and Earth Structure*, 1st edn, Blackwell Publishing, Malden, MA.
- Tarantola, A. (2005), *Inverse Problem Theory and Methods for Model Parameter Estimation*, 1st edn, Society for Industrial and Applied Mathematics, Philadelphia, PA.
- Tarantola, A. & Valette, B. (1982), ‘Inverse Problems = Quest for Information’, *Journal of Geophysics* **50**, 159 – 170.

- Townend, J. (2006), What do faults feel? Observational constraints on the stress acting on seismogenic faults, *in* 'Radiated Energy and Physics of Earthquake Faulting', AGU, New York.
- Udias, A. (1999), *Principles of Seismology*, Cambridge University Press, Cambridge, UK.
- Weir, M., Hass, J. & Giordano, F. (2005), *Thomas' Calculus*, 11th edn, Pearson Education, Boston, MA.
- Wood, A. T. A. (1987), 'The simulation of spherical distributions in the Fisher-Bingham family', *Commun. Statist. -Simula.* **16**(3), 885 – 898.
- Wood, A. T. A. (1994), 'Simulation of the von Mises Fisher Distribution', *Commun. Statist. -Simula.* **23**(1), 157 – 164.
- Zollo, A. & Bernard, P. (1991), 'Fault mechanisms from near-source data: joint inversion of S polarizations and P polarities', *Geophysics Journal International* **104**, 441 – 451.

# Appendix A

## Useful Definitions and Results

This appendix contains some useful definitions and results that are used in this project.

### A.1 Rotation

#### A.1.1 Euler Angles

*Euler Angles*  $\Phi = (\phi, \theta, \psi)$  are used to describe a rotation in three dimensions, where

$$0 \leq \phi \leq 2\pi, \quad 0 \leq \theta \leq \pi, \quad 0 \leq \psi \leq 2\pi$$

There are various definitions of Euler Angles, determined by which axes are rotated around and in which order. Here we will use the definition provided by Arnold & Townend (2007), who use a zyz convention. If  $(x, y, z)$  are the coordinates of an initial coordinate system, the Euler Angles are used to rotate the system as follows:

- Rotate the initial system anticlockwise by  $\phi$  about the  $z$  axis to obtain  $(x', y', z')$
- Rotate  $(x', y', z')$  anticlockwise by  $\theta$  about the  $y'$  axis to obtain  $(x'', y'', z'')$ .
- Rotate  $(x'', y'', z'')$  anticlockwise by  $\psi$  about the  $z''$  axis to obtain the final coordinate system  $(x''', y''', z''')$ .

These rotations can be expressed as matrices

$$B_y(\alpha) = \begin{bmatrix} \cos \alpha & 0 & \sin \alpha \\ 0 & 1 & 0 \\ -\sin \alpha & 0 & \cos \alpha \end{bmatrix} \quad B_z(\alpha) = \begin{bmatrix} \cos \alpha & -\sin \alpha & 0 \\ \sin \alpha & \cos \alpha & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

where  $B_i(\alpha)$  describes a rotation of angle  $\alpha$  about axis  $i$  (Arnold & Townend 2007). A rotation matrix formed from the Euler Angles is given by

$$\begin{aligned} R(\Phi) &= R(\alpha, \theta, \psi) = B_z(\phi)B_y(\theta)B_z(\psi) \\ &= \begin{bmatrix} \cos \phi \cos \theta \cos \psi - \sin \phi \sin \psi & -\cos \phi \cos \theta \sin \psi - \sin \phi \cos \psi & \cos \phi \sin \theta \\ \sin \phi \cos \theta \cos \psi + \cos \phi \sin \psi & -\sin \phi \cos \theta \sin \psi + \cos \phi \cos \psi & \sin \phi \sin \theta \\ -\sin \theta \cos \psi & \sin \theta \sin \psi & \cos \theta \end{bmatrix} \end{aligned}$$

Rotation matrices are orthogonal ( $R^T R = I$ ) and have  $\det R = 1$ . Any rotation matrix  $R$  has equivalent Euler angles, given by

$$\phi = \tan^{-1} \left( \frac{R_{23}}{R_{13}} \right), \quad \theta = \cos^{-1} R_{33}, \quad \psi = \tan^{-1} \left( \frac{R_{32}}{-R_{31}} \right)$$

(Arnold & Townend 2007).

Euler angles  $\Phi = (\phi, \theta, \psi)$  are equivalent to focal mechanism parameters  $\Theta = (\xi, \delta, \lambda)$ :

$$\begin{aligned}\phi &= \xi + \frac{\pi}{2} & \xi &= \phi - \frac{\pi}{2} \\ \theta &= \pi - \delta & \delta &= \pi - \theta \\ \psi &= \lambda - \frac{\pi}{2} & \lambda &= \psi + \frac{\pi}{2}\end{aligned}$$

So the rotation matrix that describes the angles  $\Theta$  is

$$\begin{aligned}R(\Theta) &= B_z(\xi + \frac{\pi}{2})B_y(\pi - \delta)B_z(\lambda - \frac{\pi}{2}) \\ &= \begin{bmatrix} \sin \xi \cos \delta \sin \lambda + \cos \xi \cos \lambda & \sin \xi \cos \delta \cos \lambda - \cos \xi \sin \lambda & -\sin \xi \sin \delta \\ -\cos \xi \cos \delta \sin \lambda + \sin \xi \cos \lambda & -\cos \xi \cos \delta \cos \lambda - \sin \xi \sin \lambda & \cos \xi \sin \delta \\ -\sin \delta \sin \lambda & \sin \delta \cos \lambda & -\cos \delta \end{bmatrix} \\ &= [\hat{\mathbf{u}} \hat{\mathbf{a}} \hat{\mathbf{n}}]\end{aligned}$$

(Arnold & Townend 2007). From this it can be seen that  $\Theta = (\xi, \delta, \lambda)$ , Euler angles  $\Phi = (\phi, \theta, \psi)$ , and the rotation matrix  $R(\Theta) = [\hat{\mathbf{u}} \hat{\mathbf{a}} \hat{\mathbf{n}}]$  are all equivalent ways of describing a focal mechanism.

