

## Silver and Chan revisited

E. Walsh,<sup>1</sup> R. Arnold,<sup>1</sup> and M. K. Savage<sup>2</sup>

Received 18 June 2013; revised 12 September 2013; accepted 25 September 2013; published 24 October 2013.

[1] Seismic shear waves emitted by earthquakes can be modeled 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 polarization, that is determined by geologic conditions. We present here a comprehensive analysis of the Silver and Chan (1991) method, used to obtain shear wave splitting parameters, comprising theoretical derivations and statistical tests of the assumptions used to construct the standard errors. We find discrepancies in the derivations of equations in their article, with the most important being a mistake in how the standard errors are calculated. Our simulations suggest that the degrees of freedom are being overestimated by this method, and consequently, the standard errors are too small. Using a set of *S* waveforms from very similar shallow earthquakes on Reunion Island, we perform a statistical analysis on the noise of these replicates and find that the assumption of Gaussian noise does not hold. Further, the properties of background noise differ substantially from the noise obtained from the shear wave splitting analysis. However, we find that the estimated standard errors for the fast polarization are comparable to the spread in the fast polarization parameters between events. Delay time errors appear to be comparable to delay time estimates once cycle skipping is accounted for. Future work using synthetic seismograms with simulated noise should be conducted to confirm this is the case for earthquakes in general.

**Citation:** Walsh, E., R. Arnold, and M. K. Savage (2013), Silver and Chan revisited, *J. Geophys. Res. Solid Earth*, 118, 5500–5515, doi:10.1002/jgrb.50386.

## 1. Introduction

### 1.1. Motivation

[2] Seismic anisotropy is one of the few methods available to detect and quantify deformation beneath the Earth's surface. Common causes for anisotropy include the alignment of crystals such as olivine in the mantle or composite structures in the crust such as horizontal layers of sediment [Babuska and Cara, 1991]. Fluid-filled microcracks can also cause anisotropy with their opening and closing being an indication of the stress applied to them [Nur and Simmons, 1969; Crampin, 1984]. Applications include analyzing present and past tectonic activity [Silver and Chan, 1991], monitoring stress changes [Teanby et al., 2004a], 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 review of crustal shear wave splitting and a detailed discussion of applications.

[3] Shear wave splitting or seismic birefringence is one of the most common methods used to measure anisotropy. The Silver and Chan (SC) method for calculating shear wave splitting parameters [Silver and Chan, 1991] has been used for many years now and includes several different program implementations, with over 815 citations according to Google Scholar (as at 21 May 2013). However, it has long been suspected that the standard error estimates for the splitting parameters made by this method are too small [e.g., Sandvol and Hearn, 1994]. There has, as yet, been no in-depth assessment of the robustness of the SC method, which makes many assumptions to estimate these standard errors. It is the goal of this paper to revisit the SC method, including rederivations of the key results, and to test the assumptions that underlie the method. There have been criticisms that scientists often opt for software that is presented alongside peer-reviewed articles from well-known authors. However, most of the software itself is never peer reviewed or independently tested [Joppa et al., 2013]. Therefore, we also examine current commonly used implementations of the SC method and software to check their accuracy and faithfulness to the published method. In order to make realistic tests, we have identified a set of 146 observations of very similar earthquakes observed at a single seismometer station. These observations therefore have very similar signal components,

<sup>1</sup>School of Mathematics, Statistics and Operations Research, Victoria University of Wellington, Wellington, New Zealand.

<sup>2</sup>Institute of Geophysics, SGEES, Victoria University of Wellington, Wellington, New Zealand.

Corresponding author: M. K. Savage, School of Geography, Environment and Earth Sciences, Victoria University Wellington, PO Box 600, Wellington 6140, New Zealand. (martha.savage@vuw.ac.nz)

©2013. American Geophysical Union. All Rights Reserved.  
2169-9313/13/10.1002/jgrb.50386

and we can use them to characterize the variability in the noise components.

