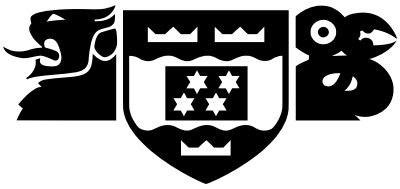
VICTORIA UNIVERSITY OF WELLINGTON
Te Whare Wananga o te Upoko o te Ika a Maui



School of Mathematics, Statistics
and Operations Research
Te Kura Mātai Tatauranga, Rangahau Pūnaha

PO Box 600
Wellington
New Zealand

School Tel: +64 4 463 5341
School Fax: +64 4 463 5045
School Email: office@msor.vuw.ac.nz

Measuring Shear Wave Splitting using the Silver and Chan Method

Ernestynne Walsh

Supervisors: Richard Arnold and Martha Savage

Submitted in fulfilment of the requirements for
Master of Science. Deposited January 2013.

Abstract

Seismic shear waves emitted by earthquakes can be modelled as plane (transverse) waves. When entering an anisotropic medium they can be split into two orthogonal components moving at different speeds. This splitting occurs along an axis, the fast direction, that is determined by the ambient tectonic stress. Shear wave splitting is thus a commonly used tool for examining tectonic stress in the Earth's interior. A common technique used to measure shear wave splitting is the Silver and Chan (1991) method. However, there is little literature assessing the robustness of this method, particularly for its use with local earthquakes, and the quality of results can vary.

We present here a comprehensive analysis of the Silver and Chan method comprising theoretical derivations and statistical tests of the assumptions behind this method. We then produce an automated grading system calibrated against an expert manual grader using multiple linear regression.

We find that there are errors in the derivation of certain equations in the Silver and Chan method and that it produces biased estimates of the errors. Further, the assumptions used to generate the errors do not hold. However, for high quality results (earthquake events where the signal is strong and the earthquake geometry is optimal), the standard errors are representative of the spread in the parameter estimates. Also, we find that our automated grading method produces grades that match the manual grades, and is able to identify mistakes in the manual grades by detecting substantial inconsistencies with the automated grades.

Acknowledgements

A big thank you to my supervisors Richard and Martha. I would still be here scratching my head if it wasn't for you guys - you were always full of ideas and able to answer all my questions, so thanks again.

I would also like to thank Victoria University Wellington for offering me a Masters by Thesis Scholarship. It was wonderful not having to worry about paying fees and it was great to receive a stipend or a youth pension as Dad calls it.

This project would not have been possible without the kindness of Elodie Rivemale and Florent Brenguier who supplied the Reunion Island data set used in this project.

A special mention goes out to Roma Klapaukh for his computer wizardry. Also, a mention to Sima for providing some nice hacks to make some of the more fancy R plots.

He mihi teenei ki aku rau rangatira me ooku hoa piripono i Wikitoria mo o koutou aawhina, manaakitanga aroha i nga waa katoa. Te reka o nga kai o Miraka me Miiere ka ki te puku, ka pai te haere o nga koorero ka pai hoki nga tuhituhi.

Lastly thanks to the Geophysics Discussion Group for being so nice to me and making me feel welcome.

Contents

1 Introduction **1**

1.1 Motivation and Objectives 1

1.2 Seismic Anisotropy and Shear Wave Splitting 2

1.2.1 Seismic Anisotropy - Definition 2

1.2.2 Seismic Anisotropy - Causes and Applications 2

1.2.3 Seismic Waves 3

1.2.4 Shear Wave Splitting 4

1.3 Obtaining Shear Wave Parameters and their Errors 5

1.3.1 Splitting Techniques 5

1.3.2 Estimating Errors 8

1.3.3 Problems with Identifying Shear Wave Splitting 9

1.3.4 Comparisons of Techniques 11

1.4 Contribution of Thesis 12

1.5 Outline 13

2 Silver and Chan (1991) Method - A Full Description **15**

2.1 The Shear Wave 15

2.2 Geophysical Properties 19

2.3 Split Wave and the Splitting Operator 20
2.4 Inverse Splitting Operator 24
2.5 Coordinate Systems and Rotations 25
2.5.1 Overview 25
2.5.2 Coordinate Systems 27
2.5.3 Shear Wave Splitting and Rotated Components 33
2.6 Covariance Matrix 36
2.7 Confidence Region 43
2.8 Degrees of Freedom 46
2.9 Estimating the Quantities Needed to Calculate the Degrees of Freedom 54
2.10 Summary 58

3 Grading System **61**

3.1 Dataset 61
3.2 Manual Grading 64
3.2.1 Mfast Output Plots 64
3.2.2 Manual Grading Criteria 66
3.2.3 Manual Grading Results 68
3.3 Automated Grading 71
3.3.1 Converting Manual Criteria into Numerical Formulae 71
3.3.2 Using the Criteria to Automate Grading 82
3.3.3 Suitability of Model 92
3.3.4 Summary of Model Suitability 99

4 Incoming Polarisation & Cycle Skipping **101**

4.1 Incoming Polarisation . . . 101

4.2 Cycle Skipping . . . 102

4.3 Checking for Cycle Skipping . . . 103

4.4 Effects on the Incoming Polarisation . . . 104

5 Testing Assumptions & Estimators **111**

5.1 Test 1: Bias of Estimators . . . 111

5.1.1 Test 1 Program . . . 111

5.1.2 Test 1 Results . . . 116

5.2 Test 2: Chi-squared Assumption . . . 121

5.2.1 Testing the Distribution of the Data . . . 121

5.2.2 Comparison to Background Noise . . . 126

5.3 Test 3: Reliability of the Standard Errors . . . 130

5.4 Summary . . . 136

6 Conclusions and Future Directions **139**

Glossary **143**

List of Symbols **149**

References **155**

A Useful Definitions and Results **159**

A.1 Matrix Exponential . . . 159

A.2 Why $\hat{\mathbf{f}}\hat{\mathbf{f}}^T + \hat{\mathbf{s}}\hat{\mathbf{s}}^T = \mathbf{I}$ in 2D . . . 159

A.3 Rotation with Euler Angles 161

A.4 Proof that the Trace is Invariant in the 2D Problem 162

A.5 Relationship between Gamma, Chi-squared and Exponential Distributions . . 163

A.6 Trapezoidal Rule 164

Figures

1.1 Single crystals in purple randomly oriented in an isotropic medium and lattice preferred orientation causing anisotropy. 2
1.2 Illustration of a P wave and the two S wave components. The SV component is inside the plane of propagation and the SH component is normal to it. . . . 4
1.3 Examples of an SKS wave travelling through the Earth. Diagram produced by Mattieu Duclos and Etienne Audoine (personal communication, 2012). . . 5
1.4 Illustration of shear wave splitting caused by microcracks. The cracks that are aligned parallel to the maximum horizontal stress remain open whilst cracks that are not parallel to the maximum horizontal stress close. The incoming shear wave splits into two orthogonal components labelled s_1 and s_2. The splitting is measured by the two splitting parameters ϕ and δt. Diagram produced by Alex Gerst (personal communication, 2012). 6
1.5 Example of a seismogram for the East, North and vertical components. The shear wave labelled S is located within the analysis window shown by the vertical red lines. 7
1.6 Illustration of cycle skipping. Cycle skipping can occur here because the dashed blue peak can in phase with either peaks or the trough of the solid black wave (Matcham, 1997). 10

2.1 An unsplit shear wave $\mathbf{u}(\omega)$ travelling towards the surface shown in propagation coordinates $(\mathcal{P})$. The wave is oriented along $\hat{\mathbf{p}}$ and travelling along $\hat{\mathbf{b}}$. 17
2.2 An initial shear wave split in two, modelled in space. The faster wave moves ahead and the slower wave lags behind. The dot represents the half way point between the two peaks and is where the wave would have been centred if it had not been split. 18
2.3 An initial shear wave is split in two with the faster (red) $u_1(\omega, t)$ wave being moved back $\delta t/2$ in time (i.e. sped up) and the slower (blue) $u_2(\omega, t)$ one moved forward in time by $\delta t/2$. 21
2.4 Diagram showing the relationship between the orthogonal unit vectors $\hat{\mathbf{f}}$, $\hat{\mathbf{s}}$ and $\hat{\mathbf{b}}$ that form the splitting coordinates $(\mathcal{S})$ and the vectors $[\hat{\mathbf{p}}\hat{\mathbf{a}}\hat{\mathbf{b}}]$ that define the propagation coordinates $(\mathcal{P})$. 23
2.5 Propagation coordinates $(\mathcal{P})$ 27
2.6 Rotation from $\mathcal{P}$ to $\mathcal{S}$ 27
2.7 Rotation from $\mathcal{P}$ to $\mathcal{E}$ 28
2.8 Rotation from $\mathcal{E}$ to $\mathcal{R}$ 29
2.9 Rotation from $\mathcal{R}$ to $\mathcal{G}$ 30
2.10 Rotation from $\mathcal{G}$ to $\mathcal{B}$ 30

2.11 Different right hand coordinate systems for a near vertically travelling wave that may be used to model the shear wave splitting process. Not drawn to scale. 32
2.12 Two wave components u_1 and u_2 that only differ by a time shift. If the faster wave u_1 is shifted back by δt then the waves look similar and the covariance between the two waves is maximised. 39
2.13 Two examples of particle motion. Top graphs show the fast and slow components. The top left plot, which is prior to correcting, shows a time lag between the two components. The top right graph, after rotating and time shifting one of the components should have both wave forms in phase. The bottom plots show the particle motion of the fast and slow components before and after correcting. The particle motion should be linear after correcting because the splitting has been reversed so the unsplit shear wave should be oriented along a single direction $\hat{\mathbf{p}}$. 40
2.14 Illustration of three different confidence regions. Minimum is denoted by the cross and the 95% confidence region by the bolded lines. 43
2.15 Example of aliasing. If we sample only at the red circles then the two waves become indistinguishable. 48
2.16 Illustration of sampling a continuous process with even spacing over the interval. 49

3.1 Map of Reunion island and the BOR station. World map produced using Google Earth (Google Earth, 2011). 62
3.2 Example of replicates of the North component from BOR station for the first 10 seconds. 0 marker represents the event origin; A, the P wave arrival; T2, the window beginning; T3, the window end and T0 the S wave arrival. 63
3.3 Flow chart showing how a subset of family 1 was chosen for analysis. 64
3.4 Example of a poor quality event (grade 5 - the worst grade). Notice the poor waveform match, non linear particle motion (bottom left), multiple 95% confidence contours (bottom right) and energy pushed outside the window (top right). 65
3.5 Example of a high quality event (grade 1 - the best grade). Notice the good waveform match, linear particle motion (bottom left) and the single bullseye style, 95% confidence contour (bottom right). 66
3.6 R generated output plot of the event corresponding to Figure 3.5. 67
3.7 Example of a grade 2 event. Notice the good fit between the waveforms (bottom left) but no bullseye style contours (bottom right). 69
3.8 Example of a grade 3 event. Notice the good fit between the waveforms (bottom left) but multiple 95% contours (bottom right). 70
3.9 Example of a grade 4 event. Notice a moderate fit between the waveforms with possible cycle skipping (bottom left) and multiple 95% contours (bottom right). 71
3.10 Plot of fast direction estimates against delay time estimates after the `jitter()` function was applied. Numbers and colours reflect the grades. 72
3.11 Plot of fast direction error estimates against delay time error estimates after the `jitter()` function was applied. Numbers and colours reflect the grades. . 73
3.12 A self-intersecting polygon with interior points defined by the even-odd rule. Numbers denote how many edges the ray passes through. 76
3.13 Area inside the polygon based on the even-odd rule. Red is the area inside, black is considered outside the polygon. 77

3.14 Illustration of how Step 3 works. A search back and forth across the rows and the columns must be made so that none of the area outside the polygon is missed. . . . 78
3.15 Area inside the polygon using an iterative method. Blue is the area inside, green is considered outside the polygon. . . . 78
3.16 Example where the area of the corrected particle motion is slightly larger than the area of the original particle motion. This is because the top of the turnip shaped particle motion expanded into 2 large petals after correcting for splitting. Compare this to Figure 3.5 which has a small particle motion area after correcting for splitting. . . . 79
3.17 Summary of what is a great result and what is a terrible result for each grading criterion. . . . 81
3.18 Eigenvalue grading criteria GC_{1a} and GC_{1b} plotted against the manual grades. The `jitter()` function is applied along the y axis to make results more visible. 83
3.19 Waveform grading criteria GC_2 and GC_{3a} plotted against the manual grades. The `jitter()` function is applied along the y axis to make results more visible. 84
3.20 Waveform grading criterion GC_{3b} and particle motion criterion GC_4 plotted against the manual grades. The `jitter()` function is applied along the y axis to make results more visible. . . . 85
3.21 Anomalous grade 1 event. Note the absence of bullseye style confidence contours. . . . 86
3.22 Example plot corresponding to a high GC_{3b}. . . . 87
3.23 Comparison of automated grades to manual grades. Red points indicate large discrepancies. Blue line indicates agreement between automated and manual grades. . . . 91
3.24 Contour plots showing discrepancies in grading. M denotes manual grade and A denotes automated grade. . . . 92
3.25 Diagnostic Plots checking normality and constant variance. . . . 94
3.26 Diagnostic plots of the jackknife residuals for each coefficient. . . . 95
3.27 Scatterplot matrix of predictors along with their correlation (r) and p-value, p. 96
3.28 Plots of leverage points and Cook's distance. Values that exceed the horizontal dashed blue line are considered leverage and influence points respectively. 97
3.29 Plots of the DFBETAs for each criterion used in the multiple linear regression. Horizontal dashed blue line is the cutoff. . . . 98

4.1 Incoming polarisation angle in geographical coordinates. . . . 101
4.2 Illustration of cycle skipping. Cycle skipping can occur here because the dashed blue peak can be matched to either peaks or the trough of the solid black wave. Adding a pre S phase means that the two waveforms are more likely to be matched correctly. . . . 102
4.3 Two contour plots of different events from group 1 (fast directions 0-45°). Note the splitting parameters differ (ϕ is similar but δt is different). . . . 104
4.4 Two wave forms of different events from group 1 (fast directions 0-45°). Note the different shifts of the slow wave annotated in red. . . . 105
4.5 Two contour plots of different events from group 2 (fast directions 90-120°). Note the splitting parameters differ. . . . 105
4.6 Two waveforms of different events from group 2 (fast directions 90-120°). Note the different shifts of the slow wave annotated in red. . . . 106

4.7 Polarisation waveforms of different events from group 1 (fast directions 0-45°). Note that the p ($\hat{\mathbf{p}}$) component of one looks like the $p \perp$ ($\hat{\mathbf{a}}$) component of the other. 106
4.8 Two wave forms of different events from group 1 (fast directions 0-45°). Note the different shifts and incoming polarisations annotated in red. The polarisation annotations show ϕ_b and α (see Equation 4.1). These two parameters are also given on the contour plot (Figure 4.3). 107
4.9 Incoming polarisation estimates vs delay time estimates for group 1 (fast directions 0-45°) events. Two clusters have formed that are $90° \pm SE(spol)$ apart. 108
4.10 Diagram illustrating that the incoming polarisation obtained from the Silver and Chan (1991) method will differ from the polarisation obtained from the focal mechanism. The focal mechanism method returns an α based on the polarisation of the original wave while the method in this chapter (Silver and Chan, 1991) returns an α based on one of the split waves. If the waveforms are split by a time greater than the wave length, then α is the fast direction for the second to last layer from the surface. 110

5.1 Example of the Silver and Chan (1991) estimated degrees of freedom using the a_n coefficients. Estimates have been both scaled and unscaled for a single smoothing width (w=0.0001), which is given in the title of the plot. 117
5.2 Example of the estimated degrees of freedom using our b_n coefficients. Estimates have been both scaled and unscaled for a single smoothing width (w=0.0001), which is given in the title of the plot. 118
5.3 Comparison of the scaled estimated degrees of freedom for different smoothing widths based on `nrep`=1,000 repetitions. Circles are based on our degrees of freedom based, which uses the b_n coefficients. Triangles denote the Silver and Chan (1991) estimate of the degrees of freedom based on the the a_n coefficients that are used in the Silver and Chan (1991) code. If the estimates were unbiased they would sit on the horizontal red line. 119
5.4 Comparison of the true degrees of freedom for various smoothing widths. . . 120
5.5 Example of the test statistic D which is the largest vertical distance between two distribution functions. 122
5.6 Grade 1 event - plots of the empirical distribution, the windowed noise in the time domain and the squared modulus fast Fourier transform of the windowed noise up to the Nyquist frequency 124
5.7 Grade 2 event - plots of the empirical distribution, the windowed noise in the time domain and the squared modulus fast Fourier transform of the windowed noise up to the Nyquist frequency. 124
5.8 Grade 3 event - plots of the empirical distribution, the windowed noise in the time domain and the squared modulus fast Fourier transform of the windowed noise up to the Nyquist frequency. 125
5.9 Power of the windowed noise and background noise for a grade 1 event. . . . 127
5.10 Power of the windowed noise and background noise for another grade 1 event. 127
5.11 Power of the windowed noise and background noise for a grade 2 event. . . . 128
5.12 Power of the windowed noise and background noise for another grade 2 event. 128
5.13 Power of the windowed noise and background noise for a grade 3 event. . . . 129
5.14 Power of the windowed noise and background noise for another grade 3 event. 129
5.15 Histograms of the standard errors and estimates using the Silver and Chan (1991) method. Number of events=146. 133

5.16 Histograms of the standard errors and estimates using the Silver and Chan (1991) method for grade 1, 2 and 3 results. Number of events=107. 134
5.17 Histograms of the standard errors and estimates using the Silver and Chan (1991) method for grade 4 and 5 results. Number of events=39. 135
5.18 Diagram illustrating the relationship between the interquartile range and the standard deviation. Vertical dash dot lines denote the lower and upper quartiles. 136

A.1 Diagram showing how to rotate a coordinate system using zyz right hand convention. 161
A.2 Diagram illustrating the trapezoidal rule. The area A under the solid curve is approximated by summing up the area of the A_n trapezia. 164

List of Tables

2.1 Summary of the different right handed coordinate systems and the vectors that define them. The angles that rotate you into these coordinate systems is also given along with the appropriate rotation. 31

2.2 Differences between the formulae given in this thesis and the ones provided in the Silver and Chan (1991) article. Differences are indicated by red symbols (ignoring the extension into 3 dimensions). Note that we have chosen the equations from our thesis that are the most similar to the ones in Silver and Chan (1991) so that comparisons can easily be made. 59

3.1 Number of Group 1 events in each grade band. 69

3.2 Summary of the grading criteria. 80

3.3 Chosen model and AIC at each iteration. 91

5.1 Summary statistics of the maximum amplitude for the two types of noise. . . 126

5.2 Comparison of median standard errors to a regular standard deviation and a standard deviation based on the interquartile range. All values rounded to 2dp. 136

Chapter 1

Introduction

1.1 Motivation and Objectives

Seismic waves are waves of energy that travel through the Earth. They occur during events such as earthquakes and travel from the location of the event (called the source) and are picked up by seismometers [1] at stations (called receivers) on the surface of the Earth. Seismic waves are very useful because they provide information about what is happening within the Earth's interior. To extract this information contained in the seismograms many methods have been developed, but the one we will consider is shear wave splitting using the Silver and Chan (1991) method (see section 1.3).

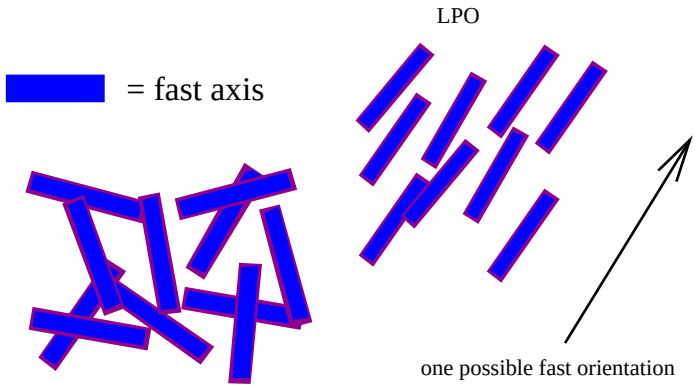
While this method is used by many and there are several software packages available implementing it, no one has conducted in-depth research to assess the robustness of it. Here we provide an in-depth analysis of the Silver and Chan (1991) method analysing the theory, implementation and the reliability of the results produced by this method.

More specifically, the objectives are

1. provide a detailed theoretical discussion of how the Silver and Chan (1991) method works
2. construct an automated method to grade the results obtained from the Silver and Chan (1991) method
3. test key assumptions that the theory relies upon and assess the impact they are having on results

The next two sections outline the basics of shear wave splitting.

[1] Green word(s) correspond to a glossary entry, page 143. These are links in the electronic copy



(a) Isotropic (b) Anisotropic

Figure 1.1: Single crystals in purple randomly oriented in an isotropic medium and lattice preferred orientation causing anisotropy.

1.2 Seismic Anisotropy and Shear Wave Splitting

1.2.1 Seismic Anisotropy - Definition

Anisotropy is a property where a medium possesses preferential orientation. A special type we focus on here is seismic anisotropy, where the velocity of a seismic wave depends on its propagation direction and its orientation or polarisation as well as the properties of the medium itself (Babuska and Cara, 1991). When a wave passes through an anisotropic medium the wave's velocity will be faster in one direction and slower in an approximately orthogonal direction. In contrast, an isotropic medium has wave velocities that are independent of direction. The magnitude of the anisotropy is expressed as a percentage as is given by the anisotropy coefficient k (Babuska and Cara, 1991).

$$k = \frac{v_{max} - v_{min}}{\bar{v}} \times 100 \tag{1.1}$$

where

$$\begin{aligned} v_{max} &= \text{velocity of the shear wave along the fast direction} \\ v_{min} &= \text{velocity of the shear wave along the slow direction} \\ \bar{v} &= \text{an average of } v_{max} \text{ and } v_{min} \end{aligned}$$

1.2.2 Seismic Anisotropy - Causes and Applications

Anisotropy can be caused by different properties. One of the most common causes of anisotropy in the mantle is preferred alignment of crystals such as olivine to create lattice preferred orientation (LPO) (Stein and Wysession, 2003) as shown in Figure 1.1. In the anisotropic medium with LPO a seismic wave will travel faster parallel to one axis of the

crystal and slower if it is perpendicular to that axis. Since olivine is one of the main minerals in the upper mantle (Stein and Wysession, 2003) the alignment of such crystals can provide information regarding strain and deformation in the upper mantle.

On the other hand, anisotropy in the crust tends to be caused by composite structures such as horizontal layers of sediments (Stein and Wysession, 2003). Fluid filled microcracks can also cause anisotropy with their opening and closing being an indication of the stress applied to them.

Anisotropy studies are important because they have many applications such as analysing present day tectonic actions (Silver and Chan, 1991), monitoring stress changes (e.g. Teanby et al. (2004*b*)) monitoring temporal changes in volcanoes (e.g. Gerst and Savage (2004)) and providing information for reservoir exploitation (Angerer et al., 2002). Crampin and Chastin (2003) provide a detailed discussion of applications of crustal shear wave splitting as well as references to other papers.

1.2.3 Seismic Waves

There are three types of waves we refer to in this thesis.

P Wave

The P wave is the fastest travelling wave. It is a longitudinal wave, i.e. it oscillates in the direction of the wave propagation (Stein and Wysession, 2003). Whilst P waves are not often used in shear wave splitting analyses, they are mentioned here since there are some comments about them in this thesis.

S Wave

The wave of interest in this thesis is the shear wave or S wave. This wave travels more slowly than the P wave and unlike the P wave it oscillates perpendicularly to the wave propagation. Both types of waves are illustrated in Figure 1.2. The S wave can be broken down into a horizontal (SH) component, and the component perpendicular to this, referred to as the SV even though it is not purely vertical. The direction that SV displaces is referred to as the radial direction. SV describes displacement that is in the vertical plane whilst SH displaces horizontally or normally to the plane (Figure 1.2). The direction that SH displaces is referred to as the transverse direction.

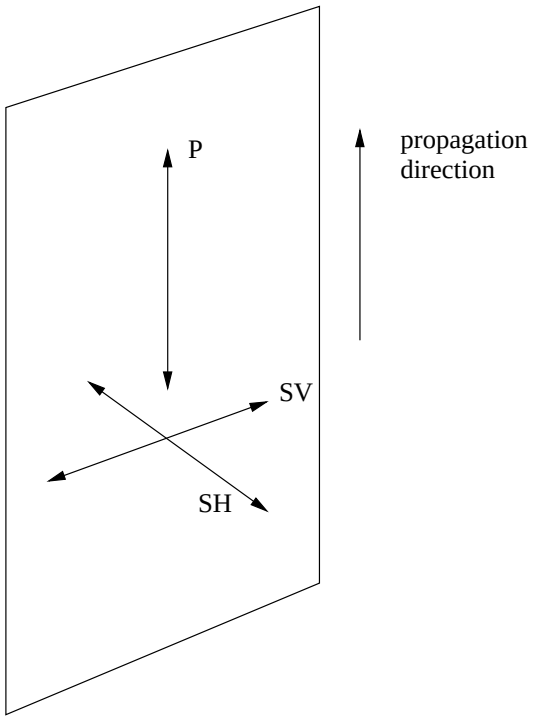


Figure 1.2: Illustration of a P wave and the two S wave components. The SV component is inside the plane of propagation and the SH component is normal to it.

SKS Wave

Both types of waves can be converted into other types of waves whilst travelling underground. One converted wave of interest is the SKS wave (Figure 1.3). S waves cannot pass through the outer core since it is liquid (Stein and Wysession, 2003) so an S wave is converted into a P wave as it enters the outer core. If it is converted back to an S wave as it leaves the core and passes back into the mantle then it is called an SKS wave. These waves are popular for the analysis of geophysical properties because they are easy to use and interpret (Savage, 1999). When the wave is converted back from a P wave to an S wave the energy is only in the vertical plane so the S wave only has an SV component (Figure 1.2). This means that the incoming polarisation (i.e. the orientation) is always known for an SKS wave and does not need to be estimated (this also applies for another type of wave, the SKKS wave). If there were energy on the SH component it can be interpreted as noise. However, not all waves that reach a station are SKS or SKKS wave so other more general S waves must be used. A general S wave has a polarisation that depends on the fault geometry and usually contains a signal on its SH component.

1.2.4 Shear Wave Splitting

Shear wave splitting is one of the most common methods used to measure anisotropy. A shear wave will split into two perpendicular components with different velocities as it passes through an anisotropic medium. The splitting is characterised by two parameters

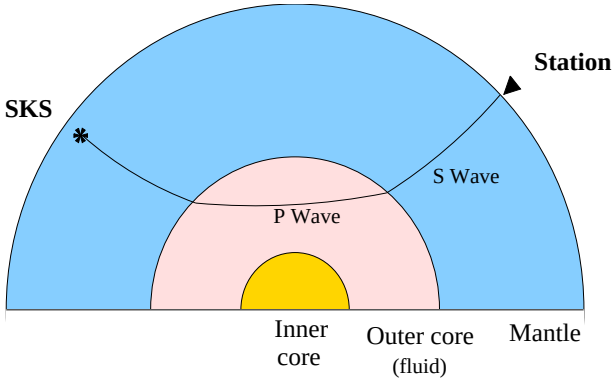


Figure 1.3: Examples of an SKS wave travelling through the Earth. Diagram produced by Mattieu Duclos and Etienne Audoine (personal communication, 2012).

defined as follows

$$\phi = \text{fast direction}$$
$$\delta t = \text{delay time}$$

The fast direction is the polarisation (orientation) of the faster travelling which is usually reported with respect to North wave whilst the delay time represents the temporal separation between the fast and slow waves when they arrive at the receiver. Figure 1.4 illustrates this. These two parameters are referred to as the splitting parameters and provide information about the anisotropic medium. For a single anisotropic layer along a shear wave's ray path the fast polarisation identifies the fast orientation (see Figure 1.1) in the anisotropic medium for that path. The delay time is a function of distance and velocity and will therefore give a measure of thickness of the anisotropic layer. The thickness of the medium can be obtained if you know or assume the percentage of anisotropy. If splitting is caused by micro cracks then the fast direction would give you an idea of how the cracks are aligned and the delay time would give you an estimate of crack density. In some cases, even when anisotropy is present it can not be detected. A null measurement is one where splitting cannot be detected for one of three reasons: there is no splitting, the incident wave is parallel to the fast direction or it is parallel to the slow direction.

The next section shows how the splitting parameters and their standard errors are obtained.

1.3 Obtaining Shear Wave Parameters and their Errors

1.3.1 Splitting Techniques

Three techniques often used to obtain splitting parameters are the transverse minimisation method, eigenvalue method and cross correlation method (Vecsey et al., 2008). These techniques are applied to a data set consisting of three seismogram components at a specified sampling rate, for a specified amount of time. The S wave arrival is identified and a window,

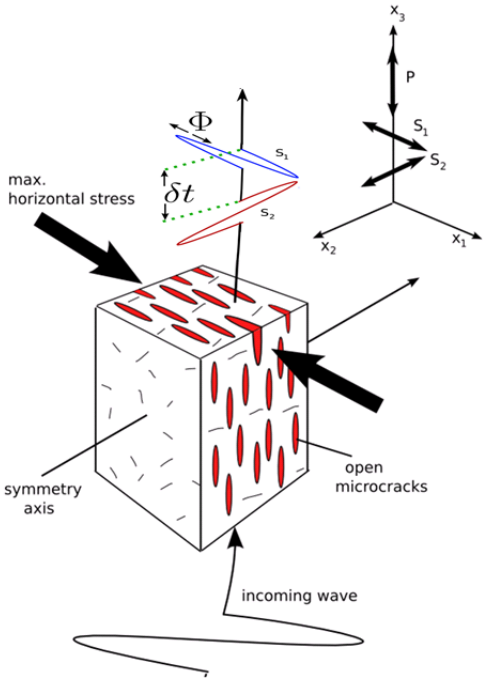


Figure 1.4: Illustration of shear wave splitting caused by microcracks. The cracks that are aligned parallel to the maximum horizontal stress remain open whilst cracks that are not parallel to the maximum horizontal stress close. The incoming shear wave splits into two orthogonal components labelled s_1 and s_2. The splitting is measured by the two splitting parameters ϕ and δt. Diagram produced by Alex Gerst (personal communication, 2012).

which is the segment of data containing the S wave, is identified for analysis purposes. Figure 1.5 shows an example of the three seismogram components. The window is illustrated by the vertical red lines while the vertical black line extending down through all three components denotes the S wave arrival.

Cross correlation Method

The cross correlation method was used early on by Fukao (1984) who used two rotated and time shifted orthogonal horizontal components to reverse the effect of splitting. The fast direction and a delay time that would lead to the greatest correlation between these two components was considered the optimal parameter set. For this method to work Fukao (1984) notes that the orthogonal components must contain similar waveforms. This method can be used at long period as well as at high frequency (Long and Silver, 2009). However, Crampin and Gao (2006) note that the method is dependent on selecting appropriate end points for the window (although Teanby et al. (2004*a*) provide a solution to this). Crampin and Gao (2006) also note that the waveforms are often not similar and that fast and slow waves are not always strictly orthogonal, making the cross correlation of these components almost impossible. However, teleseisms (such as SKS) do not usually encounter these problems since they have similar waveforms and near normal incidence, making the cross correlation technique useful for measuring anisotropy from splitting that occurs in the mantle.

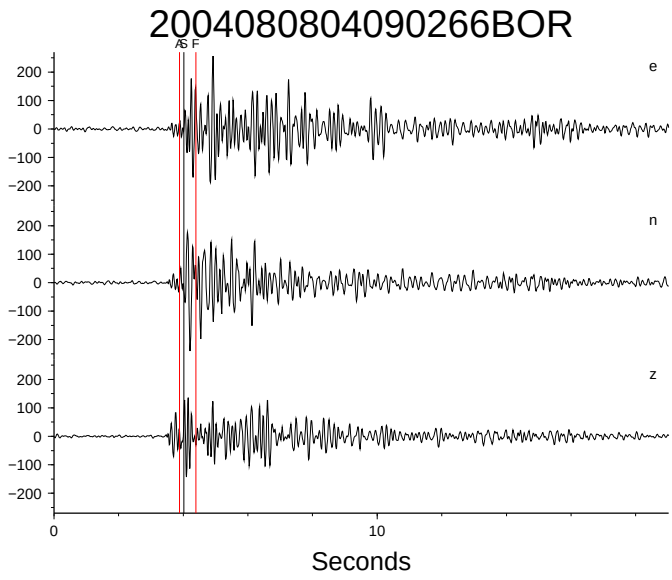


Figure 1.5: Example of a seismogram for the East, North and vertical components. The shear wave labelled *S* is located within the analysis window shown by the vertical red lines.

Transverse Minimisation Method

This method, introduced by Silver and Chan (1991), attempts to reverse the effects of splitting to retrieve the original wave (that is assumed to be propagating in a near vertical direction upwards from the source to the station) before it split by using an inverse splitting operator. To do this it performs a grid search over all possible pairs of $(\phi, \delta t)$ and selects the pair that best remove the effects of splitting. The best parameters are the ones that minimise the energy on one of the transverse components. The method assumes a single layer of anisotropy and that two of the three symmetry axes lie in the horizontal plane. However extensions were made later that relax these assumptions, for example Wolfe and Silver (1998) extend the method to include two layers that do not commute (Silver and Savage, 1994) i.e. the two splitting operators do not possess the commutativity property. These yield apparent splitting parameters, but the actual splitting parameters and useful information can be retrieved (Silver and Savage, 1994). Marson-Pidgeon and Savage (2004) extend this method to allow for known focal mechanisms. Savage (1999) provides a review of some of these extensions. This method also assumes that the movement of particles of two orthogonal components (the particle motion) caused by the incoming isotropic wave is linear and that there is initially no energy on the transverse component (Vecsey et al., 2008). As a result this method is applicable to waves such as SKS (see Section 1.2.3).

Eigenvalue Method

The eigenvalue method (Silver and Chan, 1991) is similar to the transverse minimisation method but instead of minimising the energy on the transverse component, the smaller eigenvalue of the corrected covariance matrix of two orthogonal components is minimised (Silver and Chan, 1991). This is equivalent to creating the most linear particle motion (Long and Silver, 2009). If anisotropy is completely removed the covariance matrix will have one

non-zero eigenvalue corresponding to the speed of the unsplit wave. However, if noise is present there will be two eigenvalues with the second eigenvalue being a measure of how much noise is present in the system. The best set of splitting parameters will produce the smallest second eigenvalue. Like the transverse minimisation method it assumes that the particle motion of the unsplit wave is linear and that the wave is near vertical. Unlike the transverse minimisation method, this method can be used even when the polarisation is unknown (Savage, 1999). This is the method that is investigated in this thesis, and is explained in detail in Chapter 2.

Other Methods

More recent methods have been developed such as the splitting intensity method (Chevrot, 2000) as an alternative for measuring the splitting parameters from SKS and SKKS waves. No assumptions are made regarding geological properties of the medium. However, interpreting the splitting vector requires some assumptions (see Chevrot (2000)). This method requires good coverage from various sources to the receiver (back azimuth coverage). Long and Silver (2009) suggest this as a reason why there are only a few data sets that have been analysed using this method. This method was later improved by applying a Wiener filter beforehand to minimise noise (Montellier and Chevrot, 2010).

Menke and Levin (2003) create a cross convolution method that also does not make geological assumptions regarding the anisotropy. Essentially it convolves the radial and tangential components with an impulse response from a hypothetical model and then attempts to minimise the difference between the observed and the hypothetical waveforms. This method also has the advantage of determining the number of distinct anisotropic layers by using an F test to estimate the significance of the error reduction each time. However, the degrees of freedom are constructed assuming each seismogram data point is independent, which is highly unlikely.

1.3.2 Estimating Errors

Silver and Chan (1991)

In order to calculate errors of the estimates, Silver and Chan (1991) make assumptions about the noise process having a Gaussian distribution. To calculate the degrees of freedom of the noise process, further assumptions are made about the noise process being a convolution of a filter and Gaussian noise. These are strong assumptions which are investigated further in Chapter 2 when the method is described in full with all the assumptions clearly outlined and derivations of the equations given.

Wolfe and Silver (1998)

The Silver and Chan (1991) method works best in stable continental interiors where there is low background noise. On the other hand, splitting measurements from island stations are likely to be of poor quality due to high ambient noise. To overcome this problem the Silver and Chan (1991) method was modified to incorporate stacking of the eigenvalues from a number of events. The method takes several earthquake events and sums (or stacks) the second eigenvalue for each pair of ϕ and δt to create a new error map based on these summed eigenvalues. The best splitting parameters are then estimated by finding the minimum value of the sum of the second eigenvalue. The method works best when stacking varying polarisations, but it will still work stacking like polarisations because noise properties will vary with each event. Whilst they note that their method improves the accuracy of the shear wave splitting parameters, they still assume that the sum of squared errors has a chi-squared distribution. They also note that ideally the wave forms should have a period that is less than 10 times the delay time to achieve good results.

Sandvol and Hearn (1994)

Instead of assuming that the noise process follows a Gaussian distribution, Sandvol and Hearn (1994) opt for a bootstrapping method to calculate the errors. The bootstrapping method has the advantage of not making any assumption about the errors or relying on correctly estimating the degrees of freedom. However, it does assume that the points are independent and identically distributed. To account for correlations in the noise they construct a bootstrap noise sequence in such a way that the statistical properties remain the same each time as the original noise but with different sequencing. Further randomisation is performed to remove consistent waveforms to preserve the statistical properties of the noise process.

Menke and Levin (2003)

The cross convolution method by Menke and Levin (2003) defines the error in a more general form than Silver and Chan (1991). The observed seismograms are compared to a hypothetical anisotropic model containing at least two parameters, ϕ and δt. The number of parameters increases with the complexity of the model. For instance, a model with two anisotropic layers would have four parameters $(\phi_1, \phi_2, \delta t_1, \delta t_2)$. The best set of parameters are the ones that minimise the error i.e. the normalised, squared difference between the observed and the hypothetical time series of the seismograms.

1.3.3 Problems with Identifying Shear Wave Splitting

Shear wave splitting cannot always be identified and the quality of results can vary. This section explains some of the common problems with identifying shear wave splitting and how different methods compare.

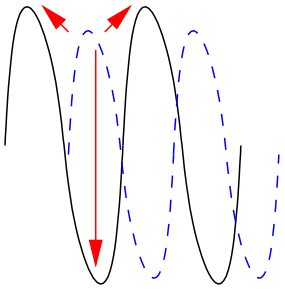


Figure 1.6: Illustration of cycle skipping. Cycle skipping can occur here because the dashed blue peak can in phase with either peaks or the trough of the solid black wave (Matcham, 1997).

Cycle Skipping

Cycle skipping is a phenomenon where two waveforms are mismatched by an integer number of half cycles of the waves' period (Figure 1.6). This results in matching the wrong peak of the wave form, or a peak with a trough, that leads to errors in the shear wave splitting parameters. Mismatching the waveforms is typical for local events when narrow band pass filters are used (Teanby et al., 2004*a*). This tends to be because local events (close to the receiver) have shorter periods so it is easier to skip a cycle. Vecsey et al. (2008) suggested damping the waveforms to a single pulse to avoid cycle skipping but this only appears to work for the cross correlation method and not the eigenvalue method. Cycle skipping can also occur if the amplitude of the first peak of the wave is small (Matcham, 1997).

Null Measurements

Null measurements can provide useful information as outlined below. For example, Sandvol and Hearn (1994) provide a method to detect whether or not splitting has occurred. They re-parametrise the problem by replacing the fast direction parameter ϕ with a polarisation direction that may refer to either the fast direction or the slow direction. The delay time may be positive (meaning that the polarisation direction is the fast direction) or negative (meaning the polarisation direction is the slow direction). This new parametrisation means that if a confidence interval for the delay time δt contains zero then splitting has not occurred.

Wustefeld and Bokelmann (2007) present a method for detecting null measurements and their quality by utilising both the Silver and Chan (1991) method and the rotation correlation method. Their method produces two values of interest: the difference between the fast direction of these two methods and the ratio between the two delay times. These can be used to detect nulls.

Wustefeld and Bokelmann (2007) also note the nulls can be used to detect the strength of anisotropy with strong anisotropy being characterised by nulls from small distinct ranges of back azimuths compared to nulls from all back azimuths for a purely isotropic medium.

Later their method for classifying nulls was revised by Wustefeld et al. (2010) to quantify the quality of nulls ranging from 0 (poor) to 1 (good null). By placing a numerical value on quality the detection of nulls can be automated. The measure of quality also allows one to distinguish between genuine nulls and measurements that appear null because of noise.

Grading Measurements

Not all splitting measurements obtained are reliable and as a result grading systems for their results are needed. This can be done manually by a person, which is quite time consuming for a large set of measurements. Furthermore, the process is subjective making consistent grading difficult to obtain (Aster et al., 1990). Alternatively, several automated grading systems currently exist. Peng and BenZion (2004) assign either a high or low grade based on 10 objective criteria that check the strength of the signal, the stability of a solution with respect to changes in the window that is used to remove the effects of splitting, small confidence regions and significant energy reduction on the perpendicular to the polarisation component. This grading system can also detect null measurements. (Evans et al., 2006) have an automated grading system based on 5 criteria. These criteria place limits on the size of the errors of ϕ and δt plus a limit on the maximum delay time. The other criteria look at the linearity of the particle motion between two orthogonal components before and after the effects of splitting have been removed. More recently, (Savage et al., 2010) created a grading system that combines aspects of both (Peng and BenZion, 2004) and (Evans et al., 2006) with the main difference being that most of the grading is based on the number of measurements and the variation within groups of like parameters called clusters. It also has a wider variety of grades and a criterion to check the number of contours enclosing the 95% confidence region. This provides an idea of the quality of the measurement, with more than 8 contours indicating a high quality measurement.

1.3.4 Comparisons of Techniques

Several authors have looked at comparing these techniques. Vecsey et al. (2008) look at the three most common splitting methods (cross correlation, eigenvalue, and transverse minimisation). The errors were estimated by creating a 95% confidence interval using the method in Silver and Chan (1991) as well as bootstrapping (Sandvol and Hearn, 1994). They found that the transverse minimisation method works best, but it can only be used when the incoming polarisation is known (i.e SKS, SKKS waves). Although if the geometric representation of the fault slip during an earthquake (i.e. the focal mechanism) is known then S waves can have a known polarisation. For a more general waveform we then have to use either the eigenvalue method or rotation correlation. They found that the eigenvalue method is more prone to cycle skipping but less susceptible to noise and thus can provide more accurate splitting parameters compared to cross correlation.

Wustefeld and Bokelmann (2007) compared the Silver and Chan (1991) method to the rotation correlation method. They find that in null directions the Silver and Chan (1991) method gives an approximate solution but the rotation correlation method has fast direction devia-

tions of around 45°. This result is used to create a null detection method (see Section 1.3.3). Synthetic tests using the three main methods in Splitlab (Wustefeld et al., 2008) show similar results. They also found that the results for the transverse minimisation and eigenvalue minimisation were similar.

Montellier and Chevrot (2010) used SKS and SKKS wave to compare Silver and Chan (1991) and Wolfe and Silver (1998) to their splitting intensity method. They find that Silver and Chan (1991) works best on a high quality records with a strong signal on the transverse component. This significantly limits the number of records available to analyse. In contrast, stacking the error maps (Wolfe and Silver, 1998) considerably improves the reliability of splitting parameters and agrees with results from their splitting intensity method.

Menke and Levin (2003) test their convolution method against cross correlation, transverse minimisation and eigenvalue methods. They find it performs similarly when there is a single anisotropic layer and performs better when there are two anisotropic layers and a two layer cross convolution is performed. This is probably because the other models make assumptions about there being only one layer of anisotropy.

1.4 Contribution of Thesis

As mentioned before, there is no in depth literature assessing the robustness of the Silver and Chan (1991) method itself. The closest assessments are the comparisons of methods mentioned in Section 1.3.4. Here we provide an in-depth, step by step break down of the Silver and Chan (1991) method. This is useful because the original Silver and Chan (1991) article only covers a few key formulas and most proofs are omitted. By doing this we are also able to identify some problems with the theory behind the Silver and Chan (1991) method as well as bugs in some of the original computer codes implementing the method, some of which have been carried over to other software using this method.

In this project we produce an automated grading method to assess the quality of the results coming from the Silver and Chan (1991) method. Unlike other automated grading methods (e.g. Peng and BenZion (2004)), which criteria to use and how to determine grades was decided statistically using the AIC criterion and multiple regression.

Many assumptions are made in order to perform the Silver and Chan (1991) method, but we know of no articles that carry out statistical tests to check these assumptions. We test some of the key assumptions in the Silver and Chan (1991) method and see what influence they have on results.

1.5 Outline

The structure of this thesis is as follows. In chapter 2 the Silver and Chan (1991) method is explained in great detail along with problems with the theory and bugs we have identified in the code. Chapter 3 discusses in detail a dataset we have used for testing, and explains how manual grades were determined. It then describes how this grading process is automated. Chapter 4 discusses cycle skipping in more detail and whether the dataset shows signs of it. In Chapter 5 we test some of the assumptions in the Silver and Chan (1991) method and investigate how they affect the method. Chapter 6 contains some concluding statements and future directions. There is also a glossary and a list of symbols. The appendix contains some useful mathematical results.

Chapter 2

Silver and Chan (1991) Method - A Full Description

The Silver and Chan (1991) method estimates the splitting parameters ϕ and δt along with their corresponding (standard) errors. These errors give an indication of how accurate the estimates of ϕ and δt are. In order to understand how this is done, this chapter carefully explains the process outlined in their article, as well as giving additional steps and explanations that the article omits. In order to obtain the standard errors of the parameter estimates, several assumptions are made. Each time an assumption is made it is clearly highlighted in a box. Differences and potential points where the method may break down are given at the end. This chapter is completely self contained and it is not necessary to have read the article in advance.

2.1 The Shear Wave

In order to carry out the Silver and Chan (1991) method we require observations of shear waves as inputs. These are usually obtained from a seismic station during an earthquake, which is sometimes referred to as an event. One way of expressing a shear wave displacement mathematically is by using the following one dimensional expression

$$u(x,t) = A\exp[i(kx-\omega t)] = A(\omega,k)\exp[-i(\omega t-kx)] \tag{2.1}$$

where

$$\begin{aligned} u(x,t) &= \text{wave at time } t \text{ and position } x \\ A &= \text{amplitude of the wave} \\ k &= \text{wavenumber=} 2\pi/\lambda \text{ where } \lambda \text{ is the wave length} \\ x &= \text{point along the x axis} \\ \omega &= \text{angular frequency=} 2\pi f \text{ where } f \text{ is the frequency} \\ t &= \text{point in time} \\ i &= \text{imaginary number, } i^2 = -1 \end{aligned}$$

The wave number measures how quickly the wave changes over space at a particular point in time. On the other hand, the angular frequency measures how quickly the wave changes over time at a particular point in space. Euler's formula shows that $\exp(ix) = \cos(x) + i\sin(x)$ so Equation 2.1 can be rewritten using trigonometric functions as follows

$$u(x,t) = A(\omega,k)\left[\cos(kx - \omega t) + i\sin(kx - \omega t)\right] \tag{2.2}$$

Silver and Chan (1991) simplify the shear wave shown in Equation 2.1 by looking at a shear wave in an isotropic medium that has travelled a distance of length L and arrives at the surface of the Earth at time T_0. As a result they define a vector that represents the shear wave as follows

$$\mathbf{u}(\omega, T_0) = \mathbf{A}(\omega)\exp(-i\omega T_0) \tag{2.3}$$

(SC 1)[1]

where

$$\begin{aligned}
\omega &= \text{angular frequency} \\
\mathbf{u}(\omega) &= \text{isotropic shear wave vector} \\
\mathbf{A}(\omega) &= w(\omega)\hat{\mathbf{A}}(\omega) \text{ complex amplitude vector} \\
w(\omega) &= \|\mathbf{A}(\omega)\| \text{ norm of the amplitude with angular frequency } \omega \\
T_0 &= \text{time at which the wave arrives at the surface}
\end{aligned}$$

The shear wave travels in a particular direction and is oriented a particular way. In order to represent this, a right handed set of orthogonal vectors $[\hat{\mathbf{p}}\hat{\mathbf{a}}\hat{\mathbf{b}}]$ are defined and given the name, propagation coordinates ($\mathcal{P}$). The three orthogonal vectors that define this coordinate system are

$$\begin{aligned}
\hat{\mathbf{b}} &= \text{propagation vector} \\
\hat{\mathbf{p}} &= \text{polarisation vector pointing in the direction of shear wave displacement} \\
\hat{\mathbf{a}} &= \text{null vector } \hat{\mathbf{b}} \times \hat{\mathbf{p}} = \hat{\mathbf{a}}
\end{aligned}$$

The vector $\hat{\mathbf{p}}$, which is normal to $\hat{\mathbf{b}}$ because it is an S wave, represents the orientation, while the vector $\hat{\mathbf{b}}$ represents the direction of travel. To make a coordinate system, a third vector is created by taking the cross product of the existing two. Silver and Chan (1991) take into account the polarisation but do so in a simplified manner. They assume that waves (for all frequencies, ω) displace in the same direction $\hat{\mathbf{p}}$. Consequently, the complex amplitude vector $\mathbf{A}(\omega) = w(\omega)\hat{\mathbf{A}}(\omega)$ where $w(\omega) = \|\mathbf{A}(\omega)\|$ can be approximated by replacing $\hat{\mathbf{A}}(\omega)$ with $\hat{\mathbf{p}}$ for all ω. Taking this into account, $\hat{\mathbf{p}}$ is substituted in place of $\hat{\mathbf{A}}(\omega)$ and Silver and Chan (1991) express Equation 2.3 alternatively as

$$\mathbf{u}(\omega, T_0) = w(\omega)\exp(-i\omega T_0)\hat{\mathbf{p}} \tag{2.4}$$

(SC 2)

with $\mathbf{A}(\omega)$ being replaced by $w(\omega)\hat{\mathbf{p}}$. Figure 2.1 illustrates a wave, $\mathbf{u}(\omega)$, travelling through an isotropic medium.

[1] SC numbering denotes equation numbering in the original Silver and Chan (1991) article

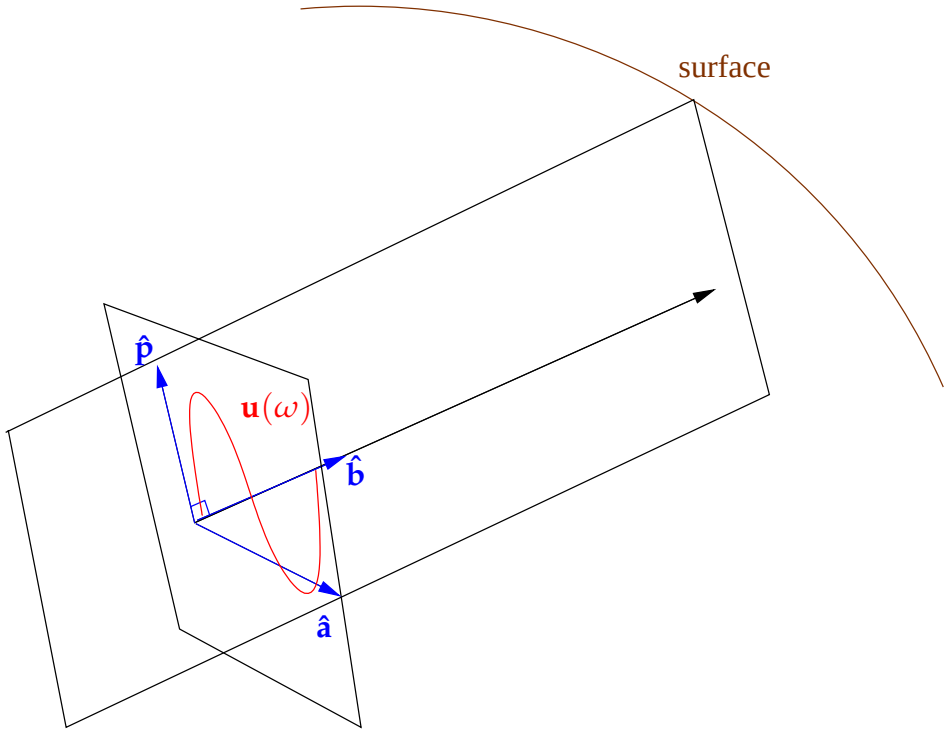


Figure 2.1: An unsplit shear wave $\mathbf{u}(\omega)$ travelling towards the surface shown in propagation coordinates $(\mathcal{P})$. The wave is oriented along $\hat{\mathbf{p}}$ and travelling along $\hat{\mathbf{b}}$.

The above case describes a shear wave in an isotropic medium. For an anisotropic medium (one with preferential orientation) we can define another coordinate system that we call the splitting coordinate system $(\mathcal{S})$. We define two new directions with respect to a propagation direction $\hat{\mathbf{b}}$. One of them is the fast direction $\hat{\mathbf{f}}$ in which a wave will travel faster and the other is the slow direction $\hat{\mathbf{s}}$ in which the wave travels more slowly. $\hat{\mathbf{f}}$ and $\hat{\mathbf{s}}$ are orthogonal to each other. The three orthogonal vectors $\left[\hat{\mathbf{f}}\hat{\mathbf{s}}\hat{\mathbf{b}}\right]$ form the splitting coordinate system $(\mathcal{S})$. We describe $\mathcal{P}$, $\mathcal{S}$ and other coordinate systems in more detail in Section 2.5.2. A shear wave will split in two when it passes into an anisotropic medium and travel at different speeds with displacements in these directions. To model the splitting caused by anisotropy on the shear wave $\mathbf{u}(\omega)$ orientated along $\hat{\mathbf{p}}$, we project $\hat{\mathbf{p}}$ onto these two directions and then by the time the wave emerges at the surface, i.e. time T_0, the faster wave arrives at a time δt before the slower wave. If T_0 is the time an unsplit wave would have arrived, then the fast wave arrives at time $T_0 - \delta t/2$ and the slow wave at time $T_0 + \delta t/2$ (Figure 2.2).

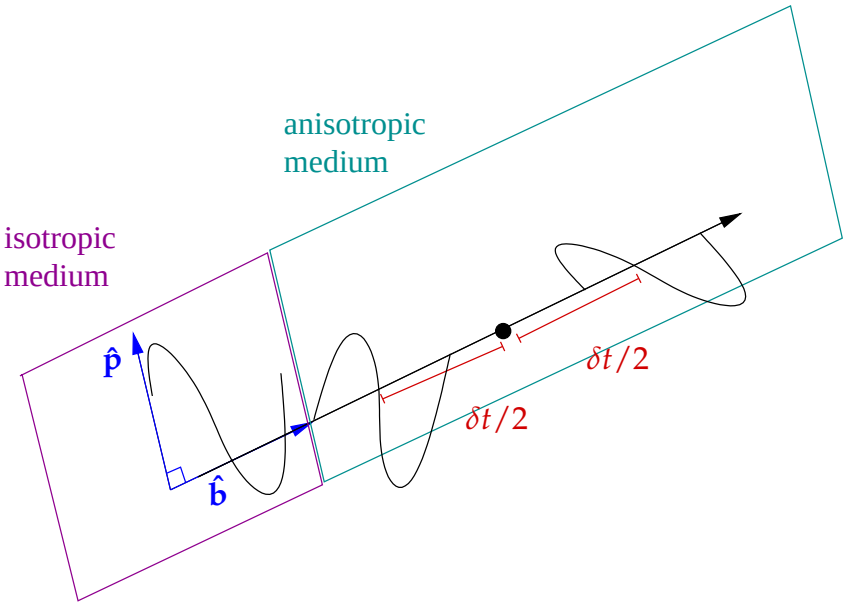


Figure 2.2: An initial shear wave split in two, modelled in space. The faster wave moves ahead and the slower wave lags behind. The dot represents the half way point between the two peaks and is where the wave would have been centred if it had not been split.

2.2 Geophysical Properties

Elastic properties described below lead to fast and slow directions $\hat{\mathbf{f}}$ and $\hat{\mathbf{s}}$ being eigenvectors of a polarisation tensor $\mathbf{V}$ (also known as the Christoffel matrix). In order to define the polarisation tensor, the concept of elastic moduli must first be explained. Elastic materials under the ground are under stress and as a result they deform (strain). Hooke's law relates stress to strain in a linearly elastic material as follows

$$\sigma_{ij} = c_{ijk\ell}\epsilon_{k\ell} \tag{2.5}$$

where

$$\begin{aligned} \sigma_{ij} &= \text{stress tensor} \\ c_{ijk\ell} &= \text{elastic moduli} \\ \epsilon_{k\ell} &= \text{strain tensor} \end{aligned}$$

The elastic moduli reflect how displacements in both space and time arise as a result of stress so anisotropy can be introduced by this tensor (see Stein and Wysession (2003) for further information).

With this in mind, we can define the polarisation tensor. Let

$$\begin{aligned} \rho &= \text{density} \\ V_{i\ell} &= \text{polarisation tensor} \\ c_{ijk\ell} &= \text{elastic coefficients} \\ b_j &= j\text{th component of the propagation vector } \hat{\mathbf{b}} \\ b_k &= \text{k}\text{th component of the propagation vector } \hat{\mathbf{b}} \end{aligned}$$

then $\mathbf{V}$ may be defined by the following relationship

$$\rho V_{i\ell} \equiv \sum_{j=1}^{3}\sum_{k=1}^{3} c_{ijk\ell} b_j b_k \tag{2.6}$$

Silver and Chan (1991) use the summation convention, i.e. when an index appears on the same side of the equation twice it is implied that you sum over that index. This leads to a more compact version of Equation 2.6

$$\rho V_{i\ell} \equiv c_{ijk\ell} b_j b_k \tag{2.7}$$

(SC 3)

Using the symmetries $c_{ijk\ell} = c_{jik\ell} = c_{ij\ell k}$ one can derive Equation 2.7. Babuska and Cara (1991) give a full proof of this equation starting with a generalised form of Hooke's law to obtain Equation 2.7. However, what's of more interest is the interpretation of this polarisation matrix. This matrix forms part of an eigenvalue problem and solving it gives $\hat{\mathbf{f}}$ and $\hat{\mathbf{s}}$ as eigenvectors corresponding to the direction of particle displacement for the split wave and their corresponding eigenvalues are shear velocities in each direction.