### A.1.2 Passive and Active Rotation

Given a coordinate system and a vector, a rotation matrix can be interpreted in two ways. *Passive rotation* is the case when the rotation matrix is thought of as rotating the coordinate system while the vector remains fixed - this was the method described in Section A.1.1.

Let the vector  $\mathbf{x}_{\mathbb{B}}$  be the representation of the vector  $\mathbf{x}$  in coordinate system  $\mathbb{B}$ . Then

$$\mathbf{x}_{\mathbb{A}} = R_{\mathbb{A}\mathbb{B}}\mathbf{x}_{\mathbb{B}}$$

gives the representation of  $\mathbf{x}$  in coordinate system  $\mathbb{A}$ . The columns of  $R_{\mathbb{A}\mathbb{B}}$  are the unit vectors of the axes of coordinate system  $\mathbb{B}$  expressed in coordinate system  $\mathbb{A}$ .

*Active rotation* is the case when the rotation matrix is thought of as rotating the vector while keeping the coordinate system fixed. A rotation of a vector  $\mathbf{x}$  about an axis to obtain  $\mathbf{x}'$  can be performed by putting

$$\mathbf{x}' = R(\phi, \theta, \psi)\mathbf{x}$$

for Euler angles  $(\phi, \theta, \psi)$ .

### A.1.3 Rotation Angles and Angular Difference

A rotation matrix  $R$  rotates a coordinate system or vector about an axis by an angle  $b$  given by

$$b = \cos^{-1} \left( \frac{\text{tr}(R) - 1}{2} \right).$$

The matrix  $R$  has one real eigenvalue, and the eigenvector corresponding to this represents the axis about which the rotation takes place (Arnold & Townend 2007).



If  $R_1$  and  $R_2$  are two rotation matrices, then the angular difference  $a$  between them is the minimum rotation about any axis needed to make the two coincide. This angle is given by

$$a = \cos^{-1} \left( \frac{\text{tr}(R_1^T R_2) - 1}{2} \right).$$

(Arnold & Townend 2007). This is a means of comparing two focal mechanism solutions obtained by different methods. Due to the existence of four equivalent focal mechanisms for a given solution (see Section 1.2.3), angular differences are calculated between  $R_2$  and the four equivalent representations of  $R_1$ , or vice versa. By convention the minimum of these four angles,  $\min(a)$ , is taken (Kagan 2007).

When  $a$  is zero, the matrices are identical. The maximum value for  $\min(a)$  varies between  $90^\circ$  and  $120^\circ$  depending on the axis about which the rotation occurs (Kagan 2007).

## A.2 Spherical Coordinates

The spherical coordinate system can be used to locate points in three dimensions. This coordinate system is useful due to the importance of directional data in this project.

Under this coordinate system, a point P in space is located using two angles and one distance. This description is based on that of Weir et al. (2005).

- $p$  is the distance from P to the origin.
- $\phi$  is the angle clockwise from the positive x-axis to P ( $0 \leq \phi \leq 2\pi$ ). In this project we take x to be positive northwards.
- $\theta$  is the angle from the positive z-axis to P ( $0 \leq \theta \leq \pi$ ). In this project we take z to be positive downwards.

All points we are interested in are located on the unit sphere. A point P is therefore able to be located using just the two angles  $\phi$  and  $\theta$ . This is shown in Figure A.1.

Spherical coordinates can be converted to Cartesian coordinates by the equations:

$$\begin{aligned} x &= \sin \theta \cos \phi \\ y &= \sin \theta \sin \phi \\ z &= \cos \theta \end{aligned}$$

Cartesian coordinates can be converted to Spherical coordinates by rearranging the above equations:

$$\begin{aligned} \phi &= \text{atan2}(y, x) \\ &= \begin{cases} \tan^{-1} \left( \frac{y}{x} \right) & y > 0 \\ \pi + \tan^{-1} \left( \frac{y}{x} \right) & y < 0 \end{cases} \\ \theta &= \cos^{-1}(z) \end{aligned}$$

## A.3 Lambert-Schmidt Projection

The stereographic projection of the focal sphere onto a circular area can be constructed using an equal-area Lambert-Schmidt projection. As mentioned in Section 1.2.1, the stereonet

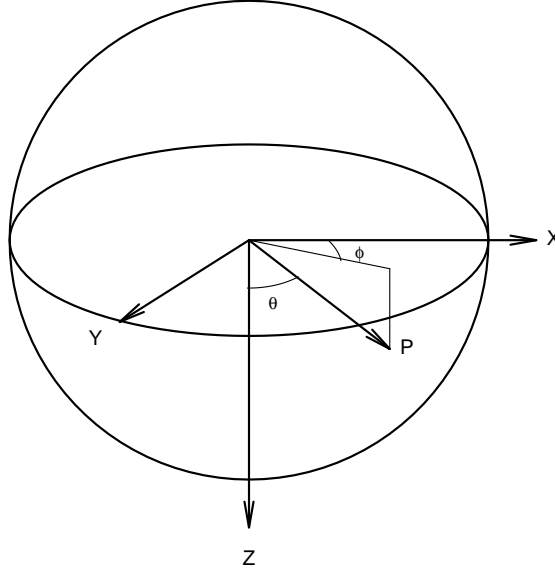


Figure A.1: Spherical polar coordinates. A point P on the sphere can be located by two angles:  $\phi$ , the clockwise angle from the positive x-axis, and  $\theta$ , the angle from the positive z-axis.

represents the lower hemisphere of the focal sphere. It is plotted using the usual mapping convention; North is upwards, East is to the right.

A point on a sphere has an azimuth  $\phi$  and a take-off angle  $\theta$ . This point is plotted on the stereonet at a distance  $r$  from the origin, at an angle  $\phi$  measured clockwise from north, where  $r$  is given by:

$$r = \sqrt{2} \sin \frac{\theta}{2}$$

This method is known as the Lambert-Schmidt projection (see Figure A.2), and is used to plot stereonets in this project.