## 1.2. Review

### 1.2.1. Measuring Shear Wave Splitting

[4] A shear wave passing through an anisotropic medium will split into two perpendicular components with different velocities. For a near-vertical seismic ray, the splitting can be modeled by two parameters: the polarization of the fast wave (the azimuth of the fast polarization) and the delay time, which is the time separation between the two waves when they arrive at the surface. There are several techniques in common use which estimate these splitting parameters, including cross correlation, transverse minimization, the eigenvalue method, and the splitting intensity method [Vecsey *et al.*, 2008; Chevrot, 2000].

[5] Some of the methods attempt to transform the observed seismograms to remove the effects of splitting, and in doing so estimate the orientation of the splitting axis, and the delay time induced between the fast and slow components. Cross correlation involves finding a set of splitting parameters that maximize the correlation between the two orthogonal horizontal components [Fukao, 1984]. It can be used at a long period as well as at a high frequency [Long and Silver, 2009], so it is good for teleseisms as well as local earthquakes. The method is sensitive to the selection of appropriate end points for the window (the interval of a seismogram used for analysis) [Crampin and Gao, 2006], although Teanby *et al.* [2004b] provide a solution to this. This method assumes that the fast and slow waveforms are similar [Fukao, 1984], which may not be the case [Crampin and Gao, 2006].

[6] The Silver and Chan (SC) method (which employs the eigenvalue and, when the polarization is known, the transverse minimization method) [Silver and Chan, 1991], attempts to reverse the effects of shear wave splitting to retrieve the original wave. This method is described in some detail in section 2 below. Briefly, a search is made over the space of splitting parameters with the optimal parameters being those that either minimize the energy on one of the transverse components or (equivalently) minimize the smaller eigenvalue of the corrected covariance matrix of two orthogonal components. This is equivalent to creating the most linear particle motion [Long and Silver, 2009]. Both methods assume (1) that there is a single layer of anisotropic material and (2) that fast and slow axes lie in the horizontal plane [Vecsey *et al.*, 2008]. It is also recommended that the delay time  $\delta t$  should be at least 10 times smaller than the dominant period  $T$  of the wave to obtain reliable results [Wolfe and Silver, 1998]. However, in our simulations on synthetic seismograms with parameters  $\delta t = 0.4$  and  $T = 0.3$ , the SC method encountered no problems correctly estimating parameters. This was the case for Gaussian noise, background noise captured from a station in central New Zealand and coda-based noise taken after the  $S$  wave [Savage *et al.*, 2011]. The transverse minimization method has a third assumption that the initial energy from the wave is in the vertical plane containing the source and the receiver, and consequently that the incoming polarization is always known. Extensions of the SC method include allowing for multiple layers of anisotropy [e.g., Ozalaybey and Savage, 1994; Silver and Savage, 1994; Wolfe and Silver, 1998] and

known earthquake focal mechanisms [Marson-Pidgeon and Savage, 2004]. A comprehensive review can be found in Savage [1999].

[7] More recently, the splitting intensity method was introduced by Chevrot [2000], which requires the delay time to be considerably smaller than the dominant period of the wave. It exploits the functional relationship between the (split) waveforms on the radial and transverse components. In this approach, no assumptions are made regarding geological properties of the medium; however, the method requires knowledge of the incoming polarization and good back azimuth coverage, which could be the reason for its low use to date [Long and Silver, 2009]. This method was later improved by applying a Wiener filter beforehand to minimize noise [Montellier and Chevrot, 2010].

[8] Menke and Levin [2003] introduced a cross convolution method that also does not make geological assumptions regarding the anisotropy. It convolves the radial and tangential components with an impulse response from a hypothetical model and then attempts to minimize 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. However, the degrees of freedom of the  $F$  test are estimated assuming each seismogram data point is independent, which is highly unlikely.