The second splitting parameter, the delay time (δt), may be expressed in terms of these shear velocities if the amount of anisotropy is relatively small. The easiest way to understand this is to think of the formula that links distance, time and velocity together. i.e. $distance = velocity \times time$. Assuming that the waves enter the anisotropic medium at time $t = 0$, both the fast wave and the slow wave will eventually reach the surface so they travel the same distance therefore

$$\begin{aligned} v_1 t_1 &= v_2 t_2 && \text{or} \\ \beta_1 t_1 &= \beta_2 t_2 && \text{using the notation in Silver and Chan (1991)} \end{aligned} \tag{2.8}$$

Let

$$\begin{aligned} L &= \text{the length of the path the wave has travelled} \\ \beta_0 &= \text{shear wave velocity in an isotropic medium} \end{aligned}$$

The waves initially travel at β_0 in an isotropic medium. At time $t = 0$ the wave enters the anisotropic medium and the wave is split. When the wave splits, the faster wave (with $\delta\beta_1 > 0$) increases speed to travel at $\beta_1 = \beta_0 + \delta\beta_1$ whilst the slower wave (with $\delta\beta_2 < 0$) decreases speed to $\beta_2 = \beta_0 + \delta\beta_2$. Rearranging and adding this to Equation 2.8 yields

$$\begin{aligned} \beta_1 t_1 &= \beta_2 t_2 \\ (\beta_0 + \delta\beta_1) t_1 &= (\beta_0 + \delta\beta_2) t_2 \\ \beta_0 t_1 + \delta\beta_1 t_1 &= \beta_0 t_2 + \delta\beta_2 t_2 \\ \beta_0 (t_1 - t_2) &= \delta\beta_2 t_2 - \delta\beta_1 t_1 \\ -\beta_0 \delta t &= \delta\beta_2 t_2 - \delta\beta_1 t_1 && \text{since } \delta t = t_2 - t_1 > 0 \\ &\approx (\delta\beta_2 - \delta\beta_1) t && \text{for small } \delta t = (t_2 - t_1) \\ &= (\delta\beta_2 - \delta\beta_1) \frac{L}{\beta_0} && t = distance/velocity \approx (t_2 + t_1)/2 \\ \Rightarrow \delta t &\approx \left(\frac{\delta\beta_1 - \delta\beta_2}{\beta_0} \right) \frac{L}{\beta_0} \end{aligned} \tag{2.9}$$

If the relative velocity $\delta\zeta$ is defined as

$$\delta\zeta = \frac{\delta\beta_1 - \delta\beta_2}{\beta_0} = \frac{\delta\beta_1 + |\delta\beta_2|}{\beta_0} \tag{2.10}$$

then for small anisotropy and consequently small delay times, δt may be defined as

$$\delta t = \frac{\delta\zeta L}{\beta_0} \tag{2.11}$$

(c.f. SC 4)

Silver and Chan (1991) use the term $\delta\hat{\beta}$ instead of $\delta\zeta$ but this might give the impression that this term has the same units as β when in fact it is dimensionless, which is why we chose to label it $\delta\zeta$.

2.3 Split Wave and the Splitting Operator

So far we have looked at a shear wave and a few of its properties, but we are yet to discuss shear wave splitting in detail. The splitting operator ($\mathbf{\Gamma}(\phi, \delta t)$) takes an unsplit wave, then

rotates and time shifts it. The splitting operator is a function of the splitting parameters and Silver and Chan (1991) define it as

$$\mathbf{\Gamma}(\phi, \delta t) \equiv \exp(i\omega\delta t/2)\hat{\mathbf{f}}\hat{\mathbf{f}} + \exp(-i\omega\delta t/2)\hat{\mathbf{s}}\hat{\mathbf{s}} \tag{2.12}$$

(SC 5)

There are two points here to make. Firstly both $\hat{\mathbf{f}}$ and $\hat{\mathbf{s}}$ are column vectors so to multiply such vectors the second one should be transposed. Secondly and more importantly, it appears that the signs on the exponentials are the wrong way round. To justify this we define the split shear wave $\mathbf{u_s}(\omega, t)$ to have two components: a component for the faster travelling wave ($u_1(\omega, t)$) and a component for the slower travelling wave ($u_2(\omega, t)$). Figure 2.3 illustrates this for a one dimensional situation. If we ignore the spatial components, then the

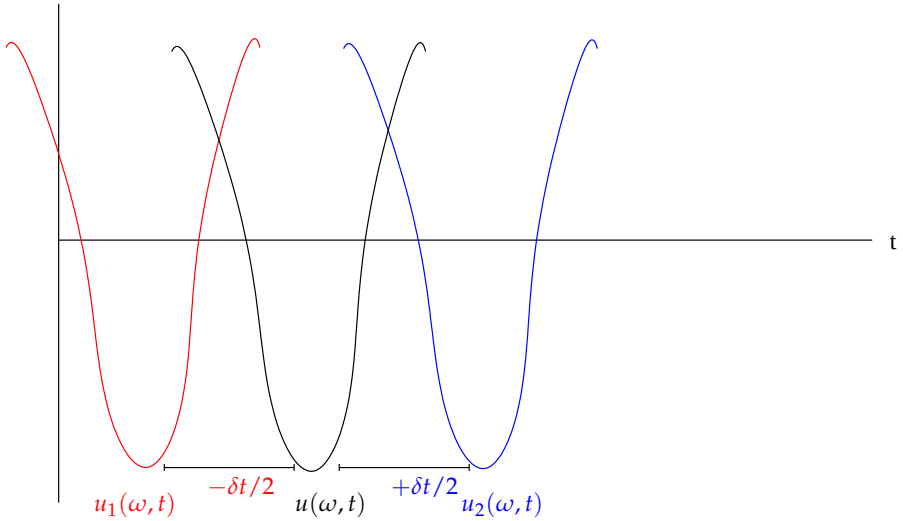


Figure 2.3: An initial shear wave is split in two with the faster (red) $u_1(\omega, t)$ wave being moved back $\delta t/2$ in time (i.e. sped up) and the slower (blue) $u_2(\omega, t)$ one moved forward in time by $\delta t/2$.

initial wave has a real component $Re[u(\omega, t)] = Re(A(\omega)e^{-i\omega t}) = A(\omega)\cos(\omega t)$. The faster wave will arrive $\delta t/2$ seconds earlier than an unsplit wave and the slower wave will arrive $\delta t/2$ seconds later compared to an unsplit wave. In other words,

$$\begin{aligned} Re[u_1(\omega, t)] &= Re[u(\omega, t)\exp(-i\omega\delta t/2)] = Re[A(\omega)\exp(-i\omega t - i\omega\delta t/2)] = A(\omega)\cos[\omega(t + \delta t/2)] \\ Re[u_2(\omega, t)] &= Re[u(\omega, t)\exp(+i\omega\delta t/2)] = Re[A(\omega)\exp(-i\omega t + i\omega\delta t/2)] = A(\omega)\cos[\omega(t - \delta t/2)] \end{aligned} \tag{2.13}$$

An alternative way to view this is to consider a wave, in a one dimensional situation, propagating in the direction x with a wave number k

$$u(x, t) = \cos(\omega t - kx)$$

Then consider two more waves propagating with a slightly higher constant $k + dk/2$ (shorter wave length) and one with a slightly lower constant $k - dk/2$ (longer wave length). The speed of the wave is given by ω/k so that the wave with higher constant $k + dk/2$ represents a slower wave while $k - dk/2$ represents a faster wave

$$\begin{aligned} u_f(x, t) &= \cos(\omega t - (k - dk/2)x) \\ u_s(x, t) &= \cos(\omega t - (k + dk/2)x) \end{aligned}$$

After travelling a distance L (as in Silver and Chan (1991)), the waveforms have the following form

$$u_f(L,t) = \cos(\omega t - (k - dk/2)L) = \cos(\omega t - (kL - Ldk/2)) = \cos(\omega t - kL + Ldk/2)$$
$$u_s(L,t) = \cos(\omega t - (k + dk/2)L) = \cos(\omega t - (kL + Ldk/2)) = \cos(\omega t - kL - Ldk/2)$$

There is a phase shift between the two waves. If we let $dt = Ldk/\omega$ we can rewrite the equations as

$$u_f(L,t) = \cos(\omega t - kL + \omega dt/2) = \cos(\omega t + \omega dt/2 - kL) = \cos(\omega(t + dt/2) - kL)$$
$$u_s(L,t) = \cos(\omega t - kL - \omega dt/2) = \cos(\omega t - \omega dt/2 - kL) = \cos(\omega(t - dt/2) - kL)$$

The two waves are identical except for a phase shift of dt. The fast wave has a shift of $(t + dt/2)$ while the slower wave has a time shift of $(t - dt/2)$. The sign of the $dt/2$ shifts agree with Equation 2.13. Silver and Chan (1991) omit the kL term because they assume the wave has travelled a distance L. Note that the same result occurs for a wave defined as $u(x,t) = \exp(i(kx - \omega t))$. The important thing to note is that the wave speed ω/k is the inversely proportional to k. Both sets of reasoning suggest that the signs on the Silver and Chan (1991) splitting operator (Equation 2.12, *(SC 5)*) are incorrect.

Silver and Chan (1991) only consider almost vertically travelling waves and thus only consider the 2D horizontal plane. We wish to give the problem a more general treatment and so extend the splitting operator to have 3 components or dimensions. Therefore, we define the splitting operator as

$$\mathbf{\Gamma}(\phi, \delta t) \equiv \exp(-i\omega\delta t/2)\hat{\mathbf{f}}\hat{\mathbf{f}}^T + \exp(i\omega\delta t/2)\hat{\mathbf{s}}\hat{\mathbf{s}}^T + \hat{\mathbf{b}}\hat{\mathbf{b}}^T \tag{2.14}$$

(c.f. SC 5)

We can define the time difference tensor as

$$\delta\mathbf{T} = \delta t/2(\hat{\mathbf{f}}\hat{\mathbf{f}}^T - \hat{\mathbf{s}}\hat{\mathbf{s}}^T) \tag{2.15}$$

(c.f. SC 7)

so that $\mathbf{\Gamma}(\phi, \delta t)$ can be written in a more compact form. Since Silver and Chan (1991) define ϕ as the angle between $\hat{\mathbf{f}}$ and $\hat{\mathbf{p}}$ (Figure 2.4), it follows that $\mathbf{\Gamma}(\phi, \delta t)$ and therefore $\delta\mathbf{T}$ are both functions of ϕ. Evaluating the dot product gives $\hat{\mathbf{f}} \cdot \hat{\mathbf{p}} = \|\hat{\mathbf{f}}\|\|\hat{\mathbf{p}}\| \cos(\phi) = \cos(\phi)$ and $\hat{\mathbf{s}} \cdot \hat{\mathbf{p}} = \|\hat{\mathbf{s}}\|\|\hat{\mathbf{p}}\| \cos(\pi/2 + \phi) = -\sin(\phi)$. This leads us to define the compact splitting operator

$$\mathbf{\Gamma}(\phi, \delta t) = \exp[-i\omega\delta\mathbf{T}(\phi, \delta t)] \tag{2.16}$$

(c.f. SC 8)

This formula and Equation 2.15 are different to the ones in Silver and Chan (1991). This follows as a result of changes in how we define the splitting operator. To show Equations

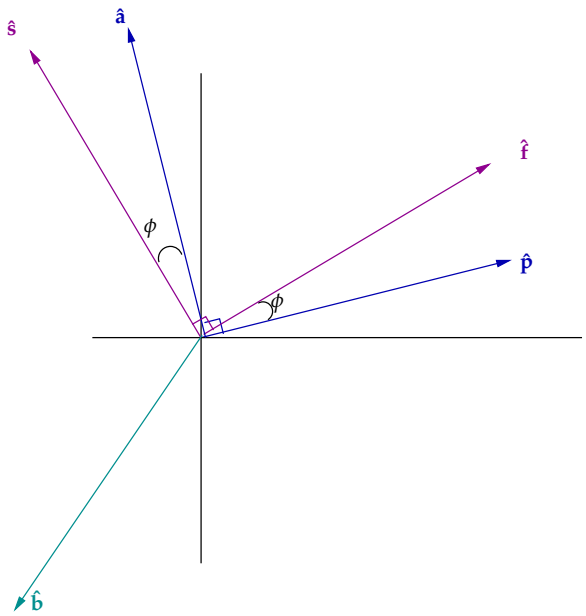


Figure 2.4: Diagram showing the relationship between the orthogonal unit vectors $\hat{\mathbf{f}}$, $\hat{\mathbf{s}}$ and $\hat{\mathbf{b}}$ that form the splitting coordinates ($\mathcal{S}$) and the vectors $[\hat{\mathbf{p}}\hat{\mathbf{a}}\hat{\mathbf{b}}]$ that define the propagation coordinates ($\mathcal{P}$).

2.16 and 2.14 are equivalent we rewrite $\mathbf{\Gamma}(\phi, \delta t)$ as a Taylor series

$$\begin{aligned}
\mathbf{\Gamma}(\phi, \delta t) &= \exp[-i\omega\delta\mathbf{T}(\phi, \delta t)] \\
&= \sum_{k=0}^{\infty} \frac{(-i\omega\delta\mathbf{T})^k}{k!} \\
&= \sum_{k=0}^{\infty} \frac{(-i\omega)^k}{k!}(\delta\mathbf{T})^k \\
&= (\delta\mathbf{T})^0 + \sum_{k=1}^{\infty} \frac{(-i\omega)^k}{k!}(\delta\mathbf{T})^k
\end{aligned} \tag{2.17}$$

$\delta\mathbf{T}$ has been isolated because of its recursive form

$$(\delta\mathbf{T})^k = (\delta t/2)^k(\hat{\mathbf{f}}\hat{\mathbf{f}}^T + (-1)^k\hat{\mathbf{s}}\hat{\mathbf{s}}^T) \qquad \text{(for } \forall \text{ integers } k : k \geq 1) \tag{2.18}$$

We will call Equation 2.18 $P(k)$ and show it is true for all $k \geq 1$ by induction.

Proof. Base case: $k = 1$

$$(\delta\mathbf{T})^1 = (\delta t/2)^1(\hat{\mathbf{f}}\hat{\mathbf{f}}^T + (-1)^1\hat{\mathbf{s}}\hat{\mathbf{s}}^T) = (\delta t/2)(\hat{\mathbf{f}}\hat{\mathbf{f}}^T - \hat{\mathbf{s}}\hat{\mathbf{s}}^T) \tag{2.19}$$

Appendices A.1 & A.2 explain why this is the case.

Inductive step: Assume $P(n)$, for some $n \geq 1$ is true i.e.

$$(\delta\mathbf{T})^n = (\delta t/2)^n(\hat{\mathbf{f}}\hat{\mathbf{f}}^T + (-1)^n\hat{\mathbf{s}}\hat{\mathbf{s}}^T) \tag{2.20}$$

and prove $P(n+1)$ is true

$$\begin{aligned}
(\delta\mathbf{T})^n &= (\delta t/2)^n(\hat{\mathbf{f}}\hat{\mathbf{f}}^T + (-1)^n\hat{\mathbf{s}}\hat{\mathbf{s}}^T) && \text{(inductive hypothesis)}\\
(\delta\mathbf{T})^n(\delta\mathbf{T}) &= \left[(\delta t/2)^n(\hat{\mathbf{f}}\hat{\mathbf{f}}^T + (-1)^n\hat{\mathbf{s}}\hat{\mathbf{s}}^T)\right]\left[(\delta t/2)(\hat{\mathbf{f}}\hat{\mathbf{f}}^T - \hat{\mathbf{s}}\hat{\mathbf{s}}^T)\right]\\
(\delta\mathbf{T})^{n+1} &= (\delta t/2)^{n+1}(\hat{\mathbf{f}}\hat{\mathbf{f}}^T + (-1)^n\hat{\mathbf{s}}\hat{\mathbf{s}}^T)(\hat{\mathbf{f}}\hat{\mathbf{f}}^T - \hat{\mathbf{s}}\hat{\mathbf{s}}^T) && \text{(then expand)}\\
&= (\delta t/2)^{n+1}(\hat{\mathbf{f}}\hat{\mathbf{f}}^T\hat{\mathbf{f}}\hat{\mathbf{f}}^T + (-1)^{n+1}\hat{\mathbf{s}}\hat{\mathbf{s}}^T\hat{\mathbf{s}}\hat{\mathbf{s}}^T) && \text{(since } \hat{\mathbf{f}} \text{ \& } \hat{\mathbf{s}} \text{ are } \perp)\\
&= (\delta t/2)^{n+1}(\hat{\mathbf{f}}\hat{\mathbf{f}}^T + (-1)^{n+1}\hat{\mathbf{s}}\hat{\mathbf{s}}^T) && \text{(inner product of unit vectors is 1)} \quad (2.21)
\end{aligned}$$

This is the same form as $P(n)$ so $P(n+1)$ is true. Therefore $P(k)$ is true for all integers $k : k \geq 1$ □

Substituting Equation 2.18 into Equation 2.17, using Appendix A.2 and splitting into fast and slow components gives,

$$\mathbf{\Gamma}(\phi,\delta t) = \underbrace{(\hat{\mathbf{b}}\hat{\mathbf{b}}^T + \hat{\mathbf{f}}\hat{\mathbf{f}}^T + \hat{\mathbf{s}}\hat{\mathbf{s}}^T)}_{(\delta\mathbf{T})^0 = \mathbf{I} \text{ (Appendix A.2)}} + \sum_{k=1}^{\infty}\frac{(-i\omega)^k}{k!}(\delta t/2)^k\hat{\mathbf{f}}\hat{\mathbf{f}}^T + \sum_{k=1}^{\infty}\frac{(-i\omega)^k}{k!}(\delta t/2)^k + (-1)^k\hat{\mathbf{s}}\hat{\mathbf{s}}^T \quad (2.22)$$

The two sums in this expression are almost the Taylor series form of the exponential but the first term is missing in each series. If we rewrite this equation as an exponential we must subtract off the first term from each of the sums to obtain

$$\mathbf{\Gamma}(\phi,\delta t) = (\hat{\mathbf{f}}\hat{\mathbf{f}}^T + \hat{\mathbf{s}}\hat{\mathbf{s}}^T) - \hat{\mathbf{f}}\hat{\mathbf{f}}^T - \hat{\mathbf{s}}\hat{\mathbf{s}}^T + \exp(-i\omega\delta t/2)\hat{\mathbf{f}}\hat{\mathbf{f}}^T + \exp(i\omega\delta t/2)\hat{\mathbf{s}}\hat{\mathbf{s}}^T + \hat{\mathbf{b}}\hat{\mathbf{b}}^T \quad (2.23)$$

Simplifying gives

$$\mathbf{\Gamma}(\phi,\delta t) = \exp(-i\omega\delta t/2)\hat{\mathbf{f}}\hat{\mathbf{f}}^T + \exp(i\omega\delta t/2)\hat{\mathbf{s}}\hat{\mathbf{s}}^T + \hat{\mathbf{b}}\hat{\mathbf{b}}^T \quad (2.24)$$

which is Equation 2.14. Therefore Equations 2.14 and 2.16 are equivalent.

The splitting operator enables the faster wave to move ahead by $-\delta t/2$ (i.e. because it is travelling faster we see the wave sooner in time) and moves the slower wave back by $+\delta t/2$ (i.e. the wave is retarded so we see the wave later in time). If we apply the splitting operator to a shear wave (Equation 2.4) we obtain a split wave.

$$\mathbf{u}_s(\omega,t) = \mathbf{\Gamma}(\phi,\delta t)\mathbf{u}(\omega,t) \quad (2.25)$$

2.4 Inverse Splitting Operator

The previous section shows how to split a shear wave. Here we look at how to desplit a wave once it has been split. The inverse splitting operator ($\mathbf{\Gamma}^{-1}(\phi',\delta t')$) for some chosen $(\phi',\delta t')$ is applied to a split shear wave $\mathbf{u}_s(\omega,t)$ to obtain a corrected shear wave $\tilde{\mathbf{u}}(\omega,t)$

$$\begin{aligned}
\tilde{\mathbf{u}}(\omega,t) &= \mathbf{\Gamma}^{-1}(\phi',\delta t')\mathbf{u}_s(\omega,t)\\
&= \mathbf{\Gamma}^{-1}(\phi',\delta t')\mathbf{\Gamma}(\phi,\delta t)\mathbf{u}(\omega,t)
\end{aligned} \quad (2.26)$$

If the seismograms contain no noise, which only occurs in simulated noise free seismograms, then $\phi' = \phi$ and $\delta t' = \delta t$, which implies $\tilde{\mathbf{u}}(\omega, t) = \mathbf{u}(\omega, t)$. In other words we have fully corrected for splitting. Equation 2.26 shows that we wish to find a pair of ϕ and δt so that the inverse splitting operator, $\mathbf{\Gamma}^{-1}$, best removes the effects of splitting. The complex conjugate transpose of $\mathbf{\Gamma}$ is

$$\begin{aligned}
\mathbf{\Gamma}^{\dagger} &= [\exp(-i\omega\delta\mathbf{T})]^{\dagger} \\
&= [\exp(i\omega\delta\mathbf{T})]^{T} \\
&= \left(\exp[i\omega\left(\delta t/2(\hat{\mathbf{f}}\hat{\mathbf{f}}^{T} - \hat{\mathbf{s}}\hat{\mathbf{s}}^{T}) + \hat{\mathbf{b}}\hat{\mathbf{b}}^{T}\right)]\right)^{T} && \text{(substituting in Equation 2.15)} \\
&= \sum_{k=0}^{\infty}\left[\frac{(i\omega)^{k}\left[\delta t/2(\hat{\mathbf{f}}\hat{\mathbf{f}}^{T} - \hat{\mathbf{s}}\hat{\mathbf{s}}^{T})^{k}) + \hat{\mathbf{b}}\hat{\mathbf{b}}^{T})\right]^{k})}{k!}\right]^{T} && \text{(Taylor series)} \\
&= \sum_{k=0}^{\infty}\left[\frac{(i\omega\delta t/2)^{k}}{k!}\right][(\hat{\mathbf{f}}\hat{\mathbf{f}}^{T} - \hat{\mathbf{s}}\hat{\mathbf{s}}^{T})^{T}]^{k} && \text{(since } (\mathbf{A}^{k})^{T} = (\mathbf{A}^{T})^{k}) \\
&= \sum_{k=0}^{\infty}\left[\frac{(i\omega\delta t/2)^{k}}{k!}\right][(\hat{\mathbf{f}}\hat{\mathbf{f}}^{T} - \hat{\mathbf{s}}\hat{\mathbf{s}}^{T})]^{k} && \text{(see Equation 2.18)} \\
&= \exp(i\omega\delta\mathbf{T}) && (2.27)
\end{aligned}$$

Note that since $(\delta\mathbf{T}) = (\delta t/2)(\hat{\mathbf{f}}\hat{\mathbf{f}}^{T} - \hat{\mathbf{s}}\hat{\mathbf{s}}^{T})$ is symmetric, it follows that $\mathbf{\Gamma}^{T} = \mathbf{\Gamma}$ so that $\mathbf{\Gamma}^{\dagger} = \mathbf{\Gamma}^{*}$ where $\mathbf{\Gamma}^{*}$ is the complex conjugate of $\mathbf{\Gamma}$. Multiplying the inverse splitting operator by the splitting operator gives

$$\begin{aligned}
\mathbf{\Gamma}^{\dagger}\mathbf{\Gamma} &= \exp[i\omega\delta\mathbf{T}(\phi, \delta t)]\exp[-i\omega\delta\mathbf{T}(\phi, \delta t)] \\
&= \exp[i\omega\delta\mathbf{T}(\phi, \delta t) - i\omega\delta\mathbf{T}(\phi, \delta t)] \\
&= \exp(\mathbf{0}) = \mathbf{I} \qquad \text{(Appendix A.1)} \qquad (2.28)
\end{aligned}$$

Therefore $\mathbf{\Gamma}^{-1} = \mathbf{\Gamma}^{\dagger} = \mathbf{\Gamma}^{*}$ and hence $\mathbf{\Gamma}$ is said to be unitary (i.e. $\mathbf{\Gamma}^{\dagger}\mathbf{\Gamma} = \mathbf{I}$).

2.5 Coordinate Systems and Rotations

2.5.1 Overview

So far we have seen two coordinate systems, the propagation coordinates ($\mathcal{P}$) defined by the vectors $[\hat{\mathbf{p}}\hat{\mathbf{a}}\hat{\mathbf{b}}]$ and the splitting coordinates ($\mathcal{S}$) defined by the vectors $[\hat{\mathbf{f}}\hat{\mathbf{s}}\hat{\mathbf{b}}]$. However, there are other coordinate systems used to model the shear wave splitting process. For each coordinate system we describe

- the vectors that define the coordinate system
- the transformation angle that rotates you into the current system from a previously described coordinate system

- diagram illustrating the vectors and the transformation angle
- equations and matrices describing the rotation

Different methods or computer programs may use different coordinate systems so this section should be useful for comparing waveforms and results expressed in another coordinate system. We then show how to express a shear wave in some of the coordinate systems. All coordinate systems are right handed.

Transforming between coordinate systems requires a rotation matrix (or a series of rotation matrices). We define three rotation matrices at angle θ anticlockwise around the x, y and z axes

$$\mathbf{R}_1(\theta) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{bmatrix} \tag{2.29}$$

$$\mathbf{R}_2(\theta) = \begin{bmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{bmatrix} \tag{2.30}$$

$$\mathbf{R}_3(\theta) = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{2.31}$$

where θ denotes an angle that rotates you from the current coordinate system to another and the subscripts 1,2 and 3 indicate this is a rotation around the x, y and z axis respectively. A rotation matrix has the following properties

$$\begin{aligned} \mathbf{R}^T(\theta) &= \mathbf{R}^{-1}(\theta) \\ det[\mathbf{R}(\theta)] &= 1 \end{aligned} \tag{2.32}$$

Therefore to rotate back again you simply apply $\mathbf{R}^T(\theta)$. The rotation matrix and its properties are discussed in detail in Evans (2001). More general rotations using more than one angle can be found in Appendix A.3.

We also choose to introduce additional notation that we believe is useful in this section. We define $\mathbf{U}_{\mathcal{XC}} = [\hat{\mathbf{c}}_1\hat{\mathbf{c}}_2\hat{\mathbf{c}}_3]$ as a matrix comprising of unit vectors of the coordinate system $\mathcal{C}$ expressed from the coordinate system X. In other words it rotates from coordinate system $\mathcal{X}$ to coordinate system $\mathcal{C}$. $\mathbf{U}_{\mathcal{XC}}$ is a rotation matrix therefore

$$\mathbf{U}_{\mathcal{XC}} = \mathbf{U}^T_{\mathcal{CX}} = (\mathbf{U}_{\mathcal{CX}})^{-1} \tag{2.33}$$

$$\mathbf{U}^T_{\mathcal{XX}} = \mathbf{I} \qquad (\forall \ \ \mathcal{X}) \tag{2.34}$$

A final matrix we define is

$$\mathbf{C}_3 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & -1 \end{bmatrix} \tag{2.35}$$

which exchanges the x and y axes and reverses the z axis. This matrix is used to perform a special rotation into a particular coordinate system that is shown in the next section.

2.5.2 Coordinate Systems

Propagation Coordinates ($\mathcal{P}$)

As mentioned earlier, the propagation coordinates ($\mathcal{P}$) are defined by the three vectors $[\hat{\mathbf{p}}\hat{\mathbf{a}}\hat{\mathbf{b}}]$. The vector $\hat{\mathbf{b}}$ points along the ray path, the S wave particle displacement is oriented along $\hat{\mathbf{p}}$ and the null vector is orthogonal to both of these, $\hat{\mathbf{a}} = \hat{\mathbf{b}} \times \hat{\mathbf{p}}$. These vectors are illustrated in Figure 2.5. The set of vectors can be expressed in our new, more general, notation

$$[\hat{\mathbf{p}}\hat{\mathbf{a}}\hat{\mathbf{b}}] = \mathbf{U}_{\mathcal{XP}} \tag{2.36}$$

This is the first coordinate system we have described in this section so we have no other system to rotate into at the moment.

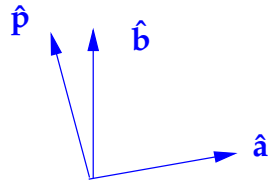


Figure 2.5: Propagation coordinates ($\mathcal{P}$)

Splitting Coordinates ($\mathcal{S}$)

The splitting coordinates ($\mathcal{S}$) are defined by the three vectors $[\hat{\mathbf{f}}\hat{\mathbf{s}}\hat{\mathbf{b}}]$. In anisotropic materials seismic velocities vary with direction so with respect to a propagation vector $\hat{\mathbf{b}}$ we define the displacement direction associated with a wave that travels fastest as the fast direction $\hat{\mathbf{f}}$. The displacement direction for a wave that travels the slowest is the slow direction $\hat{\mathbf{s}}$. The two vectors $\hat{\mathbf{f}}$ and $\hat{\mathbf{s}}$ can be non horizontal and non vertical. The rotation from $\mathcal{P}$ coordinates to $\mathcal{S}$ coordinates can be described by the angle ϕ which is the angle between $\hat{\mathbf{p}}$ and $\hat{\mathbf{f}}$ (Figure 2.6). The rotation is described below in both equation and matrix form using a rotation matrix.

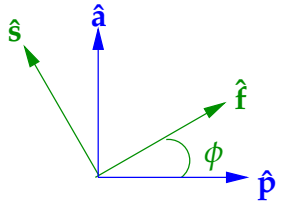


Figure 2.6: Rotation from $\mathcal{P}$ to $\mathcal{S}$

$$\begin{aligned} \hat{\mathbf{f}} &= \cos\phi\hat{\mathbf{p}} + \sin\phi\hat{\mathbf{a}} \\ \hat{\mathbf{s}} &= -\sin\phi\hat{\mathbf{p}} + \cos\phi\hat{\mathbf{a}} \\ \hat{\mathbf{b}} &= \hat{\mathbf{b}} \end{aligned} \tag{2.37}$$

$$\begin{bmatrix} \hat{\mathbf{f}}^T \\ \hat{\mathbf{s}}^T \\ \hat{\mathbf{b}}^T \end{bmatrix} = \begin{bmatrix} \cos\phi & \sin\phi & 0 \\ -\sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \hat{\mathbf{p}}^T \\ \hat{\mathbf{a}}^T \\ \hat{\mathbf{b}}^T \end{bmatrix} = \mathbf{R}_3^T(\phi) \begin{bmatrix} \hat{\mathbf{p}}^T \\ \hat{\mathbf{a}}^T \\ \hat{\mathbf{b}}^T \end{bmatrix} \tag{2.38}$$

Using our notation a rotation from any coordinate system $\mathcal{X}$ into $\mathcal{S}$ coordinates can be expressed in the following ways

$$\mathbf{U}_{\mathcal{XS}}^T = \mathbf{R}_3^T(\phi)\mathbf{U}_{\mathcal{XP}}^T \tag{2.39}$$

$$\mathbf{U}_{\mathcal{XS}} = \mathbf{U}_{\mathcal{XP}}\mathbf{R}_3(\phi) \qquad \text{(Using Equation 2.33)} \tag{2.40}$$

and a rotation from $\mathcal{P}$ coordinates into $\mathcal{S}$ coordinates can be expressed as

$$\mathbf{U}_{\mathcal{PS}} = \mathbf{U}_{\mathcal{PP}}\mathbf{R}_3(\phi) = \mathbf{R}_3(\phi) \qquad \text{(Using Equation 2.34)} \tag{2.41}$$

Emerging Ray Coordinates ($\mathcal{E}$)

Here we introduce a new coordinate system we call the emerging ray coordinates. Again, $\hat{\mathbf{b}}$ is the propagation direction. The component $\hat{\mathbf{t}}$ is perpendicular to $\hat{\mathbf{b}}$ and lies in the horizontal plane. The vector $\hat{\mathbf{q}} = \hat{\mathbf{t}} \times \hat{\mathbf{b}}$ completes the coordinate system, and together, $[\hat{\mathbf{q}}\hat{\mathbf{t}}\hat{\mathbf{b}}]$ form the emerging ray coordinates. This coordinate system is more commonly known as the $[\hat{\mathbf{q}}\hat{\mathbf{t}}\hat{\mathbf{l}}]$ system but we have chosen to be consistent with the Silver and Chan (1991) notation and use $\hat{\mathbf{b}}$ instead of $\hat{\mathbf{l}}$. The rotation from $\mathcal{P}$ coordinates to $\mathcal{E}$ coordinates can be described by the angle θ which is the angle between $\hat{\mathbf{p}}$ and $\hat{\mathbf{q}}$ (Figure 2.7). The rotation is described below in

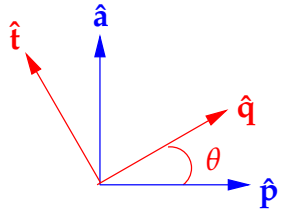


Figure 2.7: Rotation from $\mathcal{P}$ to $\mathcal{E}$

both equation and matrix form using a rotation matrix.

$$\begin{aligned} \hat{\mathbf{q}} &= \cos\theta\hat{\mathbf{p}} + \sin\theta\hat{\mathbf{a}} \\ \hat{\mathbf{t}} &= -\sin\theta\hat{\mathbf{p}} + \cos\theta\hat{\mathbf{a}} \\ \hat{\mathbf{b}} &= \hat{\mathbf{b}} \end{aligned} \tag{2.42}$$

$$\begin{bmatrix} \hat{\mathbf{q}}^T \\ \hat{\mathbf{t}}^T \\ \hat{\mathbf{b}}^T \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \hat{\mathbf{p}}^T \\ \hat{\mathbf{a}}^T \\ \hat{\mathbf{b}}^T \end{bmatrix} = \mathbf{R}_3^T(\theta) \begin{bmatrix} \hat{\mathbf{p}}^T \\ \hat{\mathbf{a}}^T \\ \hat{\mathbf{b}}^T \end{bmatrix} \tag{2.43}$$

Using our notation a rotation from any coordinate system $\mathcal{X}$ into $\mathcal{E}$ coordinates can be expressed in the following ways

$$\mathbf{U}_{\mathcal{XE}}^T = \mathbf{R}_3^T(\theta)\mathbf{U}_{\mathcal{XP}}^T \tag{2.44}$$

$$\mathbf{U}_{\mathcal{XE}} = \mathbf{U}_{\mathcal{XP}}\mathbf{R}_3(\theta) \qquad \text{(Using Equation 2.33)} \tag{2.45}$$

and a rotation from $\mathcal{P}$ coordinates into $\mathcal{E}$ coordinates can be expressed as

$$\mathbf{U}_{\mathcal{PE}} = \mathbf{U}_{\mathcal{PP}}\mathbf{R}_3(\theta) = \mathbf{R}_3(\theta) \qquad \text{(Using Equation 2.34)} \tag{2.46}$$

Ray Planar Coordinates ($\mathcal{R}$)

Another ray based coordinate system that we call ray planar coordinates ($\mathcal{R}$) is represented by the three vectors $[\hat{\mathbf{r}}\hat{\mathbf{t}}\hat{\mathbf{z}}]$. The vector $\hat{\mathbf{z}}$ is the vertical direction, $\hat{\mathbf{r}}$ is the radial direction, which is a horizontal projection of the SV displacements i.e. the horizontal projection of $\hat{\mathbf{b}}$. The final component $\hat{\mathbf{t}} = \hat{\mathbf{z}} \times \hat{\mathbf{r}}$ is the transverse component. The rotation from $\mathcal{E}$ coordinates to $\mathcal{R}$ coordinates can be described by the angle of incidence, i, which is the angle between $\hat{\mathbf{b}}$ and $\hat{\mathbf{z}}$ (Figure 2.8). In the special case of a vertically travelling wave i.e. $\hat{\mathbf{b}} = \hat{\mathbf{z}}$, $i = 0$ so the ray planar ($\mathcal{R}$) and emerging ray ($\mathcal{E}$) coordinate systems are the same. The rotation is

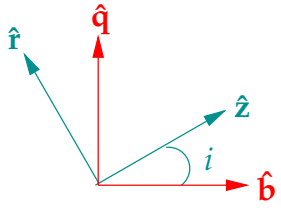


Figure 2.8: Rotation from $\mathcal{E}$ to $\mathcal{R}$

described below in both equation and matrix form using a rotation matrix.

$$\begin{aligned}\hat{\mathbf{r}} &= \cos i\hat{\mathbf{q}} - \sin i\hat{\mathbf{b}} \\ \hat{\mathbf{t}} &= \hat{\mathbf{t}} \\ \hat{\mathbf{z}} &= \sin i\hat{\mathbf{q}} + \cos i\hat{\mathbf{b}}\end{aligned} \tag{2.47}$$

$$\begin{bmatrix} \hat{\mathbf{r}}^T \\ \hat{\mathbf{t}}^T \\ \hat{\mathbf{z}}^T \end{bmatrix} = \begin{bmatrix} \cos i & 0 & -\sin i \\ 0 & 1 & 0 \\ \sin i & 0 & \cos i \end{bmatrix} \begin{bmatrix} \hat{\mathbf{q}}^T \\ \hat{\mathbf{t}}^T \\ \hat{\mathbf{b}}^T \end{bmatrix} = \mathbf{R}_2^T(i) \begin{bmatrix} \hat{\mathbf{q}}^T \\ \hat{\mathbf{t}}^T \\ \hat{\mathbf{b}}^T \end{bmatrix} \tag{2.48}$$

Using our notation a rotation from any coordinate system $\mathcal{X}$ into $\mathcal{R}$ coordinates can be expressed in the following ways

$$\mathbf{U}_{\mathcal{X}\mathcal{R}}^T = \mathbf{R}_2^T(i)\mathbf{U}_{\mathcal{X}\mathcal{E}}^T \tag{2.49}$$

$$\mathbf{U}_{\mathcal{X}\mathcal{R}} = \mathbf{U}_{\mathcal{X}\mathcal{E}}\mathbf{R}_2(i) \qquad \text{(Using Equation 2.33)} \tag{2.50}$$

and a rotation from $\mathcal{E}$ into $\mathcal{R}$ coordinates can be expressed as

$$\mathbf{U}_{\mathcal{E}\mathcal{R}} = \mathbf{U}_{\mathcal{E}\mathcal{E}}\mathbf{R}_2(i) = \mathbf{R}_2(i) \qquad \text{(Using Equation 2.34)} \tag{2.51}$$

Geographic Coordinates ($\mathcal{G}$)

Another coordinate system that exists is the one used by seismometers, geographical coordinates ($\mathcal{G}$) i.e. east ($\hat{\mathbf{e}}$) and north ($\hat{\mathbf{n}}$) and vertical ($\hat{\mathbf{z}} = \hat{\mathbf{e}} \times \hat{\mathbf{n}}$). The rotation from $\mathcal{R}$ coordinates to $\mathcal{G}$ coordinates can be described by the angle b, which is the angle between $\hat{\mathbf{r}}$ and $\hat{\mathbf{e}}$ (Figure 2.9). The rotation is described below in both equation and matrix form using a rotation matrix.

$$\begin{aligned}\hat{\mathbf{e}} &= \cos b\hat{\mathbf{r}} + \sin b\hat{\mathbf{t}} \\ \hat{\mathbf{n}} &= -\sin b\hat{\mathbf{r}} + \cos b\hat{\mathbf{t}} \\ \hat{\mathbf{z}} &= \hat{\mathbf{z}}\end{aligned} \tag{2.52}$$

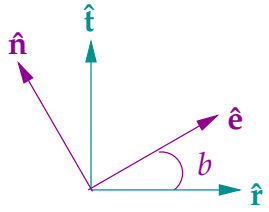


Figure 2.9: Rotation from $\mathcal{R}$ to $\mathcal{G}$

$$\begin{bmatrix} \hat{\mathbf{e}}^T \\ \hat{\mathbf{n}}^T \\ \hat{\mathbf{z}}^T \end{bmatrix} = \begin{bmatrix} \cos b & +\sin b & 0 \\ -\sin b & \cos b & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \hat{\mathbf{r}}^T \\ \hat{\mathbf{t}}^T \\ \hat{\mathbf{z}}^T \end{bmatrix} = \mathbf{R}_3^T(b) \begin{bmatrix} \hat{\mathbf{r}}^T \\ \hat{\mathbf{t}}^T \\ \hat{\mathbf{z}}^T \end{bmatrix} \tag{2.53}$$

Using our notation a rotation from any coordinate system $\mathcal{X}$ into $\mathcal{G}$ coordinates can be expressed in the following ways

$$\mathbf{U}_{\mathcal{XG}}^T = \mathbf{R}_3^T(b)\mathbf{U}_{\mathcal{XR}}^T \tag{2.54}$$

$$\mathbf{U}_{\mathcal{XG}} = \mathbf{U}_{\mathcal{XR}}\mathbf{R}_3(b) \qquad \text{(Using Equation 2.33)} \tag{2.55}$$

and a rotation from $\mathcal{R}$ coordinates into $\mathcal{G}$ coordinates can be expressed as follows

$$\mathbf{U}_{\mathcal{RG}} = \mathbf{U}_{\mathcal{RR}}\mathbf{R}_3(b) = \mathbf{R}_3(b) \qquad \text{(Using Equation 2.34)} \tag{2.56}$$

Bearing Coordinates ($\mathcal{B}$)

Another similar system to $\mathcal{G}$ is the bearing system ($\mathcal{B}$). This system is defined by the coordinates $[\hat{\mathbf{n}}\hat{\mathbf{e}}\hat{\mathbf{d}}]$ where $\hat{\mathbf{d}}$ is positive downwards ($\hat{\mathbf{d}} = -\hat{\mathbf{z}}$). Quite a few angles of interest such as the fast direction are reported clock wise from North in the bearing coordinate system ($\mathcal{B}$). Since these two systems are similar all that is required is a rotation matrix to turn the geographic coordinate system upside down and then a rotation into $[\hat{\mathbf{n}}\hat{\mathbf{e}}\hat{\mathbf{d}}]$. (Figure 2.10). The rotation is described below in matrix form using a rotation matrix

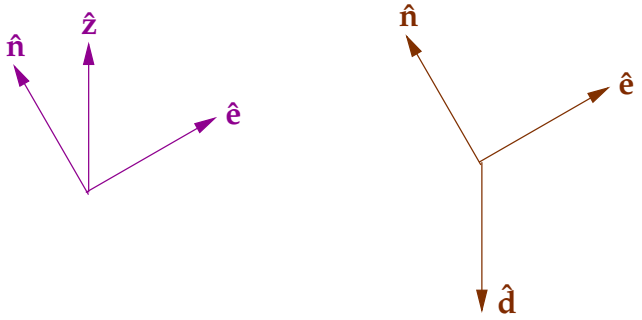


Figure 2.10: Rotation from $\mathcal{G}$ to $\mathcal{B}$

$$\begin{bmatrix} \hat{\mathbf{n}}^T \\ \hat{\mathbf{e}}^T \\ \hat{\mathbf{d}}^T \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} \hat{\mathbf{e}}^T \\ \hat{\mathbf{n}}^T \\ \hat{\mathbf{z}}^T \end{bmatrix} = \mathbf{C}_3 \begin{bmatrix} \hat{\mathbf{e}}^T \\ \hat{\mathbf{n}}^T \\ \hat{\mathbf{z}}^T \end{bmatrix} \tag{2.57}$$

Coordinate System	Vector Components	Transformation Angle	
$\mathcal{P}$ - propagation	$[\hat{\mathbf{p}}\hat{\mathbf{a}}\hat{\mathbf{b}}] = \mathbf{U}_{\mathcal{XP}}$	-	
$\mathcal{S}$ - splitting	$[\hat{\mathbf{f}}\hat{\mathbf{s}}\hat{\mathbf{b}}] = \mathbf{U}_{\mathcal{XS}} = \mathbf{U}_{\mathcal{XP}}\mathbf{R}_3(\phi)$	ϕ	$(\mathcal{P} \to \mathcal{S})$
$\mathcal{E}$ - emerging ray	$[\hat{\mathbf{q}}\hat{\mathbf{t}}\hat{\mathbf{b}}] = \mathbf{U}_{\mathcal{XE}} = \mathbf{U}_{\mathcal{XP}}\mathbf{R}_3(\theta)$	θ	$(\mathcal{P} \to \mathcal{E})$
$\mathcal{R}$ - ray planar	$[\hat{\mathbf{r}}\hat{\mathbf{t}}\hat{\mathbf{z}}] = \mathbf{U}_{\mathcal{XR}} = \mathbf{U}_{\mathcal{XE}}\mathbf{R}_2(i)$	i	$(\mathcal{E} \to \mathcal{R})$
$\mathcal{G}$ - geographic	$[\hat{\mathbf{e}}\hat{\mathbf{n}}\hat{\mathbf{z}}] = \mathbf{U}_{\mathcal{XG}} = \mathbf{U}_{\mathcal{XR}}\mathbf{R}_3(b)$	b	$(\mathcal{G} \to \mathcal{R})$
$\mathcal{B}$ - bearing	$[\hat{\mathbf{n}}\hat{\mathbf{e}}\hat{\mathbf{d}}] = \mathbf{U}_{\mathcal{XB}} = \mathbf{U}_{\mathcal{XG}}\mathbf{C}_3$		$(\mathcal{G} \to \mathcal{B})$

Table 2.1: Summary of the different right handed coordinate systems and the vectors that define them. The angles that rotate you into these coordinate systems is also given along with the appropriate rotation.

Using our notation a rotation from any coordinate system $\mathcal{X}$ into $\mathcal{B}$ coordinates can be expressed in the following ways

$$\mathbf{U}_{\mathcal{XB}}^T = \mathbf{C}_3\mathbf{U}_{\mathcal{XG}}^T \tag{2.58}$$

$$\mathbf{U}_{\mathcal{XB}} = \mathbf{U}_{\mathcal{XG}}\mathbf{C}_3 \qquad \text{(Using Equation 2.33)} \tag{2.59}$$

Summary of the Coordinate Systems

Table 2.1 summarises all the coordinate systems along with a shorthand name and the transformation angles that are used to rotate between particular coordinate systems. The coordinate systems are also summarised in Figure 2.11 with an arbitrary splitting coordinate system. The two vectors $\hat{\mathbf{f}}$ and $\hat{\mathbf{s}}$ can be non horizontal and non vertical. The summary table shows the angles that we used to rotate into different coordinate systems. Most of these angles would be used during processing to rotate into a particular coordinate systems and carry out an analysis. However, when results are reported, seismologists focus on three main angles of interest to them that are in bearing coordinates. They are

$$\begin{aligned} \phi_b &= \text{ bearing fast direction measured clockwise from } \hat{\mathbf{n}} \text{ to } \hat{\mathbf{f}} \\ \alpha &= \text{ incoming polarisation measured clockwise from } \hat{\mathbf{n}} \text{ to } \hat{\mathbf{p}} \\ \beta &= \text{ back azimuth measured clockwise from } \hat{\mathbf{n}} \text{ to } \hat{\mathbf{r}} \text{ plus } 180^\circ \end{aligned}$$

Note that ϕ_b is a true bearing for a vertical ray, otherwise this angle is not in a horizontal plane The fast direction provides information about the anisotropic medium, the incoming polarisation provides information about the orientation of the wave in an isotropic medium and the back azimuth provides information about the earthquake location.

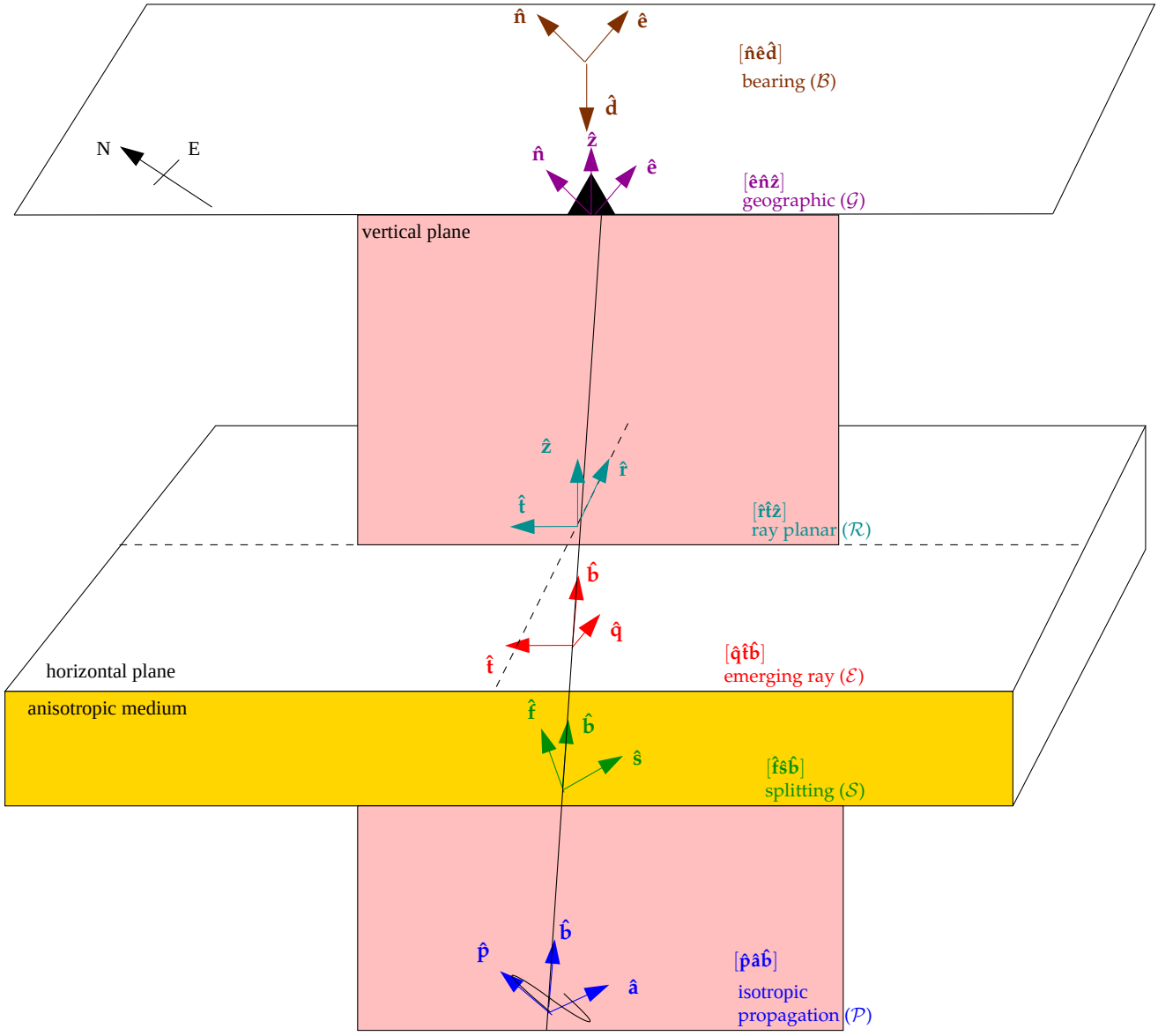


Figure 2.11: Different right hand coordinate systems for a near vertically travelling wave that may be used to model the shear wave splitting process. Not drawn to scale.

2.5.3 Shear Wave Splitting and Rotated Components

The previous section explained all the coordinate systems along with the relevant angles. Here we explain a split wave in $\mathcal{S}$ coordinates in a more general form and then look at a special case of shear wave splitting expressed in $\mathcal{R}$ coordinates

Split Wave in $\mathcal{S}$ Coordinates

Recall that a split wave can be modelled by the following equations

$$\begin{aligned}\mathbf{u}_s(\omega,t) &= \left[\exp(-i\omega\delta t/2)\hat{\mathbf{f}}\hat{\mathbf{f}}^T + \exp(i\omega\delta t/2)\hat{\mathbf{s}}\hat{\mathbf{s}}^T + \hat{\mathbf{b}}\hat{\mathbf{b}}^T\right] w(\omega)\exp(-i\omega t)\hat{\mathbf{p}} \quad \text{(Equations 2.4 \& 2.14)}\\ &= \mathbf{\Gamma}(\phi,\delta t;\omega)\mathbf{u}(\omega,t) \quad \text{(Equation 2.25)}\\ &= \exp(-i\omega\delta\mathbf{T})\mathbf{u}(\omega,t) \quad \text{(Equations 2.15 \& 2.16)}\end{aligned} \tag{2.60}$$

A split wave can also be rewritten using trigonometric functions. Remember that $\hat{\mathbf{f}}^T\hat{\mathbf{p}} = \cos\phi$ and $\hat{\mathbf{s}}^T\hat{\mathbf{p}} = \cos(\phi+\pi/2) = -\sin\phi$. First we group the common terms together and then use our definition of ϕ to simplify

$$\begin{aligned}\mathbf{u}_s(\omega,t) &= \left[\exp\left[-i\omega(t+\delta t/2)\right]\hat{\mathbf{f}}(\hat{\mathbf{f}}^T\hat{\mathbf{p}}) + \exp\left[i\omega(t-\delta t/2)\right]\hat{\mathbf{s}}(\hat{\mathbf{s}}^T\hat{\mathbf{p}})\right] w(\omega)\\ &= \left[\exp\left[-i\omega(t+\delta t/2)\right]\hat{\mathbf{f}}(\cos\phi) + \exp\left[i\omega(t-\delta t/2)\right]\hat{\mathbf{s}}(-\sin\phi)\right] w(\omega)\\ &= \left[\exp\left[-i\omega(t+\delta t/2)\right]\cos\phi\hat{\mathbf{f}} - \exp\left[i\omega(t-\delta t/2)\right]\sin\phi\hat{\mathbf{s}}\right] w(\omega)\\ &= W(\omega,t+\delta t/2)\cos\phi\hat{\mathbf{f}} - W(\omega,t-\delta t/2)\sin\phi\hat{\mathbf{s}}\end{aligned} \tag{2.61}$$

where

$$W(\omega,t) = w(\omega)\exp(-i\omega t) \tag{2.62}$$

This can be defined using our more general notation in the previous section providing we define three new vectors

$$\hat{\boldsymbol{\delta}}_1 = \begin{bmatrix}1\\0\\0\end{bmatrix}, \hat{\boldsymbol{\delta}}_2 = \begin{bmatrix}0\\1\\0\end{bmatrix}, \hat{\boldsymbol{\delta}}_3 = \begin{bmatrix}0\\0\\1\end{bmatrix} \tag{2.63}$$

which are the columns of the identity matrix i.e. $[\hat{\boldsymbol{\delta}}_1\hat{\boldsymbol{\delta}}_2\hat{\boldsymbol{\delta}}_3] = \mathbf{I}$. This allows us to access particular columns of a matrix. In the previous section we defined $\mathbf{U}_{\mathcal{XC}} = [\hat{\mathbf{c}}_1\hat{\mathbf{c}}_2\hat{\mathbf{c}}_3]$ as a matrix comprised of unit vectors of frame $\mathcal{C}$ explained from the coordinate system $\mathcal{X}$ (in other words, a rotation from $\mathcal{X}$ to $\mathcal{C}$) therefore

$$\mathbf{U}_{\mathcal{XC}}\hat{\boldsymbol{\delta}}_j = \mathbf{c}_j \qquad j = 1,2,3 \tag{2.64}$$

The vectors forming the splitting coordinates ($\mathcal{S}$) can then be rewritten in any coordinate system $\mathcal{X}$ as follows

$$\begin{aligned}\hat{\mathbf{f}} &= \mathbf{U}_{\mathcal{XS}}\hat{\boldsymbol{\delta}}_1\\ \hat{\mathbf{s}} &= \mathbf{U}_{\mathcal{XS}}\hat{\boldsymbol{\delta}}_2\\ \hat{\mathbf{b}} &= \mathbf{U}_{\mathcal{XS}}\hat{\boldsymbol{\delta}}_3\end{aligned} \tag{2.65}$$

and then we may rewrite Equation 2.61 as follows

$$\mathbf{u}_s(\omega,t) = W(\omega,t+\delta t/2)\cos\phi\mathbf{U}_{\mathcal{XS}}\hat{\boldsymbol{\delta}}_1 - W(\omega,t-\delta t/2)\sin\phi\mathbf{U}_{\mathcal{XS}}\hat{\boldsymbol{\delta}}_2 \tag{2.66}$$

We define

$$\begin{aligned} W_+ &= W(\omega,t+\delta t/2) \\ W_- &= W(\omega,t-\delta t/2) \end{aligned}$$

so that we can shorten Equation 2.66 and subsequent equations that involve these terms. Equation 2.66 can then be shortened to

$$\hat{\mathbf{u}}_s(\omega,t) = W_+\cos\phi\mathbf{U}_{\mathcal{XS}}\hat{\boldsymbol{\delta}}_1 - W_-\sin\phi\mathbf{U}_{\mathcal{XS}}\hat{\boldsymbol{\delta}}_2 \tag{2.67}$$

Expressing a split wave in this more general manner allows us to express the split wave in splitting coordinates from any coordinate system $\mathcal{X}$.