We use R to construct the stereographic projections for each earthquake. We reverse the points on the upper focal sphere (points with  $\theta \geq \frac{\pi}{2}$ ) and plot them on the lower focal sphere, to enable every ray to be shown. If a point is on the upper focal sphere with azimuth  $\phi$  and take-off angle  $\theta$ , we can transform it to a point on the lower focal sphere with azimuth  $\phi'$  and take-off angle  $\theta'$  as follows:

$$\begin{aligned}\phi' &= \phi + \pi \\ \theta' &= \pi - \theta\end{aligned}$$

Compressional points are coloured blue, while dilatations are red.

Because the azimuth and take-off angle vary for each possible hypocentre location, there are multiple points to mark (creating a cloud) for each station, corresponding to the different possible hypocentre locations. The colour scale varies according to the value of the hypocentre PDF at the point - varying from white for zero probability, to dark blue or red for high probability.

Given a focal mechanism solution  $R(\Theta) = [\hat{\mathbf{u}} \ \hat{\mathbf{a}} \ \hat{\mathbf{n}}]$ , the nodal planes can be identified and plotted on a stereonet as follows. The points where the positive and negative directions of

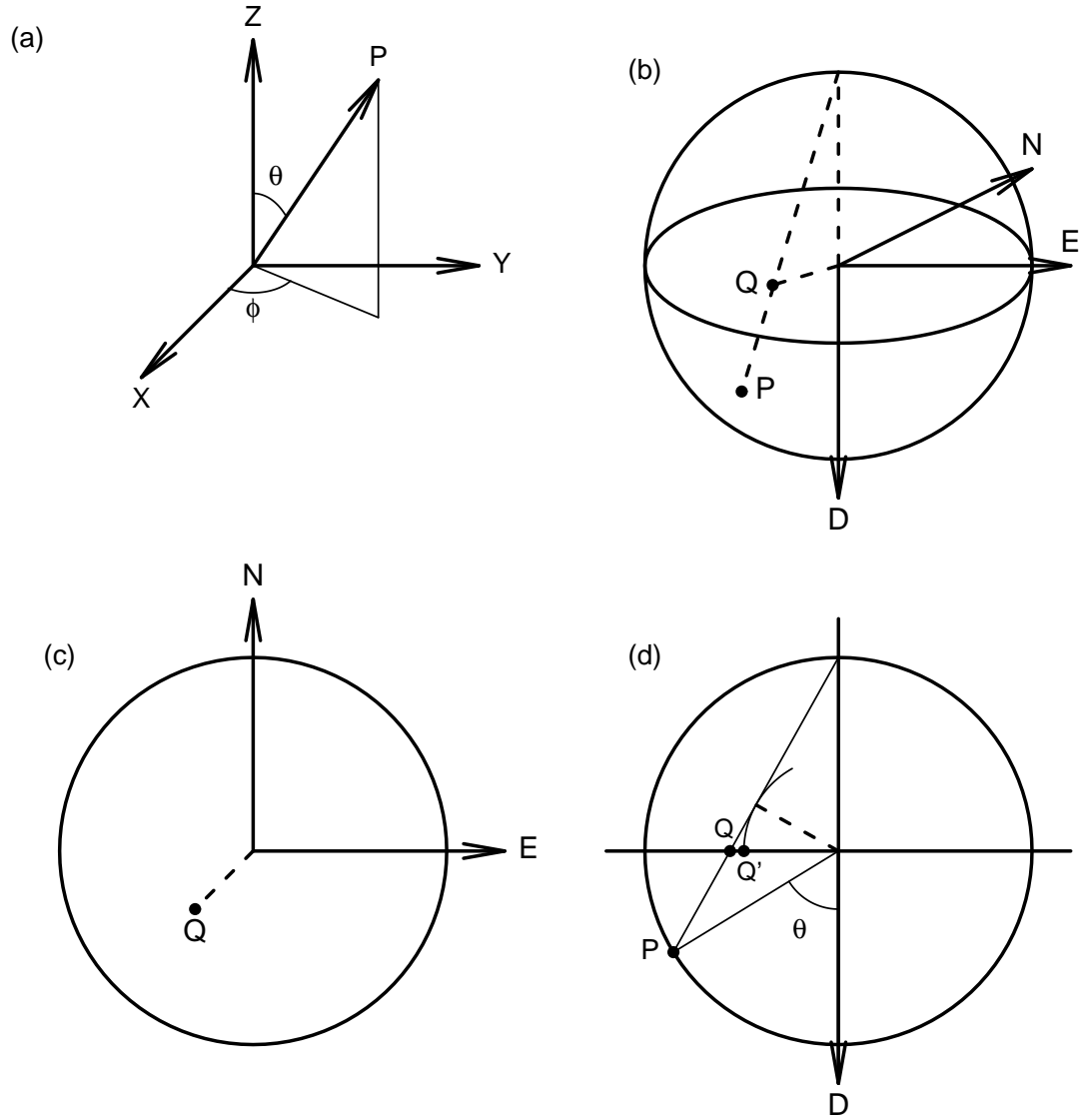


Figure A.2: Stereonet projection. (a) A point  $P$  on a sphere with azimuth  $\phi$  and take-off angle  $\theta$ . A point  $P$  on the lower half of the sphere is transformed as shown in (b) to the point  $Q$  which is plotted on the stereonet in (c). (d) The plotting position of the point under the Wulff  $Q$  and Lambert-Schmidt  $Q'$  conventions.

the vectors  $\hat{n}$ ,  $\hat{a}$  and  $\hat{u}$  intersect the focal sphere are marked. Two great circles, one passing through  $\pm\hat{n}$  and  $\pm\hat{a}$ , the other passing through  $\pm\hat{u}$  and  $\pm\hat{a}$ , give the nodal planes. Figure A.3 shows this procedure visually.