[9] Shear wave splitting measurements with sinusoidal waveforms also encounter an additional problem known as cycle skipping, whereby two waveforms are mismatched by an integer number of half cycles. This is typical of local earthquakes when narrow band-pass filters are applied [Teanby *et al.*, 2004b] and also when the first peak of the shear wave is small [Matcham, 1997]. This problem can produce delay time estimates that are out by an integer multiple of half the dominant period, and a fast polarization that has a  $90^\circ$  ambiguity. Recently, it has been shown that cycle skipping also produces a  $90^\circ$  ambiguity in the incoming polarization [Walsh *et al.*, 2012; Walsh, 2013].

### 1.2.2. Estimating Shear Wave Splitting Parameter Errors

[10] The standard errors of the parameter estimates are constructed in many different ways. In the SC method, the errors are calculated by assuming that the background noise in the seismogram is a filtered Gaussian process. Correct estimation of the residual degrees of freedom in the filtered noise process is an important part of estimating the errors. These error estimates can be improved if data from multiple earthquake events are combined to estimate splitting at a single location [Wolfe and Silver, 1998].

[11] Sandvol and Hearn [1994] employed the statistical bootstrap applied to the SC method to calculate the standard errors, thus avoiding assumptions about the error distribution and the need to estimate the degrees of freedom. However, it does assume that the observations are all independent and identically distributed. The cross-convolution method [Menke and Levin, 2003] defines the error as the normalized, squared difference between the observed and the hypothetical time series of the seismograms.

### 1.2.3. Comparison of Techniques

[12] Vecsey *et al.* [2008] compared the cross correlation, eigenvalue, and transverse minimization methods. They found that the transverse minimization method works best,

but it can only be used when the incoming polarization is known. Also, the eigenvalue method is more prone to cycle skipping where the waveforms are mismatched by an integer number of half cycles, but less susceptible to noise and thus can provide more accurate splitting parameters compared to cross correlation.

[13] *Wustefeld and Bokelmann* [2007] compared the SC method to the rotation correlation method and find systematic differences that could be used to detect null measurements. *Montellier and Chevrot* [2010] used SKS and SKKS waves to compare the *Silver and Chan* [1991] and *Wolfe and Silver* [1998] stacking method to their splitting intensity method and find that stacking the error maps [*Wolfe and Silver*, 1998] considerably improves the reliability of splitting parameters and agrees with results from their splitting intensity method.

[14] *Long and van der Hilst* [2005] compare the SC method, the rotation correlation method, and the splitting intensity method [*Chevrot*, 2000] using a noisy data set from Japan. They find that in single anisotropic layer situations (which is assumed for *Silver and Chan* [1991] and *Wolfe and Silver* [1998]), results are in agreement. However, a mixture of these methods may be required for more complex settings such as multiple layers.

[15] The above mentioned methods in this subsection all assume single layers of anisotropy. *Silver and Savage* [1994] showed that when two layers of anisotropy are present, applying the SC method would result in apparent splitting measurements of the fast polarization and delay time that depend on the polarization of the wave before entering the first anisotropic layer. Later, *Rumpker and Silver* [1998] extended the analysis to multiple layers. Because the effect is due to the process of splitting in each layer and not to the method of analysis, the same effects are expected in all measurement techniques that use single event-station paths. The methods such as *Wolfe and Silver* [1998] and *Chevrot* [2000] effectively use the coherence of the measurements at different back azimuths and cannot be applied directly with two layers, but could be extended, much like the extensions advocated by *Menke and Levin* [2003].

### 1.3. Outline

[16] In this paper, we provide a comprehensive review of the SC method and an evaluation of its reliability. We begin in section 2 with a rederivation of the main results given in the original SC paper and point out some errors in those formulas. We have examined in detail three software implementations of the SC method that are in common use: (1) the original code written by SC to accompany their paper [*Silver and Chan*, 1991], (2) the Mfast program which combines the SC method and the *Teanby et al.* [2004b] clustering method and applies multiple filters to events [*Savage et al.*, 2010; *Teanby et al.*, 2004b; *Wessel*, 2010], and (3) the Splitlab program [*Wustefeld et al.*, 2008]. We comment on differences among these codes and differences between them and the original SC paper. We also identify some small errors in these software implementations of the SC method.