Split Wave in $\mathcal{R}$ Coordinates

The split wave may also be expressed in ray planar coordinates ($\mathcal{R}$). If $\hat{\mathbf{p}}$ (and consequently $\hat{\mathbf{q}}$) is known then we can rotate into radial and transverse components. This is necessary if we wish to minimise the energy on the transverse component. Using the same more general notation as before, the three components defining this system can be expressed as

$$\begin{aligned} \hat{\mathbf{r}} &= \mathbf{U}_{\mathcal{XR}}\hat{\boldsymbol{\delta}}_1 \\ \hat{\mathbf{t}} &= \mathbf{U}_{\mathcal{XR}}\hat{\boldsymbol{\delta}}_2 \\ \hat{\mathbf{z}} &= \mathbf{U}_{\mathcal{XR}}\hat{\boldsymbol{\delta}}_3 \end{aligned} \tag{2.68}$$

To keep the equations relatively short we represent these three vectors as a single equation

$$\hat{\mathbf{r}}_j = \mathbf{U}_{\mathcal{XR}}\hat{\boldsymbol{\delta}}_j \qquad j = 1,2,3 \tag{2.69}$$

where $[\hat{\mathbf{r}}_1\hat{\mathbf{r}}_2\hat{\mathbf{r}}_3] = [\hat{\mathbf{r}}\hat{\mathbf{t}}\hat{\mathbf{z}}]$. Using this notation we can now define a split wave in $\mathcal{R}$ coordinates

$$\begin{aligned} \hat{\mathbf{u}}_{r_j} &= \hat{\mathbf{r}}_j^T\hat{\mathbf{u}}_s \\ &= \hat{\boldsymbol{\delta}}_j^T\mathbf{U}_{\mathcal{RX}}\hat{\mathbf{u}}_s \qquad \text{(Since } U_{\mathcal{XR}}^T = U_{\mathcal{RX}}\text{. Equation 2.33)} \\ &= \hat{\boldsymbol{\delta}}_j^T\mathbf{U}_{\mathcal{RX}}\left[W_+\cos\phi\mathbf{U}_{\mathcal{XS}}\hat{\boldsymbol{\delta}}_1 - W_-\sin\phi\mathbf{U}_{\mathcal{XS}}\hat{\boldsymbol{\delta}}_2\right] \qquad \text{(from Equation 2.67)} \\ &= \hat{\boldsymbol{\delta}}_j^T\mathbf{U}_{\mathcal{RX}}\mathbf{U}_{\mathcal{XS}}\left[W_+\cos\phi\hat{\boldsymbol{\delta}}_1 - W_-\sin\phi\hat{\boldsymbol{\delta}}_2\right] \\ &= \hat{\boldsymbol{\delta}}_j^T\mathbf{U}_{\mathcal{RS}}\left[W_+\cos\phi\hat{\boldsymbol{\delta}}_1 - W_-\sin\phi\hat{\boldsymbol{\delta}}_2\right] \end{aligned} \tag{2.70}$$

The matrix $\mathbf{U}_{\mathcal{RS}}$ is not defined in Section 2.5.2 but we can obtain it using the matrices (or their transposes) that are defined in that section (Equations 2.41, 2.46 and 2.51) as follows

$$\begin{aligned} \mathbf{U}_{\mathcal{RS}} &= \mathbf{U}_{\mathcal{RE}}\mathbf{U}_{\mathcal{EP}}\mathbf{U}_{\mathcal{PS}} \\ &= \mathbf{U}_{\mathcal{ER}}^T\mathbf{U}_{\mathcal{PE}}^T\mathbf{U}_{\mathcal{PS}} \\ &= \mathbf{R}_2^T(i)\mathbf{R}_3^T(\theta)\mathbf{R}_3(\phi) \end{aligned} \tag{2.71}$$

The two rotation matrices that rotate the third component can be combined into a single rotation matrix using trigonometric identities

$$\begin{aligned}
\mathbf{R}_3^T(\theta)\mathbf{R}_3(\phi) &= \begin{bmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\phi & -\sin\phi & 0 \\ \sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{bmatrix} \\
&= \begin{bmatrix} \cos\theta\cos\phi + \sin\theta\sin\phi & -\cos\theta\sin\phi + \sin\theta\cos\phi & 0 \\ -\sin\theta\cos\phi + \cos\theta\sin\phi & \sin\theta\sin\phi + \cos\theta\cos\phi & 0 \\ 0 & 0 & 1 \end{bmatrix} \\
&= \begin{bmatrix} \cos(\phi-\theta) & -\sin(\phi-\theta) & 0 \\ \sin(\phi-\theta) & \cos(\phi-\theta) & 0 \\ 0 & 0 & 1 \end{bmatrix} \\
&= \mathbf{R}_3(\phi-\theta)
\end{aligned} \tag{2.72}$$

then $\mathbf{U}_{\mathcal{RS}}$ becomes

$$\mathbf{U}_{\mathcal{RS}} = \mathbf{R}_2^T(i)\mathbf{R}_3(\phi-\theta) \tag{2.73}$$

We can then substitute this simplified version of $\mathbf{U}_{\mathcal{RS}}$ into Equation 2.70

$$\begin{aligned}
u_{r_j} &= \hat{\boldsymbol{\delta}}_j^T \mathbf{U}_{\mathcal{RS}} \left[W_+ \cos\phi \hat{\boldsymbol{\delta}}_1 - W_- \sin\phi \hat{\boldsymbol{\delta}}_2 \right] \\
&= \hat{\boldsymbol{\delta}}_j^T \mathbf{R}_2^T(i)\mathbf{R}_3(\phi-\theta) \left[W_+ \cos\phi \hat{\boldsymbol{\delta}}_1 - W_- \sin\phi \hat{\boldsymbol{\delta}}_2 \right]
\end{aligned} \tag{2.74}$$

This is the most general form describing a rotation into $\mathcal{R}$ coordinates from $\mathcal{S}$ coordinates. The Silver and Chan (1991) method looks at special case of a vertically incident ray. In other words, the angle of incidence, $i = 0$. This implies that the $\mathcal{R}$ and $\mathcal{E}$ coordinates are the same (see Figure 2.8). Another special case is when $\hat{\mathbf{p}}$ is in the vertical plane of $\hat{\mathbf{b}}$ and $\hat{\mathbf{z}}$. In other words $\theta = 0$. This implies that the $\mathcal{P}$ and $\mathcal{E}$ coordinates are the same (see Figure 2.7). If both these special cases are true then $\mathcal{P}$=$\mathcal{E}$=$\mathcal{R}$ so $[\hat{\mathbf{p}}\hat{\mathbf{a}}\hat{\mathbf{b}}] = [\hat{\mathbf{q}}\hat{\mathbf{t}}\hat{\mathbf{b}}] = [\hat{\mathbf{r}}\hat{\mathbf{t}}\hat{\mathbf{z}}]$. When both these special cases are true ($i = 0, \theta = 0$) then the radial component (the first component of the $\mathcal{R}$ coordinates) is

$$\begin{aligned}
u_{r_1} &= \hat{\boldsymbol{\delta}}_1^T \mathbf{R}_2^T(i)\mathbf{R}_3(\phi-\theta) \left[W_+ \cos\phi \hat{\boldsymbol{\delta}}_1 - W_- \sin\phi \hat{\boldsymbol{\delta}}_2 \right] \\
&= \hat{\boldsymbol{\delta}}_1^T \mathbf{R}_3(\phi) \left[W_+ \cos\phi \hat{\boldsymbol{\delta}}_1 - W_- \sin\phi \hat{\boldsymbol{\delta}}_2 \right] \\
&= W_+ \cos\phi (\hat{\boldsymbol{\delta}}_1^T \mathbf{R}_3(\phi) \hat{\boldsymbol{\delta}}_1) - W_- \sin\phi (\hat{\boldsymbol{\delta}}_1^T \mathbf{R}_3(\phi) \hat{\boldsymbol{\delta}}_2) \\
&= W_+ \cos\phi(\cos\phi) - W_- \sin\phi(-\sin\phi) \qquad \text{(By Equations 2.31 2.63)} \\
&= W_+ \cos^2\phi - W_+ \sin^2\phi \\
&= w(\omega) \exp[-i\omega(t + \delta t/2)] \cos^2\phi + w(\omega) \exp[-i\omega(t - \delta t/2)] \sin^2\phi
\end{aligned} \tag{2.75}$$

(c.f. SC 14)

and the transverse component (the second component of the $\mathcal{R}$ coordinates) when $i = 0, \theta = 0$

is

$$
\begin{aligned}
u_{r_2} &= \hat{\boldsymbol{\delta}}_2^T \mathbf{R}_2^T(i)\mathbf{R}_3(\phi-\theta)\left[W_+\cos\phi\hat{\boldsymbol{\delta}}_1 - W_-\sin\phi\hat{\boldsymbol{\delta}}_2\right] \\
&= \hat{\boldsymbol{\delta}}_2^T \mathbf{R}_3(\phi)\left[W_+\cos\phi\hat{\boldsymbol{\delta}}_1 - W_-\sin\phi\hat{\boldsymbol{\delta}}_2\right] \\
&= \left[W_+\cos\phi(\hat{\boldsymbol{\delta}}_2^T\mathbf{R}_3(\phi)\hat{\boldsymbol{\delta}}_1) - W_-\sin\phi(\hat{\boldsymbol{\delta}}_2^T\mathbf{R}_3(\phi)\hat{\boldsymbol{\delta}}_2)\right] \\
&= [W_+\cos\phi(\sin\phi) - W_-\sin\phi(\cos\phi)] \qquad \text{(By Equations 2.31 2.63)} \\
&= \sin\phi\cos\phi(W_+ - W_-) \\
&= \frac{1}{2}\sin 2\phi(W_+ - W_-) \qquad \text{(using the half angle formula)} \\
&= \frac{1}{2}\sin 2\phi w(\omega)\left(\exp[-i\omega(t+\delta t/2)] - \exp[-i\omega(t-\delta t/2)]\right)
\end{aligned}
\tag{2.76}
$$

(c.f. SC 15)

If we define $w(\omega,t) = w(\omega)\exp(-i\omega t)$ then we obtain the same equations for the radial *(SC 14)* and transverse *(SC 15)* components that are in the Silver and Chan (1991) article.

2.6 Covariance Matrix

Now that we are familiar with all the coordinate systems and what we are optimising we can search for the best splitting parameters. The best pair of ϕ and δt are the ones that produce a corrected wave that is the most similar to the original wave (Equation 2.26). Choosing the best pair is done by calculating the covariance matrix of the two orthogonal fast and slow components in the splitting coordinate ($\mathcal{S}$) system. The covariance is defined as

$$
Cov(X,Y) = \left\langle [X - \langle X\rangle][Y - \langle Y\rangle)] \right\rangle = \langle XY\rangle - \langle X\rangle\langle Y\rangle \tag{2.77}
$$

where $\langle\cdot\rangle$ denotes the expected value. The covariance matrix describes how the elements of a random vector vary with each other with the covariances being the off diagonal components and the variances along the diagonal. Two properties of the covariance matrix that we use often in this section are

$$
\text{cov}(a\mathbf{X}) = a^2\text{cov}(\mathbf{X}) \tag{2.78}
$$

$$
\text{cov}(\mathbf{AX}) = \mathbf{A}\text{cov}(\mathbf{X})\mathbf{A}^T \tag{2.79}
$$

We can use these to calculate the covariance matrix of two orthogonal components. Recall that the ground motion for a split wave and an unsplit wave are

$$
\begin{aligned}
\mathbf{u}_s(\omega,t) &= \exp(-i\omega\delta\mathbf{T})\mathbf{u}(\omega,t) = \boldsymbol{\Gamma}(\phi,\delta t)\mathbf{u}(\omega,t) \\
\mathbf{u}(\omega,t) &= w(\omega)\exp(-i\omega t)\hat{\mathbf{p}}
\end{aligned}
\tag{2.80}
$$

and that the fully corrected wave is defined as

$$
\begin{aligned}
\tilde{\mathbf{u}}(\omega,t) &= \boldsymbol{\Gamma}^{-1}(\phi,\delta t)\mathbf{u}_s(\omega,t) \\
&= \exp(i\omega\delta\mathbf{T})\mathbf{u}_s(\omega,t) \\
&= [\exp(i\omega\delta t)\hat{\mathbf{f}}\hat{\mathbf{f}}^T + \exp(-i\omega\delta t)\hat{\mathbf{s}}\hat{\mathbf{s}}^T + \hat{\mathbf{b}}\hat{\mathbf{b}}^T]\mathbf{u}_s(\omega,t) \\
&= w(\omega)\exp(-i\omega t)\hat{\mathbf{p}}
\end{aligned}
\tag{2.81}
$$

In $\mathcal{P}$ coordinates we can express the covariance of the fully corrected wave as

$$
\begin{aligned}
(\tilde{\mathbf{u}})_{\mathcal{P}} &= w(\omega)\exp(-i\omega t)\begin{bmatrix}1\\0\\0\end{bmatrix}\\
\mathrm{cov}((\tilde{\mathbf{u}})_{\mathcal{P}}) &= \mathrm{cov}\left(w(\omega)\begin{bmatrix}\exp(-i\omega t)\\0\\0\end{bmatrix}\right)\\
&= w(\omega)^2\mathrm{cov}\left(\begin{bmatrix}\mathrm{var}[\exp(-i\omega t)] & 0 & 0\\ 0 & 0 & 0\\ 0 & 0 & 0\end{bmatrix}\right) \qquad \text{(Using Equation 2.78)}
\end{aligned} \tag{2.82}
$$

When the split wave is fully corrected the covariance matrix is singular and has one eigenvector $[100]^T_{\mathcal{P}} = \hat{\mathbf{p}}$ corresponding to the polarisation of the wave. Since this is a diagonal matrix we can read the eigenvalues off the diagonal. There is only one non-zero eigenvalue $\lambda_1 = w(\omega)^2\mathrm{var}[\exp(-i\omega t)]$. In the 2D case, the other eigenvalue λ_2 is zero if we have fully corrected and in the 3D case both eigenvalues $\lambda_2 = \lambda_3 = 0$ if we have fully corrected.

The Silver and Chan (1991) method constructs a covariance matrix based on the fast and slow components in the $\mathcal{S}$ coordinate system. Therefore we want the covariance matrix of

$$
\begin{aligned}
\tilde{\mathbf{u}}(\omega,t) &= [\exp(i\omega\delta t)\hat{\mathbf{f}}\hat{\mathbf{f}}^T + \exp(-i\omega\delta t)\hat{\mathbf{s}}\hat{\mathbf{s}}^T + \hat{\mathbf{b}}\hat{\mathbf{b}}^T]\mathbf{u}_s(\omega,t)\\
&= [\hat{\mathbf{f}}\hat{\mathbf{s}}\hat{\mathbf{b}}]\begin{bmatrix}\exp(i\omega\delta t/2)\hat{\mathbf{f}}^T\\ \exp(-i\omega\delta t/2)\hat{\mathbf{s}}^T\\ 1\hat{\mathbf{b}}^T\end{bmatrix}\mathbf{u}_s(\omega,t)\\
&= [\hat{\mathbf{f}}\hat{\mathbf{s}}\hat{\mathbf{b}}]\begin{bmatrix}\exp(i\omega\delta t/2) & 0 & 0\\ 0 & \exp(-i\omega\delta t/2) & 0\\ 0 & 0 & 1\end{bmatrix}\begin{bmatrix}\hat{\mathbf{f}}^T\\ \hat{\mathbf{s}}^T\\ \hat{\mathbf{b}}^T\end{bmatrix}\mathbf{u}_s(\omega,t)\\
&= \mathbf{U}_{\mathcal{XS}}\mathbf{D}\mathbf{U}^T_{\mathcal{XS}}\mathbf{u}_s(\omega,t)
\end{aligned} \tag{2.83}
$$

where $\mathbf{D}$ is the matrix containing the exponential terms and $\mathbf{U}_{\mathcal{XS}} = [\hat{\mathbf{f}}\hat{\mathbf{s}}\hat{\mathbf{b}}]$. The covariance of this is

$$
\begin{aligned}
\mathrm{cov}\left(\tilde{\mathbf{u}}(\omega,t)\right) &= \mathrm{cov}\left(\mathbf{U}_{\mathcal{XS}}\mathbf{D}\mathbf{U}^T_{\mathcal{XS}}\mathbf{u}_s(\omega,t)\right)\\
&= \mathbf{U}_{\mathcal{XS}}\left[\mathrm{cov}\left(\mathbf{D}\mathbf{U}^T_{\mathcal{XS}}\mathbf{u}_s(\omega,t)\right)\right]\mathbf{U}^T_{\mathcal{XS}} \qquad \text{(Using Equation 2.79)}
\end{aligned} \tag{2.84}
$$

Again this more general form will allow you to define the covariance matrix in $\mathcal{S}$ coordinates using any other coordinate system $\mathcal{X}$ providing we know the rotation matrix $\mathbf{U}_{\mathcal{XS}}$. If we choose to refer everything to ray planar coordinates ($\mathcal{R}$) then we must replace each $\mathbf{U}_{\mathcal{XS}}$ with

$$
\begin{aligned}
\mathbf{U}_{\mathcal{RS}} &= \mathbf{U}_{\mathcal{RE}}\mathbf{U}_{\mathcal{EP}}\mathbf{U}_{\mathcal{PS}} \qquad \text{(Equation 2.71)}\\
&= \mathbf{R}_2^T(i)\mathbf{R}_3^T(\theta)\mathbf{R}_3(\phi) \qquad \text{(Equation 2.72)}\\
&= \mathbf{R}_2^T(i)\mathbf{R}_3(\phi-\theta) \qquad \text{(Equation 2.73)}
\end{aligned} \tag{2.85}
$$

and then the covariance matrix becomes

$$\text{cov}\left(\tilde{\mathbf{u}}_{\mathcal{R}}(\omega,t)\right) = \mathbf{U}_{\mathcal{RS}}\left[\text{cov}\left(\mathbf{D}\mathbf{U}_{\mathcal{RS}}^{T}(\mathbf{u}_s(\omega,t))_{\mathcal{R}}\right)\right]\mathbf{U}_{\mathcal{RS}}^{T} = \mathbf{U}_{\mathcal{RS}}\mathbf{c}_{\omega}(\phi,\delta t)\mathbf{U}_{\mathcal{RS}}^{T} \tag{2.86}$$

In the special case where $i = 0$ and $\theta = 0$ then $\mathbf{U}_{\mathcal{RS}} = \mathbf{U}_{\mathcal{RR}}\mathbf{R}_3^T(\phi) = \mathbf{R}_3^T(\phi)$ and the covariance matrix simplified to

$$\text{cov}\left(\tilde{\mathbf{u}}_{\mathcal{R}}(\omega,t)\right) = \mathbf{R}_3(\phi)\mathbf{c}_{\omega}(\phi,\delta t)\mathbf{R}_3^T(\phi) \tag{2.87}$$

(c.f. SC 10,11,12)

The elements of $\mathbf{c}_{\omega}(\phi,\delta t)$ are

$$\begin{aligned}[c_{\omega}(\phi,\delta t)]_{ij} = &\int_{-\infty}^{\infty}\left(\mathbf{D}\mathbf{U}_{\mathcal{RS}}^{T}(\mathbf{u}_s(\omega,t))_{\mathcal{R}}\right)_i\left(\mathbf{D}\mathbf{U}_{\mathcal{RS}}^{T}(\mathbf{u}_s(\omega,t))_{\mathcal{R}}\right)_j dt \\ &-\left[\int_{-\infty}^{\infty}\left(\mathbf{D}\mathbf{U}_{\mathcal{RS}}^{T}(\mathbf{u}_s(\omega,t))_{\mathcal{R}}\right)_i dt\right]\left[\int_{-\infty}^{\infty}\left(\mathbf{D}\mathbf{U}_{\mathcal{RS}}^{T}(\mathbf{u}_s(\omega,t))_{\mathcal{R}}\right)_j dt\right]\end{aligned} \tag{2.88}$$

$\mathbf{D}_{ii}$ is a time shift operator

$$\mathbf{D}_{ii}\mathbf{u}_s(\omega,t) = \mathbf{u}_s(\omega,t+\Delta_i) \tag{2.89}$$

$$\Delta_i = \begin{cases} -\delta t/2 & i = 1 \\ +\delta t/2 & i = 2 \\ 0 & i = 3 \end{cases}$$

Therefore

$$\begin{aligned}[c_{\omega}(\phi,\delta t)]_{ij} = &\int_{-\infty}^{\infty}\left(\hat{\boldsymbol{\delta}}_i^T\mathbf{u}_s(\omega,t+\Delta_i)\right)\left(\hat{\boldsymbol{\delta}}_j^T\mathbf{u}_s(\omega,t+\Delta_j)\right) \\ &-\int_{-\infty}^{\infty}\left(\hat{\boldsymbol{\delta}}_i^T\mathbf{u}_s(\omega,t+\Delta_i)\right)dt\int_{-\infty}^{\infty}\left(\hat{\boldsymbol{\delta}}_j^T\mathbf{u}_s(\omega,t+\Delta_j)\right)dt\end{aligned} \tag{2.90}$$

(c.f. SC12)

Since we are integrating over all of $\mathbb{R}$ applying the change of variable technique yields

$$\int_{-\infty}^{\infty} a(t+\delta)dt = \int_{-\infty}^{\infty} a(t')dt' \tag{2.91}$$

where $t' = t + \Delta$ and $dt' = dt$ means it does not matter whether we integrate over t or t'. The diagonal terms can be simplified to

$$\begin{aligned}[c_{\omega}(\phi,\delta t)]_{ii} &= \int_{-\infty}^{\infty}\left(\hat{\boldsymbol{\delta}}_i^T\mathbf{D}_{ii}\mathbf{u}_s(\omega,t)\right)^2 dt - \left(\int_{-\infty}^{\infty}\hat{\boldsymbol{\delta}}_i^T\mathbf{D}_{ii}\mathbf{u}_s(\omega,t)dt\right)^2 \\ &= \int_{-\infty}^{\infty}\left(\hat{\boldsymbol{\delta}}_i^T\mathbf{u}_s(\omega,t)\right)^2 dt - \left(\int_{-\infty}^{\infty}\hat{\boldsymbol{\delta}}_i^T\mathbf{u}_s(\omega,t)dt\right)^2 \\ &= \text{var}(\hat{\boldsymbol{\delta}}_i^T\mathbf{u}_s(\omega,t))\end{aligned} \tag{2.92}$$

(c.f. SC 10,11)

and the off diagonal terms can be expressed as

$$\begin{aligned}[c_{\omega}(\phi,\delta t)]_{ij} = &\int_{-\infty}^{\infty}\left(\hat{\boldsymbol{\delta}}_i^T\mathbf{D}_{ii}\mathbf{u}_s(\omega,t)\right)\left(\hat{\boldsymbol{\delta}}_j^T\mathbf{D}_{jj}\mathbf{u}_s(\omega,t)\right)dt \\ &-\int_{-\infty}^{\infty}\left(\hat{\boldsymbol{\delta}}_i^T\mathbf{D}_{ii}\mathbf{u}_s(\omega,t)\right)dt\int_{-\infty}^{\infty}\left(\hat{\boldsymbol{\delta}}_j^T\mathbf{D}_{jj}\mathbf{u}_s(\omega,t)\right)dt\end{aligned} \tag{2.93}$$

This is the most general from of the covariance matrix. From here, we restrict ourselves to the special case of a vertically travelling wave i.e. $i = 0$ and the back azimuth β and θ are known.

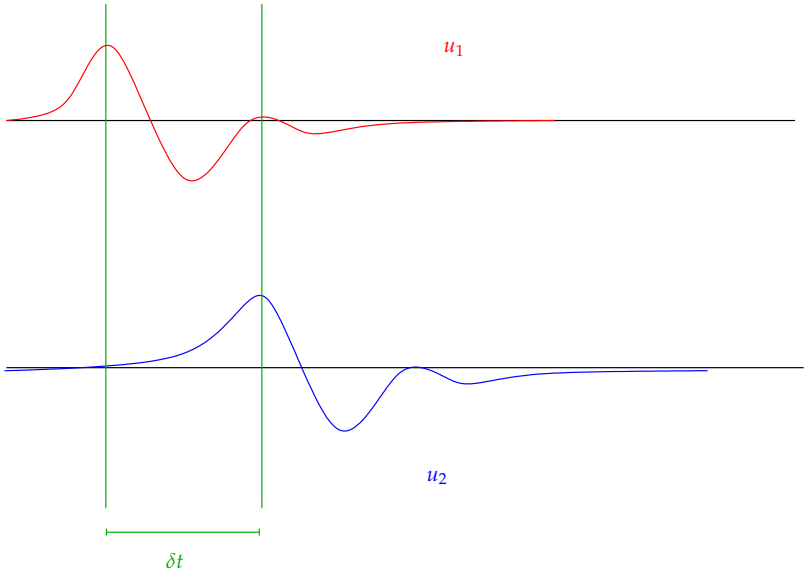
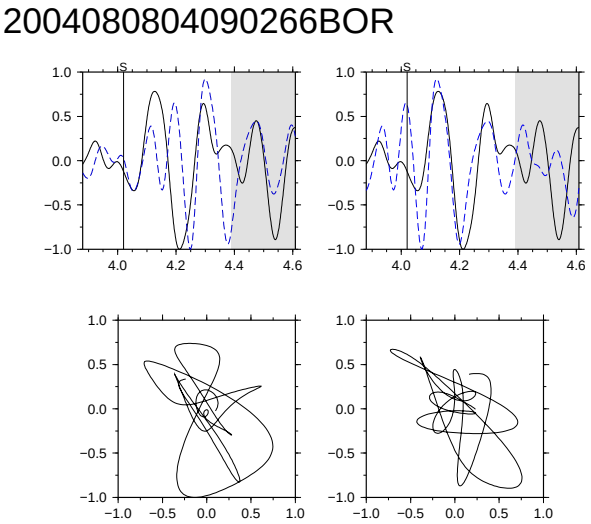


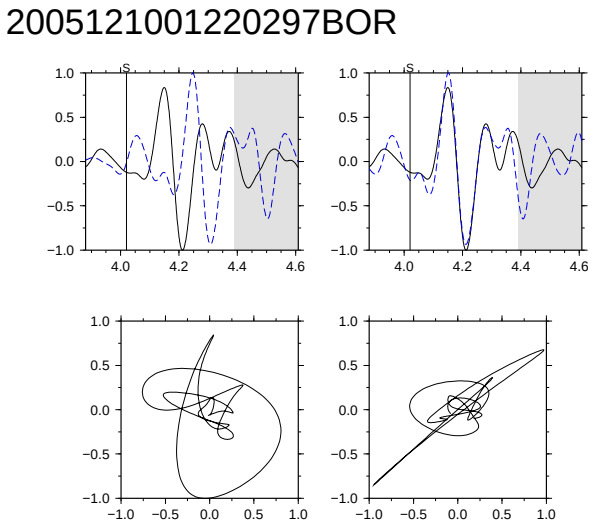
Figure 2.12: Two wave components u_1 and u_2 that only differ by a time shift. If the faster wave u_1 is shifted back by δt then the waves look similar and the covariance between the two waves is maximised.

Now that we have a definition for the elements of the covariance matrix we can provide a more detailed interpretation. The covariance matrix is used to find an inverse splitting operator that makes the two waveforms look similar (see Figure 2.12). In anisotropic cases where the shear wave has not been fully corrected (which is the case for all data from a station), the covariance matrix (in 2D i.e. ignoring the vertical component) has two non-zero eigenvalues λ_1 and λ_2 whilst in isotropic cases $\mathbf{c}$ will have only one non-zero eigenvalue λ_1. The first eigenvalue λ_1 is a measure of how well the two split waves match after correcting for splitting. A graph that helps illustrate this is the graph of particle motion. If we think of a small piece of earth as a particle, the graph of particle motion shows the movement (through the relative amplitudes) of the particle over time in particular directions. We originally start with East and North motion but we rotate by ϕ and time shift by δt to obtain $\mathcal{S}$ coordinates. Figure 2.13 shows movement in fast and slow directions with the amplitudes being normalised by the maximum value for each component. When the fast and slow components are out of phase the particle motion is elliptical. When they are in phase the particle motion is linear. The first column shows the uncorrected plots. These are the plots before the inverse splitting operator is applied. The second column shows the plots after the splitting operator has been applied using the pair of $(\phi, \delta t)$ with the smallest eigenvalue λ_2. Figure 2.13b has very linear particle motion so λ_1 will be large and λ_2 will be very small. On the other hand, Figure 2.13a still has elliptical particle motion so its λ_1 would be relatively small and its λ_2 would be relatively large.

When λ_1 is at its maximum the particles are almost linear. Therefore, to obtain the original wave before it split we wish to find the values of $(\phi, \delta t)$ via a grid search such that the



(a) Example where particle motion between the fast and slow components is not linear after applying the inverse splitting operator because of the significantly different amplitudes of the fast and slow components.



(b) Example where particle motion between the fast and slow components is linear after applying the inverse splitting operator. The waveforms are both in phase and the amplitudes are similar.

Figure 2.13: Two examples of particle motion. Top graphs show the fast and slow components. The top left plot, which is prior to correcting, shows a time lag between the two components. The top right graph, after rotating and time shifting one of the components should have both wave forms in phase. The bottom plots show the particle motion of the fast and slow components before and after correcting. The particle motion should be linear after correcting because the splitting has been reversed so the unsplit shear wave should be oriented along a single direction $\hat{\mathbf{p}}$.

covariance matrix **c** is singular, i.e.

$$\mathbf{c} = \begin{bmatrix} max(\lambda_1) & 0 \\ 0 & \lambda_2 = 0 \end{bmatrix} \tag{2.94}$$

In the 2D case, the trace of **c** does not vary with changes to either ϕ or δt so maximising λ_1 is equivalent to minimising λ_2. A proof of the invariance of trace(**c**) is given in Appendix A.4.

$$\begin{aligned} c &= \lambda_1 + \lambda_2 && \text{(Since the trace is invariant to changes in } \phi \text{ and } \delta t) \\ \max(\lambda_1) &= \min(c - \lambda_1) = \min(\lambda_2) && (\lambda_2 \text{ is at its smallest when } \lambda_1 \text{ is at its largest}) \end{aligned} \tag{2.95}$$

Silver and Chan minimise λ_2. Note that in the 3D case it is not possible to minimise λ_2 or λ_3 and instead we must find $\beta, i, \theta, \phi, \delta t$ such that λ_1 is maximised. This is because even though λ_2 or λ_3 could be minimised, there is no guarantee that λ_1 would be maximised because there are three eigenvalues. In 2D the value of $\min(\lambda_2)$ can be thought of as the residual variation. It is discussed more in the next section.

In reality we will not have a perfect situation like Figure 2.12. As mentioned in the caption, $u_1(t + \delta t) = u_2(t)$ so the energy will be all on the polarisation component after desplitting and consequently the covariance matrix will be singular. Figure 2.13 shows some more realistic waveforms. There will be residual noise along the null component so the variance on that component will be non-zero so **c** will not be singular. Instead we search for **c** that is closest to being singular. The second eigenvalue, λ_2, of **c** is used to determine the pair of ϕ and δt that best reverses the effects of splitting. The pair that produce the smallest λ_2 are the pair that get the waveforms to match the best and the particle motion to be as linear as possible (Figure 2.13b).

So far we have shown our derivations of the covariance matrix for the corrected and fully corrected waves and have provided an explanation of how this matrix determines the best splitting parameters using eigenvalues. However, the Silver and Chan (1991) equations and implementation of the covariance matrix are different. They are discussed more below.

Silver and Chan (1991) attempt to write a single line equation to represent the covariance matrix

$$\mathbf{c}_{ij}(\phi, \delta t) = \int_{-\infty}^{\infty} u_i(t) u_j(t - \delta t) dt \qquad i, j = 1, 2 \tag{2.96}$$

(SC 9)

However, if Equation 2.96 is written as a matrix it does not look like a covariance matrix

$$\begin{bmatrix} \int_{-\infty}^{\infty} u_1(t) u_1(t - \delta t) dt & \int_{-\infty}^{\infty} u_1(t) u_2(t - \delta t) dt \\ \int_{-\infty}^{\infty} u_2(t) u_1(t - \delta t) dt & \int_{-\infty}^{\infty} u_2(t) u_2(t - \delta t) dt \end{bmatrix} \tag{2.97}$$

It is likely that they have tried to write the covariance in a more compact form, which as a result looks incorrect. The covariance matrix is a symmetric matrix that gives the amount of dispersion between the the two waves on the off diagonal and the variance of each wave along the diagonal. To meet the definition of the covariance matrix using the formula given

by Silver and Chan (1991) would require that the off diagonal components were equal, which may not necessarily be the case especially for more complex waveforms. The variance should be the square of a variable but the formula in the Silver and Chan (1991) looks more like an autocovariance (e.g. $\int_{-\infty}^{\infty} u_1(t)u_1(t-\delta t)dt$). The covariance matrix defined in Equation 2.96 is in fact a cross covariance which is not a symmetric matrix (Brockwell and Davis, 2006). We suggest that the covariance matrix should be defined as follows

$$\begin{aligned}\mathbf{c}(\phi,\delta t) &= Cov[(u_1(t),u_2(t-\delta t)] \\ &= \begin{bmatrix} \int_{-\infty}^{\infty} u_1^2(t)dt & \int_{-\infty}^{\infty} u_1(t)u_2(t-\delta t)dt \\ \int_{-\infty}^{\infty} u_1(t)u_2(t-\delta t)dt & \int_{-\infty}^{\infty} u_2^2(t-\delta t)dt \end{bmatrix} \\ &= \begin{bmatrix} \int_{-\infty}^{\infty} u_1^2(t+\delta t/2)dt & \int_{-\infty}^{\infty} u_1(t+\delta t/2)u_2(t-\delta t/2)dt \\ \int_{-\infty}^{\infty} u_1(t+\delta t/2)u_2(t-\delta t/2)dt & \int_{-\infty}^{\infty} u_2^2(t-\delta t/2)dt \end{bmatrix} \end{aligned} \quad (2.98)$$

(c.f. SC 9)

To check whether Equation 2.96 was an error we compared the formula in the article (Table 2.2) to the original code from the Silver and Chan method (Silver and Chan (1991)) as well as two other popular splitting programs using their methodology; Mfast (Teanby et al. (2004*a*), Wessel (2010), Savage et al. (2010)) and Splitlab (Wustefeld et al. (2008)). The covariance matrix in Splitlab matches the form we described in Equation 2.98. However, Fortran based codes such as the original codes and Mfast have a subtle difference.

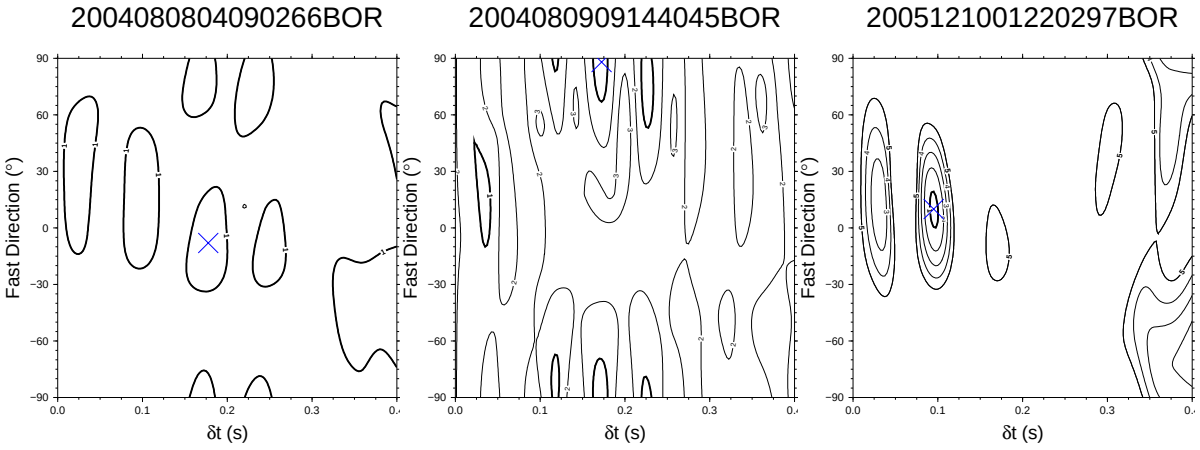
Recall the definition of the covariance is

$$Cov(X,Y) = \left\langle [X - \langle X\rangle][Y - \langle Y\rangle)] \right\rangle = \langle XY\rangle - \langle X\rangle\langle Y\rangle$$

The Fortran based codes use the following to calculate the covariance matrix

```
 c  ** calc cross correlation matrix **
        cov(1,1) = 0.
        cov(2,2) = 0.
        cov(1,2) = 0.
        cov(2,1) = 0.
        do 1 i=1,n
           cov(1,1) = cov(1,1) + x(i)**2
           cov(2,2) = cov(2,2) + y(i)**2
           cov(1,2) = cov(1,2) + x(i)*y(i)
1       continue
        cov(2,1) = cov(1,2)
```

What this code shows is that the second term in Equation 2.77 has not been subtracted off any of the entries, meaning they are assuming that the means for each component are zero. Most preprocessing involves removing the mean of the entire time series. However, the data used for analysis is the windowed time series around the S wave. The mean of the windowed data are usually not zero even if the entire time series has been demeaned. We suggest that any program that calculates the covariance matrix this way be modified so that it matches the form in Equation 2.77. Alternatively, some programming languages, such as Matlab and R, have a built in covariance function that should be used.



(a) Non-contiguous (b) Region affected by angle cyclicity (c) Contiguous

Figure 2.14: Illustration of three different confidence regions. Minimum is denoted by the cross and the 95% confidence region by the bolded lines.

2.7 Confidence Region

From here on we restrict ourselves to the special case where $i = 0$. The back azimuth and θ are known, therefore making this a 2 dimensional problem. Once the pair of ϕ and δt that lead to the smallest second eigenvalue have been identified, a measure of how accurate these estimates are can be calculated. This is done by calculating the standard errors of the estimates $\widehat{\phi}$ and $\widehat{\delta t}$. Silver and Chan (1991) do this by calculating a quarter of the width and length of the confidence region. Figure 2.14 shows examples of confidence regions for three different events. For each pair of ϕ and δt the second eigenvalue of the covariance matrix is determined and contoured to form a set of contour lines. The smallest λ_2 is marked by a cross. The 95% confidence region is indicated by a thick black line. In order to mark in the confidence region we define the sum of squares function $S(\phi, \delta t)$ which is constructed from $\lambda_2^{min}(\phi, \delta t)$ for S phases or the minimum energy on the transverse component for SKS waves. Once the ideal parameters have been identified by one of these methods and the wave has been desplit, the windowed components are rotated so that one contains the signal one contains the noise. This is discussed more in subsequent sections (particularly Section 5.2). After this process we obtain an n-point time series. The minimum value of this time series, $S_{min}(\phi, \delta t)$ is needed to construct the confidence region. It is assumed that

Assumption 1: $S_{min}(\phi, \delta t) \sim \chi^2_n$ (chi-squared distribution) where the degrees of freedom n is equal to the number of data points.

The chi-squared distribution is discussed more in Appendix A.5. This assumption would be true for Gaussian white noise, i.e. if we have n independent and identically distributed (iid) disturbances to the signal giving us the observed data. With this assumption we obtain two, independent, central chi-squared variables. The construction of the confidence region

is outlined in Jenkins and Watts (1968). The easiest way to derive this is to compare to ANOVA. ANOVA is a statistical model that tests whether at least one of the k group means is different from an overall mean that ignores groupings. This is done by looking at different sources of variation. These sources are group variation known as the sum of squares for the groups (SSG) and inherent variation known as the sum of squared errors (SSE). Adding them up gives the total sum of squares (SST). The ratio of SSG and SSE divided by their corresponding degrees of freedom yields an F statistic

$$F = \frac{SSG/(k-1)}{SSE/[(n-1)-(k-1)]} = \frac{SSG/(k-1)}{SSE/[n-k]} \tag{2.99}$$

The test statistic F is a ratio of two, independent, central chi-squared variables that have degrees of freedom $k-1$ and $n-k$ respectively. Therefore F has an F distribution with parameters $k-1$ and $n-k$ (Clark and Randal, 2011).

This can be derived as follows

$$\begin{aligned}
\frac{SSG/(k-1)}{SSE/(n-k)} &= \left(\frac{n-k}{k-1}\right)\left(\frac{SSG}{SSE}\right) \sim F_{k-1,n-k} \\
&\Leftrightarrow \left(\frac{n-k}{k-1}\right)\left(\frac{SST-SSE}{SSE}\right) \sim F_{k-1,n-k} && \text{since } SSG + SSE = SST \\
&\Leftrightarrow \underbrace{\left(\frac{n-k}{k-1}\right)\left(\frac{SST}{SSE}-1\right)}_{\text{F}} \sim F_{k-1,n-k} && (2.100)
\end{aligned}$$

where

$$\begin{aligned}
k &= \text{number of parameters excluding the variance parameter} \\
n &= \text{number of data points}
\end{aligned}$$

For a $(1-\alpha)$ confidence region we can obtain the critical value $F = F^{1-\alpha}_{k-1,n-k}$. Substituting this critical value into Equation 2.100 gives

$$\begin{aligned}
F &= F^{1-\alpha}_{k-1,n-k} \\
\left(\frac{n-k}{k-1}\right)\left(\frac{SST}{SSE}-1\right) &= F^{1-\alpha}_{k-1,n-k} \\
\frac{SST}{SSE} &= 1 + \frac{k-1}{n-k}F^{1-\alpha}_{k-1,n-k} && (2.101)
\end{aligned}$$

In ANOVA the SSG is a sum of squares excluding the overall mean so it has $k-1$ degrees of freedom. Likewise SST is a corrected sum of squares total excluding the overall mean so it has $n-1$ degrees of freedom. This implies the degrees of freedom for $SSE = SST - SSG$ are $(n-1)-(k-1) = n-k$. However, in this case we do not use the corrected sum of squares, so the corresponding degrees of freedom for this case are k, n and $n-k$ respectively.

Our parameter λ_2^{min} can be thought of as a measure of the variation inherent in the system like SSE whilst λ_2 (which is determined by ϕ and δt) can be thought of as the SSG. If we

substitute these into Equation 2.101, the confidence region is defined as

$$\frac{\lambda_2}{\lambda_2^{min}} \leq 1 + \frac{k}{n-k} F_{k,n-k}^{1-\alpha} \tag{2.102}$$

where

$$\begin{aligned} k &= \text{number of parameters, which is 2 in this case } (\phi, \delta t) \\ n &= \text{number of data points} \\ F_{k,n-k}^{1-\alpha} &= \text{inverse } F(k, n-k) \text{ distribution i.e } \mathrm{P}(F < F_{k,n-k}^{1-\alpha}) = 1-\alpha \text{ for } F \sim F_{k,n-k} \\ \lambda_2 &= \text{second eigenvalue of the covariance matrix} \end{aligned}$$

Any point below this value will be inside the $(1-\alpha)$ confidence region.

A value of interest is the boundary of the confidence region $\lambda_2^{(1-\alpha)CR}(\alpha)$. The $1-\alpha$ confidence regions is

$$(1-\alpha)\mathrm{CR} = \{(\phi, \delta t) : \lambda_2 \leq \lambda_2^{(1-\alpha)CR}(\alpha)\} \tag{2.103}$$

where

$$\lambda_2^{(1-\alpha)CR}(\alpha) = \lambda_2^{min}\left[1 + \frac{k}{n-k} F_{k,n-k}^{1-\alpha}\right] \tag{2.104}$$

This $\lambda_2^{(1-\alpha)CR}$ value is often used to normalise the contour plot, which becomes a plot of

$$g(\phi, \delta t) = \frac{\lambda_2(\phi, \delta t)}{\lambda_2^{95CR}(\alpha)} \qquad (\text{for } \alpha = 0.05) \tag{2.105}$$

so that 1 corresponds to the boundary of the 95% confidence region and any number less than one is inside the confidence region. If we assume a normal distribution then the confidence region length and width is twice the margin of error where the margin of error is $1.96 \times SE$ (SE is the standard error). Therefore, to obtain approximate standard errors for ϕ and δt we can say that these are roughly a quarter of the length and width of the ellipse enclosing the minimum. Determining the length and width for a contiguous region like Figure 2.14c can be done with a few lines of code, but problems occur for more complex regions like the ones in Figures 2.14a and 2.14b. The fast direction has a 180° ambiguity, meaning that an angle that is 180° away is in fact one and the same angle. This means for example that -90° and 90° are the same. The cyclicity of the fast direction can make it difficult to calculate the length of the confidence region. Silver and Chan (1991) acknowledge this in their codes (sic)

```
c  ** finding min max i is more difficult because of cyclicity of angles **
c  ** sweep a line over all j, set point on line equal to 1 if it falls
within the 95% convidence contour for any j. The height of the bounding
rectangle is defined by the shortest line which includes all points with a
value of line(i)=1. This line is found by searching all line lengths from
the minimum = sum_i linr(i) to maximum = np1**
```

However what this means is that for Figure 2.14a the fast direction error is reported as ±45° (quarter the length of the grid) since the shortest range of ϕ that includes all the bolded

points is the entire length of the grid, not the length of the ellipse enclosing the minimum. It is possible to define the errors this way, but the main problem with such a method is consistency. The delay time errors for Figures 2.14a and 2.14b are reported as a quarter of the width of the ellipse and not a quarter of the width of the grid. We suggest that both errors be a quarter of the width and length of the ellipse.

2.8 Degrees of Freedom

In order to construct the confidence region, the degrees of freedom must be calculated. k will always be 2 because we have 2 parameters, ϕ and δt. If we use Assumption 1 then the other degrees of freedom is the number of points in the time series, n. If Assumption 1 does not hold and $S(\phi, \delta t)$ does not have a chi-squared distribution then changes are needed and a new assumption is made by Silver and Chan (1991)

Assumption 2: If $S(\phi, \delta t)$ is not χ^2_n distributed assume $S(\phi, \delta t) \sim \chi^2_\nu$ where the degrees of freedom (ν) depend upon the noise spectrum and instrument response.

To obtain an estimate of the actual degrees of freedom, ν, we must first define the power, E, which integrates the squared modulus of a signal.

$$E = \frac{T}{2} \int_{-\infty}^{\infty} \eta^2(t) dt = \frac{T}{2} \int_{-\infty}^{\infty} |\eta(t)|^2 \tag{2.106}$$

(SC A1)

Here the signal is denoted $\eta(t)$ and is a finite sample of the noise process of length T. The noise is obtained by looking at the signal of the shear wave as well as a small segment before and after the shear wave on the fast and slow components. This section of the data are called the (analysis) window and has a time length of T. The desplit components in splitting coordinates are then rotated by $(\phi - \alpha)$ so that one component contains the signal and the other component contains the noise, $\eta(t)$. This process is outlined in more detail in Section 5.2.

The next assumption we use is

Assumption 3: After rotating and shifting the noise component, $\eta(t)$ may be written as the convolution of a filter $f(t)$ and Gaussian white noise $g(t)$.

The Convolution Theorem takes two functions and creates a third which is a combination of the two.

Theorem 2.8.1. [*Convolution Theorem*] *A convolution in the time domain may be defined as*

$$\eta(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(y)g(t-y)dy \equiv f * g \tag{2.107}$$

It can be shown that in the frequency domain this convolution is

$$\tilde{\eta}(\omega) = [f \tilde{*} g](\omega) = \tilde{f}(\omega)\tilde{g}(\omega) \tag{2.108}$$

where

$$\tilde{f}(\omega) \equiv FT[f(t)](\omega) \tag{2.109}$$

The $\sim$ above a symbol denotes the Fourier transform, which is a mathematical function that converts a function of time into a function in the frequency domain. One can also convert a function in the frequency domain back into the time domain using the inverse Fourier transform. The transform and the inverse transform are defined respectively as

$$\begin{aligned} \tilde{\eta}(\omega) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \eta(t) \exp[-i\omega t]dt \\ \eta(t) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \tilde{\eta}(\omega) \exp[i\omega t]d\omega \end{aligned} \tag{2.110}$$

or if the data are discrete they are defined as

$$\begin{aligned} \tilde{\eta}(\omega) &= \sum_{n=0}^{N} \eta(t_n) \exp[-i\omega t_n] \\ \eta(t) &= c \sum_{n=0}^{N} \tilde{\eta}(\omega_n) \exp[i\omega t_n] \end{aligned} \tag{2.111}$$

where c is a constant factor.

A result related to Fourier transforms that is useful to us is Parseval's Theorem. The Theorem has many different forms but we choose to represent it as follows

Theorem 2.8.1. [*Parseval's Theorem*] *For two functions $f(t)$ and $g(t)$ where $f(t) \neq g(t)$*

$$\int_{-\infty}^{\infty} |\tilde{f}(\omega)||\tilde{g}^*(\omega)|d\omega = \int_{-\infty}^{\infty} |f(y)||g^*(t-y)|dy \tag{2.112}$$

here $*$ denotes the complex conjugate.

For $f(t) = g(t)$

$$\int_{-\infty}^{\infty} |\tilde{f}(\omega)|^2 d\omega = \int_{-\infty}^{\infty} |f(t)|^2 dt \tag{2.113}$$

With these theorems and Assumption 3 we can rewrite the power as

$$\begin{aligned} E &\approx \int_{-\infty}^{\infty} |\eta^2(t)| dt \\ &= \int_{-\infty}^{\infty} |(f * g)(t)|^2 dt && \text{Assumption 3} \\ &= \int_{-\infty}^{\infty} |f \tilde{*} g(\omega)|^2 d\omega && \text{Parseval's Theorem} \\ &= \int_{-\infty}^{\infty} |\tilde{f}(\omega)\tilde{g}(\omega)|^2 d\omega && \text{Convolution Theorem} \\ &= \int_{-\infty}^{\infty} |\tilde{f}(\omega)|^2 |\tilde{g}(\omega)|^2 d\omega \end{aligned} \tag{2.114}$$

In all practical experimental cases $\eta(t)$ is in the form of discretely sampled data. Taking a discrete sample of a continuous process causes a loss of information in the form of aliasing. Aliasing is when two different signals become indistinguishable because of the way they were sampled. For example the two signals in Figure 2.15 are indistinguishable based on the current set of sample points. It would be impossible to reconstruct the black curve using this sample. For a given sampling frequency f_s, a waveform at frequency f can be reconstructed if

$$f \leq (1/2) f_s \tag{2.115}$$

In other words, it can be perfectly reconstructed providing the frequency of the waveform does not exceed half the sampling rate. The right hand side of the inequality is referred to as the Nyquist frequency. Imagine we have discretely sampled data comprising a sample

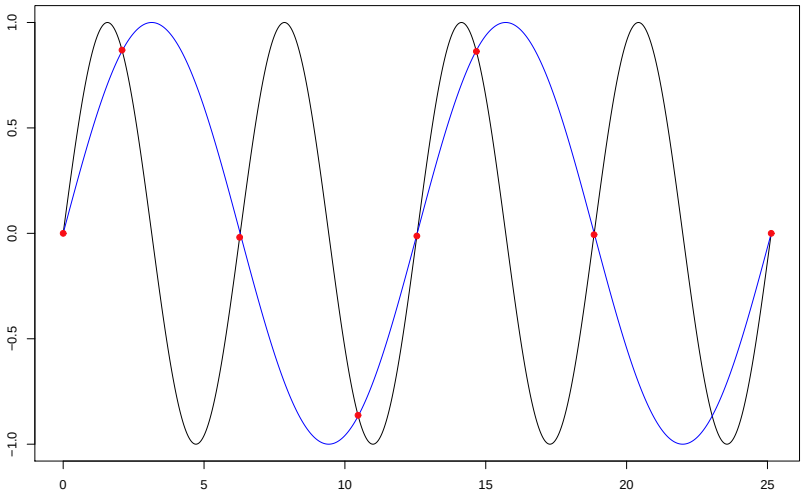


Figure 2.15: Example of aliasing. If we sample only at the red circles then the two waves become indistinguishable.

of $N+1$ data points up to the Nyquist frequency over the interval $[0, T]$ with a sampling frequency of $\frac{N}{(T-0)}$ and each point evenly spaced by $\frac{(T-0)}{(N)}$ as illustrated in Figure 2.16.

For $\eta(t)$ at points $\eta(t_0), \eta(t_1), ..., \eta(t_N)$ we can use a discrete Fourier transform to obtain

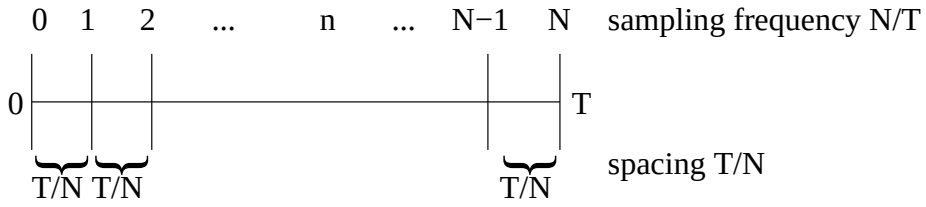


Figure 2.16: Illustration of sampling a continuous process with even spacing over the interval.

$\tilde{\eta}(\omega)$. E can then be approximated (up to a multiplicative constant) by

$$E \approx \sum_{n=1}^{N-1} |\tilde{f}_n|^2 |\tilde{g}_n|^2 + \frac{1}{2}(|\tilde{f}_0|^2 |\tilde{g}_0|^2 + |\tilde{f}_N|^2 |\tilde{g}_N|^2) \tag{2.116}$$

(SC A3)

which uses the trapezoidal rule from numerical integration (Appendix A.6). If we assume Assumption 4

Assumption 4: E is approximately distributed as a χ^2_ν scaled by a constant.

then the degrees of freedom can be calculated as follows. The expected value and variance of a scaled central chi-squared variable $Y \sim a\chi^2_\nu$ are

$$\begin{aligned} \langle Y \rangle &= a\nu \\ \text{var}(Y) &= 2a^2\nu \end{aligned} \tag{2.117}$$

where $\langle E \rangle$ is the expected value of E and $\text{var}(E)$ is the variance of E. This can be derived using moments. Rearranging the two Equations in 2.117 we can obtain method of moments estimators for a and ν respectively

$$\begin{aligned} a &= \langle Y \rangle / \nu \\ \nu &= \text{var}(Y)/2a^2 \\ &= \text{var}(Y)/2(\langle Y \rangle / \nu)^2 \\ \nu/\nu^2 &= \text{var}(Y)/2\langle Y \rangle^2 \\ 1/\nu &= \text{var}(Y)/2\langle Y \rangle^2 \\ \nu &= 2\langle Y \rangle^2 / \text{Var}(Y) \end{aligned} \tag{2.118}$$

Therefore the degrees may be estimated using the following formula

$$\nu = \frac{2\langle E \rangle^2}{\text{var}(E)} \tag{2.119}$$

(SC A4)

More specific details can be found in Jenkins and Watts (1968) .

The expected value of the power E can be calculated as follows if we make the following assumption

Assumption 5: $g_n \sim N(0, \sigma^2)$ i.e. g_n follows a normal distribution with mean 0 and variance σ^2.

Since g_n has a zero mean the second moment can be calculated quite easily. The variance and mean are related by the equation

$$\mathrm{Var}(X) = \langle X^2 \rangle - \langle X \rangle^2 \tag{2.120}$$

If the mean $\langle X \rangle$ is zero we have

$$\mathrm{Var}(X) = \langle X^2 \rangle - 0^2 = \langle X^2 \rangle \tag{2.121}$$

Starting with Equation 2.116 and using Equation 2.121 we can calculate the expected value of the power as

$$\langle E \rangle = \left\langle \sum_{n=1}^{N-1} |\tilde{f}_n|^2 |\tilde{g}_n|^2 + \frac{1}{2}(|\tilde{f}_0|^2 |\tilde{g}_0|^2 + |\tilde{f}_N|^2 |\tilde{g}_N|^2) \right\rangle$$

Since $\tilde{f}$ is a known filter there is no random variation associated with it so it can be regarded as a constant and move outside the expectation. On the other hand, $\tilde{g}$ is the noise signal so it is a stochastic process with a second moment equal to the variance, σ^2. Therefore,

$$\begin{aligned}
\langle E \rangle &= \sum_{n=1}^{N-1} |\tilde{f}_n|^2 \langle |\tilde{g}_n|^2 \rangle + \frac{1}{2}(|\tilde{f}_0|^2 \langle |\tilde{g}_0|^2 \rangle + |\tilde{f}_N|^2 \langle |\tilde{g}_N|^2 \rangle) \\
&= \sum_{n=1}^{N-1} |\tilde{f}_n|^2 \sigma^2 + \frac{1}{2}(|\tilde{f}_0|^2 \sigma^2 + |\tilde{f}_N|^2 \sigma^2) \qquad \text{since } \tilde{g}_n \sim N(0, \sigma^2), Var(\tilde{g}) = E(\tilde{g}^2) = \sigma^2 \\
&= \sigma^2 \left[\sum_{n=1}^{N-1} |\tilde{f}_n|^2 + \frac{1}{2}(|\tilde{f}_0|^2 + |\tilde{f}_N|^2) \right] \qquad \text{rearranging} \\
&\equiv \sigma^2 F_2
\end{aligned} \tag{2.122}$$

(SC A5)

where σ^2 is the variance of the noise process and F_2 is the second moment of the signal F. In order to write these expectations in a more compact form we define the following coefficients for the trapezoidal rule of integration

$$a_n = \begin{cases} 1/2 & n = 0, N \\ 1 & \text{otherwise} \end{cases}$$

so that

$$E = \sum_{n=0}^{N} a_n |\tilde{f}_n|^2 |\tilde{g}_n|^2 \tag{2.123}$$

then

$$\begin{aligned}\langle E\rangle &= \sum_{n=0}^{N} a_n|\tilde{f}_n|^2\langle|\tilde{g}_n|^2\rangle \\ &= \sum_{n=0}^{N} a_n|\tilde{f}_n|^2\sigma^2 \\ &= \sigma^2\left[\sum_{n=0}^{N} a_n|\tilde{f}_n|^2\right] \\ &\equiv \sigma^2 F_2\end{aligned} \tag{2.124}$$

and then the second moment of the power can be written as

$$\langle E^2\rangle = \left\langle \sum_{n=0}^{N} a_n|\tilde{f}_n|^2|\tilde{g}_n|^2 \sum_{m=0}^{N} a_m|\tilde{f}_m|^2|\tilde{g}_m|^2 \right\rangle \tag{2.125}$$

The easiest way to evaluate this is to split $\langle E^2\rangle$ into two cases. One where $m = n$ and the other where $m \neq n$

$$\begin{aligned}\langle E^2\rangle &= \langle \sum_{n=0}^{N}\sum_{m=0}^{N} a_n|\tilde{f}_n|^2|\tilde{g}_n|^2 a_m|\tilde{f}_m|^2|\tilde{g}_m|^2\rangle \\ &= \sum_{n=0}^{N}\sum_{m\neq n}^{N} a_n|\tilde{f}_n|^2 a_m|\tilde{f}_m|^2\langle|\tilde{g}_n|^2|\tilde{g}_m|^2\rangle + \sum_{n=0}^{N} a_n^2|\tilde{f}_n|^4\langle|\tilde{g}_n|^4\rangle\end{aligned} \tag{2.126}$$

For independent $\tilde{g}_n$ and $\tilde{g}_m$ (i.e. when $m \neq n$) we have

$$\langle|\tilde{g}_n|^2|\tilde{g}_m|^2\rangle = \langle|\tilde{g}_n|^2\rangle\langle|\tilde{g}_n|^2\rangle = \sigma^2 \times \sigma^2 \tag{2.127}$$

based on the reasoning explained in Equation 2.121, substituting this into Equation 2.126 gives

$$\langle E^2\rangle = \sum_{n=0}^{N}\sum_{m\neq n}^{N} a_n|\tilde{f}_n|^2 a_m|\tilde{f}_m|^2\sigma^2\sigma^2 + \sum_{n=0}^{N} a_n^2|\tilde{f}_n|^4\langle|\tilde{g}_n|^4\rangle \tag{2.128}$$

Equation 2.128 requires the 4th moment, $\langle|\tilde{g}_n|^4\rangle$, to be evaluated. In order to calculate this the following assumption is made

Assumption 6: The real and imaginary parts of $\tilde{g}_n$ are independent, identical and normally distributed.

The fourth moment of a standard normal variable is

$$E(z^4) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z^4 e^{-z^2/2} dz \tag{2.129}$$

The 4th moment of a normal variable, $\langle |\tilde{g}_n|^4 \rangle$ can be evaluated by calculating the 4th moment of a standard normal variable $z \sim N(0,1)$ then scaling up by the constant σ^4. To obtain this Equation 2.129 must be evaluated using integration by parts i.e.
$\int u\, dv = uv - \int v\, du$ where

$$\begin{aligned} u &= \frac{1}{\sqrt{2\pi}} z^3 & v &= -e^{-z^2/2} \\ du &= \frac{1}{\sqrt{2\pi}} 3z^2 dz & dv &= z e^{-z^2/2} dz \end{aligned}$$

Solving this leads to

$$\begin{aligned} \int u\, dv &= uv - \int v\, du \\ \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} z^3 z e^{-z^2/2} dz &= \underbrace{-\frac{1}{\sqrt{2\pi}} z^3 e^{-z^2/2} \Big|_{-\infty}^{\infty}}_{=0} - \int_{-\infty}^{\infty} -e^{-z^2/2} \frac{1}{\sqrt{2\pi}} 3z^2 dz \\ &= 3 \underbrace{\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} z^2 e^{-z^2/2} dz}_{\langle z^2 \rangle = Var(z) = 1 \text{ (Eq 2.121)}} \\ &= 3 \\ E[(\sigma z)^4] &= \sigma^4 E[z^4] \\ &= 3\sigma^4 \\ \text{and so } \langle |\tilde{g}_n|^4 \rangle &= 3\sigma^4 \end{aligned} \tag{2.130}$$

Equation 2.130 can be substituted into Equation 2.128 as follows

$$\begin{aligned} \langle E^2 \rangle &= \sum_{n=0}^{N} \sum_{m \neq n}^{N} a_n |\tilde{f}_n|^2 a_m |\tilde{f}_m|^2 \sigma^2 \sigma^2 + \sum_{n=0}^{N} a_n^2 |\tilde{f}_n|^4 \langle |\tilde{g}_n|^4 \rangle \\ &= \sum_{n=0}^{N} \sum_{m \neq n}^{N} a_n |\tilde{f}_n|^2 a_m |\tilde{f}_m|^2 \sigma^2 \sigma^2 + \sum_{n=0}^{N} a_n^2 |\tilde{f}_n|^4 (3\sigma^4) \end{aligned} \tag{2.131}$$

If we wish to adjust this so that both sums for the first term start at zero then the terms where $m = n$ will have been counted twice so they must be subtracted off as follows

$$\begin{aligned}
\langle E^2 \rangle &= \sum_{n=0}^{N} \sum_{m=0}^{N} a_n |\tilde{f}_n|^2 a_m |\tilde{f}_m|^2 \sigma^2 \sigma^2 - \sum_{n=0}^{N} a_n^2 |\tilde{f}_n|^4 \sigma^4 + 3 \sum_{n=0}^{N} a_n^2 |\tilde{f}_n|^4 \sigma^4 \\
&= \sum_{n=0}^{N} \sum_{m=0}^{N} a_n |\tilde{f}_n|^2 a_m |\tilde{f}_m|^2 \sigma^2 \sigma^2 + \sum_{n=0}^{N} 2a_n^2 |\tilde{f}_n|^4 \sigma^4 \\
&= \left(\sum_{n=0}^{N} a_n |\tilde{f}_n|^2 \sum_{m=0}^{N} a_m |\tilde{f}_m|^2 \right) \sigma^4 + \sum_{n=0}^{N} 2a_n^2 |\tilde{f}_n|^4 \sigma^4 \\
&= \sigma^4 \left(\sum_{n=0}^{N} a_n |\tilde{f}_n|^2 \right)^2 + \sigma^4 \left(\sum_{n=0}^{N} 2a_n^2 |\tilde{f}_n|^4 \right) \\
&= \sigma^4 (F_2^2 + F_4) \qquad (2.132)
\end{aligned}$$

where

$$\left(\sum_{n=0}^{N} a_n |\tilde{f}_n|^2 \right)^2 \equiv F_2^2 \qquad \text{is the second squared moment of } F \qquad (2.133)$$

$$\left(\sum_{n=0}^{N} 2a_n^2 |\tilde{f}_n|^4 \right) \equiv F_4 \qquad \text{is the fourth moment of } F \qquad (2.134)$$

If we replace a_n with their appropriate values we obtain

$$\begin{aligned}
\langle E^2 \rangle &= \underbrace{\left(\left[\frac{1}{2} |\tilde{f}_0|^2 + \sum_{n=1}^{N-1} |\tilde{f}_n|^2 + \frac{1}{2} |\tilde{f}_N|^2 \right] \right)^2}_{F_2^2} \sigma^4 \\
&\quad + \underbrace{\left(\left[\frac{1}{2} |\tilde{f}_0|^4 + \sum_{n=1}^{N-1} 2|\tilde{f}_n|^4 + \frac{1}{2} |\tilde{f}_N|^4 \right) \right]}_{F_4} \sigma^4 \\
&= \sigma^4 (F_2^2 + F_4) \qquad (2.135)
\end{aligned}$$

(SC A6)

The F_4 defined in Silver and Chan (1991) *(SC A8)* is very similar except their coefficient for $|f_n|^4$ is 1 not 2 (Equation 2.139). This is discussed further in Section 2.9.