## A.4 Averaging Angles

One must take care when averaging angles that are scattered either side of  $2\pi$ . If we take the average of such angle values, the result will incorrectly be close to  $\pi$ , since around half the angles are just  $> 0$  and half are just  $< 2\pi$  (Arnold & Townend 2007). In this section we give formulae for averaging directional and axial data of various kinds.

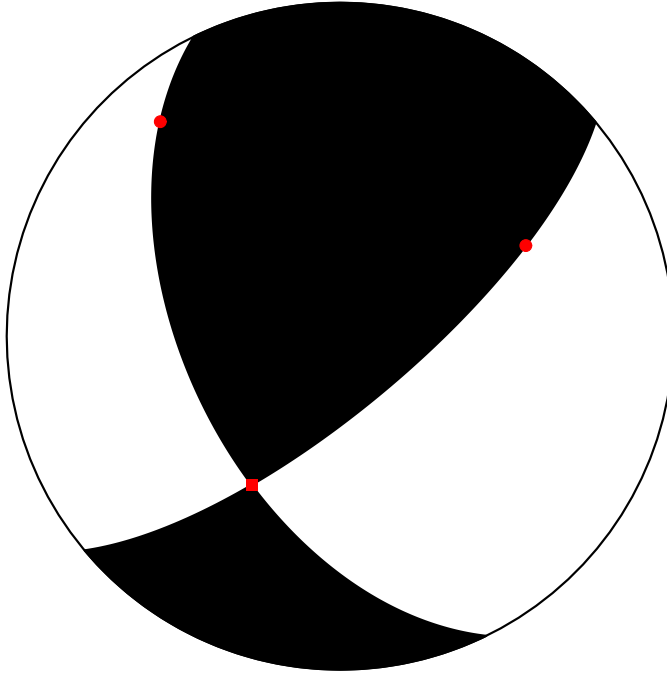


Figure A.3: Beachball diagram of a focal mechanism with an arbitrary strike, dip and rake of  $(\xi, \delta, \lambda) = (154^\circ, 52^\circ, 22^\circ)$ . The direction of the fault normal and slip vectors are shown by small circles, while the direction of the null vector is shown by the square.

#### A.4.1 Circular Data

When averaging  $n$  angles  $\{\phi_i\}$ ,  $i = 1, \dots, n$ , the necessary adjustment is

$$\bar{\phi} = \text{atan2} \left( \frac{\langle \sin \phi \rangle}{\langle \cos \phi \rangle} \right)$$

where  $\langle \dots \rangle$  represents an average or a weighted average. If the data is axial then the necessary adjustment is

$$\bar{\phi} = \frac{1}{2} \text{atan2} \left( \frac{\langle \sin 2\phi \rangle}{\langle \cos 2\phi \rangle} \right)$$

(Arnold & Townend 2007).

### A.4.2 Spherical Data

When averaging  $n$  directions specified in spherical polar coordinates by  $\{\phi_i, \theta_i\}$ ,  $i = 1, \dots, n$ , the necessary adjustment is

$$\bar{\phi} = \text{atan2} \left( \frac{\langle \sin \phi \rangle}{\langle \cos \phi \rangle} \right)$$

and

$$\bar{\theta} = \cos^{-1} \left( \frac{\langle \cos \theta \rangle}{\sqrt{\langle \sin \theta \cos \phi \rangle^2 + \langle \sin \theta \sin \phi \rangle^2 + \langle \cos \theta \rangle^2}} \right)$$

(Arnold & Townend 2007). If the data are axial then

$$\mathbf{r} = (\sin \bar{\theta} \cos \bar{\phi}, \sin \bar{\theta} \sin \bar{\phi}, \cos \bar{\theta})^T$$

is the unit vector in the mean direction (Arnold & Townend 2007).

## A.5 Change of Variable Technique

Here we will consider the change of variable technique involving a single variable. This method is based on that described by Hogg & Tanis (2001).

Given that a variable  $Y$  with pdf  $g(y)$  is a function of another variable  $X$  with pdf  $f(x)$ , i.e.  $Y = h(X)$ , and  $h$  is monotonic, how is the pdf  $g(y)$  related to  $f(x)$ ?

The function  $h$  maps a point  $x$  onto a point  $y$ . Hence the support of  $X$ , say  $x_{\min} \leq x \leq x_{\max}$  maps onto the support of  $Y$ ,  $h(x_{\min}) = y_{\min} \leq y \leq y_{\max} = h(x_{\max})$ . Thus the distribution function of  $Y$  can be written,

$$\begin{aligned} P(Y \leq y) &= P(h(X) \leq y) \\ &= P(X \leq h^{-1}(y)) \\ &= \int_{x_{\min}}^{h^{-1}(y)} f(x) dx \\ &= \int_{h^{-1}(y_{\min})}^{h^{-1}(y)} f(x) dx. \end{aligned} \tag{A.1}$$

Now, integration by substitution tells us that:

$$\int_a^b f(\phi(t)) \phi'(t) dt = \int_{\phi(a)}^{\phi(b)} f(x) dx$$

So, if we put  $\phi = h^{-1}$ ,  $b = y$ ,  $a = y_{\min}$ , and  $t = y$ , we can express Equation A.1 as:

$$\begin{aligned} \int_{h^{-1}(y_{\min})}^{h^{-1}(y)} f(x) dx &= \int_{y_{\min}}^y f(h^{-1}(y)) \frac{dh^{-1}(y)}{dy} dy \\ &= \int_{y_{\min}}^y f(x) dx \quad (\text{as } h^{-1}(y) = x) \end{aligned}$$

Therefore,

$$P(Y \leq y) = \int_{y_{\min}}^y f(x) dx \tag{A.2}$$

$$\implies \int_{y_{\min}}^y g(y) dy = \int_{y_{\min}}^y f(x) dx \tag{A.3}$$

$$\implies g(y) = f(x) \left| \frac{dx}{dy} \right| \tag{A.4}$$

So we now have a relationship between  $g(y)$  and  $f(x)$ . The reason for taking the absolute value of  $\frac{dx}{dy}$  is to ensure  $g(y)$  is non-negative when  $h(X)$  is monotonically decreasing. This term is known as a *Jacobian* (Sivia 1996).