[17] In section 3, we test key assumptions using observations of a sequence of very similar local/shallow earthquakes all observed at a single seismometer station. We treat this set of earthquakes as having uniform properties and being affected by the same anisotropic layer. In this way, we can

assess the statistical variation introduced by observational noise separately from any systematic differences that exist between stations and between earthquakes. The SC method was originally tested on SKS (teleseismic) waves; however, we have been unable to locate a sufficiently large set of identical teleseisms.

## 2. The Silver and Chan Method

[18] In this section, we rederive the splitting parameter estimation procedure presented by *Silver and Chan* [1991], with certain modifications and generalizations. We also point out certain formulas that are incorrect in that paper. One notable difference between their treatment and ours is that we use a fully three-dimensional specification [cf. *Sileny and Plomerova*, 1996], whereas SC assumed a (nearly) vertically incident ray, which restricted their development to displacement in the horizontal plane.

### 2.1. Coordinate Systems

[19] We first construct a set of six Cartesian coordinate systems (Figure 1 and Table 1) and show how to transform between them.

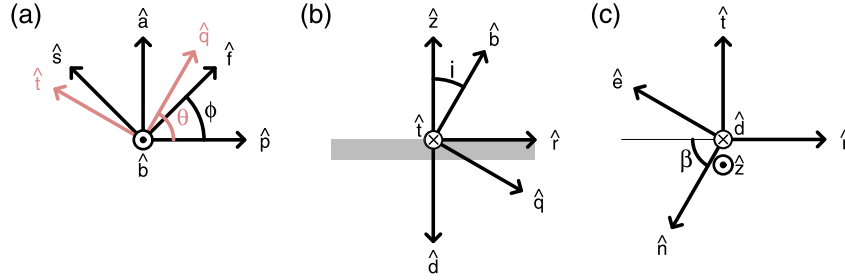
[20] The unit vectors  $\{\hat{\mathbf{a}}_1, \hat{\mathbf{a}}_2, \hat{\mathbf{a}}_3\}$  along the axes of Cartesian coordinate system  $\mathbb{A}$  are, when expressed in coordinate system  $\mathbb{B}$ , the columns of an orthogonal matrix  $\mathbf{U}_{\mathbb{B}\mathbb{A}} = [\hat{\mathbf{a}}_1 \ \hat{\mathbf{a}}_2 \ \hat{\mathbf{a}}_3]$ . An arbitrary vector  $\mathbf{x}$  can be transformed from its representation in  $\mathbb{A}$  to its representation in  $\mathbb{B}$  by  $\mathbf{x}_{\mathbb{B}} = \mathbf{U}_{\mathbb{B}\mathbb{A}} \mathbf{x}_{\mathbb{A}}$ , and we also note that  $\mathbf{U}_{\mathbb{A}\mathbb{B}} = \mathbf{U}_{\mathbb{B}\mathbb{A}}^T$  and  $\mathbf{U}_{\mathbb{A}\mathbb{A}} = \mathbf{I}$ . Moreover, the unit vector  $\hat{\mathbf{a}}_i$  along the  $i$ th axis of  $\mathbb{A}$  can be expressed in coordinate system  $\mathbb{B}$  as  $\hat{\mathbf{a}}_i = \mathbf{U}_{\mathbb{B}\mathbb{A}} \hat{\mathbf{e}}_i$ , where  $\hat{\mathbf{e}}_i$  is the  $i$ th column of the identity matrix.

[21] Consider a transverse wave propagating in a direction  $\hat{\mathbf{b}}$  through an isotropic volume of the Earth with displacement along the polarization direction  $\hat{\mathbf{p}}$ , which is orthogonal to  $\hat{\mathbf{b}}$ . The propagation and polarization vectors, together with the null vector  $\hat{\mathbf{a}} = \hat{\mathbf{b}} \times \hat{\mathbf{p}}$  form the propagation coordinate system  $\mathbb{P}$  with transformation matrix  $\mathbf{U}_{\mathbb{A}\mathbb{P}} = [\hat{\mathbf{p}} \ \hat{\mathbf{a}} \ \hat{\mathbf{b}}]_{\mathbb{A}}$  in arbitrary coordinate system  $\mathbb{A}$ .