Equation 2.135 implies the variance of E is

$$\begin{aligned}
\text{Var}(E) &= \langle E^2 \rangle - \langle E \rangle^2 \\
&= \sigma^4 (F_2^2 + F_4) - (\sigma^2 F_2)^2 \\
&= \sigma^4 F_4 \qquad (2.136)
\end{aligned}$$

(SC A7)

Placing this and Equation 2.122 in Equation 2.119 gives the following expression for the degrees of freedom

$$\nu = 2\frac{(\sigma^2 F_2)^2}{\sigma^4 F_4} = 2\frac{F_2^2}{F_4} \tag{2.137}$$

(SC A9)

2.9 Estimating the Quantities Needed to Calculate the Degrees of Freedom

In order to calculate the degrees of freedom using Equation 2.137 (*SC A9*), F_2^2 and F_4 must be estimated. These quantities can be obtained directly from the data using the observed values of E^2 and E_4. Remember, we defined F_4 as

$$F_4 = \sum_{n=1}^{N-1} 2|\tilde{f}_n|^4 + (1/2)(|\tilde{f}_0|^4 + |\tilde{f}_N|^4) \tag{2.138}$$

whereas Silver and Chan (1991) defined F_4 as

$$F_4^{(SC)} = \sum_{n=1}^{N-1} 1|\tilde{f}_n|^4 + (1/2)(|\tilde{f}_0|^4 + |\tilde{f}_N|^4) \tag{2.139}$$

(SC A8)

To determine the correct coefficients we define E_4 with unknown coefficients b_n

$$E_4 = \sum_{n=0}^{N} b_n|\tilde{f}_n|^4|\tilde{g}_n|^4 \tag{2.140}$$

The expected value of E_4 according to Silver and Chan (1991) is

$$\langle E_4 \rangle = 2\sigma^4 F_4 \tag{2.141}$$

(SC A11)

Using our definition of F_4 (Equation 2.138) we can solve for the unknown coefficients b_n to ensure that Equation 2.141 (*SC A11*) holds.

$$\begin{aligned}
\langle E_4 \rangle &= \sum_{n=0}^{N} b_n|\tilde{f}_n|^4\langle|\tilde{g}_n|^4\rangle \\
&= \sigma^4 \sum_{n=0}^{N} 3b_n|\tilde{f}_n|^4 && \text{(Using Equation 2.130)} \\
c.f. &= 2\sigma^4 F_4 && \text{(SC A11 definition)} \\
&= 2\sigma^4 \sum_{n=0}^{N} 2a_n^2|\tilde{f}_n|^4 && \text{(Using our definition of } F_4 \text{ derived in Equation 2.132)}
\end{aligned} \tag{2.142}$$

Using the expected value of E_4 from Silver and Chan (1991) along with our definition of E_4 (Equation 2.138) implies that

$$\begin{aligned}
\langle E_4 \rangle &= 2\sigma^4 F_4 \\
\Rightarrow \sigma^4 \sum_{n=0}^{N} 3b_n |\tilde{f}_n|^4 &= 2\sigma^4 \sum_{n=0}^{N} 2a_n^2 |\tilde{f}_n|^4 \\
3\left(\sigma^4 \sum_{n=0}^{N} b_n |\tilde{f}_n|^4\right) &= 4\left(\sigma^4 \sum_{n=0}^{N} a_n^2 |\tilde{f}_n|^4\right) \\
3b_n &= 4a_n^2 \\
b_n &= \frac{4}{3} a_n^2
\end{aligned} \tag{2.143}$$

The a_n^2 coefficients are

$$a_n^2 = \begin{cases} 1/4 & n = 0, N \\ 1 & \text{otherwise} \end{cases}$$

which means the b_n are

$$b_n = \begin{cases} 1/3 & n = 0, N \\ 4/3 & \text{otherwise} \end{cases}$$

Placing this back in Equation 2.140 gives

$$E_4 = \sum_{n=1}^{N-1} (4/3) |\tilde{f}_n|^4 |\tilde{g}_n|^4 + 1/3(|\tilde{f}_0|^4 |\tilde{g}_0|^4 + |\tilde{f}_N|^4 |\tilde{g}_N|^4) \tag{2.144}$$

(c.f. SC A10)

which differs from the Silver and Chan (1991) definition given in the article (Equation 2.145)

$$E_4^{(SC)} = \sum_{n=1}^{N-1} 1 |\tilde{f}_n|^4 |\tilde{g}_n|^4 + 1/3(|\tilde{f}_0|^4 |\tilde{g}_0|^4 + |\tilde{f}_N|^4 |\tilde{g}_N|^4) \tag{2.145}$$

(SC A10)

In other words the coefficients Silver and Chan (1991) have used are

$$b_n^{SC} = \begin{cases} 1/3 & n = 0, N \\ 1 & \text{otherwise} \end{cases}$$

Again we checked some of the original codes to see if the coefficients in the article matched the coefficients in the code. In a subsequent version of the codes we found an interesting comment in the degrees of freedom subroutine (note that `c` denotes a comment in the code). The following variables are given in the code

$$\begin{aligned}
\texttt{nf} &= \text{Nyquist frequency} \\
\texttt{f2} &= \text{second squared moment of } E \\
\texttt{f4} &= \text{fourth moment of } E \\
\texttt{temp} &= \text{power of sample point } i
\end{aligned}$$

Note that we do not actually know what F_2^2 and F_4 are. The quantities E^2 and E_4 are observed from the data and are estimators for those unknown quantities (see Equations 2.122 and 2.141). Therefore more informative names for the variables would have been `e2` and `e4`.

```
do i=1,nf
  temp=cabs(a(i))
          f2=f2+temp**2
          f4=f4+temp**4
c       for zero frequency and for Nyquist:
          if(i.eq.1.or.i.eq.nf)then
            f2=f2-.5*temp**2
 ccccccccccccccccccccc
c ***  note the following line is from Silver and Chan 1990.
c but this appears to give a slightly too large value for
c the number of df for a gaussian distributed time series.
c using .5 instead of (2./3) seems to give the right
c answer but I dont know why.  It depends on the fourth
c moment of the zero frequency and nyquist frequency points.
ccccccccccccccccccccc
            f4=f4-.5*temp**4
c           f4=f4-(2./3.)*temp**4
          endif
  enddo
```

The `f2` is calculating

$$\begin{aligned}
\texttt{f2} &= \sum_{n=0}^{N} |\tilde{f}_n|^2|\tilde{g}_n|^2 - 1/2|\tilde{f}_0|^2|\tilde{g}_0|^2 - 1/2|\tilde{f}_N|^2|\tilde{g}_N|^2 \qquad \text{(sum goes from 0 to } N\text{)} \\
&= \sum_{n=1}^{N-1} |\tilde{f}_n|^2|\tilde{g}_n|^2 + 1/2(|\tilde{f}_0|^2|\tilde{g}_0|^2 + |\tilde{f}_N|^2|\tilde{g}_N|^2) \qquad \text{(sum goes from 1 to } N-1\text{)} \\
&= E
\end{aligned}$$

which matches the definition of the a_n coefficients which are the same for both us and Silver and Chan (1991). They are also the same coefficients used in $\langle E \rangle$ and F_2. The `f4` is calculating

$$\begin{aligned}
\texttt{f4} &= \sum_{n=0}^{N} |\tilde{f}_n|^4|\tilde{g}_n|^4 - 1/2|\tilde{f}_0|^4|\tilde{g}_0|^4 - 1/2|\tilde{f}_N|^4|\tilde{g}_N|^4 \qquad \text{(sum goes from 0 to } N\text{)} \\
&= \sum_{n=1}^{N-1} |\tilde{f}_n|^4|\tilde{g}_n|^4 + 1/2(|\tilde{f}_0|^4|\tilde{g}_0|^4 + |\tilde{f}_N|^4|\tilde{g}_N|^4) \qquad \text{(sum goes from 1 to } N-1\text{)} \\
&= F_4
\end{aligned}$$

which is the Silver and Chan (1991) definition of F_4 (Equation 2.139 (*SC A8*)). The commented out `f4` calculates

$$\begin{aligned}
\texttt{f4} &= \sum_{n=0}^{N} |\tilde{f}_n|^4|\tilde{g}_n|^4 - 2/3|\tilde{f}_0|^4|\tilde{g}_0|^4 - 2/3|\tilde{f}_N|^4|\tilde{g}_N|^4 \qquad \text{(sum goes from 0 to } N\text{)} \\
&= \sum_{n=1}^{N-1} |\tilde{f}_n|^4|\tilde{g}_n|^4 + (1/2)(2/3)(|\tilde{f}_0|^4|\tilde{g}_0|^4 + |\tilde{f}_N|^4|\tilde{g}_N|^4) \qquad \text{(sum goes from 1 to } N-1\text{)} \\
&= \sum_{n=1}^{N-1} |\tilde{f}_n|^4|\tilde{g}_n|^4 + 1/3(|\tilde{f}_0|^4|\tilde{g}_0|^4 + |\tilde{f}_N|^4|\tilde{g}_N|^4) \qquad \text{(sum goes from 1 to } N-1\text{)} \\
&= E_4
\end{aligned}$$

which is the Silver and Chan (1991) definition of E_4 (Equation 2.145 (*SC A10*)). Again we reiterate that we do not know F_4 and so we must estimate it through E_4 therefore the code should be using the `f4` that is commented out (which is in fact E_4). (This highlights the importance of choosing meaningful variable names in programming.) However, because our definition of E_4 differs from the Silver and Chan (1991) E_4 we believe that even the commented out version of `f4` is incorrect and that the coefficients should be the ones defined in b_n not b_n^{SC}. Based on the comments in the code it appears they knew there was an error but chose a different coefficient because it supposedly lead to the "right answer". In order to test this, an analysis of the Silver and Chan (1991) code a_n coefficients and our b_n coefficients is done in Section 5.1. For now we will complete the degrees of freedom calculation. Substituting these results into Equation 2.137 we may estimate the degrees of freedom as

$$\widehat{\nu} \equiv 2\left(\frac{2E^2}{E_4} - 1\right) \tag{2.146}$$

(SC A12)

If we look at the ratio of these two estimators we obtain

$$\frac{E^2}{E_4} = \frac{\langle E^2\rangle + (E^2 - \langle E^2\rangle)}{\langle E_4\rangle + (E_4 - \langle E_4\rangle)} \tag{2.147}$$

If we define

$$\delta E^2 = \frac{E^2 - \langle E^2\rangle}{\langle E^2\rangle} \tag{2.148}$$

then this implies $\langle \delta E^2\rangle = 0$. Likewise, if we define

$$\delta E_4 = \frac{E_4 - \langle E_4\rangle}{\langle E_4\rangle} \tag{2.149}$$

then this implies $\langle \delta E_4\rangle = 0$ and so

$$\begin{aligned}
\frac{E^2}{E_4} &= \frac{\langle E^2\rangle(1 + (\frac{E^2 - \langle E^2\rangle}{\langle E^2\rangle}))}{\langle E^4\rangle(1 + (\frac{E^4 - \langle E^4\rangle}{\langle E^4\rangle}))} \\
&= \frac{\langle E^2\rangle(1 + \delta E^2)}{\langle E_4\rangle(1 + \delta E_4)} \\
&= \frac{\langle E^2\rangle}{\langle E_4\rangle}(1 + \delta E^2)(1 + \delta E_4)^{-1} \\
&\quad (1 + \delta E_4)^{-1} \text{ can be approximated by its Taylor Series} \\
&\approx \frac{\langle E^2\rangle}{\langle E_4\rangle}(1 + \delta E^2)(1 - \delta E_4 + \dots) \\
&\approx \frac{\langle E^2\rangle}{\langle E_4\rangle}(\underbrace{1 + \delta E^2 - \delta E_4}_{\text{first order terms}} - \delta E^2 \delta E_4 + \dots) \\
\left\langle \frac{E^2}{E_4} \right\rangle &= \frac{\langle E^2\rangle}{\langle E_4\rangle}(1 + \underbrace{\langle \delta E^2\rangle}_{=0} - \underbrace{\langle \delta E_4\rangle}_{=0} + \text{second order terms}) \\
\left\langle \frac{E^2}{E_4} \right\rangle &= \frac{\langle E^2\rangle}{\langle E_4\rangle} + \text{ second order terms}
\end{aligned} \tag{2.150}$$

and so this is a first order unbiased estimator of the degrees of freedom. Using the above result and Equations 2.135, 2.137 and 2.141

$$\begin{aligned}
\widehat{\nu} &= 2\left(\frac{2E^2}{E_4} - 1\right) \\
\langle\widehat{\nu}\rangle &= 2\left(\left\langle\frac{2E^2}{E_4}\right\rangle - 1\right) \\
&\approx 2\left(\frac{2\langle E^2\rangle}{\langle E_4\rangle} - 1\right) \\
&= 2\left(\frac{2[\sigma^4(F_2^2 + F_4)]}{2\sigma^4 F_4} - 1\right) \\
&= 2\left(\frac{F_2^2}{F_4} + 1 - 1\right) \\
&= 2\frac{F_2^2}{F_4} = \nu = (SC\ A9)
\end{aligned} \tag{2.151}$$

This suggests the degrees of freedom estimate is unbiased. However, with Silver and Chan (1991) accidentally using the a_n coefficients, E_4 will be too small so $\widehat{\nu} > \nu$ that is, the estimate of the degrees of freedom is positively biased. Again, this is discussed more in Section 5.1.

2.10 Summary

In this chapter, the method given in the Silver and Chan (1991) article has been explained in detail. Sign conventions as well as the need to write things in a more compact form have led to discrepancies in the formulae as explained in this chapter and listed in Table 2.2.

Discrepancies between what is written and what is implemented in the codes have also been identified. Whilst the mistake in calculating the covariance matrix is relatively minor, using the wrong coefficients for the degrees of freedom and inconsistent definitions of the standard errors are a concern.

In order to obtain the errors and a confidence region a large number of assumptions are made mainly around the the sum of squares $S(\phi, \delta t)$ having a chi-squared distribution, which relies on having uncorrelated Gaussian noise. These assumptions need to be checked because without them the errors and the confidence region are unreliable. The chi-squared assumption is tested in Section 5.2.

Even if the assumptions hold, the use of the a_n coefficients instead of the b_n coefficients means that E_4 is too small, which would lead to the degrees of freedom being too large and consequently the error bars being too small. In Section 5.1 we test these coefficients to see how reliable estimates of the errors are.

Formula	Thesis	Silver and Chan (1991) Article
Splitting Operator	$\mathbf{\Gamma}(\phi,\delta t) \equiv \exp(-i\omega\delta t/2)\hat{\mathbf{f}}\hat{\mathbf{f}}^T + \exp(i\omega\delta t/2)\hat{\mathbf{s}}\hat{\mathbf{s}}^T + \hat{\mathbf{b}}\hat{\mathbf{b}}^T$	$\mathbf{\Gamma}(\phi,\delta t) \equiv \exp(i\omega\delta t/2)\hat{\mathbf{f}}\hat{\mathbf{f}}^T + \exp(-i\omega\delta t/2)\hat{\mathbf{s}}\hat{\mathbf{s}}^T$
Compact Splitting Operator	$\mathbf{\Gamma}(\phi,\delta t) = \exp(-i\omega\delta\mathbf{T}) + \hat{\mathbf{b}}\hat{\mathbf{b}}^T$	$\mathbf{\Gamma}(\phi,\delta t) = \exp(i\omega\delta\mathbf{T})$
Radial Component	$\mathbf{u}_r(t) = [\cos^2\phi\exp[-i\omega(t+\delta t/2)] + \sin^2\phi\exp[-i\omega(t-\delta t/2)]]w(\omega)$	$\mathbf{u}_r(t) = \cos^2\phi w(t+\delta t/2) + \sin^2\phi w(t-\delta t/2)$
Transverse Component	$\mathbf{u}_t(t) = \frac{1}{2}sin(2\phi)[exp[-i\omega(t+\delta t/2)] - \exp[-i\omega(t-\delta t/2)]]w(\omega)$	$\mathbf{u}_t(t) = -\frac{1}{2}sin(2\phi)[\omega(t+\delta t/2)] - \omega[(t-\delta t/2)]]$
Covariance Matrix	$\mathbf{c} = \begin{bmatrix} \int_{-\infty}^{\infty} u_1^2(t)dt & \int_{-\infty}^{\infty} u_1(t)u_2(t-\delta t)dt \\ \int_{-\infty}^{\infty} u_1(t)u_2(t-\delta t)dt & \int_{-\infty}^{\infty} u_2^2(t)dt \end{bmatrix}$	$\mathbf{c} = \begin{bmatrix} \int_{-\infty}^{\infty} u_1(t)u_1(t-\delta t)dt & \int_{-\infty}^{\infty} u_1(t)u_2(t-\delta t)dt \\ \int_{-\infty}^{\infty} u_2(t)u_1(t-\delta t)dt & \int_{-\infty}^{\infty} u_2(t)u_2(t-\delta t)dt \end{bmatrix}$
4th Moment of F	$F_4 = \sum_{n=1}^{N-1} 2\lvert\tilde{f}_n\rvert^4 + 1/2(\lvert\tilde{f}_0\rvert^4 + \lvert\tilde{f}_N\rvert^4)$	$F_4 = \sum_{n=1}^{N-1} 1\lvert\tilde{f}_n\rvert^4 + 1/2(\lvert\tilde{f}_0\rvert^4 + \lvert\tilde{f}_N\rvert^4)$
4th Moment of E	$E_4 = \sum_{n=1}^{N-1} 4/3\lvert\tilde{f}_n\rvert^4 + 1/3(\lvert\tilde{f}_0\rvert^4 + \lvert\tilde{f}_N\rvert^4)$	$E_4 = \sum_{n=1}^{N-1} 1\lvert\tilde{f}_n\rvert^4 + 1/3(\lvert\tilde{f}_0\rvert^4 + \lvert\tilde{f}_N\rvert^4)$

Table 2.2: Differences between the formulae given in this thesis and the ones provided in the Silver and Chan (1991) article. Differences are indicated by red symbols (ignoring the extension into 3 dimensions). Note that we have chosen the equations from our thesis that are the most similar to the ones in Silver and Chan (1991) so that comparisons can easily be made.

Chapter 3

Grading System

The previous chapter explained in detail the Silver and Chan (1991) method. However, the method sometimes produces results that are unreliable. For instance, if there is not much energy on components to begin with, it is very difficult to obtain results that can be trusted. We graded 146 output plots from the Mfast (Teanby et al. (2004*a*), Wessel (2010), Savage et al. (2010) program manually using an expert, experienced grader (Prof. Martha Savage). We then developed a set of six numerical diagnostics, which appeared to capture the features used to classify the Mfast output plots. We then use these numerical diagnostics to calibrate a prediction model for the expert's manual grades, which are regarded as a gold standard. The goal is to use this prediction model to create grades automatically from the set of numerical diagnostics, therefore avoiding the need to manually grade output plots. Our dataset does not contain any null measurements so we have not incorporated null detection into this model.

3.1 Dataset

The dataset for this project is from the BOR station (courtesy of Rivemale (2012)) on Reunion Island off the East coast of Africa (Figure 3.1). The events are replicates in the sense that they are all from the BOR station, they are in close proximity temporally (end of 2004-mid 2006), all are shallow earthquakes with depths between 0.5 and 1.5km, they have similar waveforms and were all sampled at 100 samples per second. We therefore expect them to have similar fast directions (ϕ) and similar delay times (δt). Each event records activity for 19 seconds and has an S wave arrival at around 4 seconds. Figure 3.2 shows an example of the North component for 5 different events with estimated fast directions between 0 and 45°.

A dataset provided to us by Rivemale (2012) was part of a much larger dataset that comprises several families of earthquakes. She created the dataset in the following manner: the vertical component of each earthquake recorded at the BOR station was cross-correlated with every other earthquake's vertical component using the `C-PARSE` program (Ludlam,

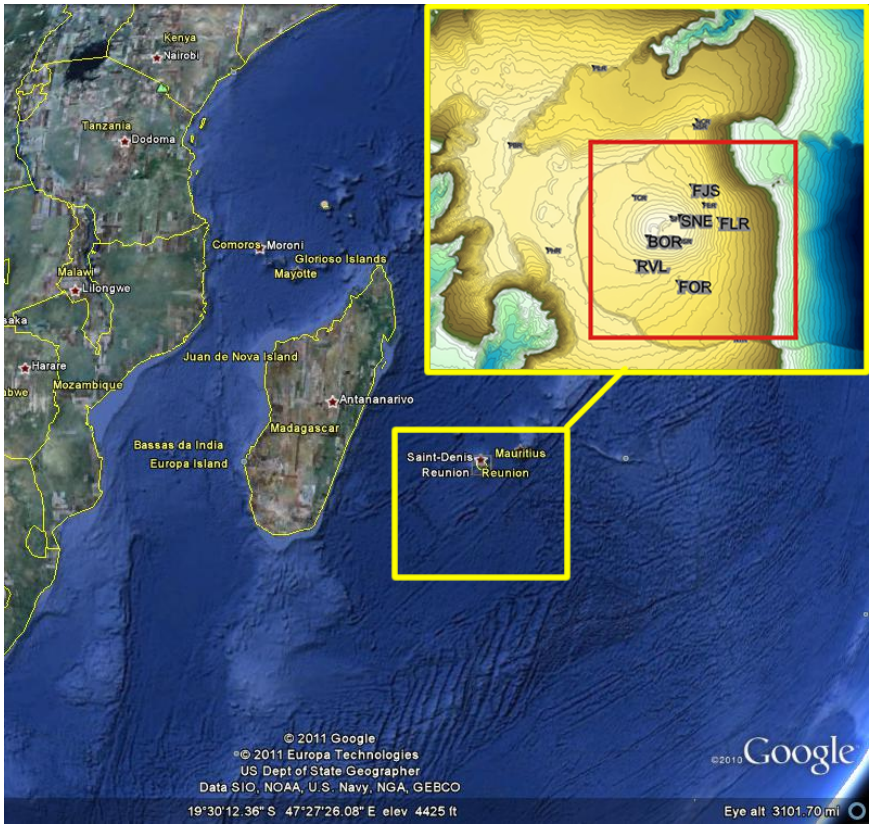


Figure 3.1: Map of Reunion island and the BOR station. World map produced using Google Earth (Google Earth, 2011).

2012). To group the earthquakes into families the coherency was calculated. The coherency C_{xy} between two signals x and y is defined as

$$C_{xy} = \frac{|G_{xy}|^2}{G_{xx}G_{yy}} \tag{3.1}$$

where G_{xy} is a cross spectral density and G_{xx} and G_{yy} are auto spectral densities. The specific details about coherency, cross and auto spectral densities can be found in Crocker (1998). The coherency must lie somewhere between 0 and 1 inclusive.

The events were grouped together if the waveforms had a coherency greater than 90%. The dataset provided originally contained 513 events that were recorded between 2004-2006. Only 510 were usable because some events were missing at least one of their East, North or vertical components and had to be discarded. The remaining 510 events were graded using Mfast (Teanby et al. (2004*a*), Wessel (2010) and Savage et al. (2010)). Prior to processing, the data are filtered to remove unwanted frequencies that interfere with the signal. The filters are applied, and Mfast selects the one that results in the strongest signal relative to the noise, in other words the one with the largest signal to noise ratio. The Mfast algorithm first determines the best filter using the signal to noise ratio around the S arrival. Then the best filter is applied to the data. A total of 202 high quality events (those with A or B grades) were observed. Using different filters can lead to different results so the filter was fixed for all of the events, rather than having a different filter for each event. The filter that had the

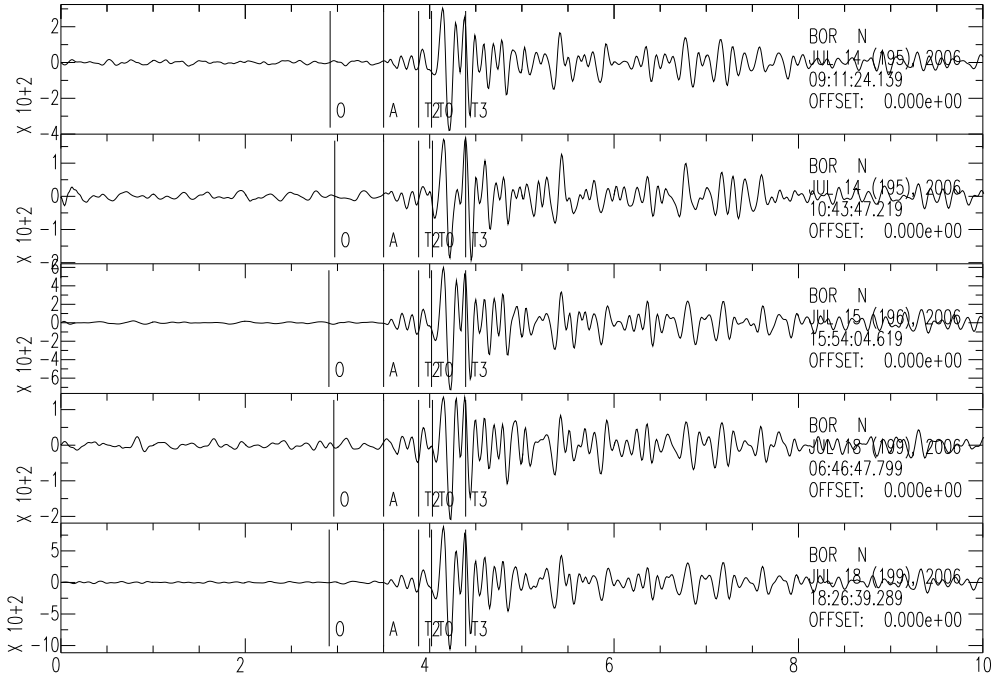


Figure 3.2: Example of replicates of the North component from BOR station for the first 10 seconds. 0 marker represents the event origin; A, the P wave arrival; T2, the window beginning; T3, the window end and T0 the S wave arrival.

most A and B grades was a 2-pole Butterworth band pass filter that had corner frequencies of 4 and 10Hz. When using the fixed filter exclusively instead of allowing Mfast to select the filter, the 4-10Hz filter obtained more A and B grades (268) than the automated method in Mfast (202 events). However, after further analysis it appeared that there were two main clusters of events. Group 1 consisted of events with fast directions between 0 and 45° (146 events). Group 2 consisted of events with fast directions between 90 and 120° (89 events). Thirty-three events that did not belong to either group were removed. Group 1 was chosen as the data set to analyse, since it has a larger number of events. Lastly, the window to be analysed during shear wave splitting (the area within $T2$ and $T3$ in figure 3.2) was fixed to prevent results that vary with the window. The 146 group 1 events with their 4-10Hz filter had an average window start time of 3.88 ± 0.002 seconds and an average window end time of 4.39 ± 0.002 seconds. Therefore, a fixed window consisting of a start time of 3.88 seconds and an end time of 4.39 seconds is used for every event. Fixing both the filter and the window means fewer sources of variation to take into account during this chapter. The group selection process is outlined in Figure 3.3.

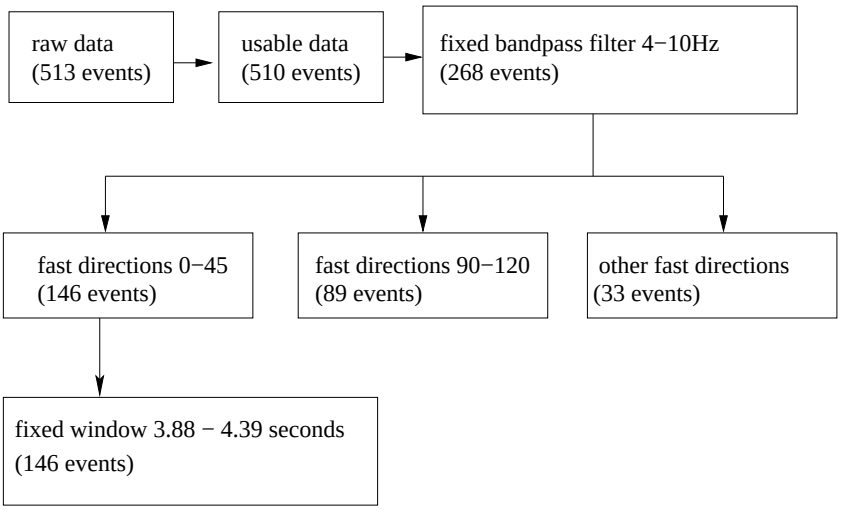


Figure 3.3: Flow chart showing how a subset of family 1 was chosen for analysis.

3.2 Manual Grading

Within group 1 there appeared to be variability in the quality of the estimates. As a result, all 146 events were manually graded with 1 being the best grade possible and 5 being the worst.

3.2.1 Mfast Output Plots

Each event was analysed using Mfast, a software package that implements the Silver and Chan (1991) method. The Mfast output plots are used to manually grade the data. Figures 3.4 and 3.5 are both examples of output plots. Understanding what each Mfast output plot consists of is outlined below:

- The top left plot contains the East (e), North (n) and vertical (z) components of the waveform (what we call $\hat{\mathbf{e}}$, $\hat{\mathbf{n}}$ and $\hat{\mathbf{z}}$ respectively). The two vertical red bars indicate the window beginning and end. The black vertical bar indicates the S wave arrival. This plot is not used in the manual grading process, except as a rough check that there are no major problems with the waveforms such as a very poor signal.
- The top right plot contains the rotated waveforms. p (what we call $\hat{\mathbf{p}}$) represents the polarisation direction determined by Mfast using the Silver and Chan (1991) method and $p\perp$ (what we call $\hat{\mathbf{a}}$) is the direction perpendicular to p for the original waveform. The corrected versions are similar except they are calculated after reversing the effects of shear wave splitting. The vertical lines are interpreted the same as the ones in the top left plot.
- The bottom left plot consists of the original fast and slow waveforms (left) and the fast and slow waveforms after the wave is desplit (right). Both waveforms are rotated into splitting coordinates ($\mathcal{S}$). The white region of these graphs represents the analysis window used for the splitting calculation. The grey section is a small segment of the

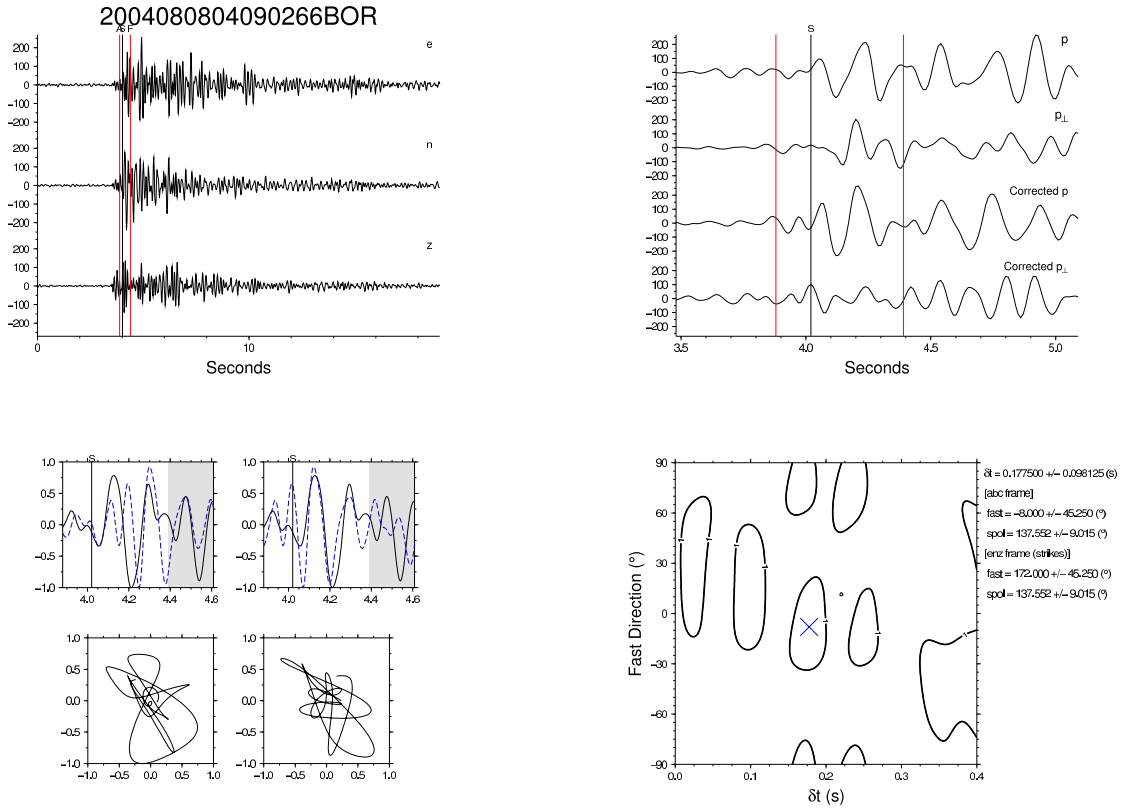


Figure 3.4: Example of a poor quality event (grade 5 - the worst grade). Notice the poor waveform match, non linear particle motion (bottom left), multiple 95% confidence contours (bottom right) and energy pushed outside the window (top right).

wave outside the analysis window. Below these plots are the particle motion in the fast and slow direction for the original (left) and corrected (right) waveforms based on the data in the window.

- The bottom right plot is a contour plot of the smallest eigenvalue of the covariance matrix as a function of the fast direction ϕ and delay time δt. The best $(\phi, \delta t)$ pair with the smallest eigenvalue is marked by a cross. The darkest contour indicates the 95% confidence region. All the contours are normalised by the 95% confidence regions boundary so that the 95% confidence region has a contour level of 1 and so on (see Equation 2.105). In a high quality event this contour is a single closed loop enclosing the cross (Figure 3.5). In lesser quality events there may be multiple disconnected 95% contour curves (Figure 3.4).

The generation of these plots, and the entire Silver and Chan (1991) method including pre-processing such as detrending the data, were rewritten in `R` to make current and future statistical analysis easier. An example of an `R` generated set of output plots is shown in Figure 3.6. Care has been taken to ensure that these closely resemble the outputs of Mfast. A legend was added to our plots to make it easier to distinguish between the fast and slow components.

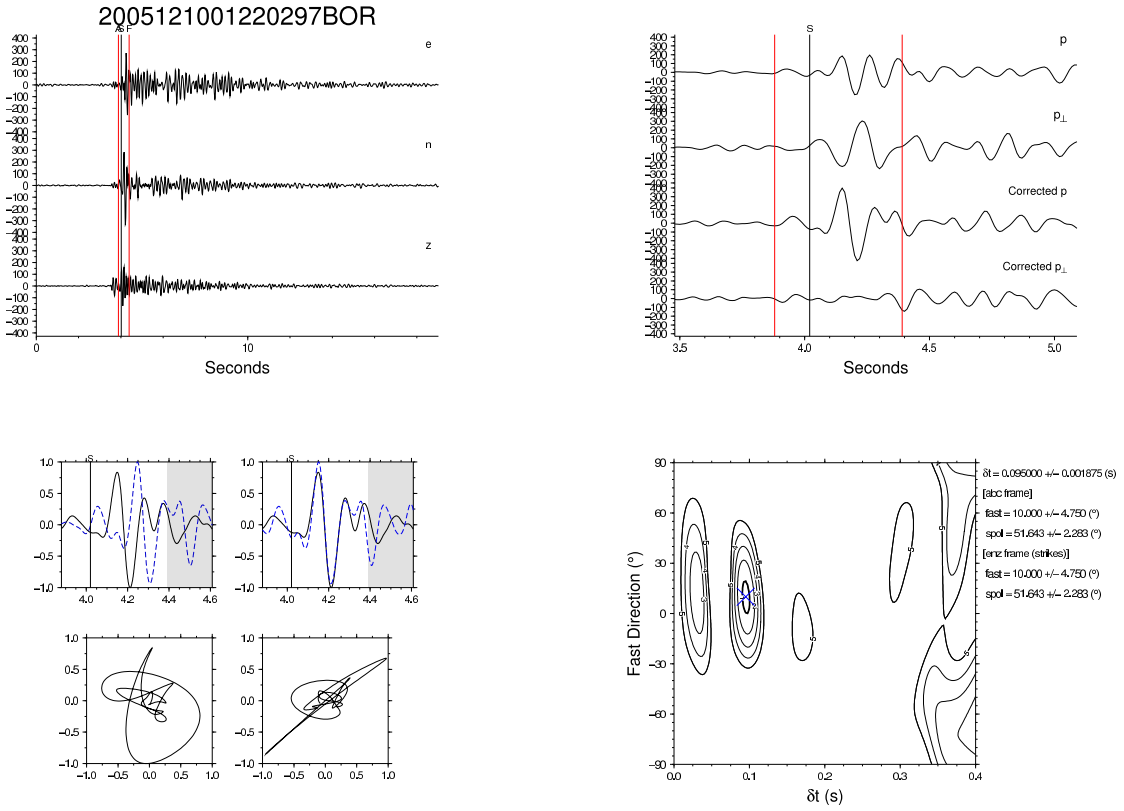


Figure 3.5: Example of a high quality event (grade 1 - the best grade). Notice the good waveform match, linear particle motion (bottom left) and the single bullseye style, 95% confidence contour (bottom right).

3.2.2 Manual Grading Criteria

To give each event a manual grade the following criteria were used:

1. Contour plot (bottom right plot on the output plot)
   (a) Count the number of closed contours enclosing the minimum. The more contour levels, the deeper the minimum, and the more accurate an estimate is.
   (b) Check for more than one closed contour corresponding to the 95% confidence region. More than one can be indicative of cycle skipping, making estimates unreliable.
2. Fitting of the curves (bottom left plot on the output plot)
   (a) Check the corrected fast and slow components are in phase within the window after splitting is reversed. The two components are out of phase by δt when they are split but should be in phase after correcting for shear wave splitting.
   (b) Check the corrected fast and slow components are in phase just outside the analysis window (grey section in the output plot) after splitting is reversed. The data outside the analysis window is not used during desplitting so if the waveforms in this area are in phase, the ϕ and δt may be considered more reliable

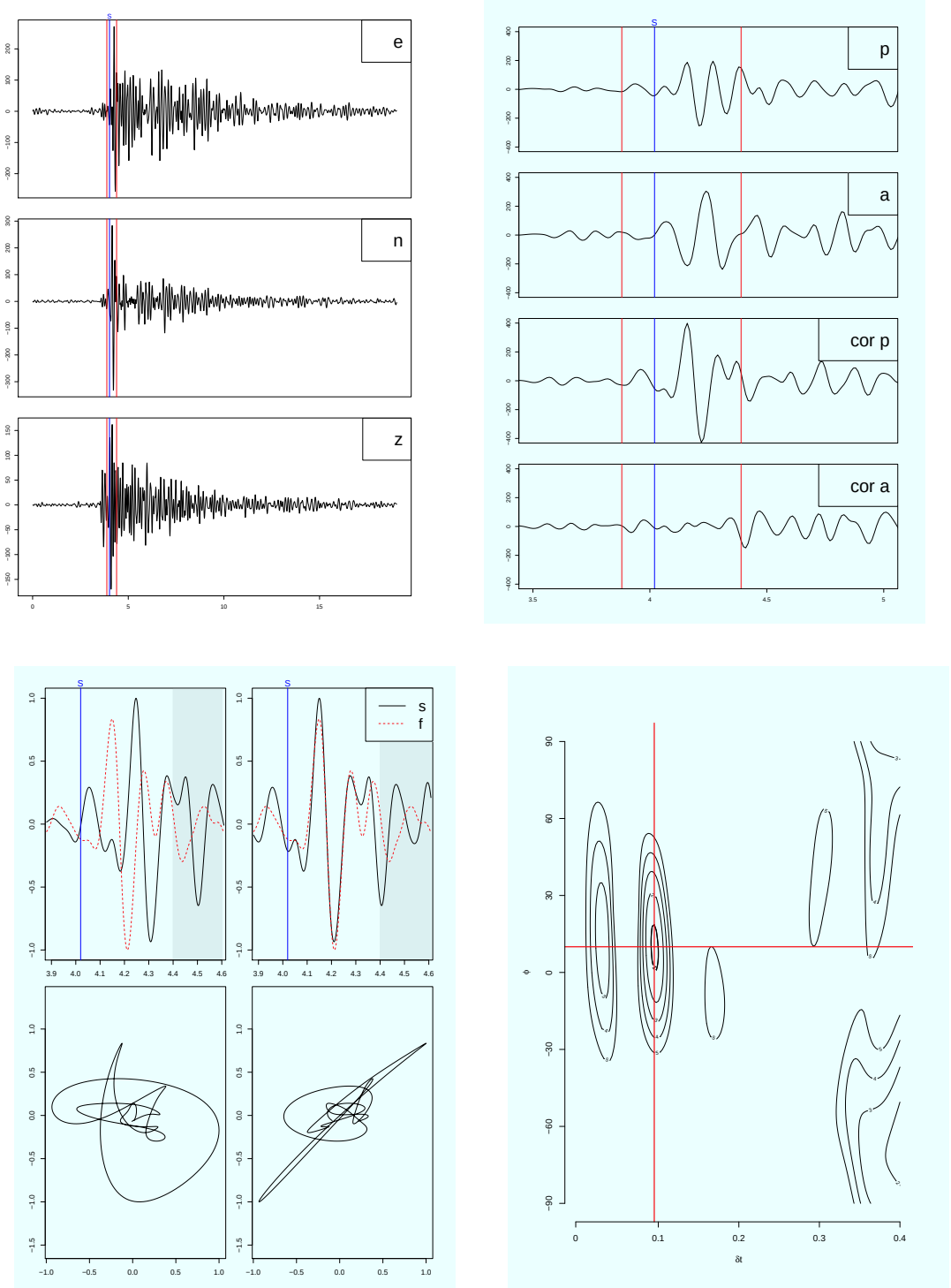


Figure 3.6: R generated output plot of the event corresponding to Figure 3.5.

3. p ($\hat{\mathbf{p}}$) and $p \perp$ ($\hat{\mathbf{a}}$) components (top right plot on the output plot)
   (a) Check how much energy has been minimised by comparing $p \perp$ ($\hat{\mathbf{a}}$) and *cor* $p \perp$ ($\hat{\mathbf{a}}$ after correcting for shear wave splitting). The more energy that has been minimised the better.
   (b) Compare the energy minimised on the *cor* $p \perp$ ($\hat{\mathbf{a}}$ after correcting for shear wave splitting) component within the analysis window and outside the analysis window. Whilst minimising the energy of the signal is desirable, minimising noise is undesirable. Minimising the noise is typically shown as an extremely flat section within the window whilst there are oscillations outside the window. In such a case we have effectively treated the noise as signal and have overcorrected the data.
4. Particle motion (bottom left plot on the output plot)
   (a) Check the area enclosed by the particles after removing splitting. Ideally, after splitting the fast and slow waves should be in phase and so the particle motion should be linear and therefore the area enclosed should be close to zero.

What typically determined each event's gold standard grading by the expert was the following:

Grade 1: good fit between fast and slow components both inside and outside the window plus bullseye style (deep) contour plots and linear particle motion (Figure 3.5)

Grade 2: good fit between fast and slow components both inside and outside the window and generally a single confidence region but not a bullseye style contour plot (Figure 3.7)

Grade 3: good fit between fast and slow components only inside the window and possible cycle skipping (Figure 3.8)

Grade 4: moderate fit and possible cycle skipping (Figure 3.9)

Grade 5: poor fit, energy pushed outside the window for *cor* $p \perp$ ($\hat{\mathbf{a}}$ after correcting for shear wave splitting), 95% confidence region over the entire search space (Figure 3.4)

Figures 3.4 (grade 5) and 3.5 (grade 1) are examples of the varying quality of estimates within the BOR dataset. In Figure 3.4 there are multiple 95% confidence regions, the wave fit is poor both inside and outside the window, there appears to be some other phase coming in on the *cor* $p \perp$ component and the particle motion is not very linear. On the other hand, Figure 3.5 has one 95% confidence region with many rings around it, the fit is very good inside the window, the signal's energy has been minimised and the particle motion is linear.

3.2.3 Manual Grading Results

Based on the filtering criteria used to obtain group 1, the fast directions of the example results should be between 0 and 45°. However, this group was determined using a 4-10Hz

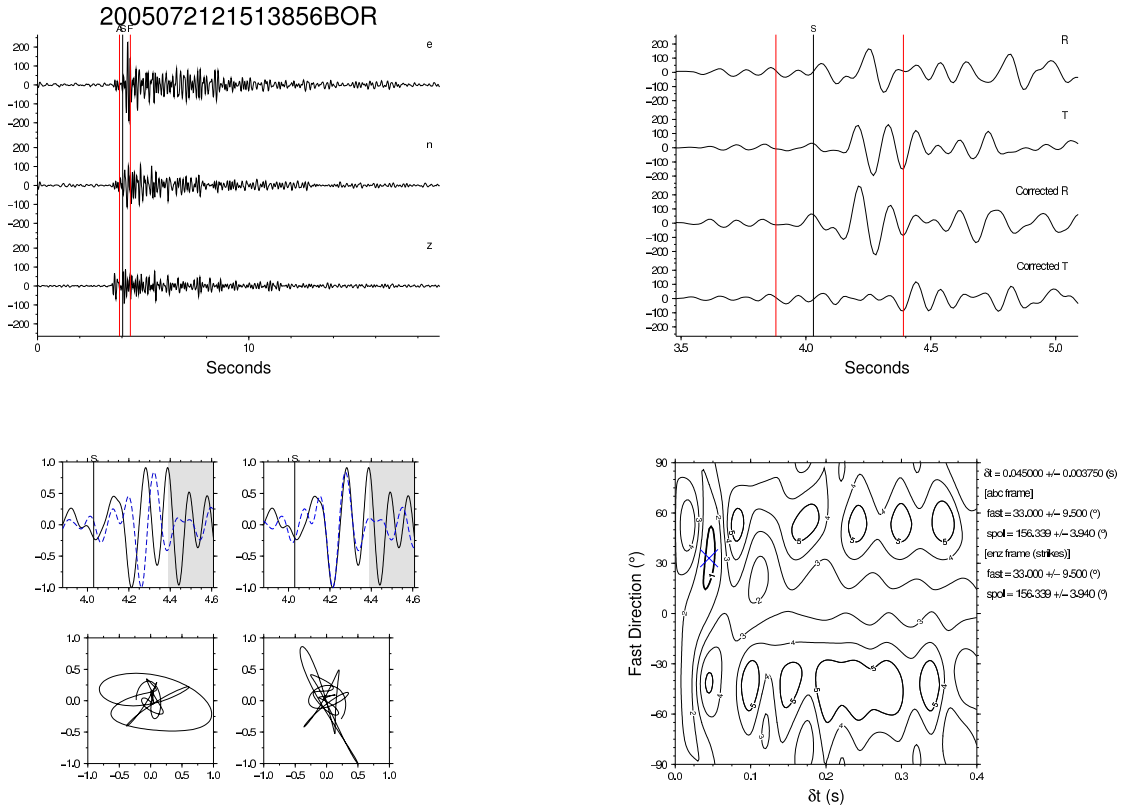


Figure 3.7: Example of a grade 2 event. Notice the good fit between the waveforms (bottom left) but no bullseye style contours (bottom right).

Grade 1	Grade 2	Grade 3	Grade 4	Grade 5	Total
7	40	60	19	20	146

Table 3.1: Number of Group 1 events in each grade band.

bandpass filter, which may not be suitable for every single event. Consequently, there may be some poor quality events with fast directions outside this range. The delay times were not used to form the groups of events so it is possible to get delay times across the entire range of the grid search which is 0-0.4 seconds.

Figures 3.10 and 3.11 show the degree of variability in the splitting estimates and their errors based on the manual grading. Many data points were clustered tightly together so R's `jitter()` command, which adds a small amount of noise to each value, was used to make them more visible. The number of events allocated a particular grade are given in Table 3.1.

Figure 3.10 shows two main clusters of grades with fast directions between 0 and 45° with some poor quality estimates (4 and 5) scattered around the outside. There are some fast

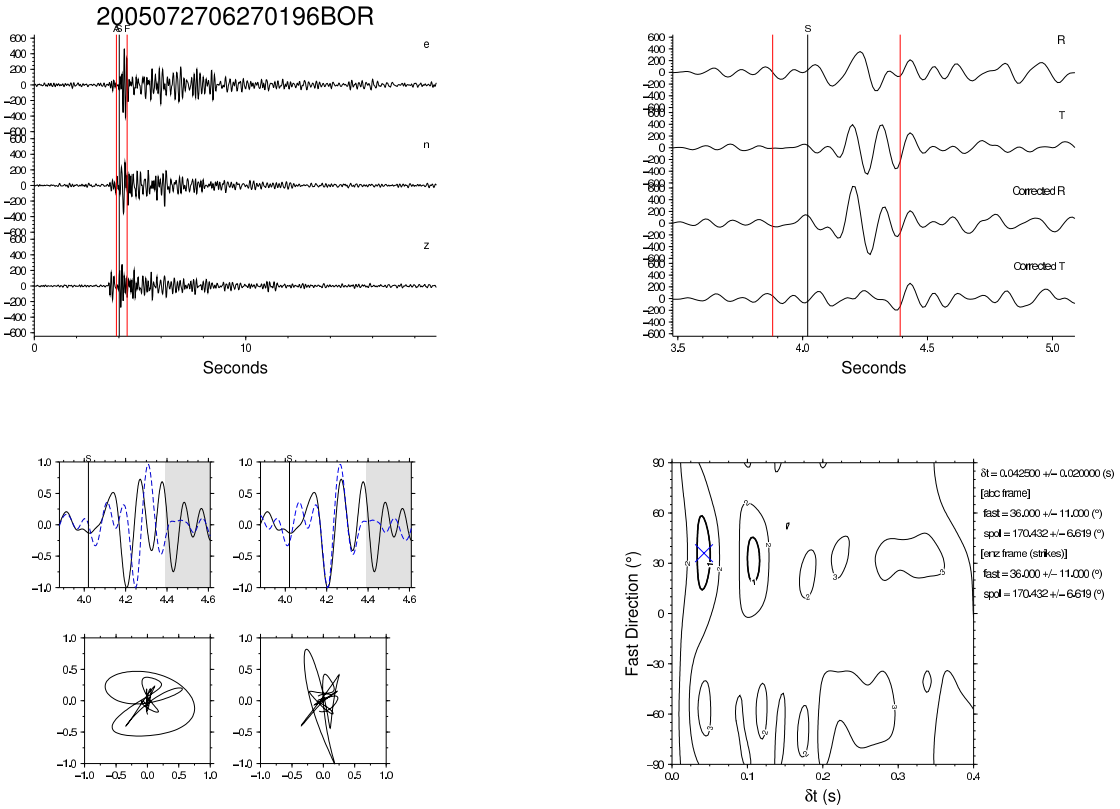


Figure 3.8: Example of a grade 3 event. Notice the good fit between the waveforms (bottom left) but multiple 95% contours (bottom right).

directions estimates outside 0-45° because the 4-10Hz bandpass filter was not be suitable for every single event. For some events the 4-10Hz filter and fixed window may not be ideal and leads to some fast directions being outside 0-45° (see Section 3.1 and Figure 3.3). This is reflected by the low grades outside this region. The two groups that have similar fast directions but different delay times could have been separated by cycle skipping. There appear to be more events in the cluster that has a higher delay time, but it also contains more lower quality estimates. The grade 1 events are scattered between both clusters, however, there are only 7 of them so it is hard to say which delay time is more likely to be correct.

Figure 3.11 shows a scatter plot of the fast direction errors and the delay time errors. Since the contour plot is used to manually grade events, the events with high errors have grades of 5 whereas events with low errors and more bullseye looking contour plots have a high quality grade of 1. The optimisation over $(\phi, \delta t)$ looks over the parameter space and takes place by evaluating the smallest eigenvalue on a regular grid of $(\phi, \delta t)$ values in units of 1° in ϕ and 0.01 seconds in δt. This creates horizontal bands of estimates of the errors in ϕ and δt, although this effect has been minimised by using `jitter()` in Figure 3.11.

The process of manually grading the 146 events in group 1 was very time consuming and not guaranteed to be consistent (for instance some 2s could be 3s and vice versa). An

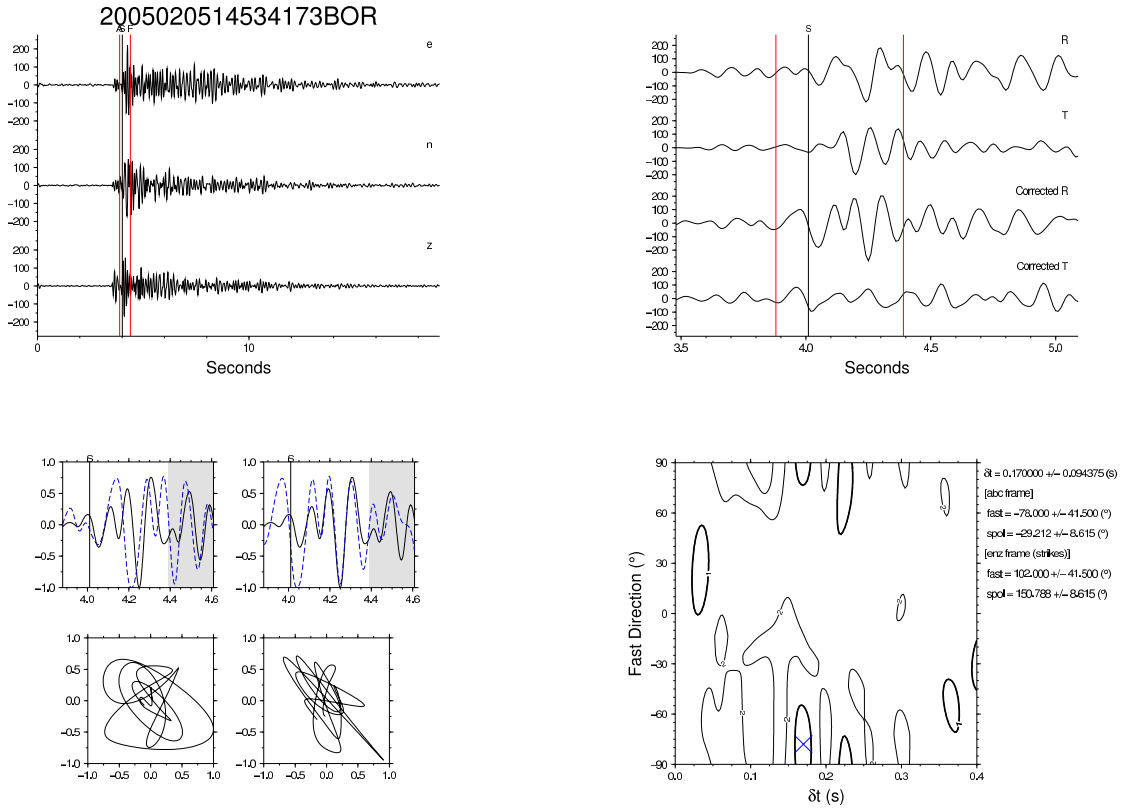


Figure 3.9: Example of a grade 4 event. Notice a moderate fit between the waveforms with possible cycle skipping (bottom left) and multiple 95% contours (bottom right).

automated, consistent, method of grading each event based on the output plots from Mfast is created in the next section.

3.3 Automated Grading

In this section we automate this process of grading by creating a series of numerical criteria and then predict an overall grade using multiple linear regression. We then check whether the regression model is suitable for predicting grades by checking that the assumptions behind the model are met and that no single event has too much influence over the model.

3.3.1 Converting Manual Criteria into Numerical Formulae

This section provides mathematical details to show how each numerical criterion was constructed. For each criterion, a description is given of what is a sign of a good grade and what is a sign of a poorer grade.

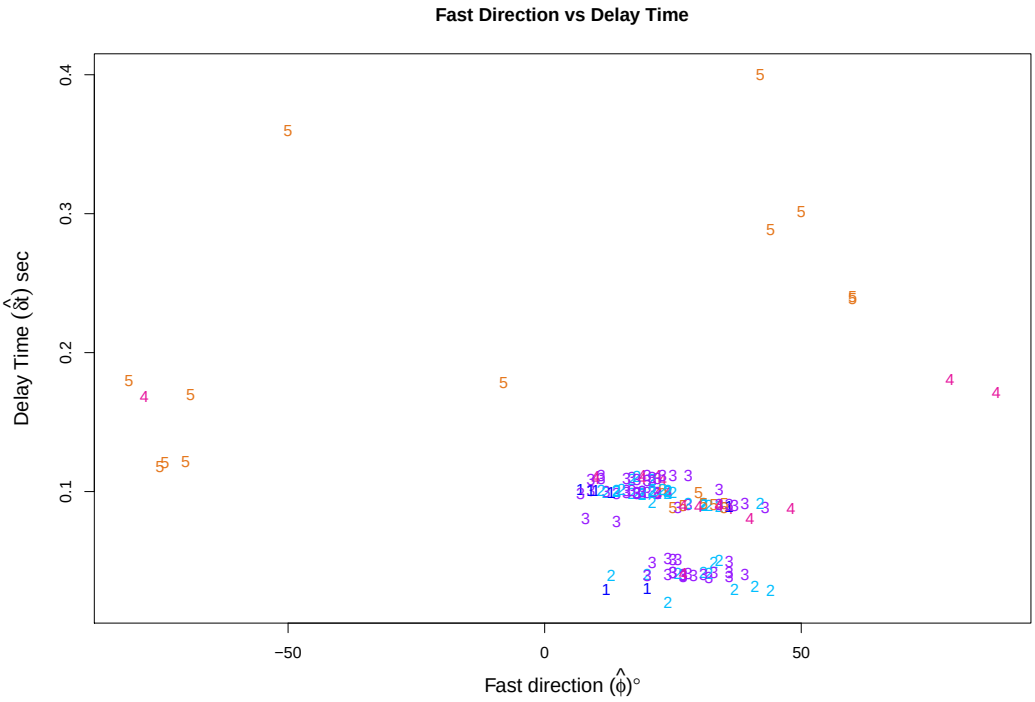


Figure 3.10: Plot of fast direction estimates against delay time estimates after the `jitter()` function was applied. Numbers and colours reflect the grades.