### A.5.1 Lighthouse Example

We can now apply this method to the lighthouse example in Section 3.2.1. Here we had a uniform prior on  $c$ :  $P(c|a, b) = \frac{1}{\pi}$ , and a relationship between  $c$  and  $x$  specified by  $b \tan(c) = x_k - a$ . We want to obtain an expression for  $P(x_k|a, b)$ , so we apply Equation A.4

$$P(x_k|a, b) = P(c|a, b) \left| \frac{dc}{dx_k} \right|$$

If we rearrange  $b \tan(c) = x_k - a$ , to make  $c$  the subject we get:

$$c = \tan^{-1} \left( \frac{x_k - a}{b} \right)$$

Deriving with respect to  $x_k$  we get:

$$\frac{dc}{dx_k} = \frac{1}{b \left[ 1 + \left( \frac{x-a}{b} \right)^2 \right]}$$

Now we can transform  $P(c|a, b)$  into  $P(x_k|a, b)$ :

$$\begin{aligned} P(x_k|a, b) &= P(c|a, b) \times \left| \frac{dc}{dx_k} \right| \\ &= \frac{1}{b\pi \left[ 1 + \left( \frac{x-a}{b} \right)^2 \right]} \\ &= \frac{b}{\pi(b^2 + (x-a)^2)} \end{aligned}$$

Which is the Cauchy pdf.

## A.6 Hypergeometric Functions of a Matrix Argument

The general form of a hypergeometric function of a positive definite symmetric  $m \times m$  matrix argument  $\mathbf{X}$  is given by Muirhead (1982):

$${}_pF_q(a_1, \dots, a_p; b_1, \dots, b_q; \mathbf{X}) = \sum_{k=0}^{\infty} \sum_{\kappa} \frac{(a_1)_{\kappa} \dots (a_p)_{\kappa}}{(b_1)_{\kappa} \dots (b_q)_{\kappa}} \frac{C_{\kappa}(\mathbf{X})}{k!}$$

where  $\kappa$  is a partition of  $k$  up to length  $m$ , i.e  $\kappa = (k_1, \dots, k_m)$  where  $k_1 \geq k_2 \geq \dots \geq k_m$ , the  $k_i$  are non-negative integers and  $\sum_i k_i = k$  (Muirhead 1982). If say  $m = 3$ , some examples of such partitions are:

$k = 1$ : There is only one partition,  $\kappa = (1) \equiv (1, 0, 0)$ .

$k = 2$ : There are two partitions,  $(2) \equiv (2, 0, 0)$  and  $(1, 1) \equiv (1, 1, 0)$ .

$k = 3$ : There are three partitions  $(3) \equiv (3, 0, 0)$ ,  $(2, 1) \equiv (2, 1, 0)$  and  $(1, 1, 1)$ .

$k = 4$ : There are four partitions  $(4) \equiv (4, 0, 0)$ ,  $(3, 1) \equiv (3, 1, 0)$ ,  $(2, 2) \equiv (2, 2, 0)$  and  $(2, 1, 1)$ , however  $(1, 1, 1, 1)$  would not be included as its length is greater than  $m = 3$ .

Further,  $C_{\kappa}(\mathbf{X})$  is known as a *zonal polynomial*, a function of the eigenvalues of  $\mathbf{X}$ , which can be expressed in terms of *monomial symmetric functions*  $M_{\kappa}(X) = x_1^{k_1} \dots x_m^{k_m}$  (all distinct reorderings) where the  $x_i$  are the eigenvalues of  $\mathbf{X}$ . For example,

$$\begin{aligned} M_{(1)}(\mathbf{X}) &= x_1^1 x_2^0 x_3^0 + x_1^0 x_2^1 x_3^0 + x_1^0 x_2^0 x_3^1 \\ &= x_1 + x_2 + x_3 \\ M_{(2)}(\mathbf{X}) &= x_1^2 + x_2^2 + x_3^2 \\ M_{(1,1)}(\mathbf{X}) &= x_1 x_2 + x_1 x_3 + x_2 x_3. \end{aligned}$$

Now,

$$C_{\kappa}(\mathbf{X}) = \chi_{[2\kappa]}(1) \frac{2^k k!}{(2k)!} Z_{\kappa}(\mathbf{X}) \quad (\text{A.5})$$

where

$$Z_{\kappa}(\mathbf{X}) = \sum_{\kappa} c_{\kappa} M_{\kappa}(X) \quad (\text{A.6})$$

and

$$\chi_{[2\kappa]}(1) = k! \frac{\prod_{i < j}^m (k_i - k_j - i + j)}{\prod_{i=1}^m (k_i + m - i)!} \quad (\text{A.7})$$

(McLaren 1976, James 1964).

Finally,  $(a)_{\kappa}$  is known as the *generalised hypergeometric coefficient*, defined as

$$(a)_{\kappa} = \prod_{i=1}^m \left(a - \frac{1}{2}(i-1)\right)_{k_i}$$

where  $(a)_k = a(a+1) \dots (a+k-1)$  (Muirhead 1982).

The hypergeometric function of a matrix argument can be calculated in MATLAB using an algorithm by Koev & Edelman (2006). We can evaluate this function from R using the R.matlab package for R (Bengtsson 2007).

## A.7 Numerical Integration

Numerical integration methods are used to calculate an approximate solution to a definite integral  $\int_a^b f(x)dx$ . Numerical integration is particularly useful in this project for evaluating the integral of functions for which we cannot find the antiderivative.

### A.7.1 Trapezium Rule

The trapezium rule is a means of approximating a definite integral  $\int_a^b f(x)dx$  by calculating the area of a number of trapezia formed by splitting the region of integration  $[a, b]$  into  $n-1$  subintervals of equal width  $\Delta x = \frac{b-a}{n-1}$ . Figure A.4 indicates how the method works.