[22] The wave enters a birefringent medium with mutually orthogonal fast polarization  $\hat{\mathbf{f}}$  and slow direction  $\hat{\mathbf{s}}$ , both orthogonal to the wave propagation vector  $\hat{\mathbf{b}}$ . The vector triple  $(\hat{\mathbf{f}}, \hat{\mathbf{s}}, \hat{\mathbf{b}})$  determines Splitting coordinates  $\mathbb{S}$  with transformation matrix  $\mathbf{U}_{\mathbb{P}\mathbb{S}} = [\hat{\mathbf{f}} \ \hat{\mathbf{s}} \ \hat{\mathbf{b}}]_{\mathbb{P}} = \mathbf{E}_3(\phi)$ , where the *fast polarization*  $\phi$  is the angle between  $\hat{\mathbf{f}}$  and  $\hat{\mathbf{p}}$  and  $\mathbf{E}_3(\phi)$  is the standard Euler matrix that rotates a vector by an angle  $\phi$  anticlockwise about the  $z$  axis (Appendix A).

[23] Where the ray meets the Earth's surface, we construct emerging ray coordinates  $\mathbb{E}$ , defined by the transformation matrix  $\mathbf{U}_{\mathbb{P}\mathbb{E}} = [\hat{\mathbf{q}} \ \hat{\mathbf{t}} \ \hat{\mathbf{b}}]_{\mathbb{P}} = \mathbf{E}_3(\theta)$ . Here  $\hat{\mathbf{t}}$  is the *transverse direction*, which is the unit vector in the horizontal plane which is orthogonal to the propagation direction  $\hat{\mathbf{b}}$ . The orientation of  $\hat{\mathbf{t}}$  is chosen so that the vector  $\hat{\mathbf{q}} = \hat{\mathbf{t}} \times \hat{\mathbf{b}}$  plunges downward. The *SV* (shear in the vertical plane) angle  $\theta$  is the angle between  $\hat{\mathbf{p}}$  and  $\hat{\mathbf{q}}$ .

[24] Ray planar coordinates  $\mathbb{R}$  resolve the ray's path into horizontal and vertical components, with transformation matrix  $\mathbf{U}_{\mathbb{E}\mathbb{R}} = [\hat{\mathbf{r}} \ \hat{\mathbf{t}} \ \hat{\mathbf{z}}]_{\mathbb{E}} = \mathbf{E}_2(i)$  from emerging ray coordinates. This coordinate system is defined by the *inclination angle*  $i$  of the ray  $\hat{\mathbf{b}}$  with respect to the vertical unit vector  $\hat{\mathbf{z}}$ . (Here  $\mathbf{E}_2(i)$  is the Euler matrix rotating by angle  $i$  around the



**Figure 1.** Coordinate systems and angle definitions. (a) The transverse plane: Propagation coordinates  $\mathbb{P}[\hat{\mathbf{p}}\hat{\mathbf{a}}\hat{\mathbf{b}}]$ , splitting coordinates  $\mathbb{S}[\hat{\mathbf{f}}\hat{\mathbf{s}}\hat{\mathbf{b}}]$ , and the emerging ray coordinates  $\mathbb{E}[\hat{\mathbf{q}}\hat{\mathbf{t}}\hat{\mathbf{b}}]$ , defining the fast polarization  $\phi$  and SV (shear in the vertical plane) direction  $\theta$ ; (b) The vertical plane containing the ray: Emerging ray coordinates and ray planar coordinates  $\mathbb{R}[\hat{\mathbf{r}}\hat{\mathbf{t}}\hat{\mathbf{z}}]$ , defining the inclination angle  $i$ . The radial  $\hat{\mathbf{r}}$  and transverse  $\hat{\mathbf{t}}$  vectors lie in the Earth's surface (shaded); (c) The horizontal plane from above: Ray planar coordinates, geographic coordinates  $\mathbb{G}[\hat{\mathbf{e}}\hat{\mathbf{n}}\hat{\mathbf{z}}]$ , and bearing coordinates  $\mathbb{B}[\hat{\mathbf{n}}\hat{\mathbf{e}}\hat{\mathbf{d}}]$ , defining the back azimuth  $\beta$ .