Eigenvalues - GC_{1a},GC_{1b}

Counting the number of contours enclosing the minimum eigenvalue is equivalent to evaluating the depth of the minimum. The first grading criterion looks at the difference between the normalised (Equation 2.105) largest eigenvalue on the contour plot and the smallest eigenvalue.

$$GC_{1a} = (\lambda_2^{max} - \lambda_2^{min})/\lambda_2^{95CR} \tag{3.2}$$

where

$$\begin{aligned} \lambda_2^{max} &= \text{largest eigenvalue} \\ \lambda_2^{min} &= \text{smallest eigenvalue} \\ \lambda_2^{95CR} &= \text{boundary of the 95\% confidence region} \end{aligned}$$

Mfast has a similar criterion where λ_2^{max} is used as a measure of quality. This measure also gives an idea of the steepness of the entire region contoured. The larger the difference, the steeper the contour region. The smaller the difference, the flatter the confidence region. Contour plots with many separate, closed 95 % confidence contours tend to have a very small GC_{1a}.

The other eigenvalue based criterion counts the number of separate closed 95% confidence contours. The `contourLines()` function returns a list of the confidence regions that contain the level for each contour line. Since the eigenvalues have been normalised, the confidence region has a level of 1. Therefore we can count how many contour lines have a level of 1. If

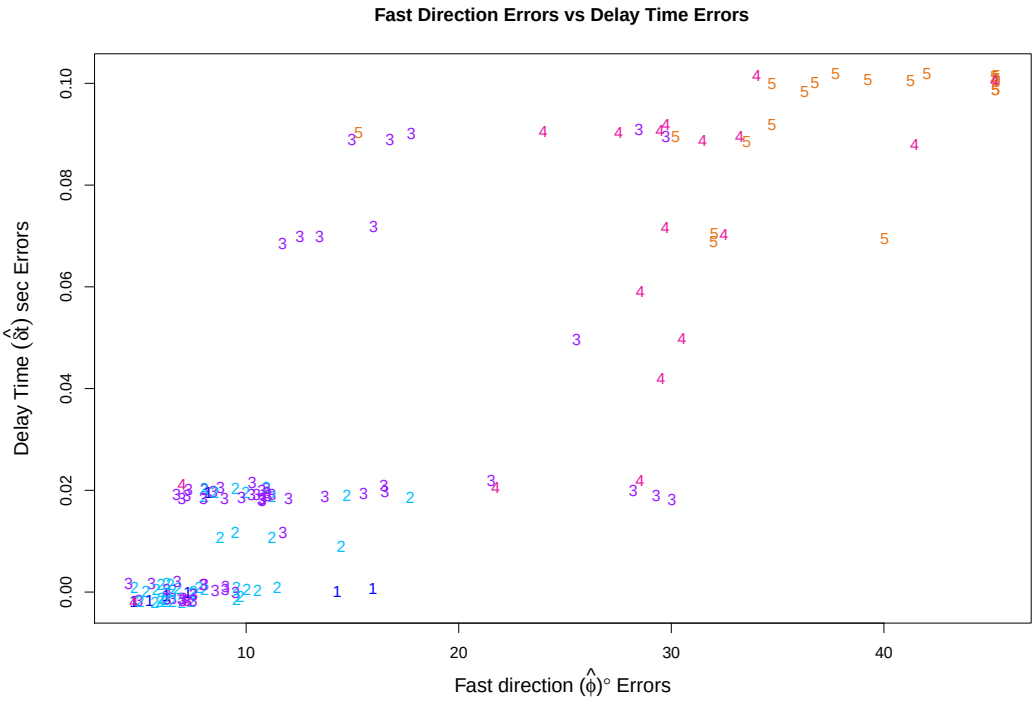


Figure 3.11: Plot of fast direction error estimates against delay time error estimates after the `jitter()` function was applied. Numbers and colours reflect the grades.

there is more than one, this can indicate cycle skipping. This is represented mathematically as

$$GC_{1b} = \sum_{x} \mathbb{1}_{95CR}(x) \tag{3.3}$$

where

$$\mathbb{1}_{95CR} = \begin{cases} 1 \text{ if the contour line } x \text{ corresponds to a 95\% confidence region} \\ 0 \text{ otherwise} \end{cases}$$

As suggested by Figure 3.5, a good quality result has one 95% confidence region. Several confidence regions (like Figure 3.4) indicate a low quality result, which should receive a low grade. Section 2.7 highlighted the cyclical nature of the fast direction (Figure 2.14). The `contourLines()` function is not aware of this cyclicity and will treat a single confidence region that lies at $\pm 90^\circ$ that should wrap around and be a single region, as two separate regions. For example Figure 2.14a has 7, 95% confidence regions but because `contourLines()` cannot directly cope with cyclicity it believes there are 9, 95% confidence regions. A method that could take into account the cyclicity and produce the correct number of 95% confidence regions would make this criterion more valuable. However, the value of this criterion is actually in its identification of a multiplicity of 95% confidence regions, the exact count being less important than the finding that there are more than just one or two.

Waveforms - GC_2, GC_{3a}, GC_{3b}

The root mean square of a vector **x** is defined as

$$\text{rms} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2} \tag{3.4}$$

where n is the length of the vector and x_i^2 is the square of the ith entry of **x**. It is a measure of the typical magnitude of x. Therefore we can calculate the rms of different waveforms and compare their magnitudes by taking ratios of the two. In R we defined a more general function called the rmse for the difference between a pair of vectors. The rmse of vectors **a** and **b** is defined as

$$\text{rmse} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(a_i - b_i)^2} \tag{3.5}$$

It is implemented in R as follows

```
rmse<- function(a, b){
 x<-sqrt(mean((a-b)^2))
 return(x)
 }
```

However, if b is zero then this is simply the rms of vector **a**.

There are three grading criteria that make use of the rms. The first criterion is one that compares the windowed slow and fast components to the slow and fast components just outside the window (i.e. the shaded area in the bottom left Mfast output plot). This can be represented mathematically as

$$GC_2 = \left[\frac{\text{rms}\left[(u_{osc}) - (u_{ofc})\right]}{\text{rms}\left[(u_{wsc}) - (u_{wfc})\right]} - 1\right]^2 \tag{3.6}$$

where

$$\begin{aligned}
\text{rms}(\mathbf{x}) &= \text{root mean square of } \mathbf{x} \\
u_{osc} &= \text{corrected slow component outside (after) the window} \\
u_{ofc} &= \text{corrected fast component outside (after) the window} \\
u_{wsc} &= \text{corrected slow component inside the window} \\
u_{wfc} &= \text{corrected fast component inside the window}
\end{aligned}$$

In the ideal case the waves will be in phase and will have similar amplitudes both inside and outside the window. This means that the ratio would be one and hence GC_2 would be zero. We expect the waveforms to be more similar inside the window because the best splitting parameters are determined using this window. Therefore the denominator should be close to 0. Poor fits outside the window should correspond to a numerator that differs from 0 and therefore the ratio term of GC_2 is either strongly negative ($\text{rms}(u_{osc}) \ll \text{rms}(u_{ofc})$) or

strongly positive ($\text{rms}(u_{osc}) \gg \text{rms}(u_{ofc})$). However, ratio term is not unidirectional because values a value too far from 0 in either direction is bad. If a criterion is not unidirectional then we cannot perform multiple regression because we require a linear relationship for each criterion. Consequently, the ratio has 1 subtracted from it and is then squared so that a small value close to 0 indicates a good results and a large value indicates a poor result.

The second criterion to use the rms function is GC_{3a}. It compares the corrected, windowed $\hat{\mathbf{a}}$ component to the corrected, windowed $\hat{\mathbf{p}}$ component

$$\text{GC}_{3a} = \frac{\text{rms}(u_{wac})}{\text{rms}(u_{wpc})} \tag{3.7}$$

where

$$\begin{aligned} \text{rms}(\mathbf{x}) &= \text{root mean square of } \mathbf{x} \\ u_{wac} &= \text{windowed, corrected null component} \\ u_{wpc} &= \text{windowed, corrected polarisation component} \end{aligned}$$

The energy on the a component should be small after correcting for splitting so a GC_{3a} that is, small is ideal while a large GC_{3a} indicates a poor result. We avoid comparing the windowed corrected null component to the original windowed null component because the the original waveforms may be polarised in such a way that leads to events being unfairly penalised. For example, if there was not much energy on the original component then GC_{3a} would be unfairly large.

While we want the energy on the null component to be small, the rms of u_{wac} should not be too small. If the waveform inside the window is too flat then we have overcorrected. In other words, we have minimised all the power including the noise rather than just minimising the signal. To overcome this problem the GC_{3b} criterion was created to ensure the waveform inside the window is not too flat relative to the waveform immediately to the left (before) of the window. We chose to compare it to the waveform before the analysis window to avoid phases that came in after the S wave. The energy before the window should contain mostly noise and the energy from the P wave but because the P wave is vertically polarised and these are not vertical components, the amount of P wave energy on the a component should be small. The GC_{3b} is defined as

$$\text{GC}_{3b} = \left[\log\left(\frac{\text{rms}(u_{wac})}{\text{rms}(u_{iac})}\right)\right]^2 \tag{3.8}$$

where

$$\begin{aligned} \text{rms}(\mathbf{x}) &= \text{root mean square of } \mathbf{x} \\ u_{wac} &= \text{windowed, corrected null component} \\ u_{iac} &= \text{corrected null component before (to the left) the analysis window} \end{aligned}$$

The ratio of this criterion (the term in the brackets) would be 1 in the ideal case where waveform inside the window is similar to the waveform outside the window. However, the ratio is not unidirectional because values smaller than one indicate overcorrecting and a value greater than one indicates undercorrecting, in other words, a value too far from

1 in either direction is bad. We make GC_{3b} unidirectional by taking the log of the ratio and squaring it so that a small value of GC_{3b} is ideal and larger values suggest under or overcorrecting. The ratio is also stored in the file so that under and overcorrection can be analysed separately if need be.

Particle Motion - GC_4

The particle motion between the fast and slow components can be thought of as a polygon. The polygon can be closed by joining the start and end points. Since the particle motion should become more linear after correcting for splitting, measuring the corrected area can be an indicator of how well the effects of splitting have been reversed.

Particle motion polygons are self-intersecting polygons meaning the edges cross over themselves so calculating the area will depend on how the interior of the polygon is defined. One way to define the interior is by using the even-odd rule. Here a ray is constructed from the point in interest to a point that is outside the polygon. If the ray passes though an odd number of edges then the point is inside the polygon. If the ray passes through an even number of edges then the point is outside the polygon. For this algorithm zero is considered an even number (Hormann and Agathos, 2001). Figure 3.12 illustrates the even-odd rule.

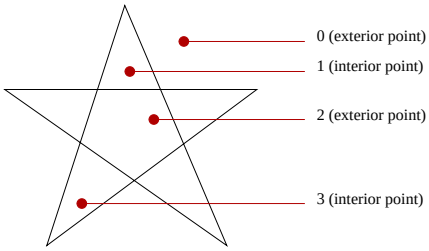


Figure 3.12: A self-intersecting polygon with interior points defined by the even-odd rule. Numbers denote how many edges the ray passes through.

It shows that some areas of the polygon can be excluded from the interior such as the inner 5 sided piece of the star. One way to correct for this would be to create a simple polygon (i.e. a polygon that does not self intersect) that covers the outer edges so that the entire area enclosed could be calculated, but it is extremely time consuming locating the set of outer points.

To see how much of the area is covered by the even-odd rule, a test example was run in R. First a square grid was set up of points ranging from -400 to 400 in steps of 1 in each direction. Next the `in.out()` function was used. This function is an implementation of the even-odd rule. It returns a boolean matrix where `true` corresponds to a point inside the polygon and `false`, a point outside the polygon. The points marked as true were plotted in red and the false points are plotted in black (Figure 3.13). Figure 3.13 shows that a large

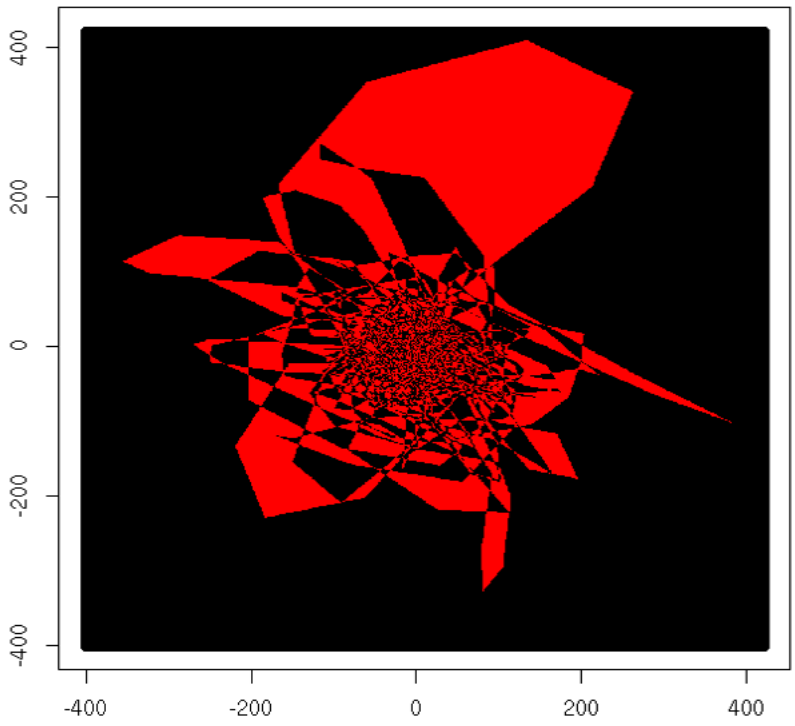


Figure 3.13: Area inside the polygon based on the even-odd rule. Red is the area inside, black is considered outside the polygon.

portion of the area is omitted based on this rule. Consequently, a second method was created to try to capture the whole area. Instead of focusing on the interior, which is difficult to identify, the method calculates the area outside the polygon as follows

1. Have `in.out()` determine whether a point is inside (denoted `true`) or outside (denoted `false`) the polygon using the even-odd rule
2. Copy the results from `in.out()` and change the `false` values to `NA`s (not available)
3. Start at the bottom left corner and check for any `NA` values, change them to `false`
4. From every point on the boundary of the region move inwards (in a direction perpendicular to the boundary), changing `NA` to `false`. Stop searching that row/column when a true value is encountered i.e. an interior point is found (Figure 3.14)
5. This produces the exterior of the polygon so the remaining `NA`s must be inside the polygon and can be changed to `true`

This method captures the area because it counts points exterior to the polygon and stops when it encounters the boundary of the polygon. The results of this method are shown in Figure 3.15 where blue corresponds to points inside the polygon. There are a few horizontal and vertical green lines extending inside the polygon. This is because the grid is not fine enough to capture all the points. However, the number of points inside the polygon that it has determined as being outside the polygon is extremely small compared to the polygon's size. Comparing these two figures for this particular example showed that the even-odd

rule based area only covered 68% of the area that the iterative method covered. It may seem inefficient to run back and forth through the rows and columns (Step 3) but if it is not done then some of the area outside the polygon would not be identified (see Figure 3.14).

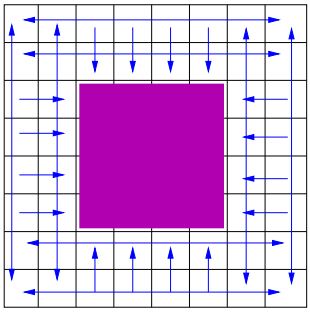


Figure 3.14: Illustration of how Step 3 works. A search back and forth across the rows and the columns must be made so that none of the area outside the polygon is missed.

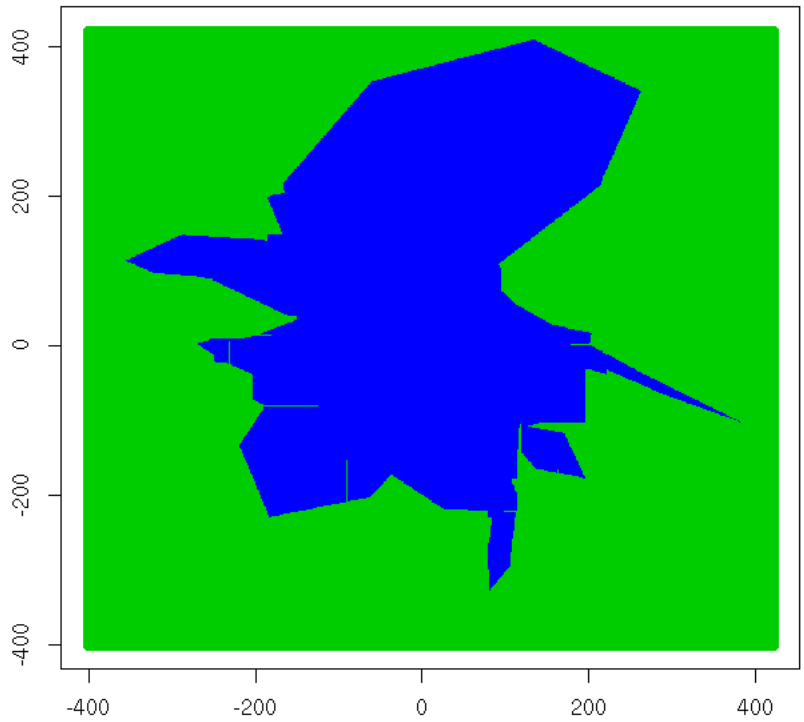


Figure 3.15: Area inside the polygon using an iterative method. Blue is the area inside, green is considered outside the polygon.

Once the number of points enclosed by the polygon is obtained, it then needs to be normalised to enable comparisons. Originally, we were going to normalise by the particle motion area of the original component. However, we found in some cases that by chance, the original particle motion area was small to begin with or that the corrected particle motion area was larger than the original (see Figure 3.16). This means the event would have a larger ratio which would lead to it being penalised with a worse (higher) grade. Instead, we chose the following grading criterion

$$\mathrm{GC}_4 = \frac{A_c/g}{\mathrm{MS}(u_{wpc})} \tag{3.9}$$

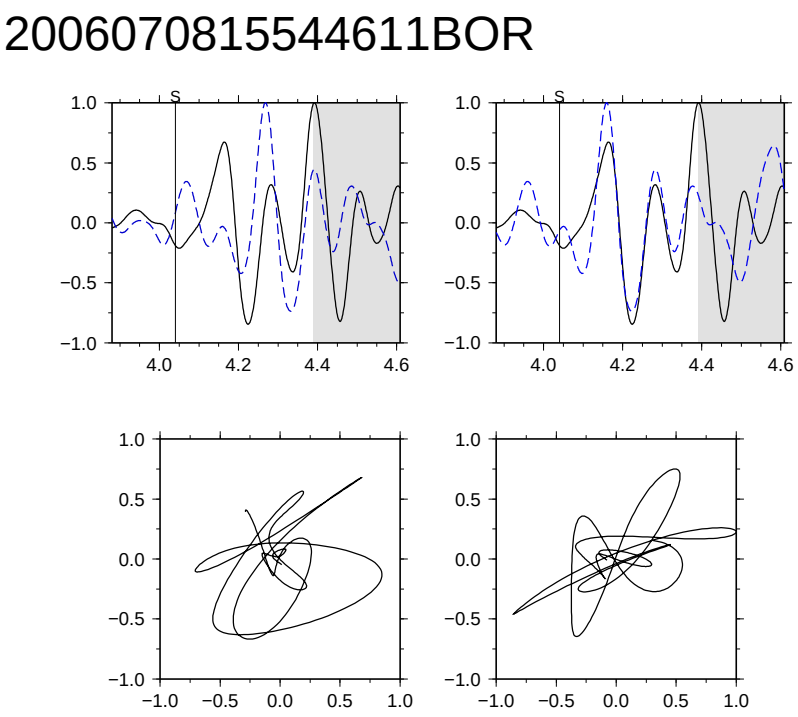


Figure 3.16: Example where the area of the corrected particle motion is slightly larger than the area of the original particle motion. This is because the top of the turnip shaped particle motion expanded into 2 large petals after correcting for splitting. Compare this to Figure 3.5 which has a small particle motion area after correcting for splitting.

where

$$\begin{aligned}
\mathrm{MS}(\mathbf{x}) &= \text{mean square} = \frac{1}{n}\sum_{i=1}^{n} x_i^2 \\
x_i &= i\text{th element of the vector } \mathbf{x} \\
n &= \text{length of the vector } \mathbf{x} \\
A_c &= \text{number of points inside the corrected particle motion polygon} \\
g &= \text{number of points in the grid} \\
u_{wpc} &= \text{the windowed, corrected } p \text{ component amplitude}
\end{aligned}$$

The grid of points is constructed from points ranging from -1 to 1 in steps of 0.002, in other words it covers the entire area of the normalised particle motion. The smaller the proportion of the area and the greater the power on the windowed, corrected polarisation component, the smaller the GC_4 ratio. The smaller this ratio, the more likely it is that this event would receive a good grade.

Figure 3.17 illustrates each criteria on a number line showing all the criteria are unidirectional. Table 3.2 provides definitions and brief justifications for each criteria, which serves as a useful reference for subsequent sections.

Criterion	Definition
GC_{1a}	the difference between the normalised maximum and minimum eigenvalues (checks how steep the contour surface is)
GC_{1b}	the number of 95% confidence regions (checks for multiple regions that typically indicate cycle skipping)
GC_2	ratio of how well the fast and slow components match outside and inside the analysis window after correction (checks to see if the waveforms are in phase after correcting for splitting)
GC_{3a}	ratio of the null component and the polarisation component after correction (a check to see how much energy has been minimised on the null component)
GC_{3b}	log ratio squared of the corrected null component before and inside the window squared (a check for under and overcorrecting)
GC_4	area of particle motion relative to the grid size and the mean squared error of the corrected $\hat{\mathbf{p}}$ component (a check to see if particle motion is linear after correcting for splitting)

Table 3.2: Summary of the grading criteria.

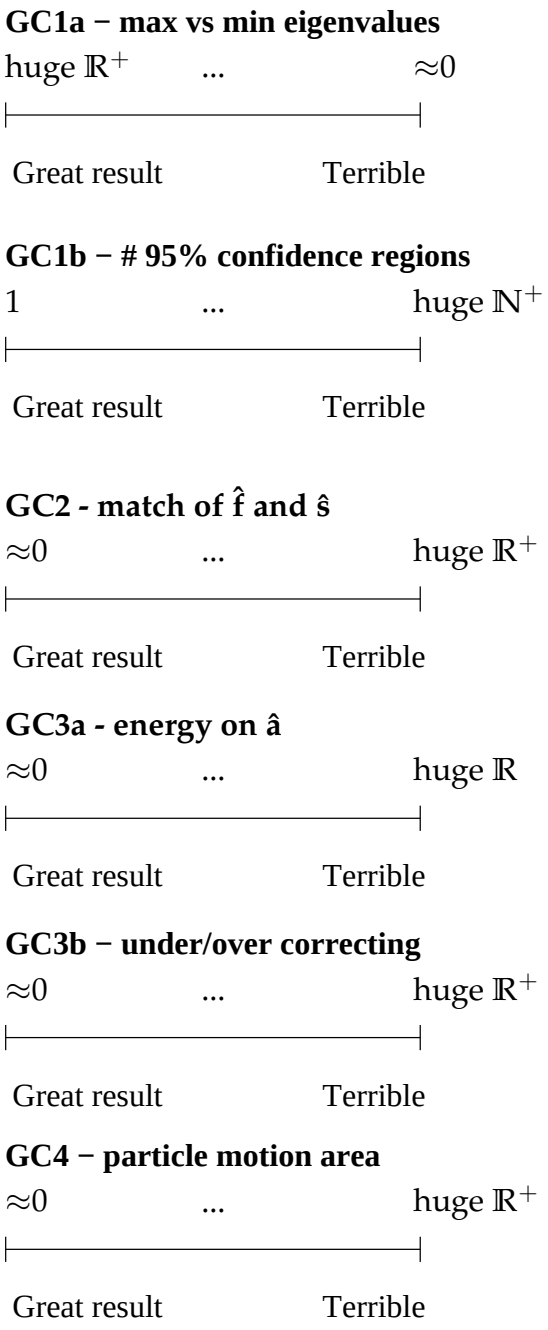


Figure 3.17: Summary of what is a great result and what is a terrible result for each grading criterion.

3.3.2 Using the Criteria to Automate Grading

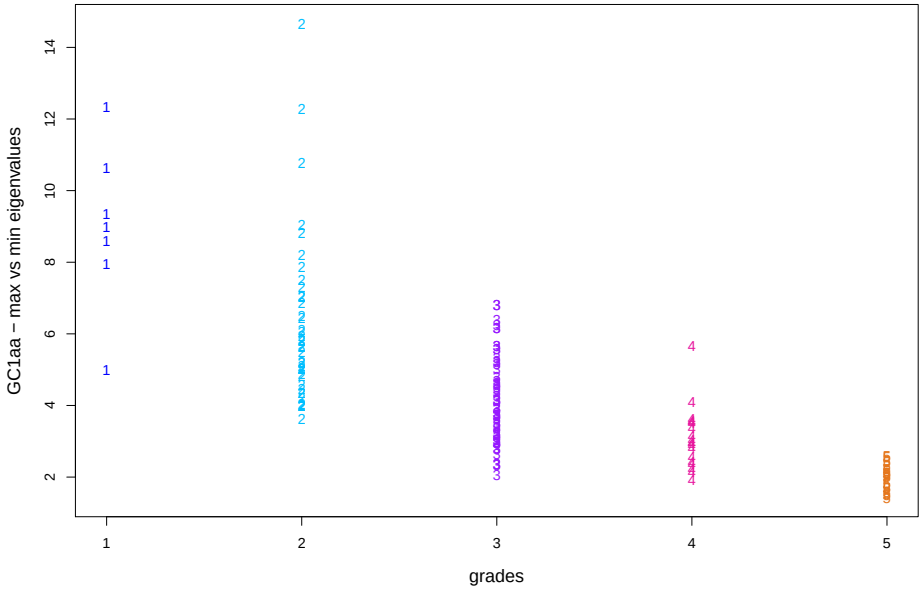
Now that we have established a set of numerical criteria we can can automate the grading process. First this section checks that the grading criteria produce results consistent with Figure 3.17. Then the grading process is automated using multiple linear regression.

Grading Criteria Plots

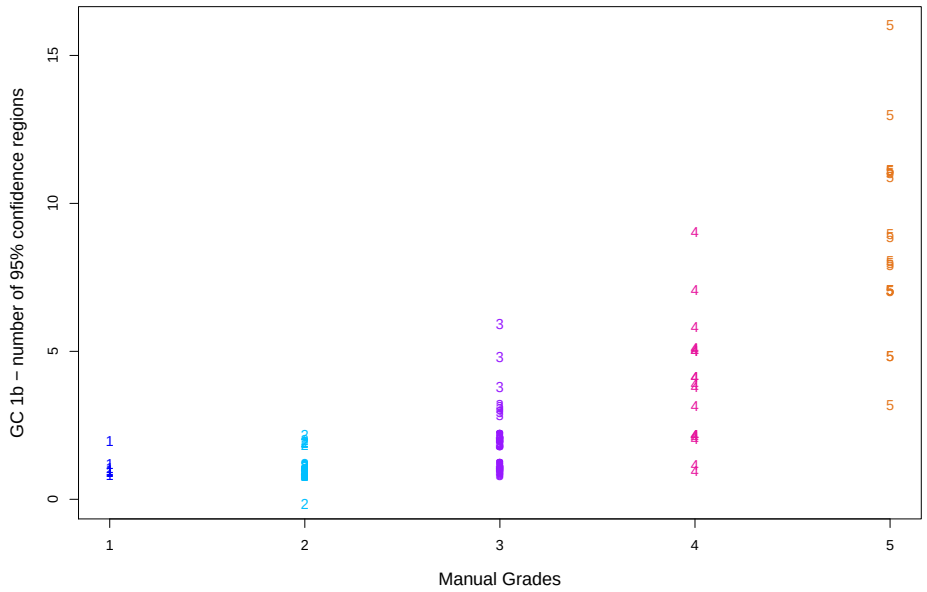
Each criterion was plotted against the manual grades (Figures 3.18, 3.19 and 3.20). Then, each plot was checked to ensure the patterns on each graph were expected. Figure 3.18 contains the two eigenvalue criteria corresponding to the depth (Figure 3.18a, GC_{1a}) and number of confidence intervals (Figure 3.18b, GC_{1b}). As mentioned earlier, we want to evaluate the difference between the maximum and the minimum eigenvalue. Higher quality events have bullseye style contour plots (Figure 3.5) so the difference between the maximum and the minimum eigenvalue should be large. For poor quality events with flat surfaces, this difference should be small. Figure 3.18a (GC_{1a}) has a clear downwards sloping trend, confirming what we expected to see.

Good grades have a single confidence region while poorer grades have many confidence regions. Figure 3.18b (GC_{1b}) has a clear upwards sloping trend confirming this. However, there was something unexpected in this diagram. There is a grade 1 event that has two confidence regions. Looking at its corresponding set of Mfast output plots (Figure 3.21), this event does have two confidence regions. Also, the corrected particle motion area is fairly large so this should not have been classified as a grade 1 result. This highlights the potential of making mistakes during manual grading, especially with a large dataset like the one used in this project.

Figure 3.19a shows the criterion GC_2. Regarding the fast and slow waveforms, we expect high quality grades to be in phase both inside and outside the analysis window. If they are exact then GC_2 is zero. Figure 3.19a shows a trend which is in the opposite direction. To explain this we look at a high quality event (Figure 3.5) and a poor quality event (Figure 3.4) and compare the two. In Figure 3.5 , the waveforms are in phase and the amplitudes are fairly similar inside the window but not outside the window. This means that the rms of the components inside the window will be very small. Since the rms waveforms outside the window is scaled by the rms waveforms inside the window, the denominator will be very small compared to the numerator, making the fit appear worse than it actually is, leading to a higher GC_2 (GC_2=1.48 (2dp)). On the other hand, the waveforms in Figure 3.4 have a poor fit both inside and outside the window. This is creating a cancelling effect when the GC_2 is calculated, leading to smaller GC_2 (GC_2=0.33(2dp)). Therefore, this criterion does not function the way we had hoped, which is outlined in Figure 3.17. We are reluctant to incorporate scaling involving the original fast and slow components because the two wave forms could have been in phase by chance.

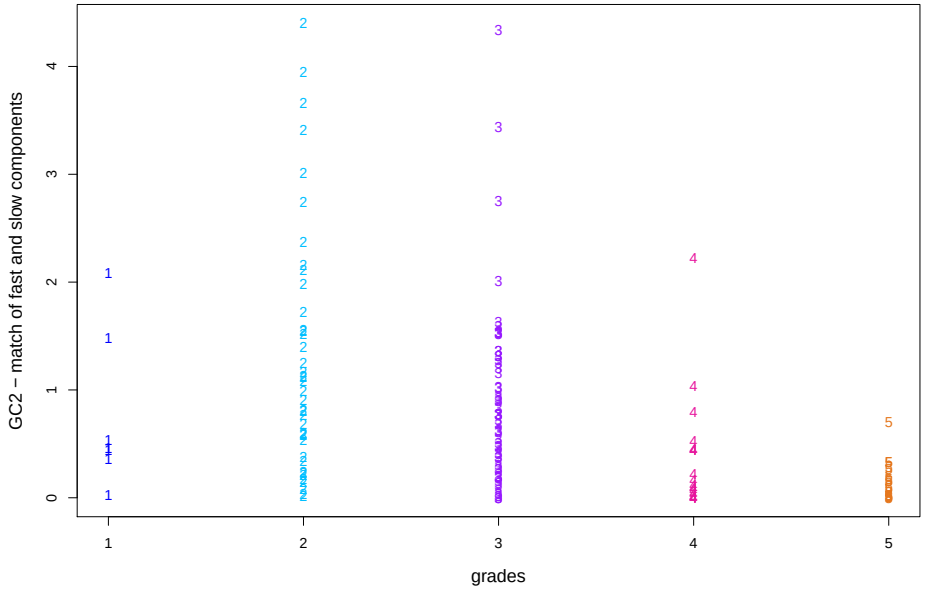


(a) GC_{1a}

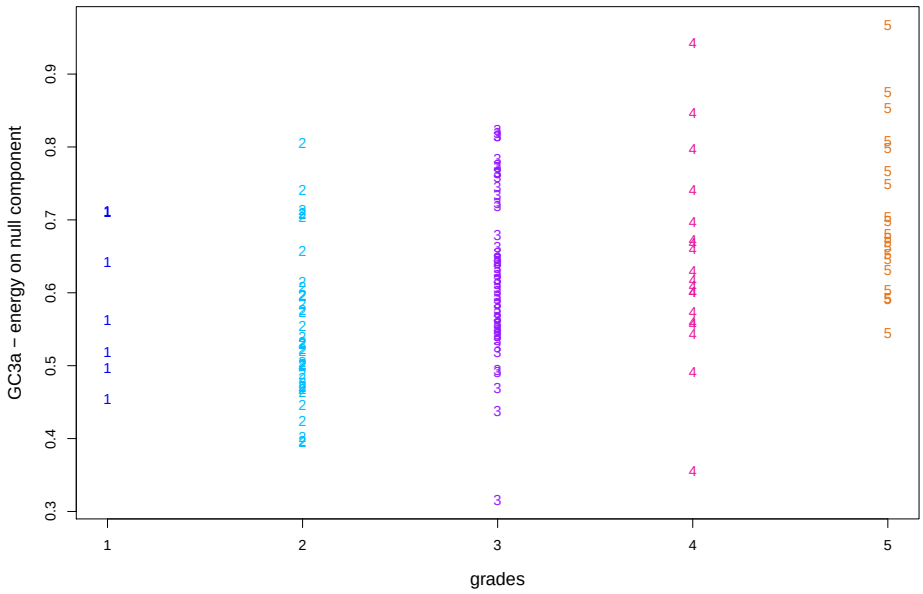


(b) GC_{1b}

Figure 3.18: Eigenvalue grading criteria GC_{1a} and GC_{1b} plotted against the manual grades. The `jitter()` function is applied along the y axis to make results more visible.

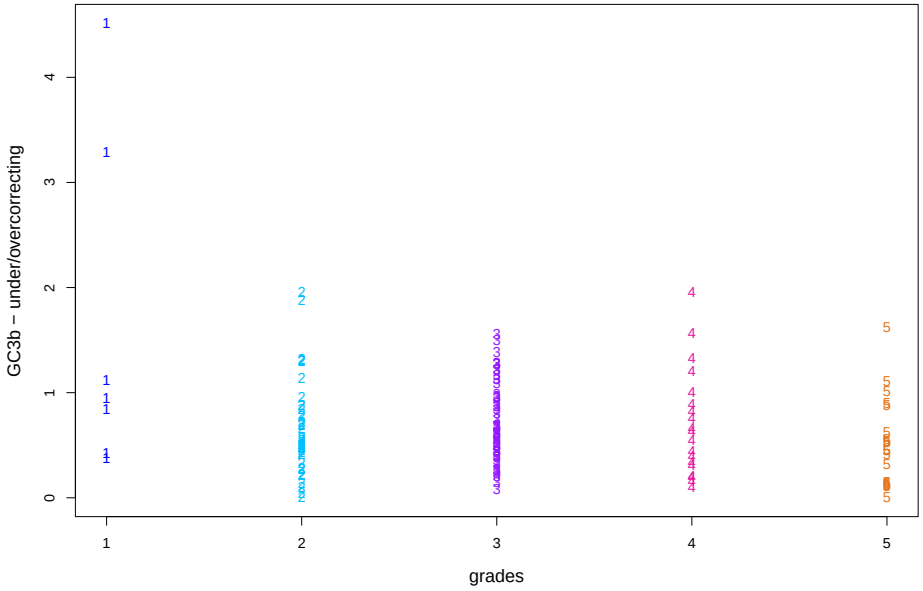


(a) GC_2

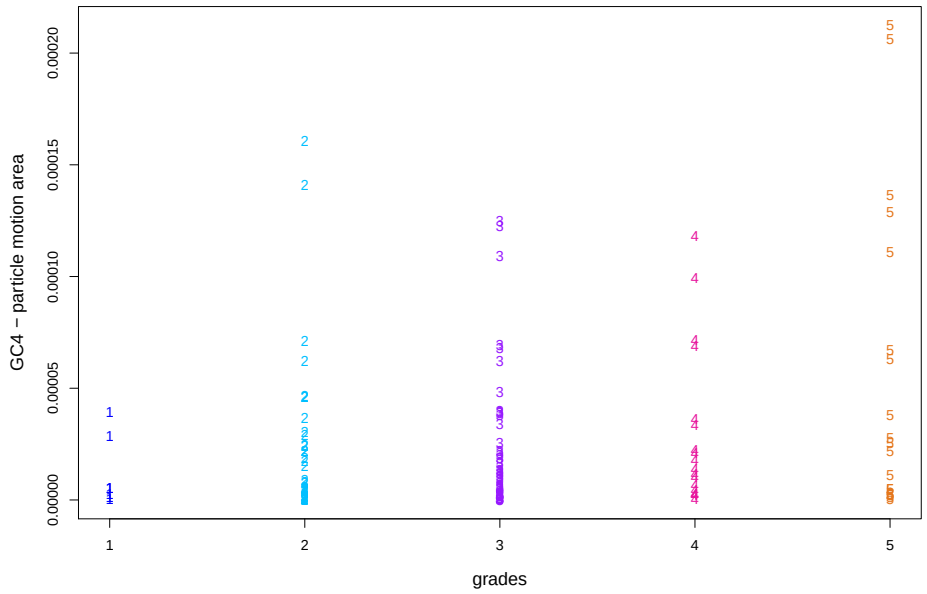


(b) GC_{3a}

Figure 3.19: Waveform grading criteria GC_2 and GC_{3a} plotted against the manual grades. The `jitter()` function is applied along the y axis to make results more visible.



(a) GC_{3b}



(b) GC_4

Figure 3.20: Waveform grading criterion GC_{3b} and particle motion criterion GC_4 plotted against the manual grades. The `jitter()` function is applied along the y axis to make results more visible.

Figure 3.19b shows the criterion GC_{3a} The amount of energy on the null component after correcting for splitting should be small. A large amount of energy suggests that some sort of signal remains instead of the null component only containing noise. High quality events have a small amount of energy on the null component while poorer events have more energy remaining on the null component. Figure 3.19b shows a slight upwards trend and is in agreement with the above reasoning.

The final plot contains GC_{3b} (Figure 3.20a) and GC_4 (Figure 3.20b) Figure 3.20a is the plot corresponding to the grading criteria that checks for under or overcorrecting. The GC_{3b} criterion involves a squared log term so ideally GC_{3b} should have a value close to zero. Large values indicate both under and overcorrecting. Figure 3.20a is sloping slightly downwards, which is opposite to what we expect. There appears to be a few grade 1 events that are creating this trend. If you ignore these two events then GC_{3b} appears constant across grades. Figure 3.22 shows an example of one of these grade 1 events. It appears that some sort of power still remains in the *corp* component (our corrected $\hat{\mathbf{a}}$), suggesting undercorrecting, which could explain the large difference. Since there are only seven grade 1 events, it is difficult to tell whether those two large values of GC_{3b} are unusual for grade 1 events.

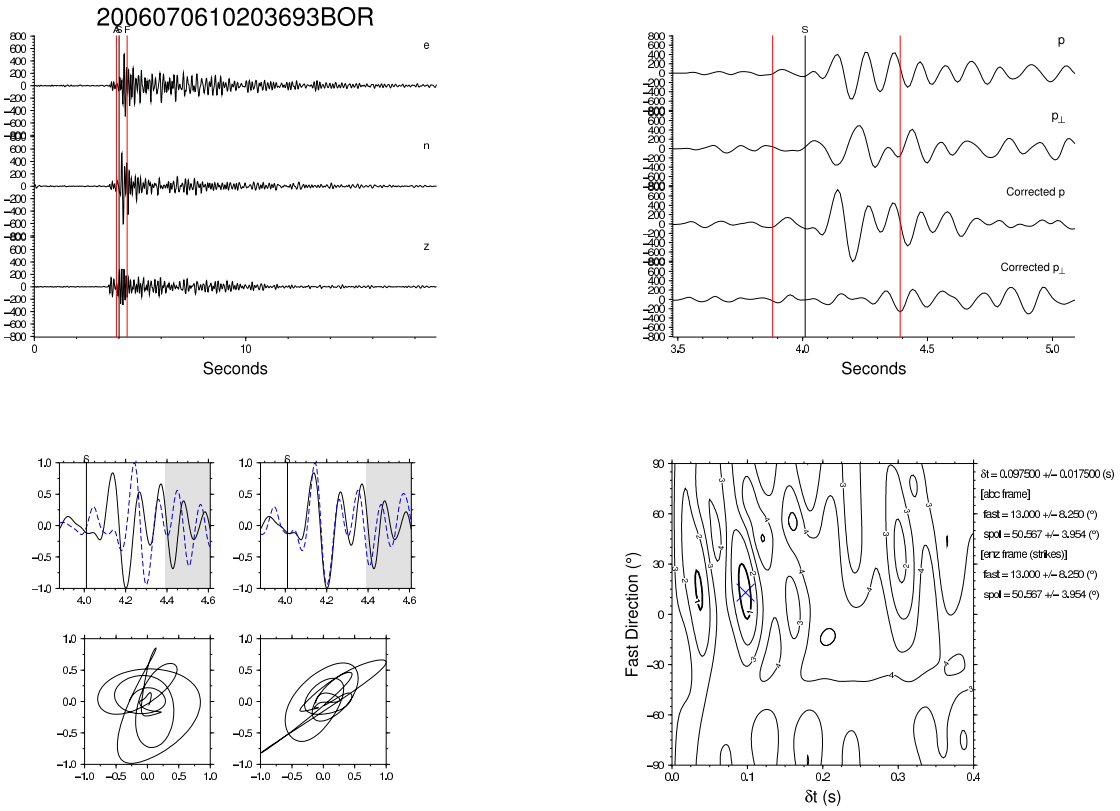


Figure 3.21: Anomalous grade 1 event. Note the absence of bullseye style confidence contours.

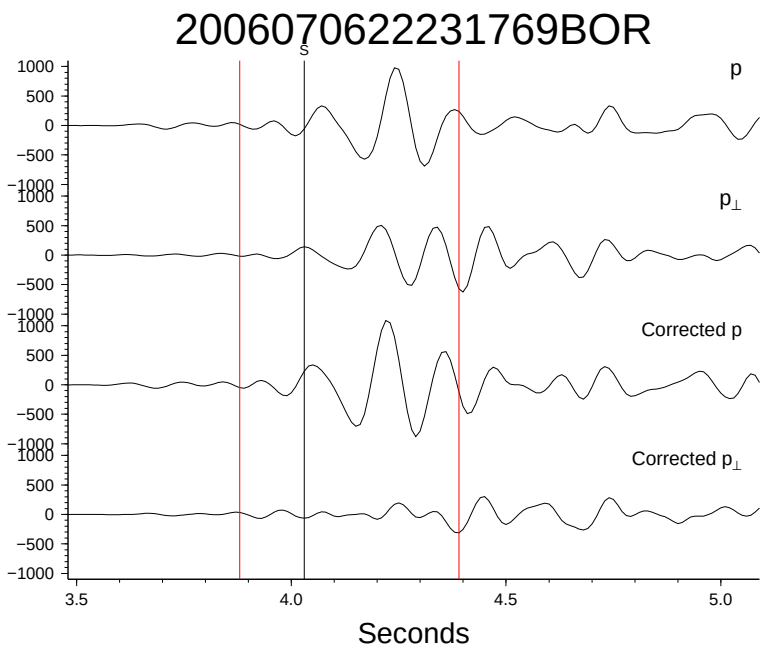


Figure 3.22: Example plot corresponding to a high GC_{3b}.

Lastly, Figure 3.20b (GC_4) shows the particle motion area criterion. High quality grades have linear particle motion after correcting for shear wave splitting so the area should be small. On the other hand, poor quality events tend to still have elliptical particle motion and therefore a larger area. Figure 3.20b confirms this with a slight upwards trend.

Another task these plots helped us to perform was determining whether individual criterion were indicative of the grade. As mentioned earlier, using original components in grading criteria such as GC_2, GC_{3a} and GC_4 is not advisable because there is a chance the particle motion area could be small to begin with, or there might not be much energy on the p component to begin with, or the fast and slow components could be in phase, making it more difficult to detect improvements in the corrected components. This became apparent when we looked at the plots relating to these criteria. Consequently, certain criteria were defined so that they did not refer to the original waveforms (i.e. GC_2, GC_{3a} and GC_4)

Multiple Regression

The previous sections showed how we calculated automated criteria and assessed whether the values from the grading criteria were what we expected. The next step is to use these criteria to assign an automated grade. This is done using multiple linear regression. This creates an automated grade that is a linear combination of the grading criteria. This can be expressed as

$$\widehat{Y} = \widehat{\beta_0} + \widehat{\beta_1} X_1 + \cdots + \widehat{\beta_p} X_p \tag{3.10}$$

where

$$\begin{aligned}\widehat{Y} &= \text{automated grade}\\ \widehat{\beta_i} &= \text{estimated regression coefficients}\\ X_i &= \text{grading criteria}\end{aligned}$$

This should be similar to the manual grade denoted, Y

$$Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p + \epsilon \tag{3.11}$$

where ϵ is a random error. This difference between the observed manual grade and the fitted automated grade is known as a residual

$$e_i = y_i - \widehat{y_i} \qquad i = 1, \cdots, n \tag{3.12}$$

There are several assumptions underpinning this model, namely:

- the sample is representative of the entire population
- the relationship between X (grading criteria) and Y (grade) is linear (which we demonstrated in the previous section)
- ϵ are normally distributed with zero mean and constant variance
- the predictors (our grading criteria) are linearly independent, meaning one criterion cannot be expressed as a linear combination of any of the others (although they may be correlated)

The unknown coefficients are determined using least squares (Rencher, 2000), which minimises the squared distances between the observed values (the manual grades) and the fitted values (the automated grades). `R` has a linear model function `lm()` that can produce these coefficients.

Selecting Criteria for Regression

Not all our grading criteria may be needed to produce a reliable automated grade. In order to decide which criteria to include in the multiple regression, different models are compared using Akaike's Information Criterion (AIC). The AIC for multiple regression (Demidenko, 2004) is defined as

$$\text{AIC} = n \log\left(\frac{\text{SSE}}{n}\right) + 2(p+1) \tag{3.13}$$

where

$$\begin{aligned}n &= \text{number of observations (events)}\\ p &= \text{number of predictors (grading criteria)}\\ \text{SSE} &= \text{sum of squared errors} = \sum_i e_i^2\end{aligned}$$

The second term in the AIC definition ensures that a model with too many parameters that does not lead to a significant reduction in the SSE is penalised accordingly. The best regression model will have the smallest AIC. The AIC is used to compare models, but we also require a method of stepping through different models. Stepwise regression is used. It moves through the models as follows.

1. Start with a model with no predictors (grading criteria)
2. Create all simple regression models with a single grading criterion
3. Select the model with the lowest AIC
4. Add or remove another predictor that results in the lowest AIC
5. Repeat steps 3 and 4 until the AIC cannot be lowered any further

R has a function that performs stepwise regression. It requires a starting model (usually one with no predictors) and a full model with all the predictors. For example

```
stepAIC(lm(Mgrade~1),Mgrade~GC1a+GC1b+GC2+GC3a+GC3b+GC4
 ,direction="both")
```

The parameter `direction=both` is needed so that a grading criterion can be added or deleted at each iteration.

Running the data and models through this function resulted in with the following final model

$$\widehat{Y} = 3.52462 + 0.19111\text{GC}_{1b} - 0.19422\text{GC}_{1a} - 0.21302\text{GC}_{3b} \tag{3.14}$$

A summary of the coefficients in the final model is given below. It provides an estimate of the coefficients and their standard errors. The t value itself is not of interest, but it is used to construct the p-value that is of interest. The p-value tells us how significantly the coefficient differs from zero. Both eigenvalue criteria have very small p-values and therefore are very important in the model. The under/overcorrecting GC_{3b} criterion is not as important as the other two criteria in the model.

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.52462    0.16969  20.771  < 2e-16 ***
GC1b         0.19111    0.01991   9.600  < 2e-16 ***
GC1a        -0.19422    0.02506  -7.751 1.59e-12 ***
GC3b        -0.21302    0.08573  -2.485   0.0141 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Next we check each coefficient to see if they make sense to us. The first coefficient in the regression model is the intercept. If no grading criteria were available then this is the automated grade allocated. Table 3.1 shows that the quality of our events is centred around 3, so the intercept coefficient makes sense. The first grading criterion in the model is the number of 95% confidence regions. The sign of the coefficient is positive so high quality events with only one region will receive a smaller contribution to the final grade. Conversely, as the events get poorer in quality, the number of confidence regions should increase, resulting in a higher contribution to the final grade (Figure 3.17). The next grading criterion in the model is the difference between the maximum and minimum eigenvalues. The sign of this coefficient is negative, so a large difference in the two eigenvalues produces a smaller final grade, while a smaller difference produces a larger final grade. The difference in eigenvalues indicates the steepness of the confidence region, so the steeper the region, the larger GC_{1a}, the smaller the grade (Figure 3.17). The third grading criterion in the model is the checking for under and overcorrecting. The negative sign is consistent with the slight downward sloping trend in Figure 3.20a, which is most likely caused by scattered phases arriving after the shear wave. All coefficients are consistent with the figures in the previous section (Figures 3.18a, 3.18b and 3.20a). This model should produce a low $\widehat{Y}$ for the high quality events and a high $\widehat{Y}$ for poor quality grades.

To summarise, the stepwise regression has chosen a model consisting of all the eigenvalue based criteria and the criterion that checks for under or overcorrection. The three criteria chosen by AIC for the model were

- number of 95% confidence regions (GC_{1b})
- difference between the maximum eigenvalue and the minimum eigenvalue (GC_{1a})
- ratio of the corrected windowed null component and the corrected null component in front of the window (GC_{3b})

Therefore there are only two key plots (contour map of the eigenvalues, corrected $\hat{\mathbf{a}}$ component) that are needed to produce an automated grade. The model does not include the following criteria

- comparing slow and fast coordinates inside and outside the window (GC_2)
- ratio of the windowed corrected a component and the windowed corrected p component (GC_{3a})
- the particle motion area measure (GC_4)

This means that when all the other grading criteria are used, adding these criteria into the model does not significantly improve correctly predicting the automated grade. Table 3.3 shows the AIC of each model. The final iteration that adds GC_{3b} to the model only results in a slightly smaller AIC so it is possible that this criterion could be removed from the final model without affecting the automated grades too much.

Model	AIC
$\widehat{y} = \beta_0$	21.47
$\widehat{y} = \beta_0 + \beta_2 \text{GC}_{1b}$	-105.16
$\widehat{y} = \beta_0 + \beta_2 \text{GC}_{1b} + \beta_1 \text{GC}_{1a}$	-153.92
$\widehat{y} = \beta_0 + \beta_2 \text{GC}_{1b} + \beta_1 \text{GC}_{1a} + \beta_5 \text{GC}_{3b}$	-158.13

Table 3.3: Chosen model and AIC at each iteration.

Finally, we compare the automated grades against the manual grades (Figure 3.23). If the model is good, the points should be close to the diagonal blue line, in other words the automated grade is similar to the manual grade. The points are roughly centred around the diagonal line, with most of the automated grades differing from the manual grades by no more than 1. A cut off of an automated grade of 3 captures almost all of the good grades and only a few that have a grade of 4. Therefore, the automated method can distinguish between good and bad results. A few points in Figure 3.23 look unusual (illustrated by red points) so they are investigated further below. Most of the grading criteria in the model are based on the contour plots, so these are used to investigate the unusual grades (Figure 3.24).

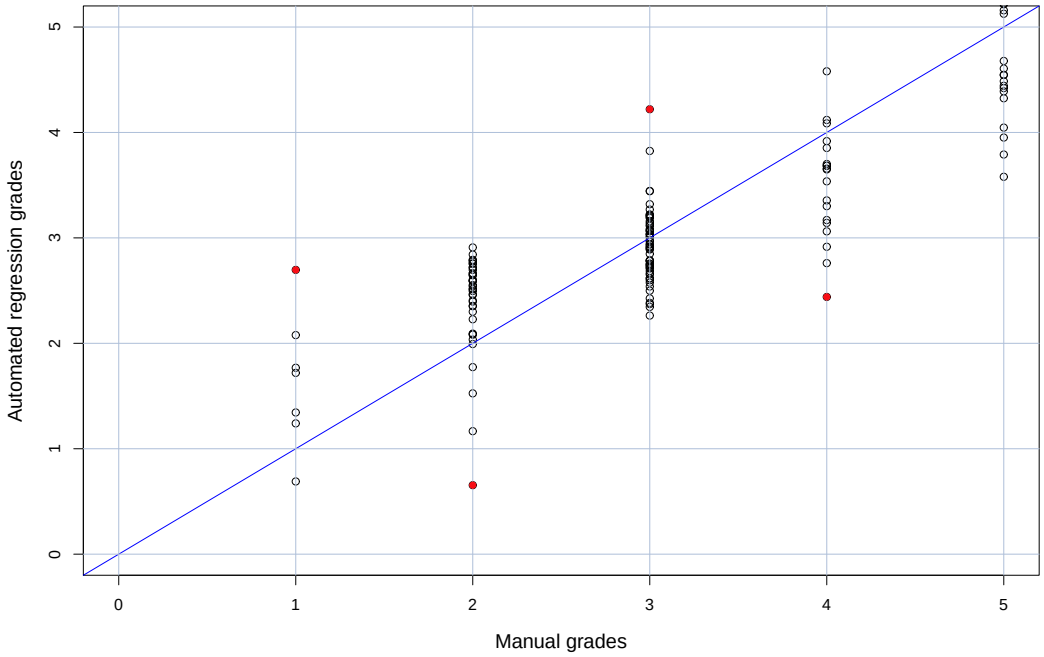
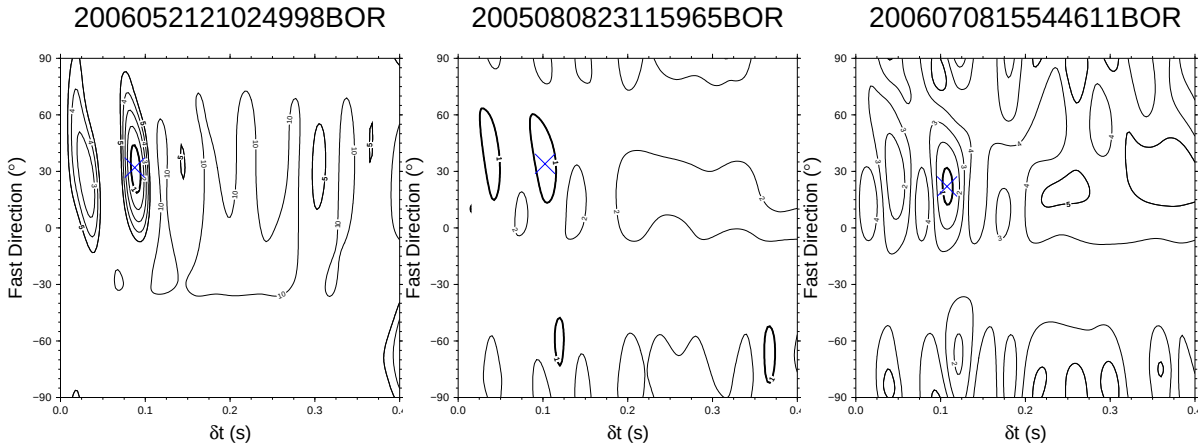


Figure 3.23: Comparison of automated grades to manual grades. Red points indicate large discrepancies. Blue line indicates agreement between automated and manual grades.



(a) M=2, A=0.82 (b) M=3, A=4.17 (c) M=4, A=2.35

Figure 3.24: Contour plots showing discrepancies in grading. M denotes manual grade and A denotes automated grade.

The first anomalous result was mentioned earlier, which was the misclassified grade 1 event (Figure 3.21). It is clearly not a grade 1 event, which is why the automated grade is so much higher than the manual grade. The other three anomalous cases have their contour maps displayed in Figure 3.24. The event corresponding to Figure 3.24a was given a very low automated grade relative to the manual grade. Looking at the contour plot, there is only 1 small 95% confidence region and there are bullseye style contours. This should be a grade 1 event as the automated grade suggests. The event corresponding to Figure 3.24b was given a very high automated grade. Looking at the contour plot shows multiple long 95% contour regions and a fairly flat surface. This should be a grade 4 event. Lastly the event corresponding to Figure 3.24c has a single small 95% confidence region and a moderately steep error surface. It has a single confidence region but not a bullseye style contour plot. It is possible that this could be a grade 2 event. Again this highlights the problem of grade misclassification when a large set of events must be manually graded.

3.3.3 Suitability of Model

The previous section generated a model that could automate grades. However, this model is based on assumptions that must be checked. Also, the events need to be checked to ensure that a single event is not having too much influence on the automated grades or the grading criteria coefficients.

Checking Assumptions

In order to test assumptions regarding the residuals we use the more robust jackknife residual. It takes into account the influence a single point can have by utilising the hat matrix. The hat matrix is defined as

$$\mathbf{H} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T \tag{3.15}$$

where $\mathbf{X}$ is the linear model matrix corresponding to Equation 3.11, our observed values for each criterion included in the model. If the manual grades are in the vector $\mathbf{y}$, then the hat matrix converts the observed values into fitted models under the linear model

$$\widehat{y} = \mathbf{H}y \tag{3.16}$$

The hat matrix is used to calculate the jackknife residual, which is defined as

$$r_{(-i)} = \frac{e_i}{\sqrt{\text{MSE}_{(-i)}(1 - h_{ii})}} \tag{3.17}$$

where

$$\begin{aligned} r_{(-i)} &= \text{jackknife residual of event } i \\ e_i &= \text{residual of observation } i \\ \text{MSE}_{(-i)} &= \text{MSE after removing event } i \\ h_{ii} &= i\text{th diagonal element of the hat matrix} \end{aligned}$$

If removing a single event dramatically decreases the MSE it results in a larger residual and indicates that the event is highly influential in the regression.

The fitted model must be checked to ensure that the assumptions mentioned in Section 3.3.2 are not violated. If they are then the results from the multiple regression are invalid. This can be done by looking at the scatter plots and quantile-quantile plots (qq plot). The qq plot will show if the errors are normally distributed by comparing the jackknife residuals to the quantiles of a standard normal distribution. If the residuals are normally distributed then the points should lie on a diagonal straight line. Scatter plots of the jackknife residuals against the fitted automatic grades or the predictors (grading criteria) can check for non constant variance and non linearity. These plots should look random and have no structure to them.

The qq plot in Figure 3.25 looks relatively straight. To be certain the jackknife residuals were consistent with a normal distribution a Shapiro-Wilk test was conducted. This is a hypothesis test that specifically tests for normality. If the p-value > 0.05 then we have no reason to doubt that the data are normally distributed. `R` has a `shapiro.test()` function that returns a p-value as shown below

```
Shapiro-Wilk normality test

data:  jkr
W = 0.9862, p-value = 0.1537
```

The Shapiro-Wilk test shows there is no reason to doubt that the residuals are normally distributed.