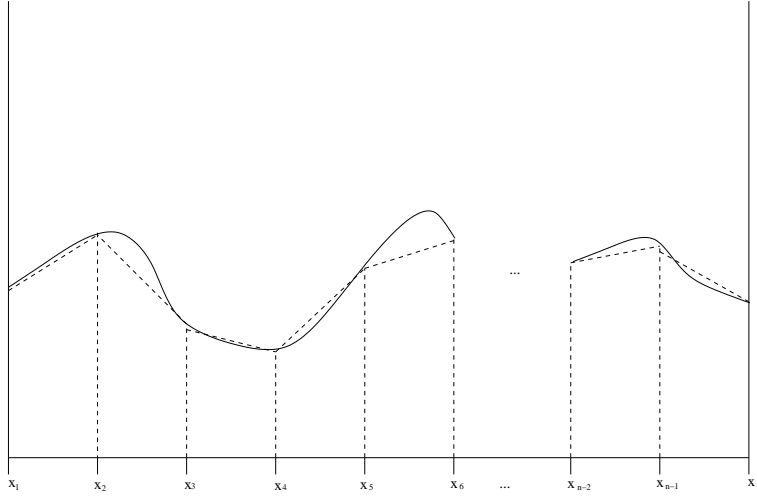


Figure A.4: Diagram illustrating the trapezium rule. The solid line is the function  $f(x)$ , which is evaluated at a number of points  $x_i$ . The dotted lines indicate the trapezia, the areas of which approximate the area under  $f(x)$ .

The approximation is given by

$$\begin{aligned}
 \int_a^b f(x)dx &\simeq \Delta x \sum_{i=1}^{n-1} \frac{1}{2} [f(x_i) + f(x_{i+1})] \\
 &= \frac{b-a}{n-1} \frac{1}{2} \left[ f(x_1) + f(x_n) + 2 \sum_{i=2}^{n-1} f(x_i) \right] \\
 &= \frac{b-a}{n-1} \left[ \sum_{i=1}^n f(x_i) - \frac{1}{2} (f(x_1) + f(x_n)) \right] \\
 &= \frac{b-a}{n-1} \sum_{i=1}^n w_i f(x_i)
 \end{aligned}$$

where

$$w_i = \begin{cases} \frac{1}{2} & i = 1 \text{ or } i = n \\ 1 & \text{otherwise} \end{cases}$$

This can be generalised to higher dimensions, e.g.

$$\int_c^d \int_a^b f(x, y) dx dy \simeq \frac{d-c}{m-1} \frac{b-a}{n-1} \sum_{i=1}^n \sum_{j=1}^m w_{ij} f(x_i, y_j)$$

while the weight function becomes  $w_{ij} = (1/2)^{p_{ij}}$ , where  $p_{ij}$  is the number of grid edges that point  $ij$  sits on.

## A.7.2 Monte Carlo Integration

Monte Carlo integration is a means of approximating a definite integral by evaluating the integrand at a random sample of points, as described in Robert & Casella (1999).

**Theorem A.1.** *[Monte Carlo integration] The integral*

$$\int_x h(x) f(x) dx$$



where  $f$  is the pdf of  $x$ , can be approximated by

$$\frac{1}{m} \sum_{j=1}^m h(x_j)$$

where  $\{x_j\}$  are  $m$  points generated from the density  $f$ .

*Proof.* The proof is simply:

$$\begin{aligned} \int_x h(x) f(x) dx &= E(h(x)) \\ &\simeq \frac{1}{m} \sum_{j=1}^m h(x_j) \end{aligned}$$

□

So in the case of Equation 3.19, we apply Monte Carlo integration to the following integral:

$$\int \left[ \prod_{i=1}^n P(Y_i | A_i^* = 2(\hat{\mathbf{p}}_i \cdot \hat{\mathbf{n}})(\hat{\mathbf{p}}_i \cdot \hat{\mathbf{u}}), \sigma_a, \pi_p) \right] P(\mathbf{x} | \{t_i\}, \omega) d\mathbf{x}$$

Here,

$$\begin{aligned} h(\mathbf{x}) &= \left[ \prod_{i=1}^n P(Y_i | A_i^* = 2(\hat{\mathbf{p}}_i \cdot \hat{\mathbf{n}})(\hat{\mathbf{p}}_i \cdot \hat{\mathbf{u}}), \sigma_a, \pi_p) \right] \\ f(\mathbf{x}) &= P(\mathbf{x} | \{t_i\}, \omega) \end{aligned}$$

where  $h(\mathbf{x})$  is a function of  $\mathbf{x}$  through  $\hat{\mathbf{p}}_i$ . The random sample from the density  $f$  are randomly sampled hypocentre locations. Therefore,

$$h(\mathbf{x}_j) = \left[ \prod_{i=1}^n P(Y_i | A_{ij}^* = 2(\hat{\mathbf{p}}_{ij} \cdot \hat{\mathbf{n}})(\hat{\mathbf{p}}_{ij} \cdot \hat{\mathbf{u}}), \sigma_a, \pi_p) \right]$$

where  $\hat{\mathbf{p}}_{ij}$  is the take-off vector for the ray traveling to station  $i$  from hypocentre location  $j$ . Hence the approximation is:

$$\begin{aligned} &\int \left[ \prod_{i=1}^n P(Y_i | A_i^* = 2(\hat{\mathbf{p}}_i \cdot \hat{\mathbf{n}})(\hat{\mathbf{p}}_i \cdot \hat{\mathbf{u}}), \sigma_a, \pi_p) \right] P(\mathbf{x} | \{t_i\}, \omega) d\mathbf{x} \\ &\propto \sum_{j=1}^m \left[ \prod_{i=1}^n P(Y_i | A_{ij}^* = 2(\hat{\mathbf{p}}_{ij} \cdot \hat{\mathbf{n}})(\hat{\mathbf{p}}_{ij} \cdot \hat{\mathbf{u}}), \sigma_a, \pi_p) \right] \end{aligned}$$

Where  $m$  is the number of hypocentre points sampled from the pdf  $P(\mathbf{x} | \{t_i\}, \omega)$ . Here we have dropped  $\frac{1}{m}$  into the normalisation constant, which we do not need to evaluate exactly.