$y$  axis, see Appendix A.) The *radial direction*  $\hat{\mathbf{r}}$  is the unit vector along the horizontal component of the ray.

[25] Geographic coordinates  $\mathbb{G}$  rotate the transverse and radial components into the standard East  $\hat{\mathbf{e}}$  and North  $\hat{\mathbf{n}}$  directions, with  $\mathbf{U}_{\mathbb{G}\mathbb{R}} = [\hat{\mathbf{e}} \ \hat{\mathbf{n}} \ \hat{\mathbf{z}}]_{\mathbb{G}} = \mathbf{E}_3(\beta - 3\pi/2)$ . The *back azimuth*  $\beta$  is the angle between North  $\hat{\mathbf{n}}$  and  $-\hat{\mathbf{r}}$ , which is the bearing from which the ray arrives. Bearing coordinates  $\mathbb{B}$  are the standard alternative to geographic coordinates with unit vectors along North, East, and Down ( $\hat{\mathbf{d}} = -\hat{\mathbf{z}}$ ), and  $\mathbf{U}_{\mathbb{B}\mathbb{G}} = [\hat{\mathbf{n}} \ \hat{\mathbf{e}} \ \hat{\mathbf{d}}]_{\mathbb{B}} = \mathbf{E}_2(\pi)\mathbf{E}_3(\pi/2)$ .

## 2.2. Shear Wave Splitting

[26] We assume that for all frequencies  $\omega$ , the displacement  $\mathbf{u}(t, \mathbf{x}|\omega)$  caused by the wave at a distance  $x$  from the source and at time  $t$  is parallel to a single polarization vector  $\hat{\mathbf{p}}$  so that

$$\mathbf{u}(t, x|\omega, k) = A(\omega)e^{i(kx - \omega t)}\hat{\mathbf{p}}, \quad (1)$$

where  $k(\omega)$  is the wave number. The total displacement is

$$\mathbf{u}(t, x) = \int_0^\infty \mathbf{u}(t, x|\omega, k(\omega)) d\omega. \quad (2)$$

[27] In the following development, we suppress the  $x$  dependence and assume we are observing the displacement at a single seismometer location at the Earth's surface. We therefore write

$$\mathbf{u}(t, \omega) = w(\omega)e^{-i\omega t}\hat{\mathbf{p}}, \quad (3)$$

with  $w(\omega) = A(\omega)e^{ik(\omega)x}$  (Compare SC (2)), and

$$\mathbf{u}(t) = \int_0^\infty w(\omega)e^{-i\omega t} d\omega \hat{\mathbf{p}} = u(t)\hat{\mathbf{p}}. \quad (4)$$

[28] By the time the wave reaches the surface, the component of the wave that displaces parallel to the fast polarization  $\hat{\mathbf{f}}$  arrives at a time  $\delta t$  before the component parallel to the slow direction  $\hat{\mathbf{s}}$ . The displacement of the split wave can be constructed using the splitting operator

$$\mathbf{\Gamma}(\phi, \delta t; \omega) = e^{-i\omega \delta t/2} \hat{\mathbf{f}} \hat{\mathbf{f}}^T + e^{+i\omega \delta t/2} \hat{\mathbf{s}} \hat{\mathbf{s}}^T + \hat{\mathbf{b}} \hat{\mathbf{b}}^T = e^{-i\omega \delta \mathbf{T}(\phi, \delta t)}. \quad (5)$$