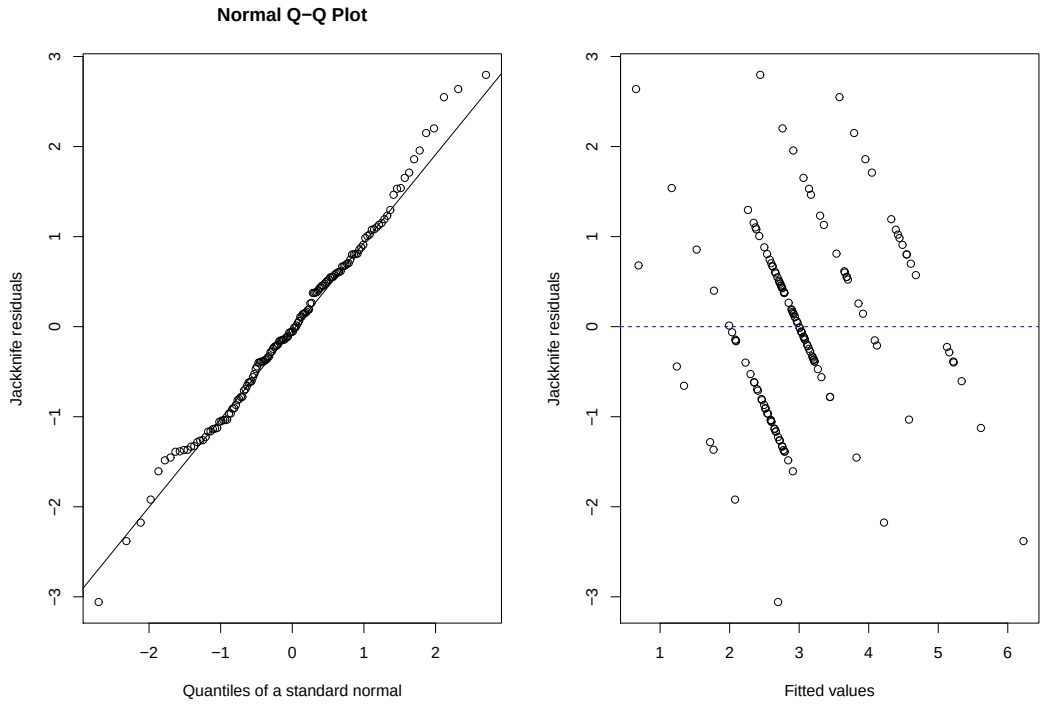


Figure 3.25: Diagnostic Plots checking normality and constant variance.

Plots of the residuals are used to confirm linearity and constant variance. The jackknife residuals plotted against the fitted values (automated grades) (Figure 3.25) have no structure (i.e. an even spread) and no outliers (extreme values). The points lie on one of five diagonal lines. This is because the manual grades are integers (1,2,3,4,5). Figure 3.26 shows plots of the jackknife residuals against each predictor. All of these plots have an even spread, so the variance is constant. However, there do appear to be a few extreme points that look quite different from others. These are investigated properly in the next section.

Finally, we check how correlated the variables are. When multicollinearity is present the grading criteria are highly correlated with each other. Figure 3.27 shows the scatter plots for each criterion, compared to all the other criteria. Correlations and p-values are given in the lower diagonal. Labels are down the main diagonal indicating which plot has which criteria on its x and y axis and what the correlation coefficient is for those two criteria. The figure shows that all criteria are only moderately correlated, with the strongest correlation being between GC_2 and GC_{3a} (r=-0.57). Since all the criteria are only moderately correlated, multicollinearity is not a problem.

Checking for Influential Observations

The final regression model chosen to produce automated grades is based on the 146 events. To check the suitability of the model we should ensure that the final model is not being overly influenced by a few events. The following are checked to ensure the final model has not been heavily influenced:

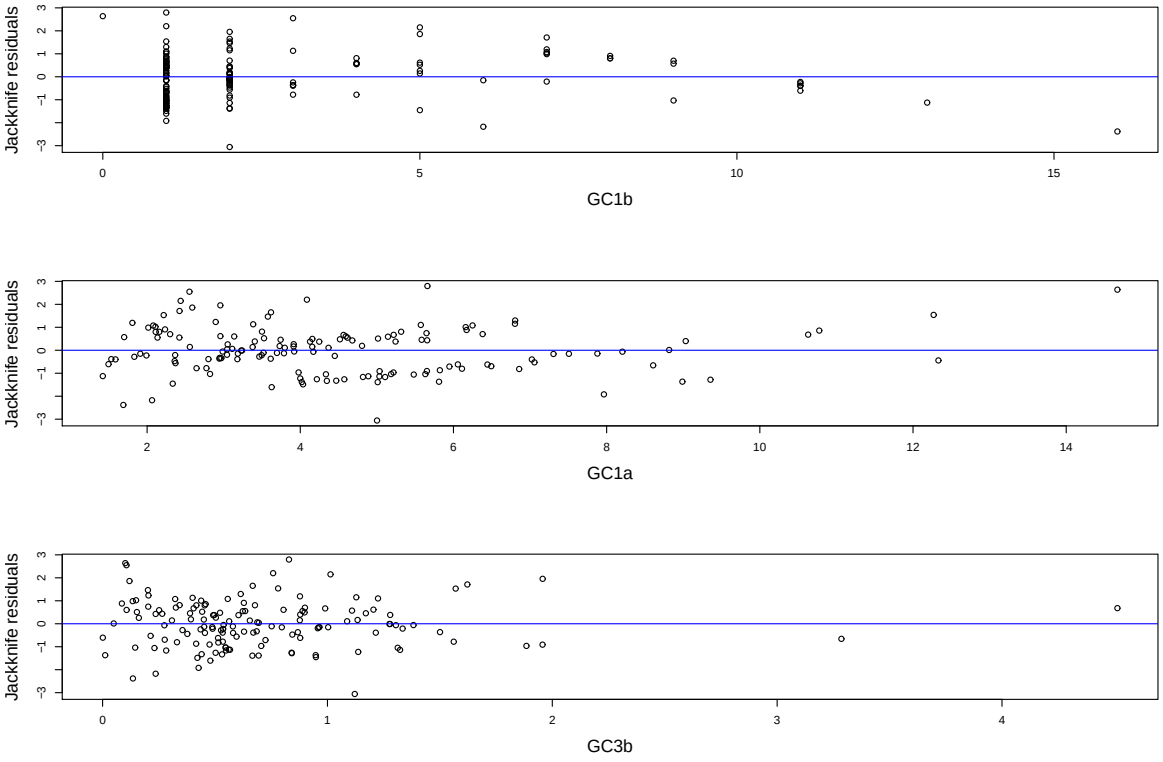


Figure 3.26: Diagnostic plots of the jackknife residuals for each coefficient.

- influence on the fitted values (automated grades)
- influence on the coefficients
- influence on summary statistics

First we check for potential influence of (the automated grades). This can be done by using Cook's Distance, which measures the change in an automated grade when an event is deleted and the final regression model using the chosen grading criteria is recalculated (which may lead to different β coefficients for each grading criterion). This measure is defined as

$$D_i = \frac{\sum_{k=1}^{n} (\widehat{y}_k - \widehat{y}_{k(i)})^2}{(p+1)\text{MSE}} \tag{3.18}$$

where

$$\begin{aligned} D_i &= \text{Cook's Distance for the } i\text{th observation (event)} \\ \widehat{y}_k &= \text{the fitted value for observation } k \\ \widehat{y}_{k(i)} &= \text{the fitted value for observation } k \text{ when observation } i \text{ is removed} \\ p &= \text{number of predictors} \\ \text{MSE} &= \text{mean squared error} \end{aligned}$$

Based on this formula, one can see that the bigger the numerator, the bigger the difference in fitted values when that particular event is removed and therefore, the more influential that

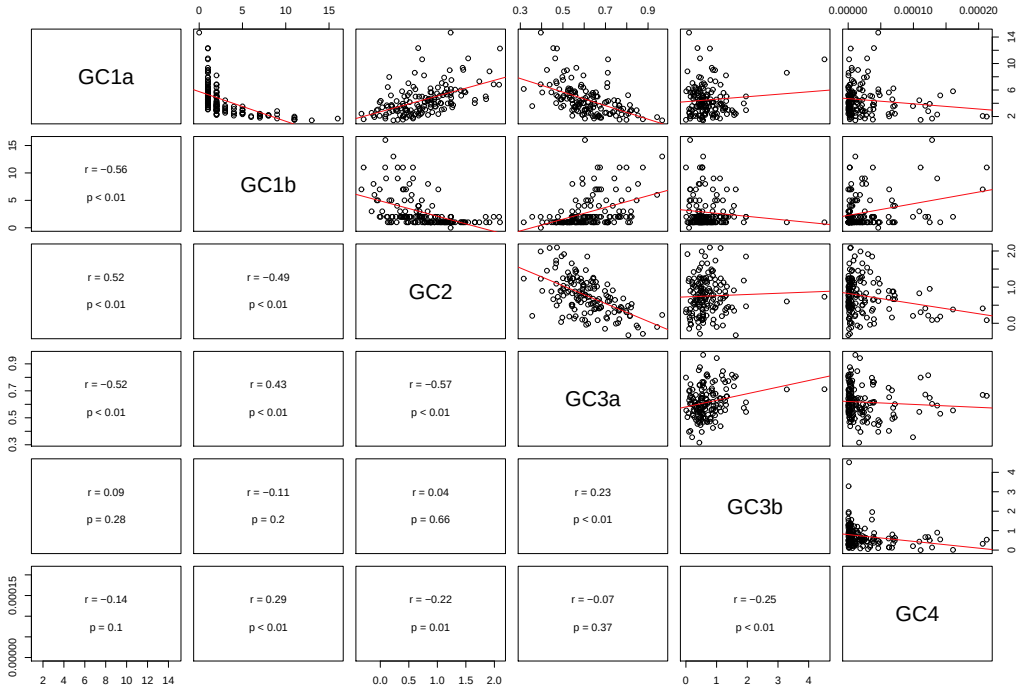


Figure 3.27: Scatterplot matrix of predictors along with their correlation (r) and p-value, p.

point is. The spread in $\hat{y}$ should be similar to the MSE. The cut off for what constitutes an influential point is

$$D_i > F^{\alpha=0.5}_{(p+1,n-p-1)} \tag{3.19}$$

In other words, any point with a D_i that exceeds the 50th percentile of that F distribution. The F distribution comes from looking at joint confidence regions and is explained in detail in Rawlings et al. (1998).

Figure 3.28 displays the results for Cook's Distance (bottom plot). The horizontal dashed line is the cut off. Since no values exceed this, none of the events significantly influence the automated grades.

Next we check whether a few observations are influencing the β coefficients (Equation 3.10). This can be done by calculating the DFBETA of each observation as follows

$$\text{DFBETAS}_{j,i} = \frac{\hat{\beta}_j - \hat{\beta}_{j(i)}}{\sqrt{c_{jj}\text{MSE}_{(i)}}} \tag{3.20}$$

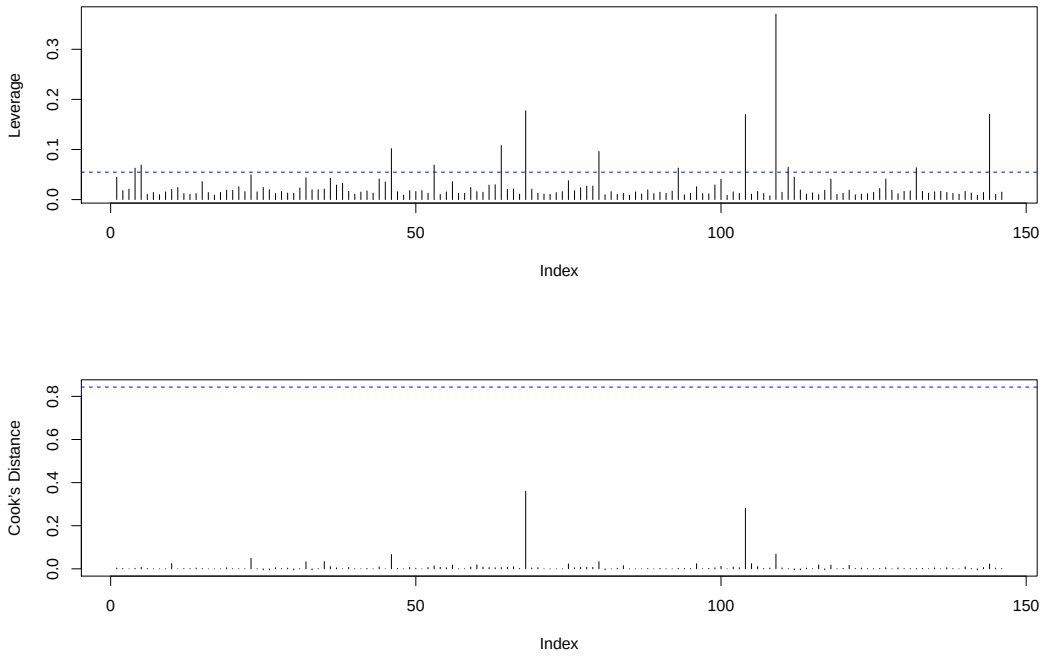


Figure 3.28: Plots of leverage points and Cook's distance. Values that exceed the horizontal dashed blue line are considered leverage and influence points respectively.

where

$$\begin{aligned}
\text{DFBETAS}_{j,i} &= \text{DFBETAS for event } i \text{ and coefficient } j \\
\widehat{\beta}_j &= j\text{th coefficient using all observations} \\
\widehat{\beta}_{j(i)} &= j\text{th coefficient calculated by using all but the } i\text{th observation} \\
c_{jj} &= j\text{th diagonal component of the matrix } (\mathbf{X}^T\mathbf{X})^{-1} \\
\text{MSE} &= \text{mean squared error}
\end{aligned}$$

This measures the standardised change in a grading criterion coefficient when an event is removed. The larger the DFBETA, the more influence the event has on a particular coefficient. The denominator was chosen so that the same proportion of points will be classified as influential regardless of the size of the dataset (Belsley et al., 2004). Any absolute DFBETA that exceeds

$$|\text{DFBETAS}_{j,i}| > 2/\sqrt{n} \tag{3.21}$$

is considered influential.

Figure 3.29 shows the DFBETAs for the three chosen grading criteria. Since no points exceed the horizontal dashed line, GC_{1b}, GC_{1a} and GC_{3b} have no influential points so the grading criteria coefficients are not heavily influenced by a few events.

Finally we check if there are any points affecting the summary statistics. These are called leverage points. In order to identify these we use the hat matrix (Equation 3.15). The trace

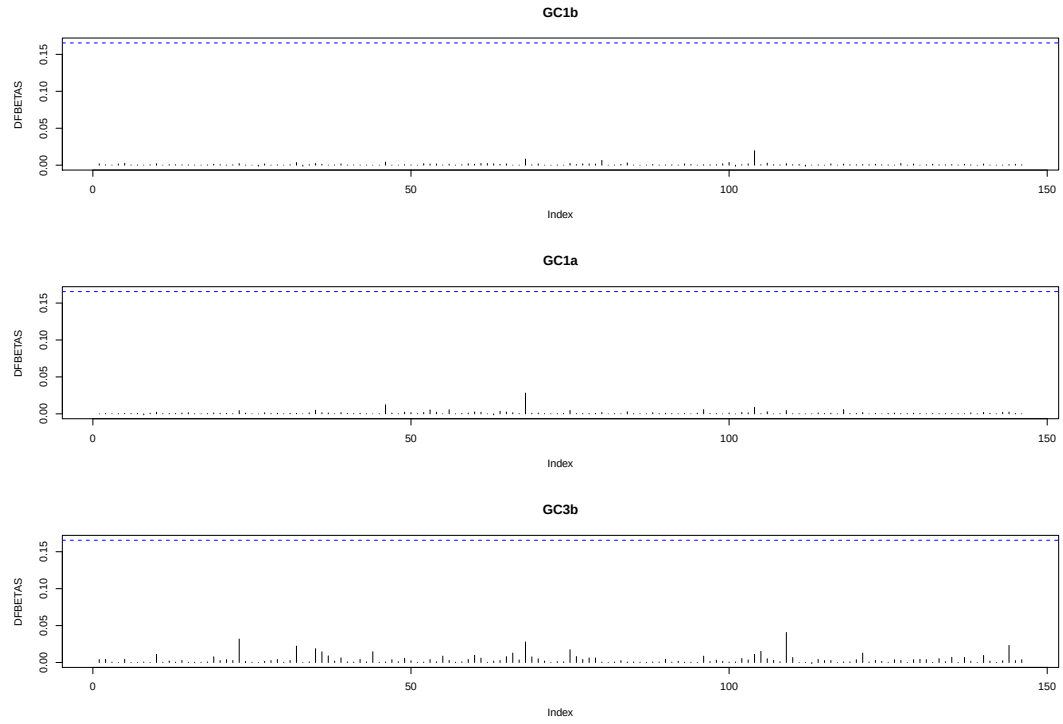


Figure 3.29: Plots of the DFBETAs for each criterion used in the multiple linear regression. Horizontal dashed blue line is the cutoff.

of this matrix **H** has a special property

$$tr(\mathbf{H}) = \sum_{i=1}^{n} h_{ii} = p + 1 \tag{3.22}$$

If each event contributed equally to the regression results, then based on the above equation each h_{ii} should be

$$h_{ii} = \frac{p+1}{n} \tag{3.23}$$

where

$$\begin{aligned} n &= \text{number of events} \\ p &= \text{number of grading criteria} \\ h_{ii} &= i\text{th diagonal component of the hat matrix} \end{aligned}$$

Hoaglin and Welsch (1978) argued any diagonal entry of the hat matrix that exceeds

$$h_{ii} > \frac{2(p+1)}{n} \tag{3.24}$$

corresponds to the ith event being a leverage point. Figure 3.28 shows the leverage plot (top plot) shows. There are 13 values that exceed the horizontal line so these are events are considered leverage points. Again some of these are the misclassified events such as the events corresponding to Figure 3.24a.

Extreme data points, known as outliers, can also affect results. Checking for outliers can be done by checking for points that have a residual more than 3 standard deviations away. For jackknife residuals, any residual that exceeds

$$r_{(-i)} > 3 \times \sqrt{\frac{1}{(n-p-1)-1} \sum_{i=1}^{n} r_{(-i)}^2} \tag{3.25}$$

where

$$\begin{aligned} n &= \text{number of events} \\ p &= \text{number of grading criteria} \\ r_{(-i)} &= \text{jackknife residual of event } i \end{aligned}$$

is considered an outlier. Using this rule none of the events are considered outliers.

3.3.4 Summary of Model Suitability

The assumptions behind the multiple linear regression model appear to hold, so the automated grades are valid. No outliers have been identified and based on Cook's Distance, there are no events that significantly influence the automated grades. However, there are a few leverage points that can influence summary statistics.

Even with this problem, the automated method works very well. The bulk of the automated grades are consistent with the manual grades. The automated method was able to identify incorrect manual grades by producing large discrepancies between the automated and the manual grades. Using a cutoff automated grade of 3 the automated method captures most of the good grades and only a couple of bad grades. Most of the automated grades are no more than 1 grade away from the manual grade. Our dataset does not contain any null measurements so we have not incorporated null detection into this model, but this could be incorporated in the future by creating additional criteria that can detect null measurements.

Chapter 4

Incoming Polarisation & Cycle Skipping

Results from the previous chapter suggested that some events in our dataset may have parameter estimates affected by cycle skipping. In this chapter we investigate whether our dataset shows signs of cycle skipping and how this affects the fast direction and delay times. We also explore what effect cycle skipping has on the incoming polarisation.

4.1 Incoming Polarisation

The incoming polarisation (α) is an angle related to the wave displacements before the wave splits. In all the implementations of the Silver and Chan (1991) method, angles are measured clockwise from North in bearing coordinates. Consequently, $\hat{\mathbf{n}} \cdot \hat{\mathbf{p}} = \cos(\alpha)$ (Figure 4.1). It is also referred to as the S polarisation and source polarisation. Silver and Chan (1991) do

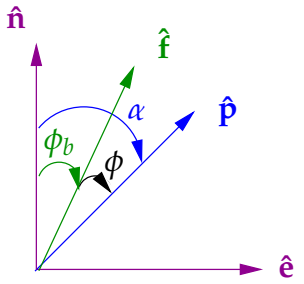
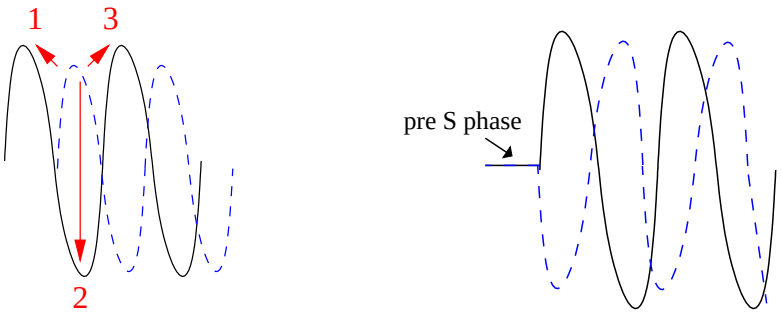


Figure 4.1: Incoming polarisation angle in geographical coordinates.

not mention the incoming polarisation angle in their article but it is mentioned here because it is used to calculate the degrees of freedom (see Section 5.2 for more details). It is also useful to study various types of geophysical parameters. Figure 4.1 shows that the incoming polarisation for a near vertical ray can be calculated by

$$\alpha = \phi_b + \phi \tag{4.1}$$



(a) Potential cycle skipping

(b) No cycle skipping

Figure 4.2: Illustration of cycle skipping. Cycle skipping can occur here because the dashed blue peak can be matched to either peaks or the trough of the solid black wave. Adding a pre S phase means that the two waveforms are more likely to be matched correctly.

The grid search for the best two parameters involves the fast direction relative to North (i.e. ϕ_b). The other ϕ is obtained by calculating the arc tangent of the two components of the first eigenvector of the covariance matrix (Equation 2.88) $\mathbf{x}_1 = [x_{11}x_{12}x_{13}]^T$ (which is $\hat{\mathbf{p}}$ if there is no noise) corresponding to the eigenvalue λ_1

$$\alpha = \phi_b + \text{atan2}(x_{11}, x_{12}) \tag{4.2}$$

where atan2 is a computer implementation of the arc tangent which returns a value between $-\pi$ and π (in radians) except when both parameters are zero, in which case atan2 is undefined (Yoshikawa, 1990).

4.2 Cycle Skipping

As mentioned in Section 1.3.3, cycle skipping is a phenomenon caused by misfitting peaks of a waveform with different peaks or a trough (see Figure 4.2). One way to reduce this problem is to include a segment of the wave prior to the S arrival so that the waves can be matched correctly (see Figure 4.2b). However, in real data it may not always be clear when the shear wave arrives.

There are several observations that can indicate cycle skipping. Firstly, mismatching the two waveforms can produce delay times that are incorrect by some integer multiple of half the dominant period (T) of the waveform i.e.

$$\delta t_{CS} = \delta t \pm nT/2 \tag{4.3}$$

where

$$\begin{aligned} \delta t &= \text{delay time} \\ \delta t_{CS} &= \text{cycle skipped delay time} \\ T &= \text{dominant period} \\ n &= \text{positive integer} \end{aligned}$$

This can be seen as several troughs in the contour plot that are half a dominant period apart. It can also be seen in the wave forms. For an odd n, a trough and peak have been matched and for an even n, an entire cycle has been skipped so the incorrect peaks have been matched. Secondly, the contour plots may have fast directions that are 90° apart (i.e. $\hat{\mathbf{s}}$ and $\hat{\mathbf{f}}$ are flipped) with δt being half a period earlier (or later) or a whole period later if they have the same fast direction indicating an entire period has been skipped (Matcham, 1997). Thirdly, if events are identical, as is the case here, substantially different fast directions and delay times can indicate cycle skipping.

4.3 Checking for Cycle Skipping

To assess whether some of the events in this dataset suffer from cycle skipping, the three indicators mentioned previously are checked.

All 146 events in our test dataset are supposedly identical so the splitting parameters $(\phi, \delta t)$ should be similar. Figure 3.10 shows two main clusters of splitting parameters along with a few others that do not belong to either. The two clusters have different delay times, which could indicate cycle skipping. Since the group consists of fast directions between 0 and 45°, this graph does not provide information about whether there are other measurements in which the fast directions have a 90° ambiguity. This is because Figure 3.10 only contains fast directions between 0 and 45°. Such occurrences are expected to be in the group of events (group 2) we excluded from the analysis (see Figure 3.3).

Next a set of output plots is taken from each cluster and their contour plots and waveforms are compared (Figures 4.3 and 4.4 respectively). Figure 4.3 shows the contour plots have two 95% confidence regions that look roughly half a period apart. This can be checked by calculating the dominant period using a script. For both events the dominant frequency is 8.19847. This means the dominant period T is

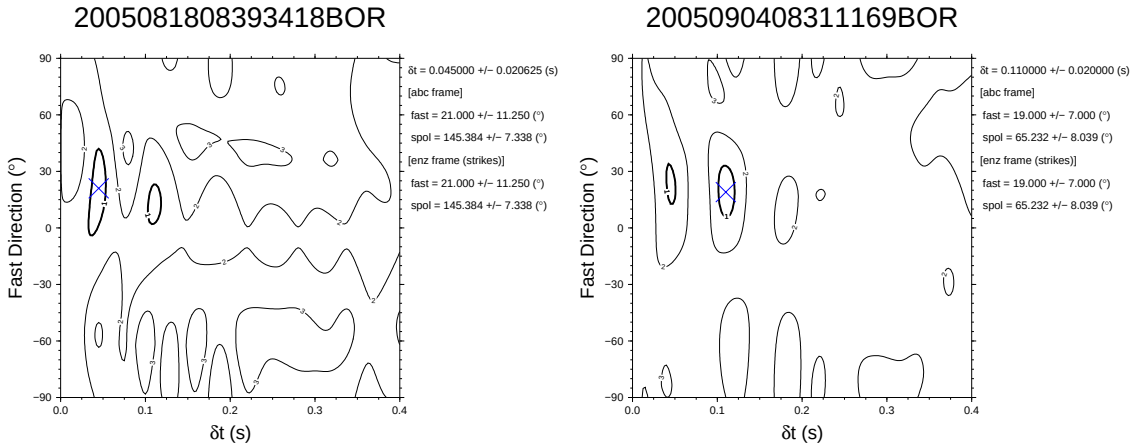
$$T = 1/f = 1/8.19847 = 0.122 \text{ sec (3dp)} \tag{4.4}$$

Substituting this into Equation 4.3 gives

$$\begin{aligned} \delta t_{CS} &= \delta t + nT/2 && (4.5)\\ &= 0.045 + (1) \times 0.121974/2 \qquad \text{(for } n = 1\text{)} && (4.6)\\ &= 0.105 \text{ (3dp)} && (4.7) \end{aligned}$$

The second event has a delay time of 0.11. This delay time and the other one both have errors of at least ± 0.02 so the second event has a delay time that is roughly half the dominant period larger. This suggests that at least one of the events suffers from cycle skipping.

Next the waveforms are compared. Figure 4.4 indicates the time shifts with red arrows. Figure 4.4a matches the peak of a slow wave to the trough of the fast wave. Figure 4.4b has matched two peaks. When a trough and peak are matched, Mfast flips the sign of the



(a) Contour plot - small δt

(b) Contour plot - large δt

Figure 4.3: Two contour plots of different events from group 1 (fast directions 0-45°). Note the splitting parameters differ (ϕ is similar but δt is different).

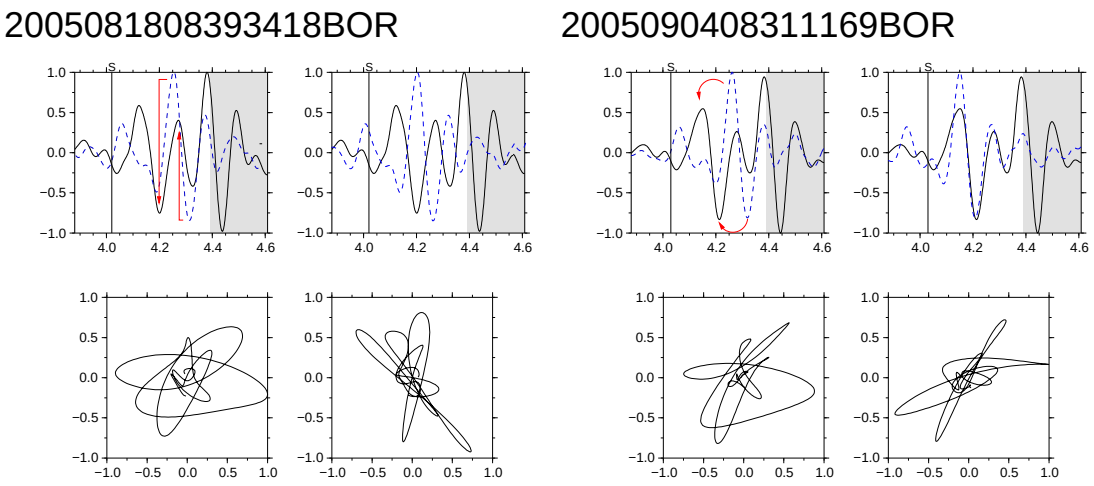
slow wave displacements so that the match becomes more obvious. This gives the illusion that two peaks have been matched. However, if the particle motion is pointing down to the bottom right corner of the graph (like Figure 4.4a) then a trough and peak have been matched. To avoid confusion, the waveform plot has been edited to show the matching of a peak to a trough. The mismatched waveforms suggests that cycle skipping is occurring.

Since the effect on the fast direction cannot be checked for with this group, events from group 2 which consist of fast directions between 90-120° are analysed. Figures 4.5 and 4.6 show the contour plots and waveforms for two events from group 2 respectively. Figure 4.5a has two 95% confidence regions with similar fast directions and different delay times. Figure 4.6a shows that a peak has been matched to a peak that is 2 periods away. Figure 4.5b is similar. The peaks have been matched to peaks which are two periods away, which explains the larger delay times compared to group 1. The existence of fast directions that are 90° away also suggests cycle skipping.

4.4 Effects on the Incoming Polarisation

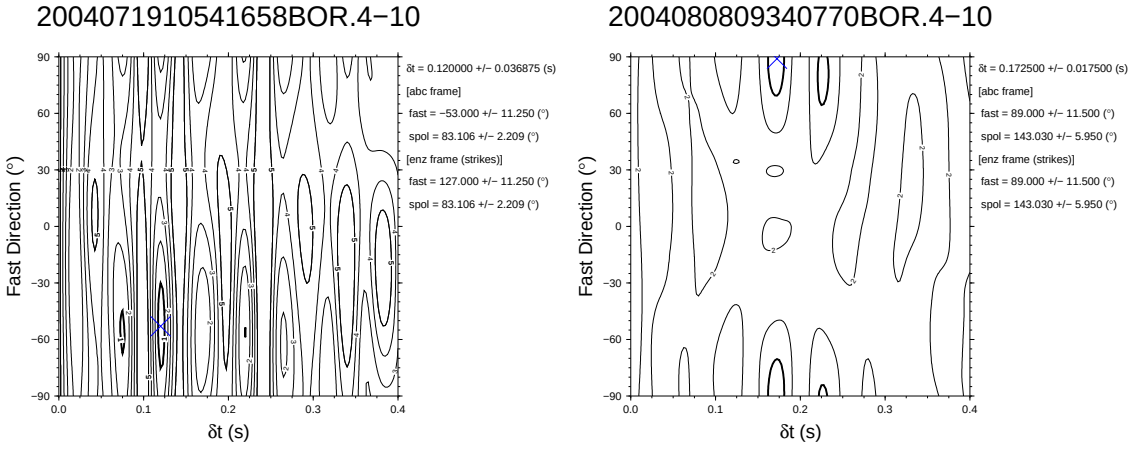
Cycle skipping causes jumps in the delay times and fast directions with a 90° ambiguity. Figure 4.7 suggests that there is also an effect on the incoming polarisation. The angle α (illustrated schematically in Figure 4.1) can be seen in the plots of particle motion (Figure 4.8).

Figures 4.4 and 4.6 show the original plots of particle motion for two group 1 and two group 2 events respectively. Figure 4.8 shows particle motion plots for the two group 1 events with annotations highlighting the similar fast directions but different incoming polarisations. From the plots of particle motion it appears that mismatching the waveforms can also



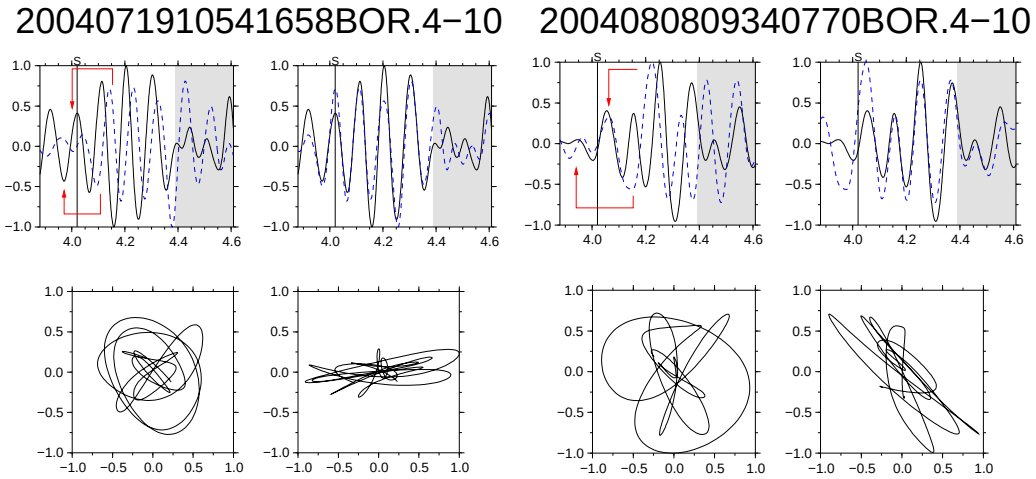
(a) Waveforms shift by $T/2$ (b) Waveforms shift by T

Figure 4.4: Two wave forms of different events from group 1 (fast directions 0-45°). Note the different shifts of the slow wave annotated in red.



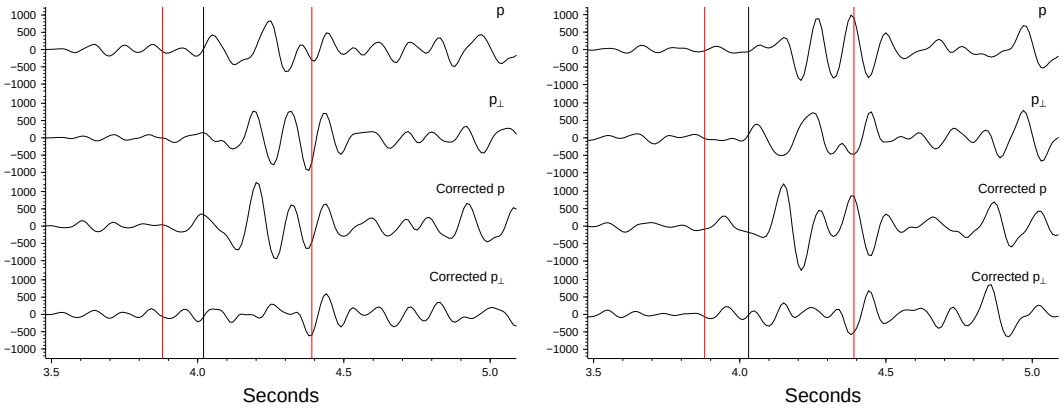
(a) Contour plot - small δt (b) Contour plot - large δt

Figure 4.5: Two contour plots of different events from group 2 (fast directions 90-120°). Note the splitting parameters differ.



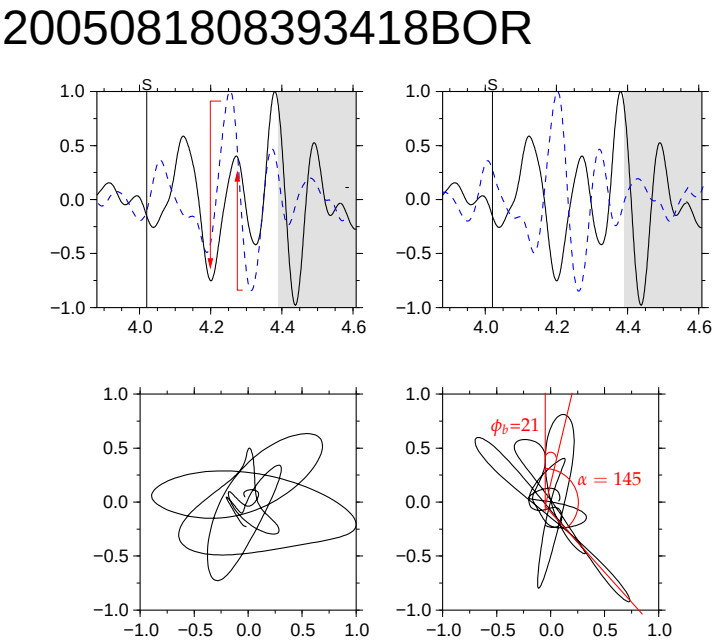
(a) Waveforms shift by $4 \times T/2$ (b) Waveforms shift by $4 \times T/2$

Figure 4.6: Two waveforms of different events from group 2 (fast directions 90-120°). Note the different shifts of the slow wave annotated in red.

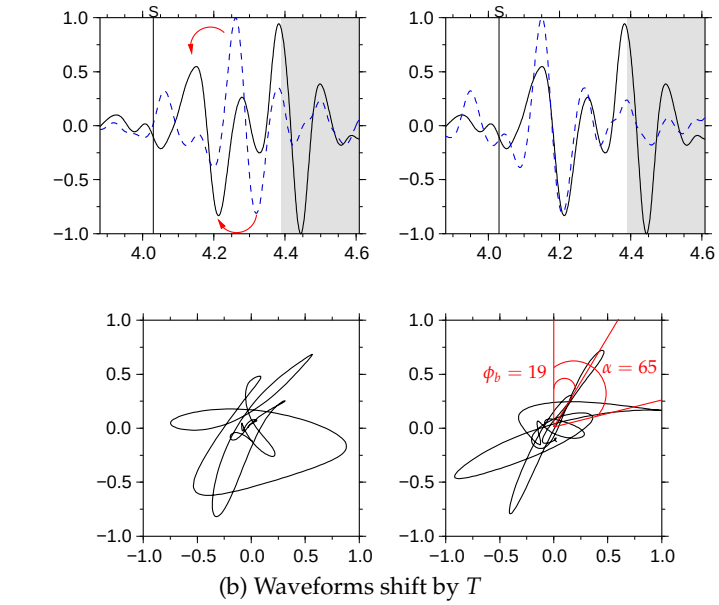


(a) Event 2005081808393418 (b) Event 2005090408311169

Figure 4.7: Polarisation waveforms of different events from group 1 (fast directions 0-45°). Note that the p ($\hat{\mathbf{p}}$) component of one looks like the $p \perp$ ($\hat{\mathbf{a}}$) component of the other.



(a) Waveforms shift by $T/2$



(b) Waveforms shift by T

Figure 4.8: Two wave forms of different events from group 1 (fast directions 0-45°). Note the different shifts and incoming polarisations annotated in red. The polarisation annotations show ϕ_b and α (see Equation 4.1). These two parameters are also given on the contour plot (Figure 4.3).

create 90° jumps in α (i.e. $\hat{\mathbf{p}}$ and $\hat{\mathbf{a}}$ are flipped). The incoming polarisation for event 2005081808393418BOR (Figure 4.8a) is 145.384 ± 7.338 while the incoming polarisation for event 2005090408311169BOR (Figure 4.8b) is 65.232 ± 8.039. These are roughly 80° apart but if you take into account the errors on both of these estimates ($7.338 + 8.039 = 15.377$) then it is likely that these two estimates of the incoming polarisation are 90° apart

$$\begin{aligned} 65.232 + 90 = 155.232 \in 145.384 \pm 15.377 \\ 145.384 - 90 = 55.384 \in 65.232 \pm 15.377 \end{aligned} \tag{4.8}$$

This switch is also shown in Figure 4.7 where the p ($\hat{\mathbf{p}}$) component of one event looks like the $p \perp$ ($\hat{\mathbf{a}}$) component of the other event.

Figure 4.9 displays the α and δt parameters for all 146 events. There are two clear clusters highlighting the effect of cycle skipping on the incoming polarisation

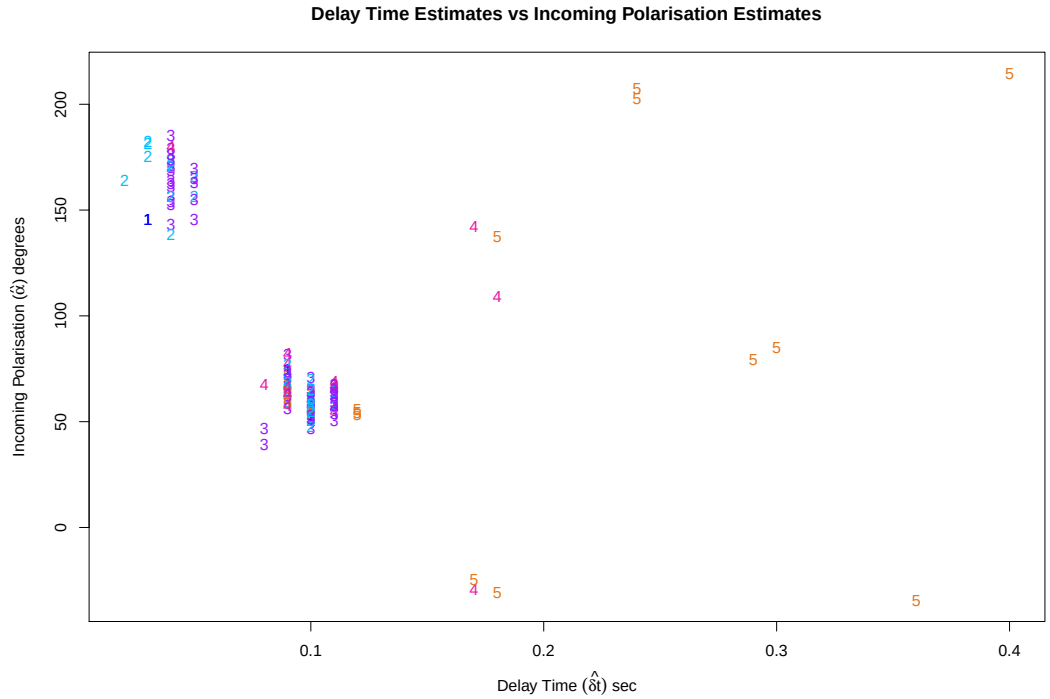


Figure 4.9: Incoming polarisation estimates vs delay time estimates for group 1 (fast directions 0-45°) events. Two clusters have formed that are $90° \pm SE(spol)$ apart.

So far we have taken two or more events and compared them. Here we take a single event that has two 95% confidence regions (Figures 4.3a and 4.3b) and see what effect shifting from one 95% confidence region to another has. File `./images/rep16dtshiftanim.gif`[1] shows an animation of moving the supposed minimum to a point that is $+T/2$ away in the other 95% confidence region. Another file, `./images/rep23dtshiftanim.gif` shows the effect of moving $-T/2$ into the other 95% confidence region. These animations make it easier to see the waveforms shifting and what effect this has. The following was observed:

[1] Pink font indicates a link to an external file. The animations are also on CD if there are viewing problems

- when the selected point shifts to the other 95% confidence region the particle motion flips and points in the opposite direction
- corrected fast and slow waveforms show what appears to be cycle skipping (matching another peak/trough)
- corrected slow waveform flips at the half way point. As mentioned before, the Mfast codes flip the sign of the slow waveform displacements when a peak is matched to a trough to make the match more obvious
- null component $p \perp (\hat{\mathbf{a}})$ appears to have more energy on it when the minimum is shifted to the other 95% confidence region
- the corrected p ($\hat{\mathbf{p}}$) component still has its first peak but the second peak starts to shrink. This waveform can also be seen slowly moving by δt

The 90° ambiguity of the incoming polarisation appears to be a new phenomenon that has not been mentioned in articles before. As a result of this 90° ambiguity, the incoming polarisation using this method will differ from the incoming polarisation obtained from the focal mechanism (see Aki and Richards (2002) for details on this method). For a single layer, cycle skipping causes a 90° ambiguity between the two methods. For multiple layers of anisotropy the polarisations using the two methods will differ but not necessarily by 90°. This is because the polarisation is being calculated from two different waves. Figure 4.10 illustrates this for two layers of anisotropy. Consequently, there may be incorrect inferences about the geometry of the fault if it is based on an incoming polarisation that is used by Silver and Chan (1991) outlined in this chapter.

It is also possible that an incorrect incoming polarisation would affect detecting null measurements (where $\phi - \alpha \approx 0$). A criterion used to define null measurements (Savage et al. (2010), Wessel (2010), Peng and BenZion (2004)) is

$$20^\circ \leq |\phi_b - \alpha| \leq (90^\circ - 20^\circ) \tag{4.9}$$

This equations tells us that if the incoming polarisation is polarised too closely to either the fast or slow direction then it is a null measurement. If the incoming polarisation has a 90° ambiguity then there is no problem because the incoming polarisation will instead be too close to the other (fast or slow) component. However, for incoming polarisations with large errors, an event may be incorrectly classified because Equation 4.9 only applies to the parameters and not their errors. For example, Figure 4.8 shows an example where the difference between two incoming polarisations is 80°.

To summarise, it appears that some of the events have contour plots with two 95% confidence regions that are $T/2$ apart, the splitting parameters have two main clusters of events that differ by a delay time of $T/2$ and some events have fast directions that are 90° away from other fast directions. This strongly suggests that some of the events in this dataset suffer from cycle skipping. It appears that cycle skipping also has an effect on the incoming polarisation α that has not been mentioned in any of the literature before. Cycle skipping also causes 90° jumps in the incoming polarisation, which results in mistaking the p component for the a component. Consequently, the incoming polarisation determined using the

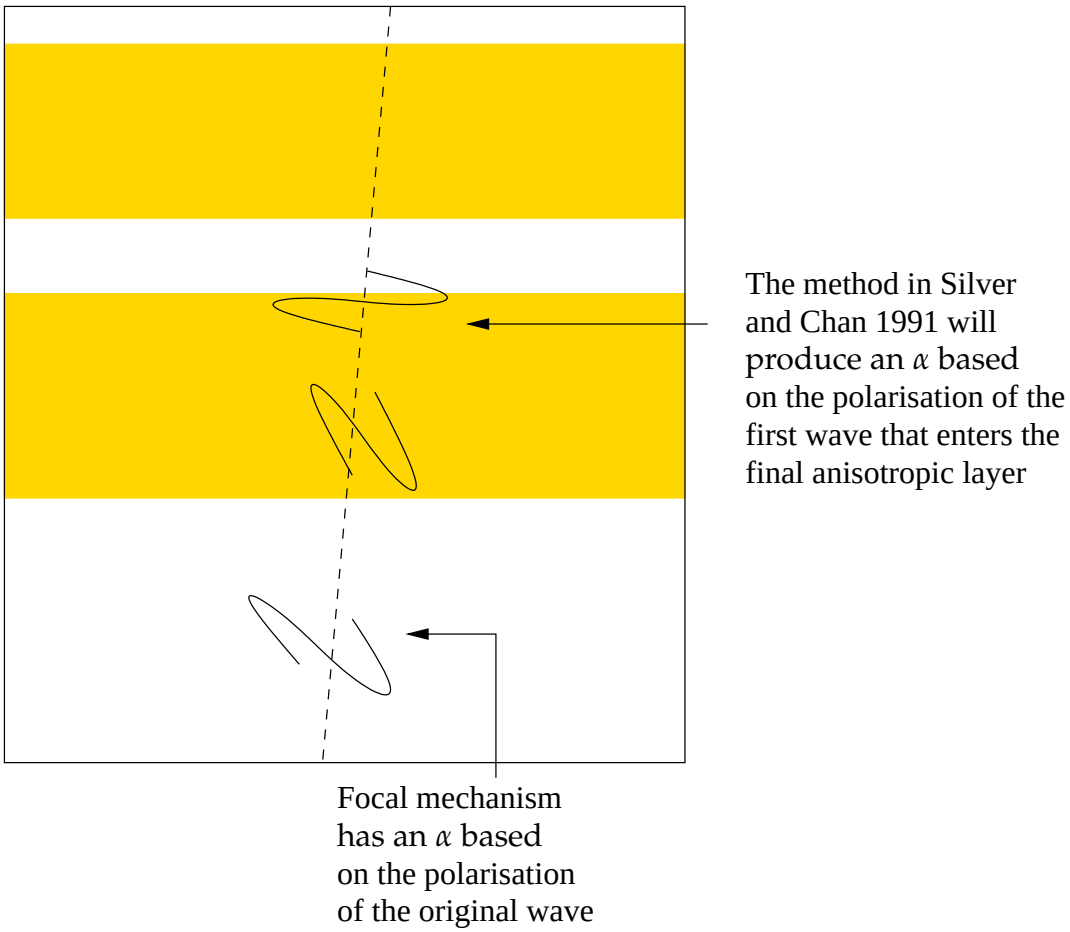


Figure 4.10: Diagram illustrating that the incoming polarisation obtained from the Silver and Chan (1991) method will differ from the polarisation obtained from the focal mechanism. The focal mechanism method returns an α based on the polarisation of the original wave while the method in this chapter (Silver and Chan, 1991) returns an α based on one of the split waves. If the waveforms are split by a time greater than the wave length, then α is the fast direction for the second to last layer from the surface.

method above may differ from the incoming polarisation obtained from the focal mechanism. This can effect inferences about the fault geometry and also the classification of null grades.

Chapter 5

Testing Assumptions & Estimators

In order to test the validity of the error estimates obtained using the Silver and Chan (1991) method and its derivatives we have formulated three tests. The first involves investigating the bias of the degrees of freedom estimate using both sets of coefficients discussed in Section 2.9. The second determines whether the time series used to calculate the errors has a chi-squared distribution. The final test determines how reliable the standard errors calculated by Silver and Chan (1991) are.

5.1 Test 1: Bias of Estimators

5.1.1 Test 1 Program

This program, which was written in `R`, compares the b_n coefficients to the Silver and Chan (1991) a_n that are used to estimate the degrees of freedom using the power E (Section 2.9). We wish to see which set of coefficients (if any) produce an unbiased estimate of the degrees of freedom. The program can be broken down into 4 steps.

Recall that we have digital data comprising a sample of $N+1$ data points over the interval $[0, T]$ with a sampling frequency of $\frac{N}{(T-0)}$ and each point evenly spaced by $\frac{(T-0)}{(N)}$ (Figure 2.16) and that we assume the sample data can be written as a convolution of a Gaussian filter and Gaussian noise (Assumption 3). In other words,

$$\eta(t) = (f * g)(t) \tag{5.1}$$

where f denotes the filter and g denotes the noise. The Gaussian filter induces correlations between the set of data points according to the width of the filter. The first step is to create the Gaussian filter $f(t)$.

Step 1: Create the filter, `f` (Note that comments in `R` are denoted by #)

```
# Parameters
sigma <- 1
tmax <- 10
N <- 1000
N1 <- N+1
# our N1 discrete points with index n
tvec <- (0:N)/N*tmax
w <- 0.05

# Filter
f <- exp(-0.5*((tvec/tmax)/w)^2)
f <- f + rev(f)
f <- f/sum(f)*2/N # scale f
```

The first set of lines define the parameters needed for the filter

$$\begin{aligned}
\texttt{sigma} &= \text{standard deviation } \sigma \\
\texttt{tmax} &= \text{maximum } t \text{ over the interval i.e. } T \\
\texttt{N} &= \text{the number of data points minus one} \\
\texttt{N1} &= \text{the number of data points} \\
\texttt{tvec} &= N+1 \text{ data points spread evenly over the interval } [0, T] \\
\texttt{w} &= \text{smoothing width } 0 < w \leq 1
\end{aligned}$$

The only parameter not seen in Section 2.9 is the smoothing width, which is the proportion of the data to be correlated. The higher the smoothing width, the greater the range of points that are related to each other, creating a smoother curve. The variable σ^2 is the variance of the Gaussian noise process.

Using these parameters we create a Gaussian filter, f_n

$$f_n = \exp\left[-1/2\left(\frac{t_n}{wT}\right)^2\right] \qquad n = 0, \cdots, N \tag{5.2}$$

We form $f_n \propto \left[\exp[-1/2\left(\frac{((t-0)/T)}{w}\right)^2\right] + \left[\exp[-1/2\left(\frac{((t-T)/T)}{w}\right)^2]\right)$. Since the range of the simulated data are [0,T) we shift the points from $(-T/2, 0)$ forward to $(T/2, T)$ so that it's centred on zero and scale f by twice the sum of all the points in f divided by N. Only the N points $n = 0, \cdots, N-1$ are retained because the fast Fourier transform treats the function as a cycle with period T. Next we fast Fourier transform f_n. `R` has an `fft()` function, which applies the following operation to f

$$\tilde{f}(\omega) = \sum_{n=0}^{N} f(t_n) \exp[-i\omega t_n]$$

Both the filter and the data are periodic over the interval $[0, T), [T, 2T)$.... In other words $x_0 = x_T$ so the last point is deleted in order to calculate the transform correctly.

Step 2: transform the filter into the frequency domain by using the fast Fourier transform

```
# f(omega)
ff <- fft(f[-N1])
```

The third step involves calculating the exact degrees of freedom. Remember to obtain the degrees of freedom we need

$$\left(\sum_{n=0}^{N} a_n |\tilde{f}_n|^2\right)^2 \equiv F_2^2 \qquad \text{(the second squared moment)}$$

$$\left(\sum_{n=0}^{N} 2a_n^2 |\tilde{f}_n|^4\right) \equiv F_4 \qquad \text{(the fourth moment)}$$

when these two quantities are known we can calculate the degrees of freedom exactly using

$$\nu = 2\frac{(\sigma^2 F_2)^2}{\sigma^4 F_4} = 2\frac{F_2^2}{F_4}$$

(SC A9)

The a_n coefficients are used to calculate the second squared moment of F and are defined as

$$a_n = \begin{cases} 1/2 & n = 0, N \\ 1 & \text{otherwise} \end{cases}$$

The a_n^2 coefficients are used to calculate the fourth moment of F and are defined as

$$a_n^2 = \begin{cases} 1/4 & n = 0, N \\ 1 & \text{otherwise} \end{cases}$$

These two sets of coefficients are called `avec` and `avec2` respectively in our code. Using these coefficients we can calculate the true degrees of freedom

Step 3: Calculate the true degrees of freedom

```
# Calculate nu true df
# SC coefficents
avec <- rep(1,n)
avec[c(1,n)] <- 0.5
avec2<- rep(1,n)
avec2[c(1,n)] <- 0.25
f2 <- sum(avec * (Mod(ff))^2)    # second moment F_2
f4 <- 2*sum(avec2 * (Mod(ff))^4) # fourth moment F_4
nu <- 2*f2^2/f4                  # true degrees of freedom
```

Equation 2.137 (*SC A9*) has been used to calculate the true degrees of freedom. We can calculate the true degrees of freedom because here we know exactly how $f(t)$ was constructed.

When analysing actual data we do not know the correlational properties of the data. Instead we can use the following equation to estimate the degrees of freedom

$$\widehat{\nu} \equiv 2\left(\frac{2E^2}{E_4} - 1\right)$$

(SC A12)

Recall both we and Silver and Chan (1991) define E^2 as

$$E^2 = \left(\sum_{n=0}^{N} a_n |\tilde{f}_n|^2 |\tilde{g}_n|^2\right)^2$$

which also has a_n coefficients like F_2^2

$$a_n = \begin{cases} 1/2 & n = 0, N \\ 1 & \text{otherwise} \end{cases}$$

We define E_4 as

$$\begin{aligned} E_4 &= \sum_{n=1}^{N-1} \left(\frac{4}{3}\right) |\tilde{f}_n|^4 |\tilde{g}_n|^4 + \left(\frac{1}{3}\right) (|\tilde{f}_0|^4 |\tilde{g}_0|^4 + |\tilde{f}_N|^4 |\tilde{g}_N|^4) \\ &= \sum_{n=0}^{N} b_n |\tilde{f}_n|^4 |\tilde{g}_n|^4 \end{aligned}$$

(c.f. SC A10)

Therefore the b_n coefficients are

$$b_n = \begin{cases} 1/3 & n = 0, N \\ 4/3 & \text{otherwise} \end{cases}$$

Silver and Chan (1991) define E_4 as

$$\begin{aligned} E_4^{SC} &= \sum_{n=1}^{N-1} 1 |\tilde{f}_n|^4 |\tilde{g}_n|^4 + \left(\frac{1}{3}\right) (|\tilde{f}_0|^4 |\tilde{g}_0|^4 + |\tilde{f}_N|^4 |\tilde{g}_N|^4) \\ &= \sum_{n=0}^{N} b_n^{SC} |\tilde{f}_n|^4 |\tilde{g}_n|^4 \end{aligned}$$

(SC A10)

$$b_n^{SC} = \begin{cases} 1/3 & n = 0, N \\ 1 & \text{otherwise} \end{cases}$$

The Silver and Chan (1991) codes accidentally define F_4^{SC} instead of E_4^{SC} and use it to estimate the degrees of freedom

$$F_4^{SC} = \sum_{n=1}^{N-1} 1 |\tilde{f}_n|^4 |\tilde{g}_n|^4 + \left(\frac{1}{2}\right) (|\tilde{f}_0|^4 |\tilde{g}_0|^4 + |\tilde{f}_N|^4 |\tilde{g}_N|^4) \tag{5.3}$$

Therefore Silver and Chan (1991) estimate the degrees of freedom using the a_n coefficients, which are the coefficients for F_4^{SC} (and coincidentally E^2) not E_4^{SC}. We estimate the degrees of freedom using the b_n coefficients, which are the coefficients for our E_4. Therefore we compare the estimates of the degrees of freedom using the coefficients in the original Silver and Chan (1991) codes (a_n) and our coefficients (b_n).

The experiment is repeated `nrep` times. Each time the estimated degrees of freedom (using Equation 2.146 (*SC A12*)) is calculated using both sets of coefficients, which requires two different E_4 values to be calculated: E_4 =`e4` with our b_n coefficients and the Silver and Chan (1991) E_4^{SC} =`e4e`, which accidentally uses the a_n coefficients instead of the b_n^{SC} coefficients. Since the experiment has the same filter each time, there is no need to recalculate it. However, the Gaussian noise (`fg`=$\tilde{g}(\omega)$) is different each time so the convolution, `fu`= $|\tilde{f}(\omega)|^2|\tilde{g}(\omega)|^2$ needs to be calculated each time. Thaller (2000) shows that the Fourier transform of a Gaussian with mean zero and variance σ^2 is a Gaussian with zero mean and variance $1/\sigma^2$. Therefore we simply create $\tilde{g}(\omega)$ directly in the Fourier domain. Since $\sigma^2 = 1$ no scaling of the variance is required either.

Step 4: Calculate the estimated degrees of freedom using both a_n (Silver and Chan (1991) code) and b_n (ours) coefficients

```
# Our coefficients
bvec <- 4/3*avec2
nrep <- 1000                          # repeat nrep times
nu.ests <- t(sapply(1:nrep,
       function(i) {
          fg <- rnorm(n,0,1/sigma)  # g(omega)
          fu <- fg*ff               # g(omega)f(omega)
          mfu2 <- (Mod(fu))^2       # squared modulus
          e <- sum( avec*mfu2 )     # power
          e4 <- sum( bvec*mfu2^2 )  # our E_4
          e4e <- sum( avec*mfu2^2 ) # SC code E_4 which is actually F_4
          nu <- 2*(2*e^2/e4 - 1)    # our estimated df
          nu.e <- 2*(2*e^2/e4e -1 ) # SC estimated df
          c(nu,nu.e,e,e4,e4e)
       }))

dimnames(nu.ests)[[2]] <- c("nu","nu.e","e","e4","e4e")
```

The code given above gives results for a single smoothing width. The code can be wrapped with a `for` loop that contains a vector of smoothing widths such as

```
ws <- c(0.00001,0.0001,0.001,0.01,0.05,0.1,0.15,0.2,0.25)
for(w in ws){
  [steps 1 to 4 here]
```

```
}
```

so that multiple smoothing windows can be analysed.