## A.8 Tensors

A *tensor* can be thought of as a generalisation of the concept of vectors and matrices. Multiplying a vector by a scalar changes the magnitude but leaves the direction unchanged. If we wish to change the direction of the vector as well we need to utilise a different type of entity (Kolecki 2002).

Tensors can be classified by their rank - the number of array indices used to describe the object. For example, a tensor of rank 0 is a scalar, a tensor of rank 1 is a vector, and a tensor of rank 2 is a matrix (Arfken 1985). However, the converse is not true, not all scalars are tensors of rank 0, not all vectors are tensors of rank 1, and so on – only those that are *coordinate independent* are tensors (Kolecki 2002). For example, for any two coordinate systems whose origins differ, the position vectors  $\mathbf{v}$  and  $\mathbf{v}^*$  from the origin in the respective systems to a point  $P$  will be different – thus a position vector is not a tensor. However if there are two points  $P_1$  and  $P_2$ , with position vectors  $\mathbf{v}_1$  and  $\mathbf{v}_2$  in the first coordinate system and position vectors  $\mathbf{v}_1^*$  and  $\mathbf{v}_2^*$  in the second, then  $\mathbf{v}_2 - \mathbf{v}_1 = \mathbf{v}_2^* - \mathbf{v}_1^*$ , and thus the difference between two position vectors is a tensor of rank 1 (Kolecki 2002).

### A.8.1 Moment Tensor

The seismic moment tensor  $\mathbf{M}$  is a quantity that depends on source strength and fault orientation (Aki & Richards 2002). The moment tensor is a representation of the earthquake source by *equivalent body forces*, that is, the forces that would yield the observed seismic wave radiation pattern. Thus equivalent body forces are a model of the real faulting process (Stein & Wysession 2003).

These forces are described by force *couples*, which are two forces acting together. These two forces are offset by a distance  $d$ , either in the direction of the force or normal to the direction of the force. In 3D space and with three possible force directions there are nine possible couples, and these make up the components of the moment tensor (Aki & Richards 2002, Stein & Wysession 2003).

$$\mathbf{M} = \begin{bmatrix} M_{xx} & M_{xy} & M_{xz} \\ M_{yx} & M_{yy} & M_{yz} \\ M_{zx} & M_{zy} & M_{zz} \end{bmatrix}$$

The equivalent body forces that describe an earthquake are double-couples, so if the fault and slip directions are oriented along the coordinate axes, the moment tensor will be of the form

$$\mathbf{M} = \begin{bmatrix} 0 & M_0 & 0 \\ M_0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = M_0 \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

where  $M_0$  is the *seismic moment*, a measure of the magnitude of the earthquake. However, in general, the fault will not be oriented along the axes, and the moment tensor is given in terms of the fault normal and slip vector,

$$M_{ij} = M_0(n_i u_j + n_j u_i)$$

and thus the tensor is symmetric.

The pressure (or P-) axis, which is parallel to  $\hat{\mathbf{u}} - \hat{\mathbf{n}}$ , and the tensional (or T-) axis, parallel to  $\hat{\mathbf{u}} + \hat{\mathbf{n}}$  (Arnold & Townend 2007, Aki & Richards 2002), are the eigenvectors of the moment tensor.

## A.9 Dirac Delta Function

The Dirac delta function  $\delta(x)$  is defined by the following three properties:

$$\begin{aligned}\delta(x) &= 0 & x &\neq 0 \\ \int_{-\infty}^{\infty} \delta(x) dx &= 1 \\ \int_{-\infty}^{\infty} f(x) \delta(x) dx &= f(0)\end{aligned}$$

Thus  $\delta(x)$  is an infinite spike at  $x = 0$ , and only makes sense as part of an integrand (Arfken 1985). An important property used in this project is

$$\int_{-\infty}^{\infty} f(x) \delta(x - x_0) dx = f(x_0)$$

(Arfken 1985).



# Appendix B

## Raukumara earthquake hypocentres

Table B.1: Raukumara earthquake hypocentres, as located by NonLinLoc

CUSPID	Date	Origin time	Latitude	Longitude	Depth (km)	Mag. ( $M_L$ )
635146	25/07/1994	15:23:56.63	-38.22	178.12	26.66	3.5
635767	31/07/1994	05:58:40.01	-38.36	177.87	23.73	3.0
636036	03/08/1994	15:47:23.60	-38.52	177.85	33.40	2.8
636120	07/08/1994	22:28:12.33	-38.18	177.46	51.46	3.3
636149	09/08/1994	09:53:27.12	-38.30	177.68	35.84	3.7
637373	16/08/1994	20:19:24.35	-38.48	177.83	35.06	3.2
642468	19/08/1994	00:27:22.18	-38.60	177.88	21.00	2.9
640980	21/08/1994	13:36:52.95	-38.20	178.14	18.02	2.8
642506	21/08/1994	23:23:15.60	-38.95	177.70	20.80	3.2
642225	22/08/1994	03:17:20.45	-38.66	177.38	34.28	2.9
639865	22/08/1994	03:19:45.99	-38.95	177.71	20.61	4.1
658523	24/08/1994	07:54:36.18	-38.85	177.32	33.40	2.8
644710	24/08/1994	11:25:24.56	-38.26	178.22	21.97	2.8
659187	25/08/1994	01:47:20.93	-38.04	177.93	32.13	3.1
644854	25/08/1994	04:31:14.37	-37.88	178.32	18.85	3.2
646569	25/08/1994	10:14:00.77	-37.97	177.96	34.28	2.9
641198	25/08/1994	13:56:26.21	-37.90	177.79	46.29	3.0
642627	30/08/1994	18:06:26.47	-38.41	177.55	36.82	3.5
646630	01/09/1994	09:19:42.86	-38.47	178.10	25.88	2.8
646638	01/09/1994	17:05:10.03	-38.22	178.18	20.61	2.9
639642	01/09/1994	20:27:02.49	-37.94	178.05	28.42	2.9
645188	02/09/1994	07:01:22.55	-38.61	177.90	19.78	2.9
645191	03/09/1994	10:46:50.14	-38.15	178.36	18.26	3.0
645987	05/09/1994	07:05:40.39	-37.77	178.32	27.54	2.9
653977	07/09/1994	08:08:38.69	-38.62	177.79	30.66	2.8
652423	08/09/1994	01:25:23.77	-38.12	178.28	24.90	3.2
653984	09/09/1994	16:17:08.41	-38.69	177.97	27.54	2.8
652892	14/09/1994	01:14:32.50	-38.39	177.88	21.83	3.0
656806	15/09/1994	18:37:50.02	-38.42	177.84	29.74	4.9
653020	15/09/1994	18:42:45.48	-38.42	177.84	29.15	3.1
654104	16/09/1994	09:11:51.33	-37.96	178.02	30.96	2.8
652903	17/09/1994	15:55:53.38	-37.78	178.29	28.42	3.0
653521	18/09/1994	02:23:56.78	-38.27	178.13	24.37	3.0
653233	19/09/1994	02:17:21.79	-38.47	177.85	33.50	2.9