(Compare SC (5) and (8), and note that our signs are reversed compared to SC, to ensure that the slow wave lags rather

than leads the fast wave.) The compact exponential expression of  $\mathbf{\Gamma}(\phi, \delta t)$  is made possible by defining (Compare SC (7))

$$\delta \mathbf{T}(\phi, \delta t; \omega) = (\delta t/2)(\hat{\mathbf{f}} \hat{\mathbf{f}}^T - \hat{\mathbf{s}} \hat{\mathbf{s}}^T). \quad (6)$$

[29] The dependence of  $\delta \mathbf{T}$  on  $\phi$  is through the matrix  $\mathbf{U}_{\mathbb{S}\mathbb{P}}(\phi)$  that transforms between propagation and splitting coordinates. The splitting operator  $\mathbf{\Gamma}(\phi, \delta t; \omega)$  is manifestly symmetric and unitary, so that  $\mathbf{\Gamma}^{-1}(\phi, \delta t; \omega) = \mathbf{\Gamma}^*(\phi, \delta t; \omega) = \exp(+i\omega \delta \mathbf{T})$ , where  $*$  indicates complex conjugation. In splitting coordinates, the splitting operator is independent of  $\phi$  and can be expressed explicitly as

$$\mathbf{\Gamma}_{\mathbb{S}}(\delta t; \omega) = e^{-i\omega \delta t/2} \hat{\mathbf{e}}_1 \hat{\mathbf{e}}_1^T + e^{+i\omega \delta t/2} \hat{\mathbf{e}}_2 \hat{\mathbf{e}}_2^T + \hat{\mathbf{e}}_3 \hat{\mathbf{e}}_3^T = \mathbf{D}(\delta t; \omega), \quad (7)$$

where  $\mathbf{D}(\delta t; \omega) = \text{diag}(e^{-i\omega \delta t/2}, e^{+i\omega \delta t/2}, 1)$ . Hence, in ray planar coordinates, the splitting operator is

$$\mathbf{\Gamma}_{\mathbb{R}}(\phi, \delta t; \omega) = \mathbf{U}_{\mathbb{R}\mathbb{S}}(\phi) \mathbf{D}(\delta t; \omega) \mathbf{U}_{\mathbb{S}\mathbb{R}}(\phi). \quad (8)$$

[30] The transformation matrix  $\mathbf{U}_{\mathbb{R}\mathbb{S}}(\phi)$  can be constructed by the series of rotation matrices

$$\mathbf{U}_{\mathbb{R}\mathbb{S}}(\phi) = \mathbf{U}_{\mathbb{R}\mathbb{E}} \mathbf{U}_{\mathbb{E}\mathbb{P}} \mathbf{U}_{\mathbb{P}\mathbb{S}} = \mathbf{E}_2^T(i) \mathbf{E}_3(\phi - \theta) \quad (9)$$

with the angles  $(\phi, \theta, i)$  defined in Table 1. We suppress explicit dependence on  $\theta$  and  $i$  since these angles are usually assumed to be known from the geometry of the earthquake