5.1.2 Test 1 Results

To analyse the results from the final steps, we plot histograms showing the true degrees of freedom in the form of a vertical red line and the estimates using both sets of coefficients (Figures 5.1a and 5.2a). A second histogram was created with the estimates being divided by the true degrees of freedom (Figures 5.1b and 5.2b). In other words for each replicate r_j we have

$$r_j = \frac{\widehat{\nu}_j}{\nu} \qquad\qquad j = 1, ..., \texttt{nrep} \tag{5.4}$$

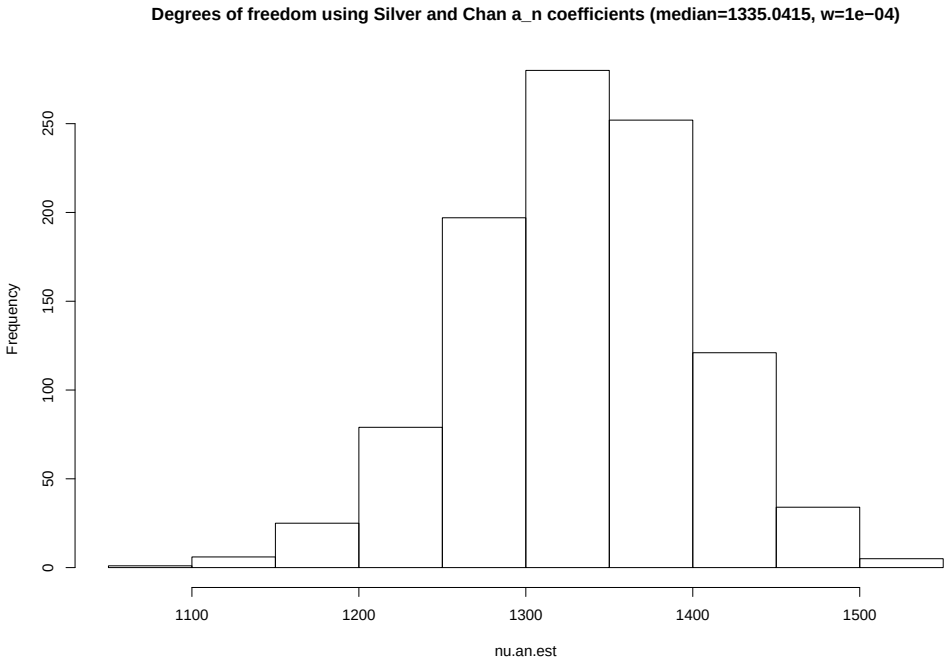
where

$$\begin{aligned} r_j &= \text{scaled degrees of freedom for the } j\text{th replicate} \\ \widehat{\nu} &= \text{estimate of the degrees of freedom} \\ \nu &= \text{true degrees of freedom} \\ \texttt{nrep} &= \text{total number of replicates} \end{aligned}$$

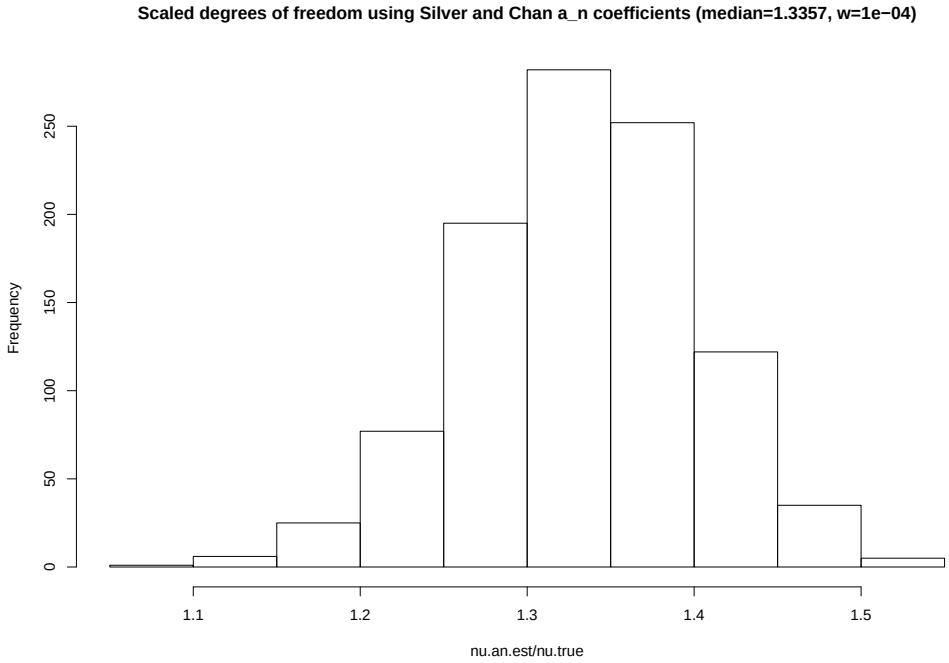
Therefore, unbiased estimators would have a scaled degree of freedom of 1 (on average). The scaled versions make comparisons across different smoothing widths possible.

Figures 5.1b and 5.2b show the estimated degrees of freedom for a smoothing width of almost zero (w=0.0001) i.e. virtually uncorrelated points in the time series. Figure 5.1b shows $\widehat{\nu}_{SC}$ based on the a_n coefficients used by Silver and Chan (1991) in their codes is a positively biased estimator with the degrees of freedom being over estimated by a factor of 1.33$\approx$4/3. For the b_n coefficients that are used in our estimate of the degrees of freedom (Figure 5.2b), it appears the degrees of freedom are unbiased.

Figure 5.3 shows a summary of the effect of different smoothing windows (i.e. differing amounts of correlation between the data points) on both sets of coefficients. All median estimates of the degrees of freedom are scaled by the true degrees of freedom to enable comparison. Both sets of degrees of freedom converge to constant values after around w=0.5. One can see that the greater the correlation, the degrees of freedom based on our b_n coefficients is unbiased. On the other hand, for smoothing widths greater than 50%, the degrees of freedom based on the Silver and Chan (1991) a_n coefficients is negatively biased (i.e. too small). As mentioned earlier, for virtually uncorrelated data, the b_n coefficients produce unbiased degrees of freedom whilst the a_n coefficients overestimate the degrees of freedom by a factor of around 4/3. Therefore, if the noise was Gaussian our estimate of the degrees of freedom using the b_n coefficients would be correct. There is an interesting bump around $0 < w < 0.4$. During this time the b_n coefficients produce a degrees of freedom estimate that is positively biased while the a_n coefficients lead to a degrees of freedom estimate that is positively biased to begin with but eventually becomes negatively biased. If the amount

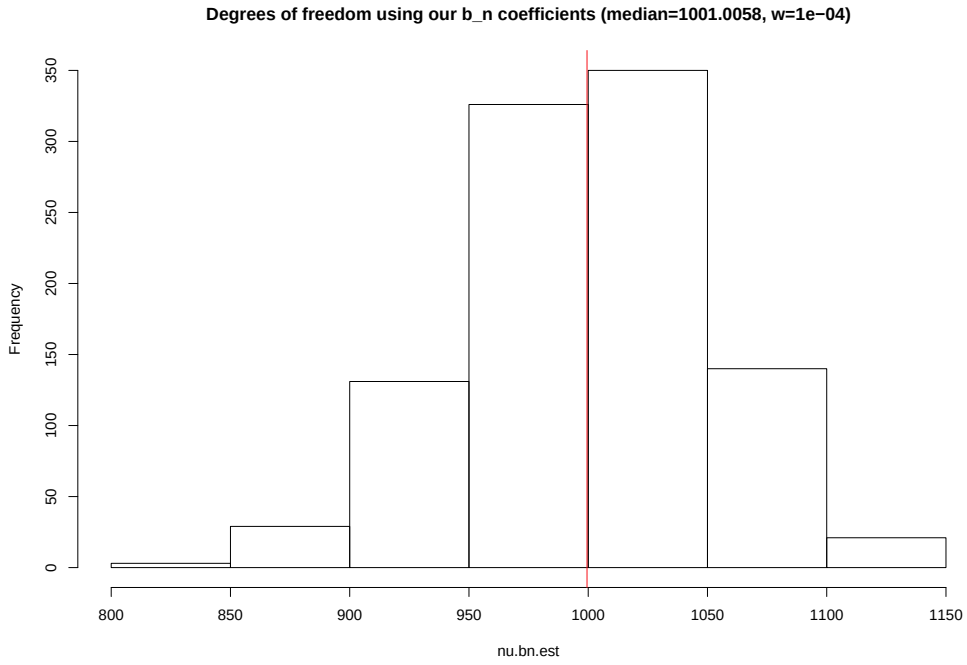


(a) Unscaled degrees of freedom with a_n coefficients

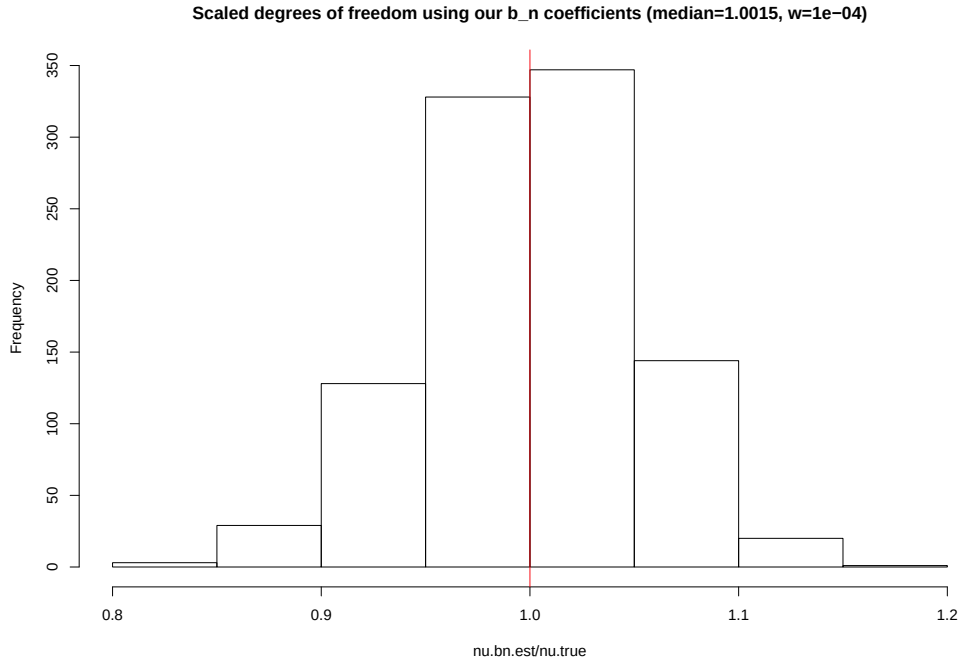


(b) Scaled degrees of freedom with a_n coefficients

Figure 5.1: Example of the Silver and Chan (1991) estimated degrees of freedom using the a_n coefficients. Estimates have been both scaled and unscaled for a single smoothing width (w=0.0001), which is given in the title of the plot.



(a) Unscaled degrees of freedom with b_n coefficients



(b) Scaled degrees of freedom with b_n coefficients

Figure 5.2: Example of the estimated degrees of freedom using our b_n coefficients. Estimates have been both scaled and unscaled for a single smoothing width (w=0.0001), which is given in the title of the plot.

of correlation between data points in the power were known, then comments about which coefficients were more appropriate could be made.

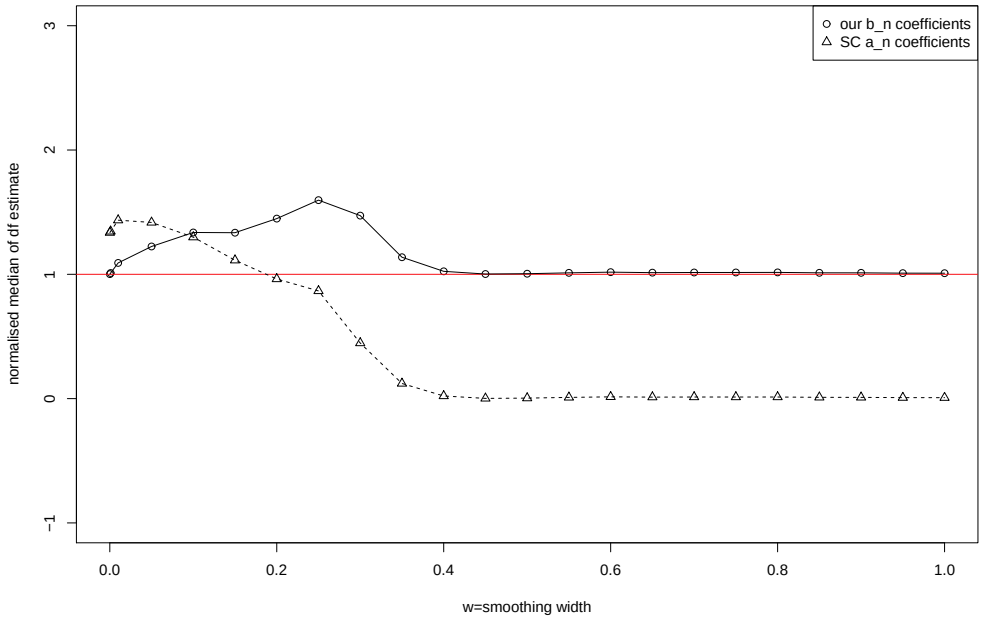


Figure 5.3: Comparison of the scaled estimated degrees of freedom for different smoothing widths based on `nrep`=1,000 repetitions. Circles are based on our degrees of freedom based, which uses the b_n coefficients. Triangles denote the Silver and Chan (1991) estimate of the degrees of freedom based on the the a_n coefficients that are used in the Silver and Chan (1991) code. If the estimates were unbiased they would sit on the horizontal red line.

The final plot, Figure 5.4 shows the true degrees of freedom over various smoothing widths ranging from approximately zero to 20% of the window length. The diagrams show that for virtually uncorrelated data the degrees of freedom is almost 1000, the number points in our time series. This drops sharply to roughly 1 degree of freedom once the smoothing width exceeds 0.05.

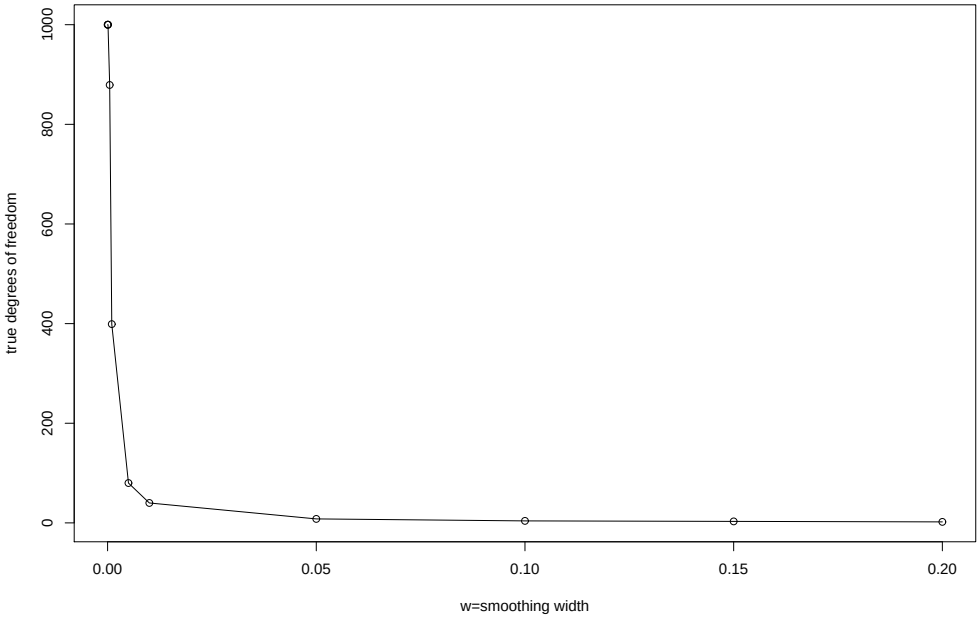


Figure 5.4: Comparison of the true degrees of freedom for various smoothing widths.

5.2 Test 2: Chi-squared Assumption

The next test involves checking whether the noise has a Gaussian distribution, which would imply that the sum of squares, $S_{min}(\phi, \delta t)$, (Section 2.7) has a chi-squared distribution. This is done by performing a Kolmogorov-Smirnov test and also by analysing plots of the empirical distribution, the noise in the time domain and the squared modulus of the Fourier transformed noise. We then compare this noise to background noise to see how similar the two types of noise are.

5.2.1 Testing the Distribution of the Data

The Kolmogorov-Smirnov test is used to determine whether one continuous distribution matches another. As a result, we can use this to test Assumptions 1 and 2 that state that the power has a chi-squared distribution. The test is a non-parametric one, meaning that it does not make any assumptions about the distribution of the data. The test has the following hypotheses

H_0: the two distributions are the same
H_A: the two distributions are not the same

To determine whether H_0 is accepted or rejected, the test statistic D is calculated, which is the largest vertical distance between the two cumulative distribution curves. The idea behind this measure is that similar distributions should have a similar shape and thus a small D. The more different the distributions are, the larger the value of D. Figure 5.5 illustrates the test statistic.

The two distributions to be compared are the distribution of the data and the chi-squared distribution More details on this test can be found in Feldman and Valdez-Flores (2010). The Kolmogorov-Smirnov test requires all parameters to be known in advance i.e. they cannot be obtained from the data, which is how we obtain the estimate of the degrees of freedom. However, we have a process involving randomisation that should let us use the Kolmogorov-Smirnov test. This is explained later. The sample data are obtained as follows

1. desplit the wave using the best pair of ϕ and δt to produce the corrected waveforms. The two components are now in splitting coordinates

2. calculate the incoming polarisation

3. rotate the windowed corrected waveforms by $(\phi - \alpha)$ to obtain the noise. The first component contains the signal and the second component contains the noise. This gives $\eta(t) = \hat{\boldsymbol{\delta}}_2^T \mathbf{R}_3(\phi - \alpha)\tilde{\mathbf{u}}_s(\omega, t)$ (Note the $\sim$ here denotes the corrected wave and not a Fourier transformed shear wave)

4. fast Fourier transform the noise in the window, $|\tilde{\eta}(\omega)|$

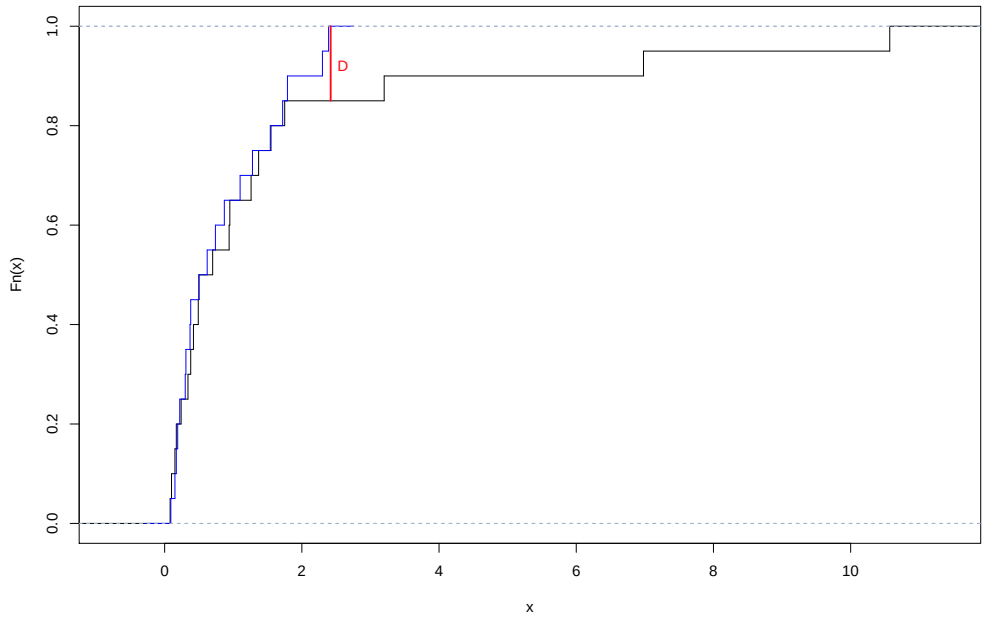


Figure 5.5: Example of the test statistic D which is the largest vertical distance between two distribution functions.

5. calculate the squared modulus of the fast Fourier transform, $|\tilde{\eta}(\omega)|^2$ up to the Nyquist frequency

6. scale $|\tilde{\eta}(\omega)|^2$ by the variance of the noise

The scaling is needed because a chi-squared distribution is the sum of squared standard normals (Appendix A.5). This dataset is compared to a chi-squared (χ^2) distribution. This distribution has a single parameter, the degrees of freedom. Silver and Chan (1991) first assume the degrees of freedom is n, the number of points in $|\tilde{\eta}(\omega)|^2$ (Assumption 1). If this assumption does not hold it is assumed that $|\tilde{\eta}(\omega)|^2$ is approximately χ^2 with $\widehat{\nu}_{SC}$ degrees of freedom (Assumption 2). We think the derivation of ν_{SC} was incorrect and that the degrees of freedom is in fact smaller. We denote our degrees of freedom $\widehat{\nu}$. The data set is compared to each of these three chi-squared distribution with $n, \widehat{\nu}_{SC}, \widehat{\nu}$ degrees of freedom respectively.

Before the Kolmogorov-Smirnov test is carried out the distribution of the data and the chi-squared distribution are plotted. The empirical distribution is the cumulative distribution of a sample and is defined as

$$F_n(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{x_i \leq t\} \tag{5.5}$$

where

$$
\begin{aligned}
n &= \text{number of data points} \\
t &= \text{time} \\
\mathbb{1} &= \begin{cases} 1 & x_i \leq t \\ 0 & \text{otherwise} \end{cases}
\end{aligned}
$$

This can be thought of as taking a $1/n$ step at each of the n data points to create a staircase looking plot like Figure 5.5. For each of the 146 events the empirical distribution was plotted against a chi-squared distribution with $n, \widehat{\nu}_{SC}$ and $\widehat{\nu}$ degrees of freedom. All of the plots showed that $|\tilde{\eta}(\omega)|^2$ was definitely not chi-squared with $n, \widehat{\nu}_{SC}$ or $\widehat{\nu}$ degrees of freedom. Because it was so obvious from the plot, there was no point displaying the results of the Kolmogorov-Smirnov test which produced p-values of zero (5dp) for all degrees of freedom. Therefore there is no need to explain or proceed with the randomisation test. Instead a few of these plots are presented along with some comments.

Figures 5.6, 5.7 and 5.8 show plots for a grade 1, 2 and 3 event respectively. Each plot shows the empirical distribution plotted against a chi-squared distribution with $\widehat{\nu}$ degrees of freedom (top plot). Since the plots with $\widehat{\nu}_{SC}$ and n degrees of freedom look the same they are not displayed. The bottom left shows the windowed noise component $\eta(t)$ while the bottom right plot shows $|\tilde{\eta}(\omega)|^2$ up to the Nyquist frequency. All three figures consist of a chi-squared distribution with a jump that appears to be soon after zero. This is because the range of x is so incredibly large, which gives the illusion of the of the chi-squared variable having a cumulative distribution of 1 soon after $x = 0$. We applied a 4-10Hz band pass filter to our data so there is a large amount of power inside this range, which attenuates rapidly outside this range (bottom right plot). The bottom left plot shows the noise in the time domain. There appears to be too much structure here to be Gaussian noise.

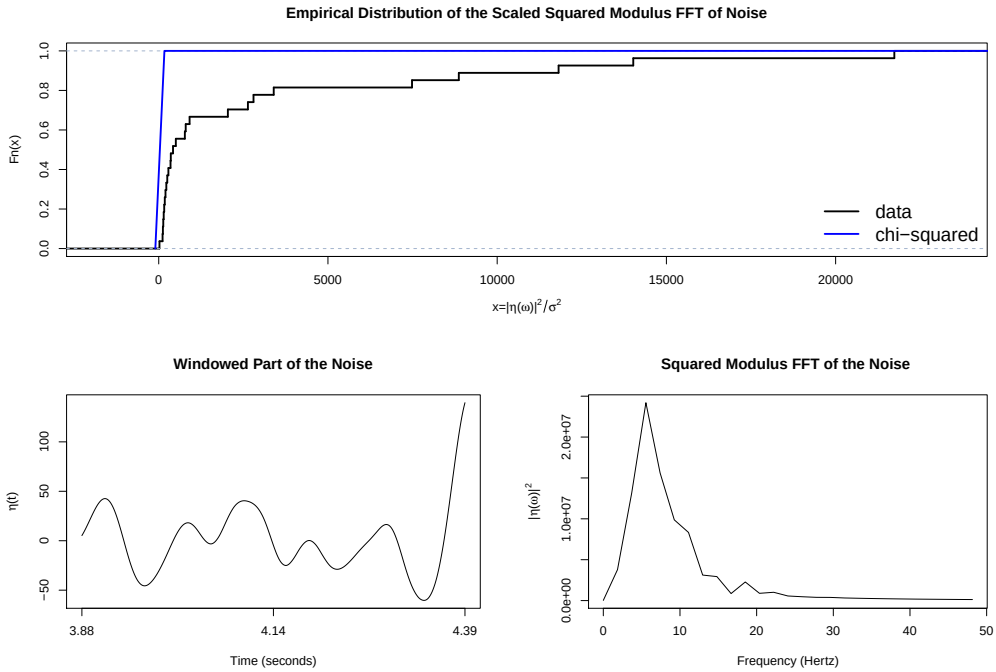


Figure 5.6: Grade 1 event - plots of the empirical distribution, the windowed noise in the time domain and the squared modulus fast Fourier transform of the windowed noise up to the Nyquist frequency

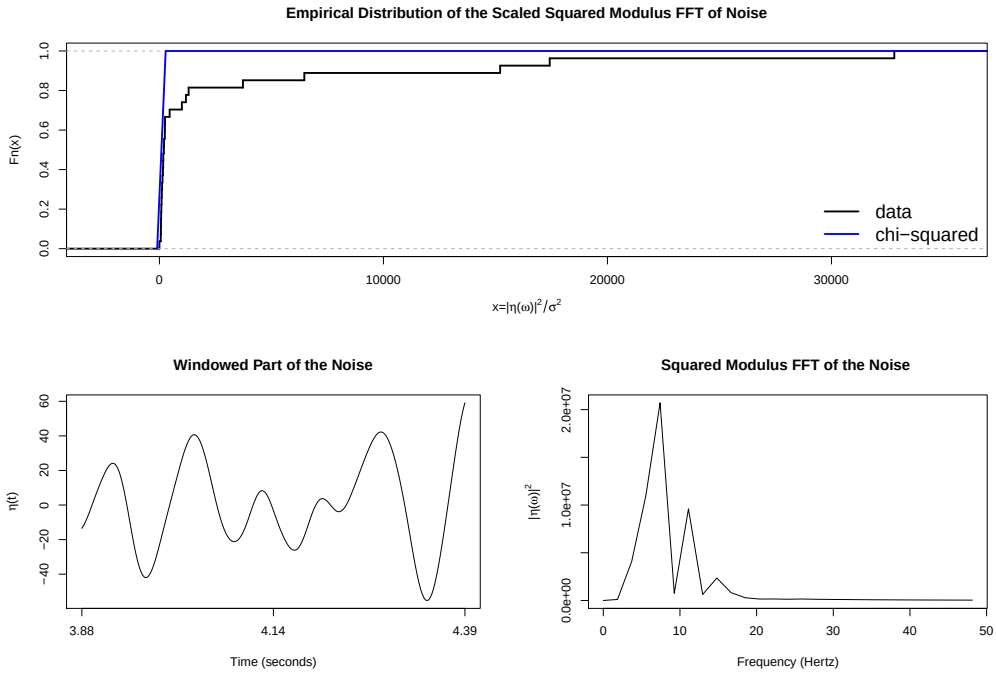


Figure 5.7: Grade 2 event - plots of the empirical distribution, the windowed noise in the time domain and the squared modulus fast Fourier transform of the windowed noise up to the Nyquist frequency.

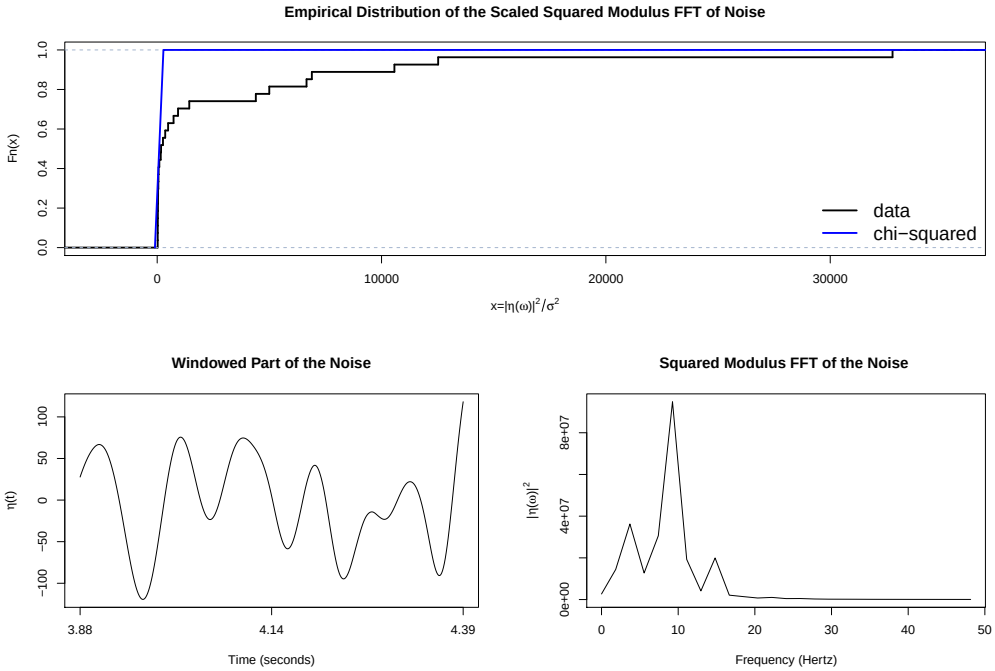


Figure 5.8: Grade 3 event - plots of the empirical distribution, the windowed noise in the time domain and the squared modulus fast Fourier transform of the windowed noise up to the Nyquist frequency.

	Max power - windowed noise	Max power - background noise
Min	3,112,000	19,440
LQ	18,800,000	132,800
Median	77,230,000	337,400
Mean	400,200,000	846,500
UQ	320,500,000	808,600
Max	17,280,000,000	905,3000

Table 5.1: Summary statistics of the maximum amplitude for the two types of noise.

5.2.2 Comparison to Background Noise

The chi-squared assumption clearly does not hold so the errors produced by the Silver and Chan (1991) method are not valid. Our previous findings suggested that there was too much structure in the windowed noise for it to be Gaussian noise. We finish this section by comparing the noise from the window, to noise on the east component before the S wave arrival (background noise) to see if the two have a similar structure. Our background noise consists of a sample of the data prior to the P-wave and event start time. We compare the windowed noise to this to see if they have a similar frequency content and amplitudes. To do this we look at six power plots and create summary statistics for all events. The plots of the windowed noise are constructed with 27 points of data up to the Nyquist frequency. The plots of the background noise are constructed with 175 points up to the Nyquist frequency. The background noise consists of more points so the fluctuations are more visible compared to the windowed noise which has a smooth looking power plot. Figures 5.9 and 5.10 show power plots for grade 1 events, Figures 5.11 and 5.12 show power plots for grade 2 events and Figures 5.13 and 5.14 show power plots for the grade 3 events. Both types of noise consist of filtered (4-10Hz) data so the power attenuates rapidly outside this range. In all the plots there is a large spike near the lower frequencies because of the filter, but the amplitudes for the windowed noise are much larger so it is clear that there is more power in the windowed noise than in the background noise. This could be caused by some sort of coherent signal contained in the windowed noise - possibly S wave power from scattered rays. Table 5.1 summarises the maximum amplitude for both the windowed noise and the background noise. The different mean and medians highlight that the distribution of maximum power for both types of noise is skewed. The table reconfirms that the power is a lot larger for windowed noise than the background noise with the medians differing by a factor of roughly 200.

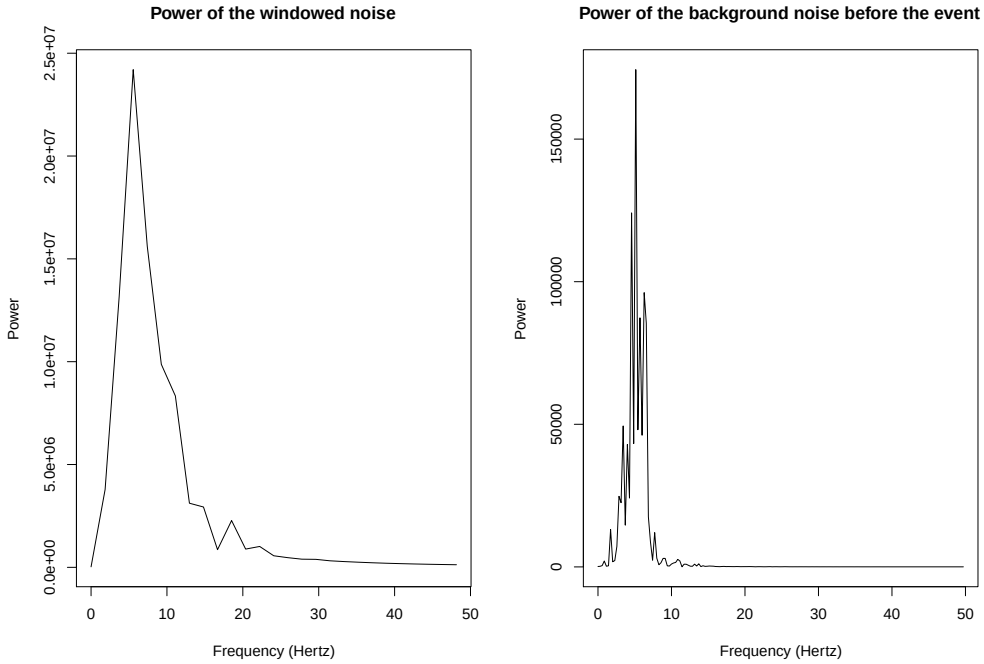


Figure 5.9: Power of the windowed noise and background noise for a grade 1 event.

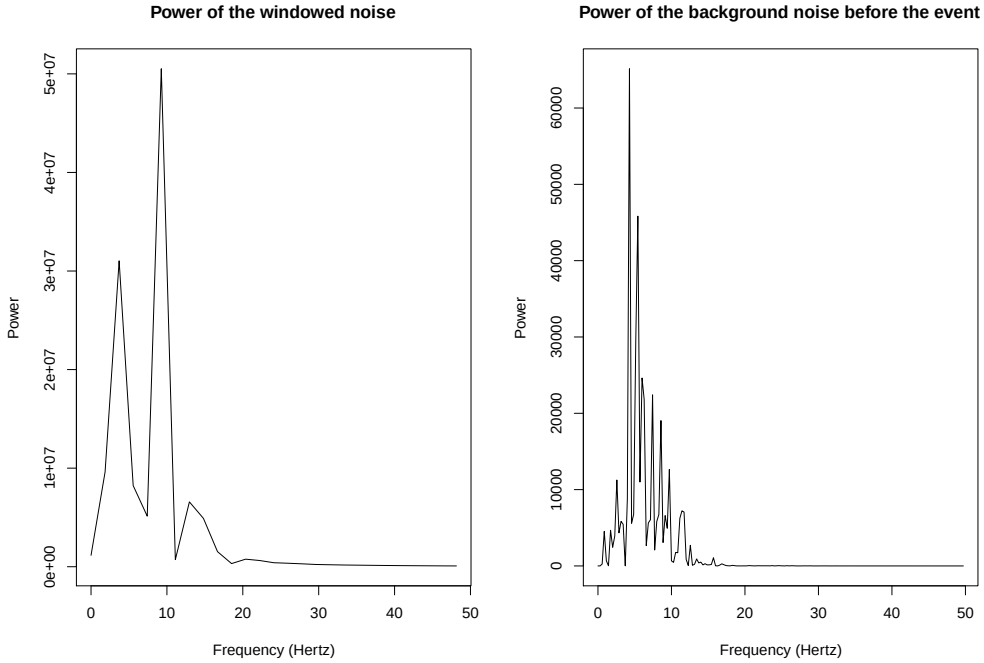


Figure 5.10: Power of the windowed noise and background noise for another grade 1 event.

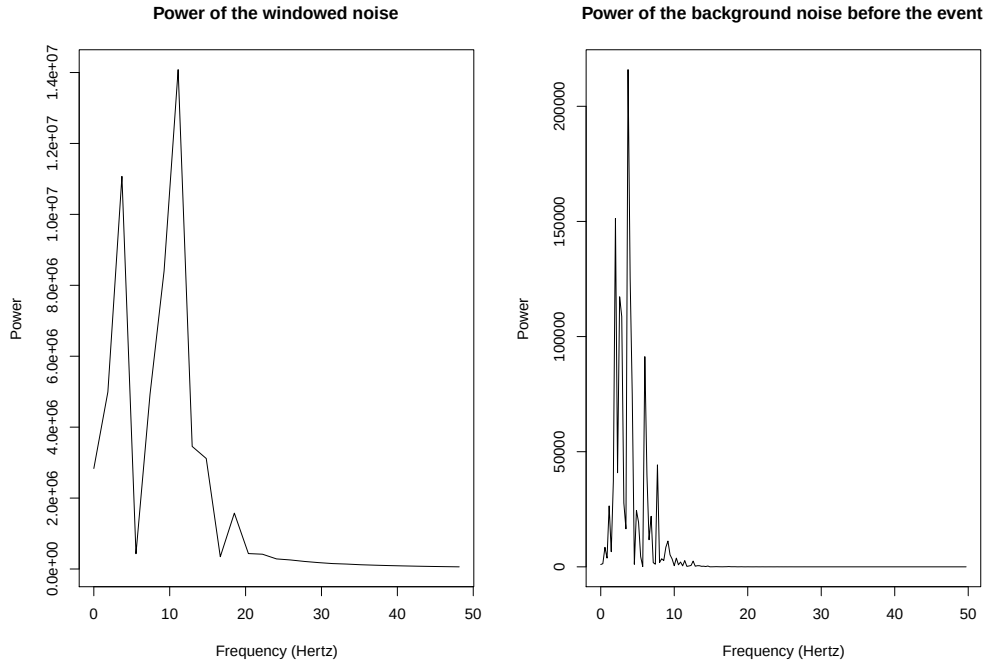


Figure 5.11: Power of the windowed noise and background noise for a grade 2 event.

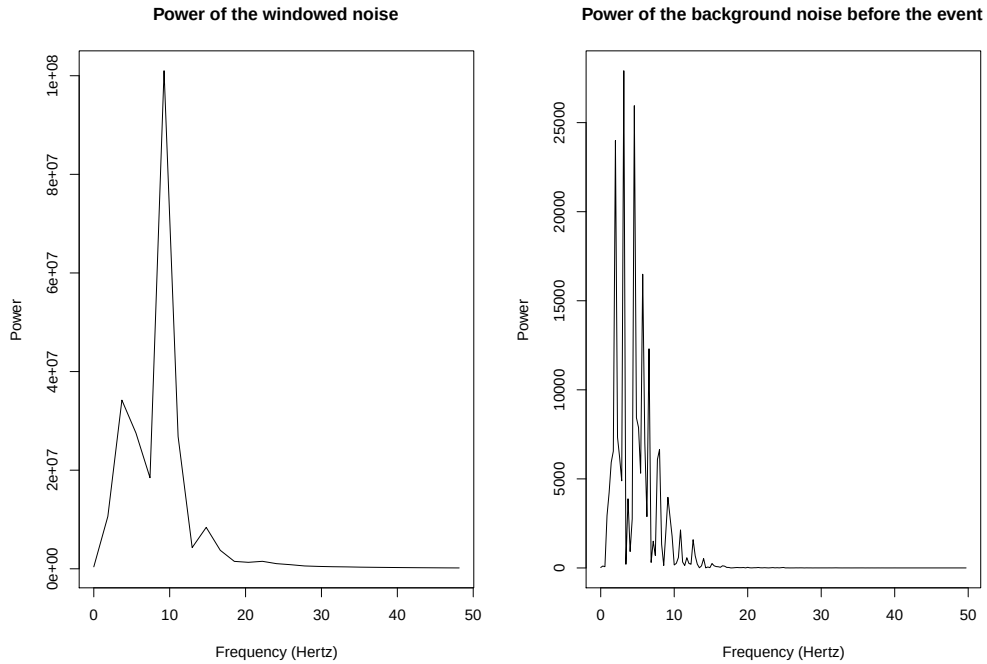


Figure 5.12: Power of the windowed noise and background noise for another grade 2 event.

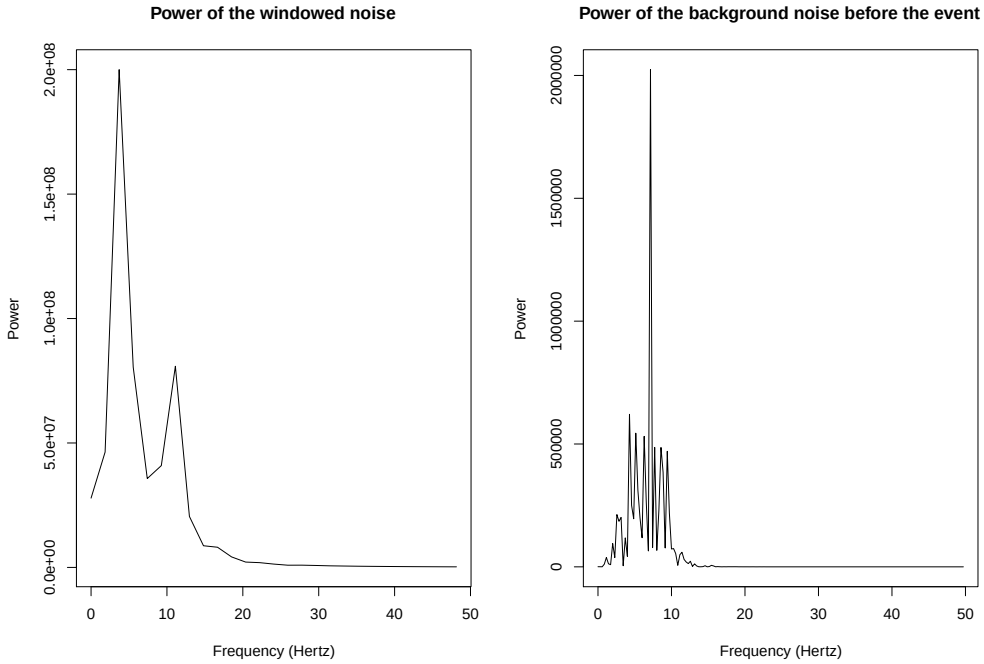


Figure 5.13: Power of the windowed noise and background noise for a grade 3 event.

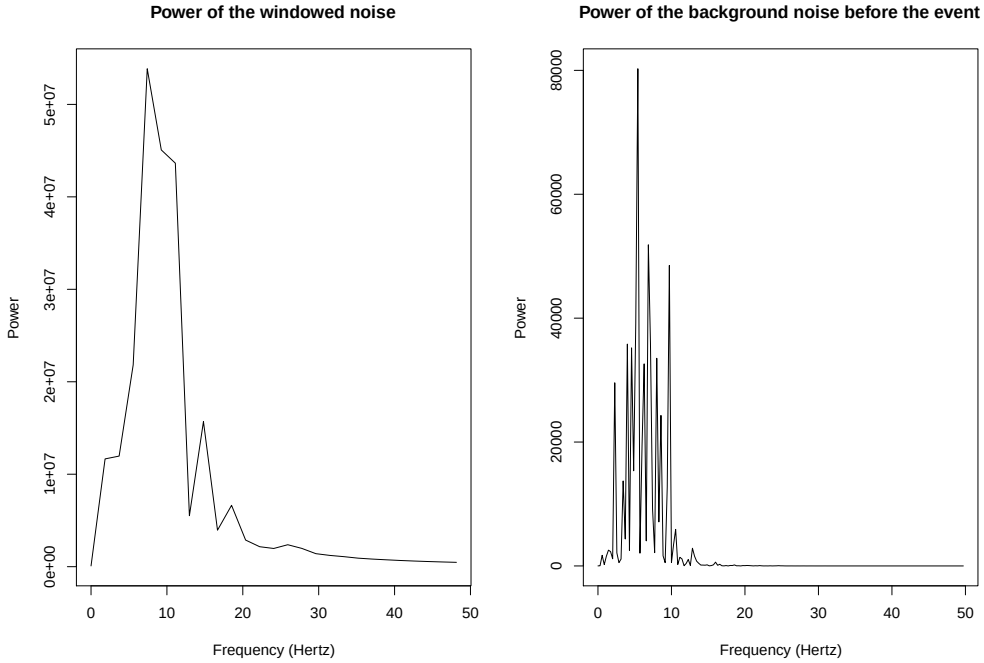


Figure 5.14: Power of the windowed noise and background noise for another grade 3 event.

5.3 Test 3: Reliability of the Standard Errors

The final test is designed to see how reliable the (standard) errors obtained from the Silver and Chan (1991) method are. This is done by creating plots of these errors and marking on these plots three different measures of spread.

Figure 5.15, 5.16 and 5.17 contain examples of the plots. Each shows a histogram of the corresponding standard errors obtained for each estimate and the corresponding parameter estimates. The top left diagram shows the standard errors based on the Silver and Chan (1991) method for the fast direction, bottom left are the fast direction estimates ($\widehat{\phi}$), top right are the standard errors based on the Silver and Chan (1991) method for the delay time, bottom right are the delay time estimates ($\widehat{\delta t}$). On each of the two standard error diagrams there are three vertical lines corresponding to three different measures of spread. The three measures are listed below. An explanation of what to expect if the errors are reliable is given after the measures are described.

1. The dashed blue line represents the median of the *standard errors* i.e. median$\left(\widehat{SE}(\phi_1), \ldots, \widehat{SE}(\phi_n)\right)$ and median$\left(\widehat{SE}(\delta t_1), \ldots, \widehat{SE}(\delta t_n)\right)$

2. The solid red line represents the standard deviation of the *parameter estimates*, which are displayed at the bottom of each figure i.e. SD$(\widehat{\phi}_1, \ldots, \widehat{\phi}_n)$ and SD$(\widehat{\delta t}_1, \ldots, \widehat{\delta t}_n)$

3. The dash-dot yellow line represents a robust estimate of the standard deviation of the *parameter estimates* based on the interquartile range i.e. $SD_{IQR} = (UQ_{\widehat{\phi}} - LQ_{\widehat{\phi}})/(2 \times .6745)$ and $SD_{IQR} = (UQ_{\widehat{\delta t}} - LQ_{\widehat{\delta t}})/(2 \times .6745)$ (see Equations 5.6 and 5.7 plus the explanation below)

The solid red line (measure 2: SD$(\widehat{\phi}_1, \ldots, \widehat{\phi}_n)$ and SD$(\widehat{\delta t}_1, \ldots, \widehat{\delta t}_n)$) will give us an idea of the spread of the estimates ϕ and δt. The more dispersed they are, the higher the SD. The blue dashed line (measure 1: median$\left(\widehat{SE}(\phi_1), \ldots, \widehat{SE}(\phi_n)\right)$ and median$\left(\widehat{SE}(\delta t_1), \ldots, \widehat{SE}(\delta t_n)\right)$) will give us an idea of what the Silver and Chan (1991) method believes the errors are. If the SC errors have been calculated correctly then you would expect the solid red (measure 2) and dashed blue lines (measure 1) to be close together.

If the distribution of the estimates is skewed then the standard deviation of the parameter estimates (solid red line) may be unfairly large, making the Silver and Chan (1991) estimates of the errors appear worse than they actually are. In order to overcome this problem, another standard deviation, robust to skewness and outliers, is calculated based on the interquartile range (IQR). The interquartile range shows how spread out the middle 50% of the data are and is calculated by using the upper (UQ) and lower quartiles (LQ) (see Figure 5.18).

$$IQR = UQ - LQ \tag{5.6}$$

The mid section boundaries correspond to a standard deviation of $\pm 0.6745\sigma$ (Figure 5.18). Using this we can calculate the standard deviation based on the IQR as

$$\begin{aligned} IQR &= 2 \times 0.6745\sigma \\ \sigma &\simeq \frac{IQR}{2 \times 0.6745} \\ \sigma &\simeq \frac{(UQ - LQ)}{(2 \times 0.6745)} \end{aligned} \tag{5.7}$$

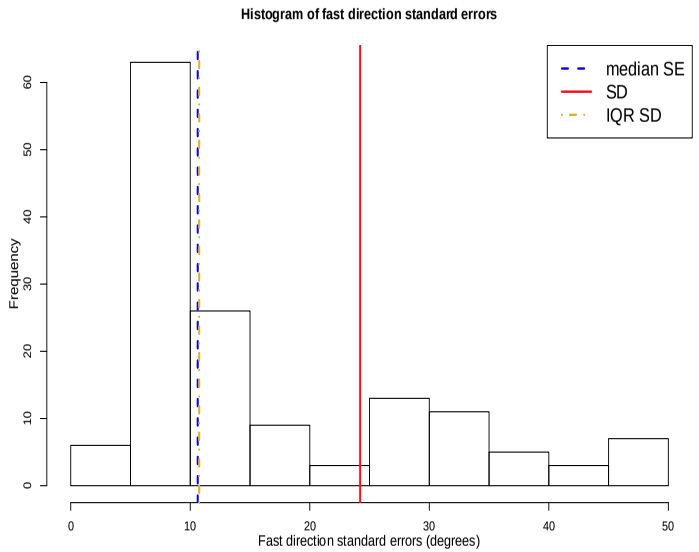
Now that we have detailed explanations of each measure of spread we can expand on how this test works. The test looks at the reliability of the errors by comparing the regular standard deviation (measure 2) of the estimates to the standard errors of the estimates (measure 1). If the two values represented are the same, then the errors obtained from the Silver and Chan (1991) method are unbiased. If the standard deviation (measure 2 - the solid red line) exceeds the median standard error (measure 1 - the dashed blue line) then the Silver and Chan (1991) errors have been underestimated. If the standard errors exceed the standard deviation then the Silver and Chan (1991) method has overestimated the errors. If the IQR standard deviation (measure 3 - dash-dot yellow line) is significantly larger than the standard deviation (measure 2 -solid red line) then the standard deviation may be unfairly too large.

A total of three Figures have been created. Figure 5.15 shows the standard errors when grading has been ignored. The last two Figures 5.16 and 5.17 show the standard errors when they are grouped into usable results (grades 1, 2 and 3) and unusable results (grades 4 and 5) respectively. Numerical values of the three measures of spread are in Table 5.2. Some general observations are made

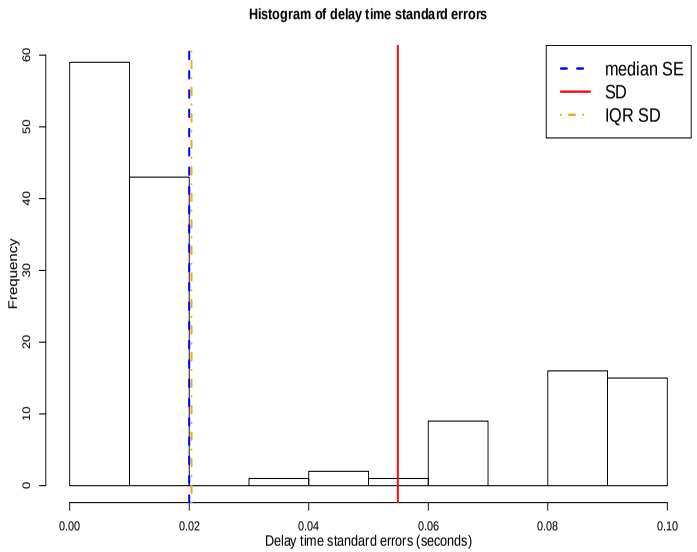
- In Figure 5.15 both dashed blue lines (measure 1) and both solid red lines (measure 2) are quite far apart. This indicates that when grades are ignored both the errors for the fast direction and the delay time have been underestimated by the Silver and Chan (1991) method by a factor of roughly 2. However, this could be caused by the poorer grade results being quite different from the high quality grades. The yellow dash dot line is not similar to the solid red line so the two types of standard deviations differ. The plots of the estimates are skewed, confirming this.
- In Figure 5.16a, which only displays the usable grades 1 2 and 3, the solid red (measure 2) and dashed blue lines (measure 1) are indistinguishable so the fast direction error estimates from Silver and Chan (1991) are consistent with the spread of the estimates The dash-dot yellow line (measure 3) also agrees with these two measures. On the other hand, Figure 5.16b has the dashed blue (measure 1) and solid red lines (measure 2) far apart. The delay time estimates themselves (Figure 5.16d) are clearly bimodal with a mix of grades in each cluster, which would explain the large regular standard deviation. This is most likely caused by the cycle skipping factors discussed in Chapter 4. This effect is not seen in the fast directions because group 1 consists of fast directions that are between 0-45°. This bimodal shape is causing the large difference between the regular standard deviation and the IQR based standard deviation, so the regular standard deviation may be unfairly large.

- In Figure 5.17, which displays the unusable grades, the estimates of the errors are very large compared to usable grades. The solid red line and the dashed blue line are distinct. The Silver and Chan (1991) method appears to underestimate the fast direction errors and slightly over estimate the delay time errors. The dashed dot yellow line and the solid red line are distinct, which is reflected in the skewed plots in Figures 5.17c and 5.17d.

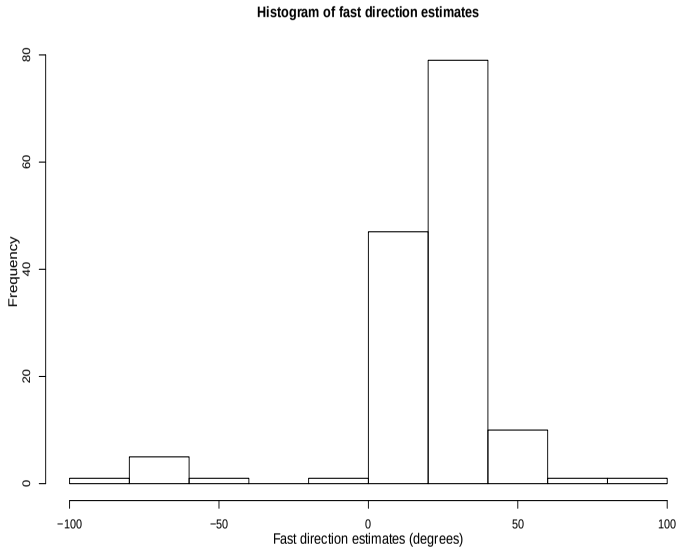
Finally, Table 5.2 summarises the three measures of spread shown in the three figures. The table shows that the usable grades have fast directions errors that are fairly similar to the spread of the parameter estimates. However, the delay times errors are different, which is caused by the cycle skipping. The standard deviation and the IQR based standard deviation are slightly different. This highlights the skewed distribution of parameter estimates.



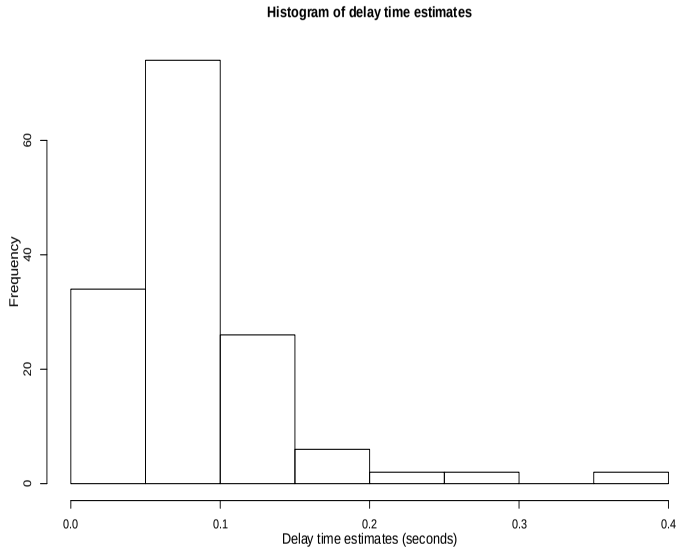
(a) Fast direction standard errors



(b) Delay time standard errors

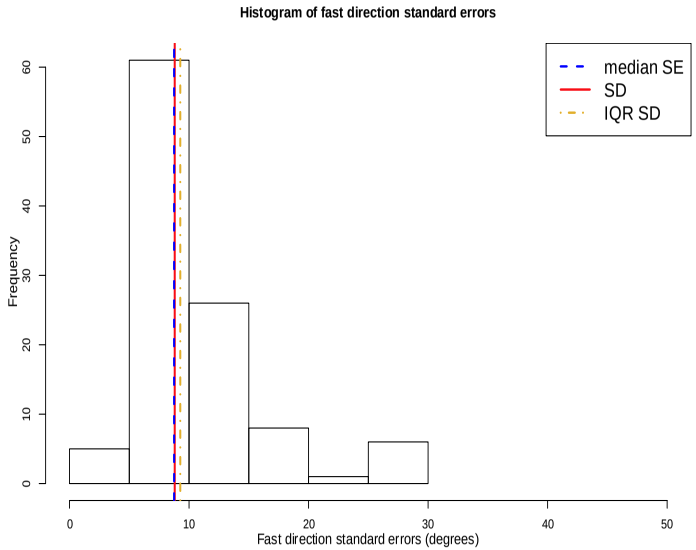


(c) Fast direction estimates

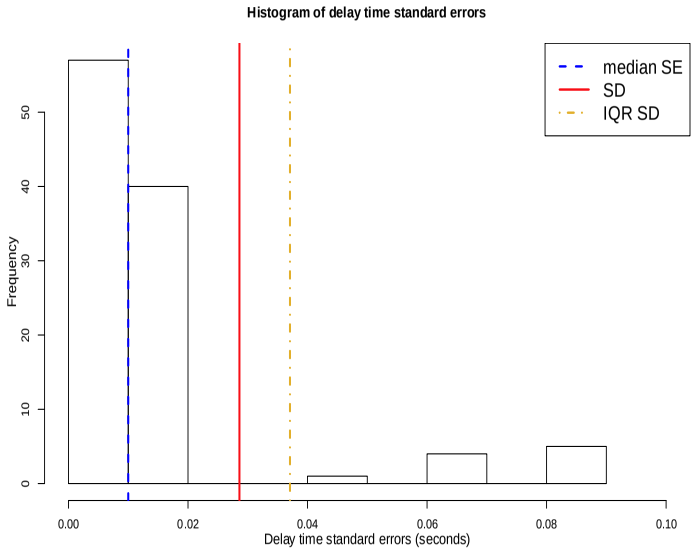


(d) Delay time estimates

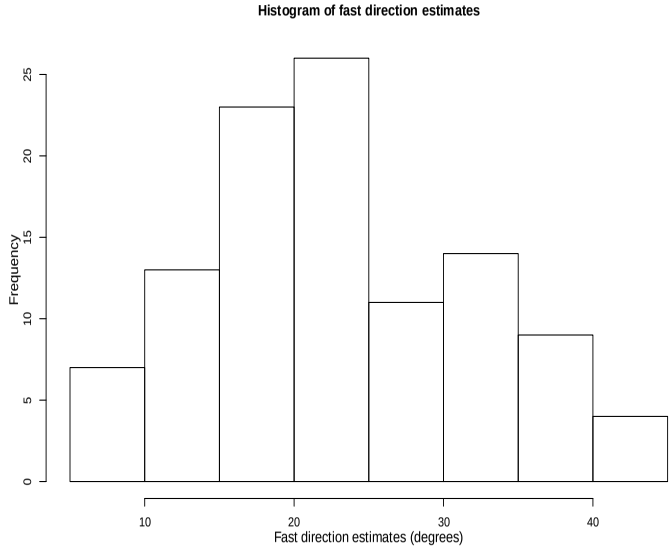
Figure 5.15: Histograms of the standard errors and estimates using the Silver and Chan (1991) method. Number of events=146.