*Continued on next page*

Continued from previous page

CUSPID	Date	Origin time	Latitude	Longitude	Depth (km)	Mag. ( $M_L$ )
655865	21/09/1994	14:17:40.82	-38.52	178.11	23.93	2.9
655881	22/09/1994	13:12:22.13	-38.80	177.84	9.33	2.8
658508	23/09/1994	07:46:53.00	-37.79	178.11	32.91	2.9
660226	24/09/1994	03:52:03.05	-37.88	177.88	41.31	3.6
660234	24/09/1994	08:24:09.37	-37.90	177.87	38.38	3.0
660254	24/09/1994	18:21:08.84	-38.60	177.87	19.04	2.8
660615	25/09/1994	15:40:31.14	-38.67	178.04	22.56	3.0
660625	26/09/1994	06:57:47.16	-37.86	177.80	76.17	3.5
655486	26/09/1994	14:55:08.83	-38.04	177.98	20.95	2.8
655951	26/09/1994	14:59:41.39	-38.12	178.05	35.64	2.8
654658	28/09/1994	04:21:49.82	-38.23	178.59	28.42	3.7
667323	29/09/1994	01:18:27.86	-38.52	177.83	28.81	2.8
668273	02/10/1994	22:38:48.96	-37.87	178.08	30.96	2.9
669233	03/10/1994	20:51:10.99	-38.54	177.81	26.12	3.0
667842	04/10/1994	04:48:30.11	-38.41	177.83	31.35	2.8
675244	05/10/1994	22:03:53.16	-38.35	177.99	24.41	3.1
665887	06/10/1994	02:16:48.18	-38.69	177.89	12.55	3.2
665710	06/10/1994	10:44:18.11	-38.69	177.89	12.84	3.3
674383	06/10/1994	18:30:33.25	-38.35	178.03	25.73	3.1
668637	08/10/1994	02:42:06.54	-38.42	178.16	23.54	2.9
668882	08/10/1994	12:02:51.74	-38.43	178.18	25.68	2.8
665895	09/10/1994	11:34:02.22	-38.55	178.06	20.90	3.0
668888	09/10/1994	14:18:11.18	-38.63	177.81	32.23	2.8
671614	10/10/1994	23:33:05.53	-38.50	178.01	20.56	3.2
671618	13/10/1994	02:14:17.54	-38.42	177.84	29.59	3.0
672299	14/10/1994	08:36:42.06	-38.34	178.20	14.89	3.2
672060	16/10/1994	09:43:48.61	-37.83	178.38	10.45	2.9
673014	20/10/1994	06:44:09.59	-38.03	177.94	28.22	2.9
672690	21/10/1994	14:31:06.11	-38.16	178.29	16.75	3.2
672691	22/10/1994	03:39:16.27	-38.10	178.22	13.48	2.9
671833	22/10/1994	04:50:16.83	-38.35	177.72	48.54	3.3
675146	24/10/1994	01:18:43.58	-38.54	178.10	25.49	2.9
679082	25/10/1994	03:07:15.74	-38.52	177.92	24.71	2.8
679409	25/10/1994	07:29:49.47	-38.52	177.91	23.63	2.8
679418	26/10/1994	23:45:51.21	-38.59	177.90	18.41	2.9
689619	29/10/1994	20:51:33.63	-38.67	178.05	22.46	3.8
693503	29/10/1994	20:53:11.54	-38.68	178.02	22.85	2.9
678369	29/10/1994	22:17:06.97	-38.51	177.83	29.25	3.8
680800	04/11/1994	22:40:47.84	-38.38	177.42	5.37	3.1
680829	05/11/1994	13:56:53.33	-38.50	177.87	45.02	3.4
683290	10/11/1994	06:43:55.66	-38.35	178.14	16.11	2.8
689175	11/11/1994	16:10:47.68	-38.10	177.94	4.05	2.6
683331	13/11/1994	05:59:38.99	-38.51	177.84	37.70	2.8
683333	13/11/1994	18:35:26.71	-38.54	177.85	26.22	3.6
688300	15/11/1994	10:53:47.71	-37.71	177.59	80.37	3.7
694887	21/11/1994	04:47:38.10	-38.18	178.16	16.94	2.8
695581	23/11/1994	15:11:39.25	-38.23	178.19	20.85	3.0

Continued on next page

*Continued from previous page*

CUSPID	Date	Origin time	Latitude	Longitude	Depth (km)	Mag. ( $M_L$ )
694945	25/11/1994	16:37:47.73	-38.33	177.97	24.27	2.8
694964	28/11/1994	17:14:45.21	-38.51	177.91	24.76	2.8
696995	30/11/1994	03:36:49.75	-38.54	177.97	25.10	3.7
697029	05/12/1994	08:37:07.84	-38.88	177.55	34.08	3.5
703722	06/12/1994	18:08:59.09	-38.23	178.19	21.14	3.3
704459	14/12/1994	03:36:37.98	-37.88	178.00	62.89	3.8