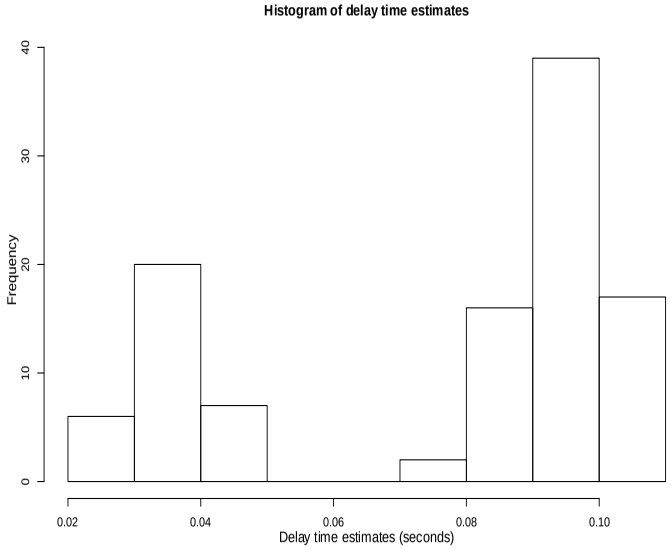
(a) Fast direction standard errors



(b) Delay time standard errors

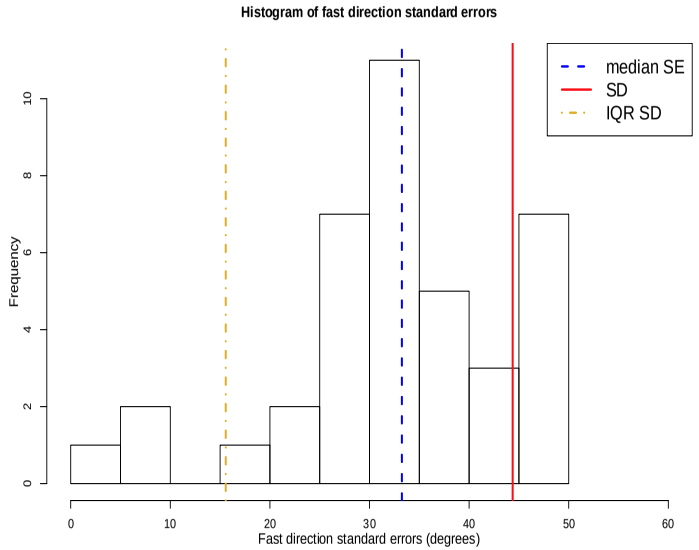


(c) Fast direction estimates

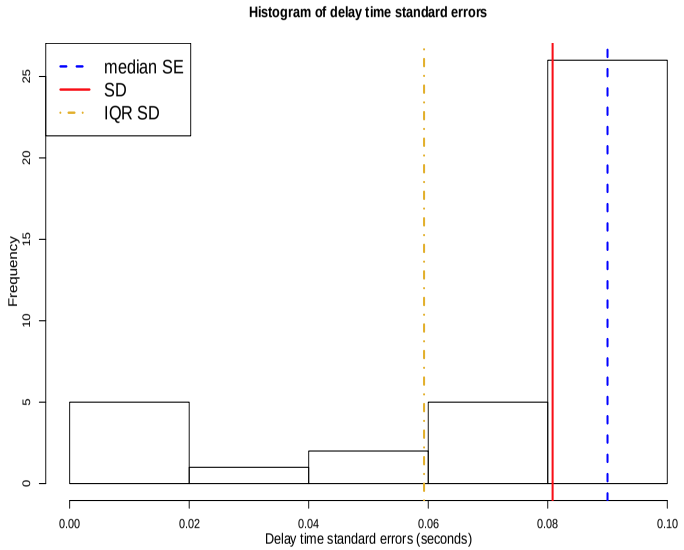


(d) Delay time estimates

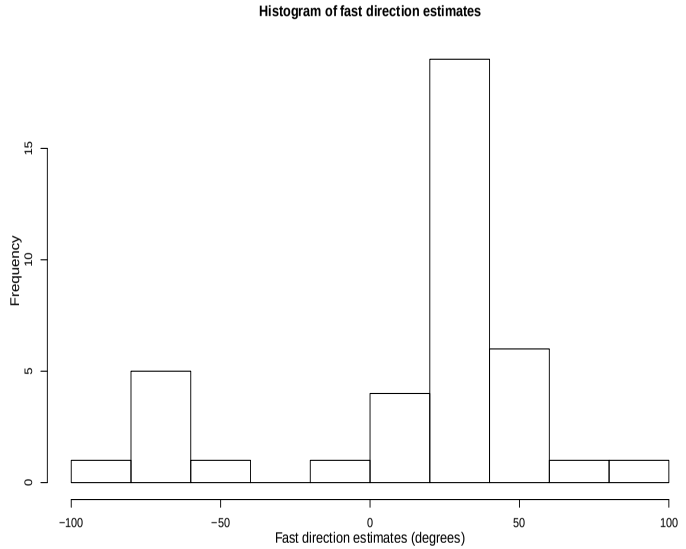
Figure 5.16: Histograms of the standard errors and estimates using the Silver and Chan (1991) method for grade 1, 2 and 3 results. Number of events=107.



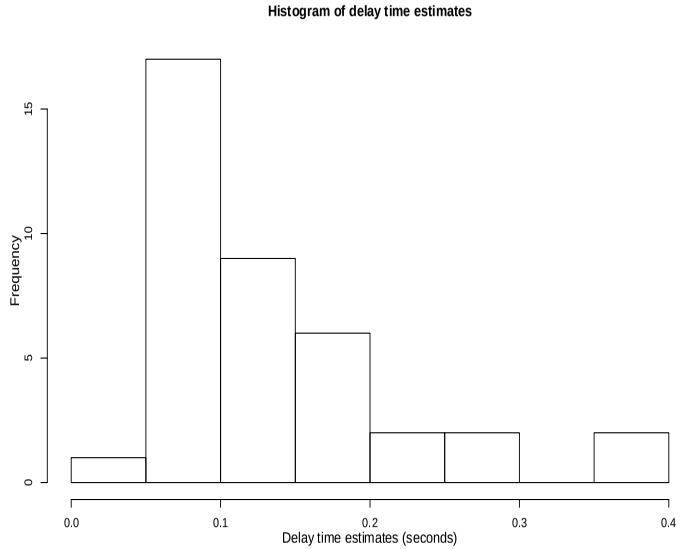
(a) Fast direction standard errors



(b) Delay time standard errors



(c) Fast direction estimates



(d) Delay time estimates

Figure 5.17: Histograms of the standard errors and estimates using the Silver and Chan (1991) method for grade 4 and 5 results. Number of events=39.

Events	Median error	Standard deviation	IQR Standard deviation
All ϕ	10.62	24.22	10.75
All δt	0.02	0.05	0.02
Usable (1,2,3) ϕ	8.75	8.80	9.27
Usable (1,2,3) δt	0.01	0.03	0.04
Unusable (4,5) ϕ	33.25	44.37	15.57
Unusable (4,5) δt	0.09	0.08	0.06

Table 5.2: Comparison of median standard errors to a regular standard deviation and a standard deviation based on the interquartile range. All values rounded to 2dp.

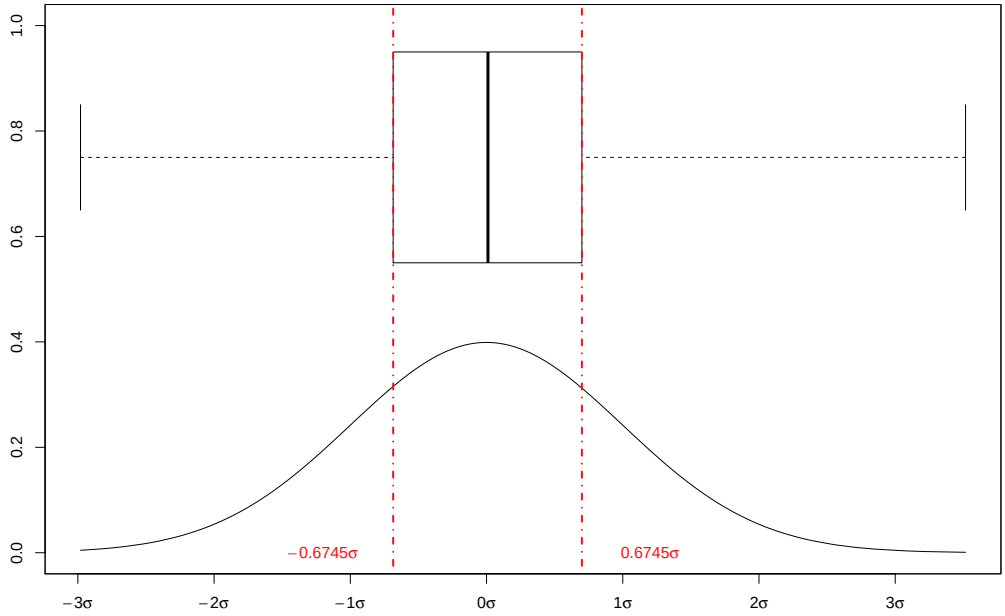


Figure 5.18: Diagram illustrating the relationship between the interquartile range and the standard deviation. Vertical dash dot lines denote the lower and upper quartiles.

5.4 Summary

To summarise, our simulations in Section 5.1 show that our b_n coefficients produce an unbiased estimate of the degrees of freedom for large smoothing widths i.e. large amounts of correlation. On the other hand, the Silver and Chan (1991) code a_n coefficients produce a negatively biased estimate of the degrees of freedom for large smoothing widths. In the case when the errors are uncorrelated our coefficients produce an unbiased estimate while the Silver and Chan (1991) code coefficients produce a positively biased estimate of the degrees of freedom. If the amount of correlation between data points in the power was known, then comments about which coefficients were more appropriate could be made.

Our diagnostic plots from Section 5.2 reveal that assumptions used to generate the degrees of freedom and consequently the errors do not hold. Therefore the windowed noise is not Gaussian and therefore the sum of squares $S(\phi, \delta t)$ (Section 2.7) used to construct the degrees of freedom does not follow a chi-squared distribution (which is the sum of squared standard normal variables (Appendix A.5)). There appears to be too much structure in the signal. We also find the structure of the noise is quite different to background noise. The power of the windowed noise is significantly larger than the power of the background noise, suggesting that the windowed noise still contains some coherent signal possibly from scattering.

Even though the assumptions used to generate the errors do not hold, our last set of diagnostic plots from Section 5.3 reveal that for good quality events, the fast direction errors from the Silver and Chan (1991) method (using the a_n coefficients) match the spread of the fast direction estimates. However, the delay time errors from the Silver and Chan (1991) method appear to be too small. This is probably caused by the bimodal delay time estimates that have been caused by cycle skipping.

Chapter 6

Conclusions and Future Directions

In this thesis we have performed a detailed analysis of the Silver and Chan (1991) method that is used to measure shear wave splitting. We find that there are discrepancies between our derivations, the Silver and Chan (1991) article and the Silver and Chan (1991) computer codes. Sign conventions as well as the need to write things in a more compact form have led to discrepancies in the article (see Table 2.2). Errors have been found in the original codes such as the covariance matrix being calculated incorrectly. Consequently there may be minor differences in the splitting parameters if this bug was corrected. More importantly, the fourth moment of the power (E_4), which is used to calculate the degrees of freedom and consequently estimates of the errors, has been incorrectly derived and also incorrectly implemented. Therefore, the estimates of the errors produced by the Silver and Chan (1991) method should be treated with caution.

We have also automated the manual grading process by creating numerical criteria that an expert grader used to classify events and incorporate them into a multiple linear regression model. We find that the automated method works well and that the automated grades are consistent with the manual grades with most of the automated grades being no more than 1 grade away from the manual grade. Using a cutoff automated grade of 3 captures most of the good grades and only a couple of the bad grades. The automated method was also able to identify incorrect manual grades by producing large discrepancies between the automated and manual grades. Our dataset does not contain any null measurements so we did not incorporate null detection into our automated method.

Additionally, we have explored the effects of cycle skipping. Jumps in delay times and fast directions for our set of near identical events is indicative of cycle skipping. Further analysis shows that cycle skipping also has an effect on the incoming polarisation. Consequently, inferences about the fault geometry based on the Silver and Chan (1991) incoming polarisation may be incorrect. Also, the classification of null events could be incorrect because the incoming polarisation is often used to determine null measurements.

We also test some of the assumptions behind the Silver and Chan (1991) method and investigate whether the degrees of freedom are unbiased and what effect this has on the errors themselves. Our simulations in Section 5.1 show that our b_n coefficients produce an unbiased estimate of the degrees of freedom for large smoothing widths. On the other hand, the Silver and Chan (1991) code coefficients (a_n) produce a negatively biased estimate of the degrees of freedom. In the case of Gaussian noise, our coefficients produce an unbiased estimate while the Silver and Chan (1991) code coefficients produce a positively biased estimate of the degrees of freedom. If the amount of correlation between data points in the power was known, then comments about which coefficients were more appropriate could be made.

Our second test in Section 5.2 reveal that assumptions used to generate the degrees of freedom and consequently the errors do not hold. Therefore the windowed noise is not Gaussian and therefore the sum of squares $S(\phi, \delta t)$ (Section 2.7) used to construct the degrees of freedom does not follow a chi-squared distribution (which is the sum of squared standard normal variables (Appendix A.5)). There appears to be too much structure in the signal. We also find the structure of the noise is quite different to background noise. The power of the windowed noise is significantly larger than the power of the background noise suggesting that the windowed noise still contains some coherent signal possibly from scattering.

Even though the assumptions used to generate the errors do not hold, our last set of diagnostic plots from Section 5.3 reveal that for good quality events, the fast direction errors from the Silver and Chan (1991) method match the spread of the fast direction estimates. However, the delay time errors from the Silver and Chan (1991) method appear to be too small. Although, this could be caused by the bimodal delay time estimates that have been caused by cycle skipping.

As indicated by the concluding statements, this thesis has made many new discoveries. However, there is still room for improvement and other areas to explore. For instance, the automated grading produces a grade for a single event. Other automated programs such as Mfast (Teanby et al. (2004*a*), Wessel (2010), Savage et al. (2010)) do grading using a combination of individual events and clustering (see Section 1.3.3). One could use our automated method to determine the clusters or explore an automated system that utilises both. With regard to our automated method, it would also be useful to test it on a less homogeneous dataset that has been manually graded. It would be interesting to see whether it can predict the manual grades of a non-homogeneous dataset and how much the regression coefficients and model change by when another dataset is used. Incorporating criteria to detect null measurements would also be beneficial.

We conducted an analysis of the errors produced by the Silver and Chan (1991) method. The assumptions did not hold so the errors from the Silver and Chan (1991) method and other variants (e.g. Wolfe and Silver (1998)) can not be trusted. Finding a more reliable method of calculating the errors would be beneficial.

To conclude, we have provided the first in-depth analysis of the Silver and Chan (1991). Our research shows that the estimates of the errors are not correct and the assumptions behind them do not hold. This highlights the need for a new method of estimating the errors of shear wave splitting parameters.

Glossary

AIC Akaike information criterion, which is a measure of how well a model fits the data. 12, 88

aliasing an effect where two different signals become indistinguishable because one of them has not been sampled enough. 48

anisotropic medium a medium that possesses anisotropy. 2, 17

anisotropy dependence on direction. 2, 17, 109

ANOVA a statistical test for whether a group of means are all equal or at least one mean differs. 44

back azimuth measured from a seismometer, the clockwise angle between North and the source of seismic energy. 8, 31

background noise noise contained in the wave signal caused by external factors such as the wind or the ocean. 9, 126

band pass filter filter that allows frequencies within a certain band to pass through. 10, 63, 123

BOR seismic station on Reunion Island. 61

chi-squared distribution statistical ditribution that is defined by a sum of squared standard normal variables and has a single parameter called the degrees of freedom. 43, 111

coherency amount of correleation (as determined by the cross correlation function) between two time series at each frequency. 62

contourLines() an R function that calculates the contour lines of a data set and can also calculate contour lines for a particular level. 72

convolution a function that transforms two separate functions into a new single function at different lag times between the function. 8, 46, 111

Cook's Distance a measure of whether particular observations are influencing the regression analysis. 95

covariance measure of how two variables vary together. 7, 36, 65, 139

cross correlation method shear wave splitting method where the correlation between two components is maximised. 5

cycle skipping an error in the shear wave splitting parameters caused by matching the wrong peak of the wave form or a peak with a trough, which leads to a cycle being skipped. 10, 66, 101, 137, 139

damping reduction of the ampltidues of a signal. 10

delay time the time separation between the fast and slow wave components. 5, 20, 65, 101, 130, 139

DFBETA measure of how much influence observations have on the coefficients of the multiple linear regression model. 96

eigenvalue method shear wave splitting method where the best parameters are found by minimising the smaller eigenvalue of the corrected covariance matrix of two orthogonal components. 5

empirical distribution step function that is the cumulative distribution (a function that describes the probabilities) of the sample. 122

F test a statistical test where the test statistic has an F distribution if the null hypothesis is true. 8

fast direction the direction in an anisotropic medium where the waves velocity is the fastest. Technically, it is an orientaiton or a polarisation but informally it is referred to as a direction. 5, 17, 61, 101, 130, 139

fitted values an output value produced by a model (in this thesis the multiple linear regression). 88

focal mechanism the geometric representation of the fault slip during an earthquake. 7, 11, 109

Fourier transform a transform that converts a function in the time domain to a function in the frequency domain. 47, 112

Gaussian noise random uncorrelated noise. 8, 58, 111, 112

hat matrix a matrix that maps the observed data to the fitted values under the linear model. 92

Hooke's law a rule relating force and displacement for elastic materials. 19

hypothesis test process of testing a claim about a parameter based on the sample. 93

in.out() an R function that returns true if a point is inside a polygon which is determined by using the even odd rule. 76

incoming polarisation orientation of particle motion before the wave passes through an anisotropic medium. 4, 101, 139

influential point observation that strongly affects the fitted values or the regression coefficients. 96

interquartile range the middle 50% of the data that is between the lower and upper quartiles. 130

inverse splitting operator operator that reverses the splitting operation. 7, 24

isotropic medium a medium where velocities are independent of direction. 2, 16

jackknife residual a residual where the function that provides an estimate of the observation is based not on the entire data set but the data set with one observation removed i.e. it is a delete one jackknife. 92

jitter() an R function that adds a small amount of random noise to each observation. 69

Kolmogorov-Smirnov test a nonparametric test that tests whether two distributions are the same. 121

lattice preferred orientation (LPO) a property of rocks that have a preferred distribution of the crystallographic axis. 2

leverage point observations that have little effect on the regression coefficients but can influence the summary statistics. ix, 97

lm() An R function that fits linear models such as multiple linear regression. 88

multicollinearity the occurrence of explanatory variables that are highly correlated. 94

multiple linear regression a model of the linear relationship between a single dependent variable and multiple explanatory variables. 71, 139

negatively biased phenomenon where an estimator consistently underestimates the parameter. 116

noise any unwanted component of the signal. 4, 25, 62, 102

null measurement a measurement that cannot detect shear wave splitting because of one of three reasons: splitting has not occurred or the initial polarisation is parallel to either the fast or slow direction. 5, 10, 61, 109, 139

Nyquist frequency half the sampling rate. 48, 122

outlier extreme observation in the data set. 94

P wave longitudinal seismic wave, which travels faster than the S wave. viii, 3, 63

p-value the probability of obtaining the text statistic or a more extreme value if the null hypothesis is true. 89

particle motion movement of particles during wave propagation. 7, 39, 65, 104

positively biased phenomenon where an estimator consistently overestimates the parameter. 58, 116

power rate at which energy is transferred which is $u(t) = |u(t)|^2$ for a signal $u(t)$. 46, 111, 119

predictor a variable used to predict the fitted value. 88

probability density function a function that shows the relative likelihood of a random variable. 163

qq plot a plot comparing two distributions. If these distributions are the same then the qq plot lies on a diagonal line. 93

radial horizontal component direction that is parallel but in the opposite direction to the source-receiver back azimuth. 3, 29

randomisation a process used to induce randomness. 9

regression coefficient the coefficient in front of each explanatory variable (predictor). 88, 140

residual the difference between an observation and the estimate of that observation. 88

rotation matrix a matrix that rotates a coordinate system about an axis by a given angle. 26, 35

S wave transverse seismic wave, which travels more slowly than the P wave. 3, 27, 61, 126

seismometer device that measures ground motions. 1, 29

self-intersecting polygon a polygon that has at least a pair of edges crossing. 76

SH S wave displacements that are horizontal or normal to the plane. 3

Shapiro-Wilk test a statistical test that checks whether the data is normally distributed. 93

shear wave see S wave. 2, 15, 90, 102

shear wave splitting phenomenon where a wave splits when it encounters an anisotropic medium, creating two roughly orthogonal waves with one travelling faster than the other. 1, 63, 139

signal to noise ratio a measure of signal strength, the level of the signal relative to the level of background noise. 62

simple polygon a polygon where none of the edges cross each other. 76

SKKS wave similar to an SKS wave but it bounces off the outer core boundary before it leaves the core. 4

SKS wave a seismic wave that is initially an S wave but transforms into a P wave as it passes through the liquid core before returning to an S wave. 4, 43

smoothing width the proportion of points used to create correlations between the data. 112

splitting intensity the ratio of the amplitude of the time component against the time derivative of the radial component of a seismogram. 8

splitting operator operator that splits a shear wave into two quasi shear waves. 20

splitting parameters the two parameters that define shear wave splitting (the fast direction and the delay time). 5, 15, 74, 103

standard deviation measure of the spread of the data. 99, 112

standard error standard deviation of an estimate. 43, 89, 111

stepwise regression the process of selecting a multiple regression model, which is done iteratively by selecting a new explanatory variable or removing one on each iteration. 89

summary statistics statistics that summarise a dataset, for example the standard deviation. 95, 126

SV S wave displacement that is in the vertical plane and also the horizontal projection of $\hat{\mathbf{b}}$. 3

transverse horizontal component that is perpendicular to the radial direction. 3, 29

transverse minimisation method shear wave splitting method that finds the best parameters by minimising the energy on the transverse component. 5

trapezoidal rule a numerical integration method that approximates the area under a curve by summing up the area of small trapezia. 49, 164

window the section of data used for analysis that contains the shear wave. 5, 46, 63, 121

List of Symbols

k	anisotropy coefficient
x	point along the x axis
$\eta^2(t)$	power of the noise
$f(t)$	filter in the time domain
f_n	filter in the time domain indexed by the number of points
$g(t)$	Gaussian noise in the time domain
g_n	Gaussian noise in the time domain indexed by the number of points
σ^2	variance of the Gaussian noise process
$\tilde{\cdot}(\omega)$	fast Fourier transform (fft) of $\cdot$
c	constant
$\tilde{g}^*(\omega)$	complex conjugate of the fft of $g(t)$
f_s	sampling frequency
f	frequency
N	total number of data points
$\langle E \rangle$	first moment of the power
$\langle E^2 \rangle$	second moment of the power
$\langle z^4 \rangle$	fourth moment of a standard normal variable
a_n	coefficients to calculate the second moment of F and E
b_n	coefficients to calculate the fourth moment of E
b_n^{SC}	Silver and Chan coefficients to calculate the fourth moment of E
$\widehat{\nu}$	estimate of the degrees of freedom ν based on b_n coefficients
$\widehat{\nu}_{SC}$	estimate of the degrees of freedom ν based on a_n coefficients
δE^2	a variable related to E^2 and its mean with an expected value of zero
ω	angular frequency

δE_4	a variable related to E_4 and its mean with an expected value of zero
C_{XY}	coherency between x and y
G_{XY}	cross spectral density between x and y
$G_{..}$	auto spectral density of $\cdot$
T_2	analysis window start time
T_3	analysis window end time
T_0	shear wave arrival time
p	Mfast symbol for what we call $\hat{\mathbf{p}}$
$p \perp$	Mfast symbol for what we call $\hat{\mathbf{a}}$
$spol$	Mfast symbol for what we call α
t	point in time
$\widehat{\phi}$	estimate of the bearing fast direction
$\widehat{\delta t}$	estimate of the delay time
$\widehat{\alpha}$	estimate of the incoming polairsation
GC.	grading criteria $\cdot$
A_c	number of points in the corrected particle motion polygon
u_{wpc}	windowed corrected $\hat{\mathbf{p}}$ component
λ_2^{max}	largest second eigenvalue
$\mathbb{1}_{95CR}$	indicator function for the 95% confidence region
rms	root mean square
u_{wac}	windowed corrected $\hat{\mathbf{a}}$ component
i	imaginary number, $i^2 = -1$
u_{iac}	corrected $\hat{\mathbf{a}}$ component before the window
$\beta_0, \cdots, \beta_p$	regression coefficients
$\widehat{Y}_i$	automated grades
X_i	grading criteria
Y_i	manual grades
ϵ	normally distributed error term
e_i	residuals
AIC	Akaike Information Criterion
p	number of preditors (grading criteria)

$r_{(-i)}$	jackknife residual
$\mathbf{u}(\omega)$	isotropic shear wave vector
MSE	mean squared error
$\mathbf{H}$	hat matrix
κ	constant
D_i	Cook Distance
$\hat{y}_k$	fitted values for observation k
$\hat{y}_{k(i)}$	fitted values for observation k when observation i is removed
$\hat{\beta}_{j(i)}$	jth regression coefficient when obseration i is removed
c_{jj}	jth diagonal of $(\mathbf{X}^T\mathbf{X})^{-1}$
$\hat{\mathbf{x}}_1$	eigenvector corresponding to the eigenvalue λ_1
δt_{CS}	cycle skipped delay time
$\mathbf{A}(\omega)$	complex amplitude vector
T	dominant period
w	smoothing width in seconds
r_j	scaled degrees of freedom for the jth replicate
`nrep`	number of replicates
SE	standard error
SD	standard deviation
LQ	lower quartile
UQ	upper quartile
IQR	interquartile range
Φ	Euler angles
$w(\omega)$	norm of the amplitude with angular frequency ω
G	Gaussian function
T_0	time at which the wave arrives at the surface
$\hat{\mathbf{b}}$	propagation vector
v_{max}	velocity of the shear wave along the fast direction
$\hat{\mathbf{p}}$	polarisation vector
$\hat{\mathbf{a}}$	null vector
$\mathcal{P}$	propagation coordinates $[\hat{\mathbf{p}}\hat{\mathbf{a}}\hat{\mathbf{b}}]$

$\hat{\mathbf{f}}$	fast direction
$\hat{\mathbf{s}}$	slow direction
$\mathcal{S}$	splitting coordinates $[\hat{\mathbf{f}}\hat{\mathbf{s}}\hat{\mathbf{b}}]$
σ_{ij}	stress tensor
$c_{ijk\ell}$	elastic moduli
$\epsilon_{k\ell}$	strain tensor
ρ	density
v_{min}	velocity of the shear wave along the slow direction
$V_{i\ell}$	polarisation vector
$C_{ijk\ell}$	elastic coefficients
b_j	j*th* component of the propagation vector $\hat{\mathbf{b}}$
b_k	k*th* component of the propagation vector $\hat{\mathbf{b}}$
v_1	velocity of the faster travelling wave
v_2	velocity of the slower travelling wave
t_1	time it takes the faster travelling wave to reach the surface
t_2	time it takes the slower travelling wave to reach the surface
L	the length of the path the wave has travelled
β_0	shear wave velocity in an isotropic medium
$\bar{v}$	average velocity
$\delta\beta_1$	increase in speed relative to β_0
$\delta\beta_2$	decrease in speed relative to β_0
$\delta\zeta$	unitless relative velocity
$\mathbf{\Gamma}$	splitting operator
$\delta\mathbf{T}$	time difference tensor
$\mathbf{u}_s(\omega, t)$	split shear wave
$\tilde{\mathbf{u}}_(\omega, t)$	corrected shear wave
$\mathbf{\Gamma}^{-1}$	inverse splitting operator
$\mathbf{\Gamma}^{\dagger}$	complex conjugate transpose of $\mathbf{\Gamma}$
$\mathbf{R}_i(\cdot)$	rotation matrix at angle $\cdot$ anticlockwise around the ith axis
ϕ	fast direction

$\mathbf{U}_{\mathcal{X}\mathcal{C}}$	matrix of unit vectors of frame $\mathcal{C}$ explained in the coordinate system $\mathcal{X}$
$\mathbf{C}_3$	shift matrix to convert geographic coordinates into bearing coordinates
$\hat{\mathbf{t}}$	transverse component
$\hat{\mathbf{q}}$	component orthogonal to $\hat{\mathbf{t}}$ and $\hat{\mathbf{b}}$
$\mathcal{E}$	emerging ray coordinates $[\hat{\mathbf{q}}\hat{\mathbf{t}}\hat{\mathbf{b}}]$
θ	angle between $\hat{\mathbf{p}}$ and $\hat{\mathbf{q}}$
$\hat{\mathbf{r}}$	radial component
$\hat{\mathbf{z}}$	vertical component
$\mathcal{R}$	ray planar coordinates $[\hat{\mathbf{r}}\hat{\mathbf{t}}\hat{\mathbf{z}}]$
i	angle between $\hat{\mathbf{b}}$ and $\hat{\mathbf{z}}$
δt	delay time
$\hat{\mathbf{e}}$	east component
$\hat{\mathbf{n}}$	north component
$\mathcal{G}$	geographical coordinates $[\hat{\mathbf{e}}\hat{\mathbf{n}}\hat{\mathbf{z}}]$
b	angle between $\hat{\mathbf{e}}$ and $\hat{\mathbf{r}}$
$\hat{\mathbf{d}}$	downwards component $\hat{\mathbf{d}} = -\hat{\mathbf{z}}$
$\mathcal{B}$	bearing coordinates $[\hat{\mathbf{q}}\hat{\mathbf{t}}\hat{\mathbf{b}}]$
$\hat{\boldsymbol{\delta}}_1$	column vector $[100]^T$
$\hat{\boldsymbol{\delta}}_2$	column vector $[010]^T$
$\hat{\boldsymbol{\delta}}_3$	column vector $[001]^T$
$W(\omega, t)$	$w(\omega)\exp(-i\omega t)$
$u(x, t)$	wave amplitude at time t and position x
W_+	$W(\omega, t + \delta t/2)$
W_-	$W(\omega, t - \delta t/2)$
$[\hat{\mathbf{r}}_1\hat{\mathbf{r}}_2\hat{\mathbf{r}}_3]$	alternative symbols for $[\hat{\mathbf{r}}\hat{\mathbf{t}}\hat{\mathbf{z}}]$
$\mathbf{D}$	diagonal matrix with entries $\exp(i\omega\delta t/2)$ $\exp(-i\omega\delta t/2)$ and 1
Δ_{ii}	diagonal time shift matrix
ϕ_b	bearing fast direction, clockwise angle between $\hat{\mathbf{n}}$ and $\hat{\mathbf{f}}$
α	bearing polarisation, clockwise angle between $\hat{\mathbf{n}}$ and $\hat{\mathbf{p}}$
β	bearing back azimuth, clockwise angle between $\hat{\mathbf{n}}$ and $\hat{\mathbf{r}}$ + 180°
u_1	fast wave

u_2	slow wave
A	maximum amplitude of the wave
$\mathbf{c}$	covariance matrix
$\langle \cdot \rangle$	expected value of $\cdot$
λ_1	largest eigenvalue
λ_2	smallest eigenvalue in the 2D case
λ_3	smallest eigenvalue in the 3D case
$S(\phi, \delta t)$	sum of squares
$\lambda_2(\phi, \delta t)$	smallest eigenvalue for each pair of ϕ and δt
n	number of data points/events
$S_{min}(\phi, \delta t)$	minimum value of the sum of squares
χ^2	chi-squared distribution
k	wave number
SSG	sum of squares for the group
SSE	sum of squared errors
SST	total sum of squares
k	degrees of freedom = number of splitting parameters = 2
$F_{k,n-k}^{1-\alpha}$	inverse F distribution with k and $n-k$ degrees of freedom
λ_2^{95CR}	second eigenvalue corresponding to the boundary of the 95% confidence region
λ_2^{min}	smallest second eigenvalue
ν	reduced degrees of freedom
E	power
T	period

References

Aki, K. and Richards, P., eds (2002), *Quantitative Seismology*, 2 edn, University Science Books, CA. Chapter 4. 109

Angerer, E., Crampin, S., Li, X. and Davis, T. (2002), 'Processing, modelling and predicting time-lapse effects of overpressured fluid-injection in a fractured reservoir', *Geophysical Journal International* **149**, 267–280. 3

Aster, R., Shearer, P. and Berger, J. (1990), 'Quantitative measurements of shear wave polarizations at the anza seismic network, southern california: Implications for shear wave splitting and earthquake prediction', *Journal of Geophysical Research* **95**, 12449–12473. 11

Babuska, V. and Cara, M. (1991), *Seismic Anisotropy in the Earth*, Kluwer Academic Publishers, Netherlands. 2, 19

Belsley, D., Kuh, E. and Welsch, R. (2004), *Regression Diagnostics*, John Wiley and Sons Inc, New York. Section 2.1. 97

Brockwell, P. and Davis, R. (2006), *Time Series: Theory and Methods*, Springer, New York. Chapter 11. 42

Chevrot, S. (2000), 'Multichannel analysis of shear wave splitting', *Journal of Geophysical Research* **105**, 21579–21590. 8

Clark, M. and Randal, J. (2011), *A first course in applied statistics*, Pearson, New Zealand. 44

Crampin, S. and Chastin, S. (2003), 'A review of splitting in the crack critical crust', *Geophysical Journal International* **155**, 221–240. 3

Crampin, S. and Gao, Y. (2006), 'A review of splitting techniques for measuring shear-wave splitting above small earthquakes', *Physics of the earth and planetary interiors* **159**, 1–14. 6

Crocker, M. (1998), *Handbook of Accoustics*, John Wiley and Sons Inc, New York. Chapter 81, Sections 4 and 5. 62

Demidenko, E. (2004), *Mixed Models: Theory and Applications*, John Wiley and Sons Inc, New Jersey. Chapter 1. 88

Evans, M., Kendall, J. and Willemann, R. (2006), 'Automated sks splitting and upper-mantle anisotropy beneath canadian seismic stations', *Geophysical Journal International* **165**, 253–274. 11

Evans, P. (2001), 'Rotations and rotation matrices', *Acta Crystallographica Section D* **57**, 1355–1359. 26

Feldman, R. and Valdez-Flores, C. (2010), *Applied Probability and Stochastic Processes*, Springer, New York. Chapter 3. 121

Fukao, Y. (1984), 'Evidence from core-reflected shear waves for anisotropy in the earth's mantle', *Nature* **309**, 695–698. 6

Gerst, A. and Savage, M. (2004), 'Seismic anisotropy beneath ruapehu volcano: A possible eruption forecasting tool', *Science* **306**, 1,543–1,547. 3

Google Earth (2011), 'Reunion island 19°30'12.36"s 47°27'26.08"e elev 4425ft'. Accessed March 2012.
URL: *http://www.google.com/earth/index.html* viii, 62

Hoaglin, D. and Welsch, R. (1978), 'The hat matrix in regression and anova', *American Statistician* **32**, 17–22. 98

Hormann, K. and Agathos, A. (2001), 'The point in polygon problem for arbitrary polygons'.
URL: *http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.19.5578* 76

Jenkins, G. and Watts, D. (1968), *Spectral Analysis and Its Applications*, Holden-Day, San Francisco. Chapter 5. 44, 49

Long, M. and Silver, P. (2009), 'Shear wave splitting and mantle anisotropy: measurements, interpretations and new directions', *Surveys in Geophysics* **30**, 407–461. 6, 7, 8

Ludlam, E. (2012), *C-Parse for Emacs*, CEDET.
URL: *http://cedet.sourceforge.net/cparse.shtml* 61

Marson-Pidgeon, K. and Savage, M. (2004), 'Modelling shear wave splitting observations from wellington, new zealand', *Geophysical Journal International* **157**, 853–864. 7

Mason, M. (2001), *Mechanics of Robotic Manipulation*, MIT Press, MA, USA. Chapter 3. 161

Matcham, I. (1997), Seismic anisotropy in the wellington region from loval events recorded at the iris station snzo, Master's thesis, Victoria University of Wellington, New Zealand. vii, 10, 103

Menke, W. and Levin, V. (2003), 'The cross-convolution method for interpreting sks splitting observations, with application to one and two-layer anisotropic earth models', *Geophysical Journal International* **154**, 379–392. 8, 9, 12

Montellier, V. and Chevrot, S. (2010), 'How to make robust splitting measurements for single-station analysis and three-dimensional imaging of seismic anisotropy', *Geophysical Journal International* **182**, 311–328. 8, 12

Peng, Z. and BenZion, Y. (2004), 'Systematic analysis of crustal anisotropy along the karadere-düzce branch of the north anatolian fault', *Geophysical Journal International* **159**, 253–274. 11, 12, 109

Rawlings, J., Pantula, S. and Dickey, D. (1998), *Applied Regression Analysis: A Research Tool. 2ed.*, Springer, New York. Chapter 11. 96

Rencher, A. (2000), *Linear Models in Statistics*, John Wiley and Sons Inc, New York. Chapters 7 and 9. 88

Rivemale, E. (2012), Microseismicity studies at Piton de la Fournaise volcano, PhD thesis, l'université Joseph Fourier, Grenoble. 61

Sandvol, E. and Hearn, T. (1994), 'Bootstrapping shear-wave splitting errors', *Bulletin of the Seismological Society of America* **84**, 1,971–1,977. 9, 10, 11

Savage, M. (1999), 'Seismic anisotropy and mantle deformation: What have we learned from shear wave splitting?', *Reviews of Geophysics* **37**. 4, 7, 8

Savage, M., Wessell, A., Teanby, N. and Hurst, A. (2010), 'Automatic measurement of shear wave splitting and applications to time varying anisotropy at mount ruapehu volcano, new zealand', *Journal of Geophysical Research* **115**. 11, 42, 61, 62, 109, 140

Silver, P. and Chan, W. (1991), 'Shear wave splitting and subcontinental mantle deformation', *Journal of Geophysical Research* **96**(10), 16,429–16,454. x, xi, xiii, 1, 3, 7, 8, 9, 10, 11, 12, 13, 15, 16, 19, 20, 21, 22, 28, 35, 36, 37, 41, 42, 43, 45, 46, 53, 54, 55, 56, 57, 58, 59, 61, 64, 65, 101, 109, 110, 111, 114, 115, 116, 117, 119, 122, 126, 130, 131, 132, 133, 134, 135, 136, 137, 139, 140, 141

Silver, P. and Savage, M. (1994), 'The interpretation of shear-wave splitting parameters in the presence of two anisotropic layers', *Geophysical Journal International* **119**, 949–963. 7

Stein, S. and Wysession, M. (2003), *An Introduction to Sesimology, Earthquakes and Earth Structure*, Blackwell Publishing, MA, USA. 2, 3, 4, 19

Strang, G. (2003), *Introduction to linear algebra*, Wellesley-Cambridge Press, MA, USA. Chapter 4. 161

Teanby, N., Kendall, J. and van der Baan, M. (2004*a*), 'Automation of shear-wave splitting measurements using cluster analysis', *Bulletin of the Seismological Society of America* **94**, 453–463. 6, 10, 42, 61, 62, 140

Teanby, N., Kendall, J. and van der Baan, M. (2004*b*), 'Stress-induced temporal variations in seismic anisotropy observed in microseismic data', *Geophysical Journal International* **156**, 459–466. 3

Thaller, B. (2000), *Visual Quantum Mechanics: Selected Topics with Computer-Generated Animations of Quantum Mechanical Phenomena*, Springer, New York. Chapter 2, Section 7. 115

Vecsey, L., Plomerova, J. and Babuska, V. (2008), 'Shear-wave splitting measurements - problems and solutions', *Tectonophysics* **462**, 178–196. 5, 7, 10, 11

Wessel, A. (2010), Automatic shear wave splitting measurements at mt. ruapehu volcano, new zealand, Master's thesis, Victoria University of Wellington, New Zealand. 42, 61, 62, 109, 140

Wolfe, C. and Silver, P. (1998), 'Seismic anisotropy of oceanic upper mantle: Shear wave splitting methodologies and observations', *Journal of Geophysical Research* **103**(B1), 749–771. 7, 9, 12, 140

Wustefeld, A., Al-Harrasi, O., Verdon, J., Wookey, J. and Kendall, J. (2010), 'A strategy for automated analysis of passive microseismic data to image seismic anisotropy and fracture characteristics', *Geophysical Prospecting* pp. 1–19. 11

Wustefeld, A. and Bokelmann, G. (2007), 'Null detection in shear-wave splitting measurements', *Bulletin of the Seismological Society of America* **97**(4), 1204–1211. 10, 11

Wustefeld, A., Bokelmann, G., Zaroli, C. and Barruol, G. (2008), 'Splitlab: A shear-wave splitting environment in matlab', *Computers and Geosciences* **34**, 515–528. 12, 42

Yoshikawa, T. (1990), *Foundations of Robotics: Analysis and Control*, MIT Press, MA, USA. Appendix 1. 102

Appendix A

Useful Definitions and Results

A.1 Matrix Exponential

The definition of a matrix exponential of a square matrix $\mathbf{A}$ is

$$\exp(\mathbf{A}) = \Sigma_{k=0}^{\infty} \frac{\mathbf{A}^k}{k!} \qquad \text{where } \mathbf{A}^0 \equiv \mathbf{I} \tag{A.1}$$

A useful property of the matrix exponential is that

$$\exp(\mathbf{A}) \exp(\mathbf{B}) = \exp(\mathbf{A} + \mathbf{B}) \tag{A.2}$$

if $\mathbf{AB} = \mathbf{BA}$ i.e. if the matrices commute.

A.2 Why $\hat{\mathbf{f}}\hat{\mathbf{f}}^T + \hat{\mathbf{s}}\hat{\mathbf{s}}^T = \mathbf{I}$ in 2D

Theorem A.2.1. *If* $\hat{\mathbf{f}}$ *is defined as*

$$\hat{\mathbf{f}} = \begin{bmatrix} \cos\phi \\ \sin\phi \end{bmatrix}$$

and $\hat{\mathbf{f}} \cdot \hat{\mathbf{s}} = 0$ *then*

$$\hat{\mathbf{s}} = \begin{bmatrix} -\sin\phi \\ \cos\phi \end{bmatrix}$$

Proof.

$$\hat{\mathbf{s}} = \begin{bmatrix} a \\ b \end{bmatrix}$$

We know that $\hat{\mathbf{s}}$ is a unit vector so $\|\hat{\mathbf{s}}\| = 1$ and $a^2 + b^2 = 1$. Since $\hat{\mathbf{s}}$ and $\hat{\mathbf{f}}$ are orthogonal their dot product is zero

$$\begin{aligned}
\hat{\mathbf{s}} \cdot \hat{\mathbf{f}} = 0 &= a\cos\phi + b\sin\phi \\
&= a\cos\phi + \sqrt{1-a^2}\sin\phi && \text{rearrange and substituting } a^2 + b^2 = 1 \\
-a\cos\phi &= \sqrt{1-a^2}\sin\phi \\
a^2\cos^2\phi &= (1-a^2)\sin^2\phi \\
a^2(\cos^2\phi + \sin^2\phi) &= \sin^2\phi \\
a^2 &= \sin^2\phi && \sin^2\phi + \cos^2\phi = 1 \\
a &= \pm\sin\phi \\
\Rightarrow b &= \pm\sqrt{1-a^2} = \pm\cos\phi
\end{aligned} \tag{A.3}$$

If $a = -\sin\phi$ then

$$\begin{aligned}
b\sin\phi &= -a\cos\phi \\
&= \sin\phi\cos\phi \\
\Rightarrow b &= \cos\phi
\end{aligned} \tag{A.4}$$

$$\begin{aligned}
\hat{\mathbf{f}}\hat{\mathbf{f}}^T + \hat{\mathbf{s}}\hat{\mathbf{s}}^T &= \begin{bmatrix} \cos\phi \\ \sin\phi \end{bmatrix} \begin{bmatrix} \cos\phi & \sin\phi \end{bmatrix} + \begin{bmatrix} -\sin\phi \\ \cos\phi \end{bmatrix} \begin{bmatrix} -\sin\phi & \cos\phi \end{bmatrix} \\
&= \begin{bmatrix} \cos^2\phi & \cos\phi\sin\phi \\ \sin\phi\cos\phi & \sin^2\phi \end{bmatrix} + \begin{bmatrix} \sin^2\phi & -\sin\phi\cos\phi \\ -\cos\phi\sin\phi & \cos^2\phi \end{bmatrix} \\
&= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \mathbf{I}
\end{aligned} \tag{A.5}$$

□

The above explains the case for two dimensions. More generally, if $\mathbf{R} = \hat{\mathbf{u}}_1, \cdots, \hat{\mathbf{u}}_m$ is an $m \times m$ matrix of orthogonal unit vectors then

$$\mathbf{R}^T\mathbf{R} = \begin{bmatrix} \hat{\mathbf{u}}_1^T \\ \cdots \\ \hat{\mathbf{u}}_m^T \end{bmatrix} \begin{bmatrix} \hat{\mathbf{u}}_1^T \cdots \hat{\mathbf{u}}_m^T \end{bmatrix} = (\hat{\mathbf{u}}_i^T \hat{\mathbf{u}}_j) = \mathbf{I} \tag{A.6}$$

and it follows that

$$\mathbf{R}\mathbf{R}^T = \begin{bmatrix} \hat{\mathbf{u}}_1^T \cdots \hat{\mathbf{u}}_m^T \end{bmatrix} \begin{bmatrix} \hat{\mathbf{u}}_1^T \\ \cdots \\ \hat{\mathbf{u}}_m^T \end{bmatrix} = (\hat{\mathbf{u}}_i \hat{\mathbf{u}}_j^T) = \mathbf{I} \tag{A.7}$$

See Strang (2003) for more details.

A.3 Rotation with Euler Angles

A rotation matrix can perform a rotation using three Euler angles ($\Phi = (\phi, \theta, \psi)$) to describe a 3D rotation, where

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

The rotations are shown in Figure A.1 using a zyz right handed system convention (Mason, 2001)

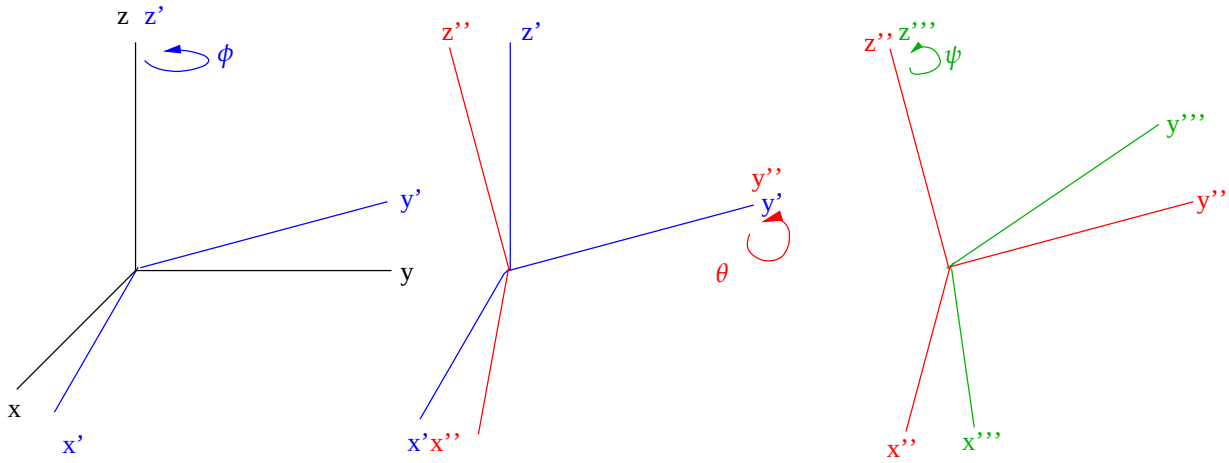


Figure A.1: Diagram showing how to rotate a coordinate system using zyz right hand convention.

The diagram illustrates an anticlockwise rotation by ϕ about the z axis, then an anticlockwise rotation by θ about the y' axis and finally an anticlockwise rotation by ψ about the z'' axis hence the name zyz convention.

This rotation can be expressed with the following matrix using the same notation that is in Section 2.5.2

$$\mathbf{R}(\Phi) = \mathbf{R}_3(\phi)\mathbf{R}_2(\theta)\mathbf{R}_3(\psi) =$$
$$\begin{bmatrix} \cos\phi\cos\theta\cos\psi - \sin\phi\sin\psi & -\cos\psi\sin\phi - \cos\phi\cos\theta\sin\psi & \cos\phi\sin\theta \\ \cos\phi\sin\psi + \cos\theta\cos\psi\sin\phi & \cos\phi\cos\psi - \cos\theta\sin\phi\sin\psi & \sin\phi\sin\theta \\ -\cos\psi\sin\theta & \sin\theta\sin\psi & \cos\theta \end{bmatrix}$$

In special cases such as a vertically travelling ray ($\theta = 0$ or $\theta = \pi$) the rotation matrix simplifies.

A.4 Proof that the Trace is Invariant in the 2D Problem

Recall that

$$\text{cov}(\tilde{\mathbf{u}}(\omega, t)) = \mathbf{U}_{\mathcal{XS}} \left(\text{cov}(\mathbf{D}\mathbf{U}_{\mathcal{XS}}\mathbf{u}_s(\omega, \delta t)\right) \mathbf{U}_{\mathcal{XS}}^T \tag{A.8}$$

where $\mathbf{U}_{\mathcal{XS}} = [\hat{\mathbf{f}}\hat{\mathbf{s}}\hat{\mathbf{b}}]$. The covariance matrix has the variances down the main diagonal and the covariances off the diagonal. The trace of a matrix is the sum of the elements down the main diagonal therefore

$$\begin{aligned}
\text{tr}\left[\text{cov}(\tilde{\mathbf{u}}(\omega, t))\right] &= \text{tr}\left[\mathbf{U}_{\mathcal{XS}} \left(\text{cov}(\mathbf{D}\mathbf{U}_{\mathcal{XS}}\mathbf{u}_s(\omega, \delta t)\right) \mathbf{U}_{\mathcal{XS}}^T\right] \\
&= \text{tr}\left[\left(\text{cov}(\mathbf{D}\mathbf{U}_{\mathcal{XS}}\mathbf{u}_s(\omega, \delta t)\right) \mathbf{U}_{\mathcal{XS}}^T\mathbf{U}_{\mathcal{XS}}\right] \\
&= \text{tr}\left[\left(\text{cov}(\mathbf{D}\mathbf{U}_{\mathcal{XS}}\mathbf{u}_s(\omega, \delta t)\right)\right] \qquad \text{(Equation 2.34)} \\
&= \sum_i \text{var}\left(\left[\mathbf{D}\mathbf{U}_{\mathcal{XS}}\mathbf{u}_s(\omega, \delta t)\right]_{ii}\right) \\
&= \sum_i \text{var}\left(\left[\mathbf{D}_{ii}\hat{\boldsymbol{\delta}}_i\mathbf{u}_s(\omega, \delta t)\right]_{ii}\right) \qquad \text{(where } [\hat{\boldsymbol{\delta}}_1\hat{\boldsymbol{\delta}}_2\hat{\boldsymbol{\delta}}_3] = [\hat{\mathbf{f}}\hat{\mathbf{s}}\hat{\mathbf{b}}]\text{)} \\
&= \sum_i \text{var}\left(\left[\hat{\boldsymbol{\delta}}_i{}^T\mathbf{D}_{ii}\mathbf{u}_s(\omega, \delta t)\right]_{ii}\right) \\
&= \sum_i \left(\int_{-\infty}^{\infty} [\hat{\boldsymbol{\delta}}_i{}^T\mathbf{D}_{ii}\mathbf{u}_s(\omega, \delta t)]^2 dt - \left[\int_{-\infty}^{\infty} \hat{\boldsymbol{\delta}}_i{}^T\mathbf{D}_{ii}\mathbf{u}_s(\omega, \delta t) dt\right]^2\right)
\end{aligned} \tag{A.9}$$

$\mathbf{D}_{ii}$ is a time shift operator. Since we are integrating over all of t we can drop this term. This is the change of variable technique mentioned in Section 2.6. Simplifying yields,

$$\begin{aligned}
\text{tr}\left[\text{cov}(\tilde{\mathbf{u}}(\omega, t))\right] &= \sum_i \left(\int_{-\infty}^{\infty} [\hat{\boldsymbol{\delta}}_i{}^T\mathbf{u}_s(\omega, \delta t)]^2 dt - \left[\int_{-\infty}^{\infty} \hat{\boldsymbol{\delta}}_i{}^T\mathbf{u}_s(\omega, \delta t) dt\right]^2\right) \\
&= \sum_i \text{var}\left(\hat{\boldsymbol{\delta}}_i{}^T\mathbf{u}_s(\omega, \delta t)\right) \\
&= \text{tr}\left[\text{cov}(\mathbf{U}_{\mathcal{XS}}^T\mathbf{u}_s(\omega, \delta t)\right] \\
&= \text{tr}\left[\mathbf{U}_{\mathcal{XS}}^T\text{cov}\left(\mathbf{u}_s(\omega, \delta t)\right)\mathbf{U}_{\mathcal{XS}}\right] \\
&= \text{tr}\left[\text{cov}\left(\mathbf{u}_s(\omega, \delta t)\right)\mathbf{U}_{\mathcal{XS}}\mathbf{U}_{\mathcal{XS}}^T\right] \\
&= \text{tr}\left[\text{cov}\left(\mathbf{u}_s(\omega, \delta t)\right)\mathbf{U}_{\mathcal{XX}}\right] \\
&= \text{tr}\left[\text{cov}\left(\mathbf{u}_s(\omega, \delta t)\right)\mathbf{I}\right] \\
&= \text{tr}\left[\text{cov}\left(\mathbf{u}_s(\omega, \delta t)\right)\right]
\end{aligned} \tag{A.10}$$

The final term has the covariance of the split wave rather than the corrected wave. This does not involve the time shift matrix or the the splitting coordinate system $\mathcal{S}$ so the trace is invariant with respect to changes in ϕ and δt.

A.5 Relationship between Gamma, Chi-squared and Exponential Distributions

The gamma distribution has the probability density function

$$f(x;k,\theta) = \frac{1}{\theta^k \Gamma(k)} x^{k-1} e^{-x/\theta} \qquad x \geq 0 \tag{A.11}$$

where

$k =$ a positive value that defines the shape of the distribution
$\theta =$ a positive value known as the scale parameter

The chi-squared distribution has the probability density function that depends on the parameter ν also known as the degrees of freedom.

$$f(x;\nu) = \frac{1}{2^{\nu/2}\Gamma(\nu/2)} x^{\nu/2-1} e^{-x/2} \; x \geq 0 \tag{A.12}$$

so it is a special case of the gamma distribution with $k = \nu/2$ and $\theta = 2$. The sum of k squared standard normal variables is also chi-squared with k degrees of freedom. In other words,

$$X = \sum_{i=1}^{k} Z_i^2 \sim \chi_k^2 \tag{A.13}$$

When the degrees of freedom are 2 the chi-squared distribution is equivalent to an exponential with $\theta = 2 = 1/\lambda$ where λ is the rate parameter.

$$\begin{aligned} f(x;\nu) &= \frac{1}{(1/\lambda)^{2/2}\Gamma(2/2)} x^{(2/2)-1} e^{-x/(1/\lambda)} \; x \geq 0 \\ &= \frac{\lambda}{\Gamma(2/2)} e^{-\lambda x} \; x \geq 0 \end{aligned} \tag{A.14}$$

For integers the gamma function can be evaluated as $\Gamma(n+1) = n!$ so $\Gamma(1) = 1$ and we obtain the probability density function of the exponential

$$f(x;\lambda) = \lambda e^{-\lambda x} \; x \geq 0 \tag{A.15}$$

therefore the exponential distribution is also a special case of the gamma distribution with $k = 1$ and $\theta = 1/\lambda$.

A.6 Trapezoidal Rule

The trapezoidal rule is a numerical integration technique used to approximate a definite integral $\int_a^b f(x)dx$ by summing up the area of $N-1$ trapezia all with equal width $w = \frac{b-a}{N-1}$. Figure A.2 illustrates this.

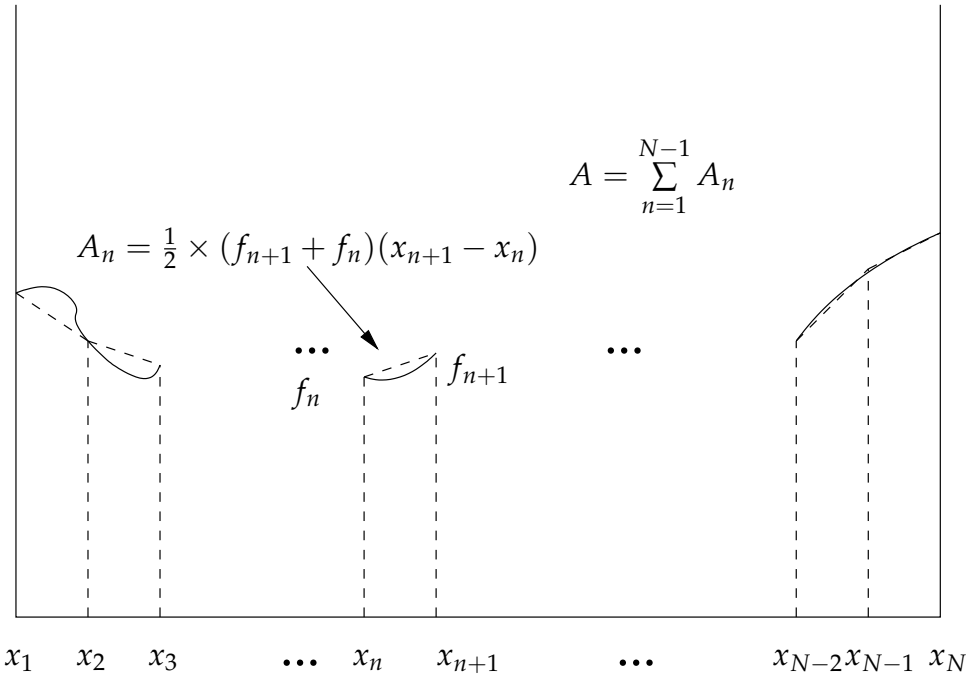


Figure A.2: Diagram illustrating the trapezoidal rule. The area A under the solid curve is approximated by summing up the area of the A_n trapezia.

The approximation is given by

$$\int_a^b f(x)dx \approx \frac{1}{2}\left[\sum_{i=1}^{N-1} w(f(x_{i+1} + f(x_i))\right] \tag{A.16}$$

where w is the width of the trapezia. If the trapezia are not of equal width then the area can be approximated by

$$\int_a^b f(x)dx \approx \frac{1}{2}\left[\sum_{i=1}^{N-1} (x_{i+1} - x_i)(f(x_{i+1} + f(x_i))\right] \tag{A.17}$$