**Table 1.** Coordinate Systems and Transformation Matrices

Coordinate System	Transformation Matrix	Angle
$\mathbb{P}$ Propagation	$\mathbf{U}_{\mathbb{A}\mathbb{P}} = [\hat{\mathbf{p}} \ \hat{\mathbf{a}} \ \hat{\mathbf{b}}]_{\mathbb{A}}$	
$\mathbb{S}$ Splitting	$\mathbf{U}_{\mathbb{P}\mathbb{S}} = [\hat{\mathbf{f}} \ \hat{\mathbf{s}} \ \hat{\mathbf{b}}]_{\mathbb{P}} = \mathbf{E}_3(\phi)$	Fast polarization $\phi$
$\mathbb{E}$ Emerging ray	$\mathbf{U}_{\mathbb{P}\mathbb{E}} = [\hat{\mathbf{q}} \ \hat{\mathbf{t}} \ \hat{\mathbf{b}}]_{\mathbb{P}} = \mathbf{E}_3(\theta)$	SV angle $\theta$
$\mathbb{R}$ Ray planar	$\mathbf{U}_{\mathbb{E}\mathbb{R}} = [\hat{\mathbf{r}} \ \hat{\mathbf{t}} \ \hat{\mathbf{z}}]_{\mathbb{E}} = \mathbf{E}_2(i)$	Inclination angle $i$
$\mathbb{G}$ Geographic	$\mathbf{U}_{\mathbb{G}\mathbb{R}} = [\hat{\mathbf{e}} \ \hat{\mathbf{n}} \ \hat{\mathbf{z}}]_{\mathbb{G}} = \mathbf{E}_3(\beta - 3\pi/2)$	Back azimuth $\beta$
$\mathbb{B}$ Bearing	$\mathbf{U}_{\mathbb{B}\mathbb{G}} = [\hat{\mathbf{n}} \ \hat{\mathbf{e}} \ \hat{\mathbf{d}}]_{\mathbb{B}} = \mathbf{E}_2(\pi)\mathbf{E}_3(\pi/2)$	

and station. In the absence of observational noise, the split wave  $\mathbf{u}_s$  in ray planar coordinates is

$$\begin{aligned}\mathbf{u}_{s\mathbb{R}}(t, \omega; \phi, \delta t) &= \mathbf{\Gamma}_{\mathbb{R}}(\phi, \delta t; \omega) \mathbf{u}_{\mathbb{R}}(t, \omega) \\ &= \mathbf{U}_{\mathbb{R}\mathbb{S}} \mathbf{D}(\delta t; \omega) \mathbf{U}_{\mathbb{S}\mathbb{R}} w(\omega) e^{-i\omega t} \mathbf{U}_{\mathbb{R}\mathbb{P}} \hat{\mathbf{e}}_1 \\ &= \mathbf{U}_{\mathbb{R}\mathbb{S}} \mathbf{D}(\delta t; \omega) \mathbf{U}_{\mathbb{S}\mathbb{P}} \hat{\mathbf{e}}_1 w(\omega) e^{-i\omega t} \\ &= \mathbf{E}_2^T(i) \mathbf{E}_3(\phi - \theta) [W_+(\delta t; t, \omega) \cos \phi \hat{\mathbf{e}}_1 \\ &\quad - W_-(\delta t; t, \omega) \sin \phi \hat{\mathbf{e}}_2],\end{aligned}$$

where

$$W_{\pm}(\delta t; t, \omega) = w(\omega) e^{-i\omega(t \pm \delta t/2)}. \quad (10)$$

[31] The radial and transverse components of the split wave are the first and second components of  $\mathbf{u}_s(t, \omega; \phi, \delta t)$  in ray planar coordinates. In the special case of a vertical ray ( $i = 0$ ), the SV angle  $\theta$  is arbitrary and can be set to zero, in which case the radial and transverse components are the same as given by SC

$$\begin{aligned}u_{sr}(t, \omega; \phi, \delta t) &= W_+(t, \omega, \delta t) \cos^2 \phi + W_-(t, \omega, \delta t) \sin^2 \phi \\ u_{st}(t, \omega; \phi, \delta t) &= [W_+(t, \omega, \delta t) - W_-(t, \omega, \delta t)] \sin \phi \cos \phi.\end{aligned} \quad (11)$$

(Compare SC (14) and (15), although the sign of the transverse component is reversed compared to SC.)

[32] The  $i$ th diagonal element of  $D(\delta t, \omega)$  acts as a time shift operator on the  $i$ th component of the wave in splitting coordinates, so that

$$D_{ii}(\delta t; \omega) u_{\mathbb{S}i}(t, \omega) = u_{\mathbb{S}i}(t + \Delta_i, \omega), \quad (12)$$

where

$$\Delta_i(\delta t) = \begin{cases} +\delta t/2 & i = 1 \\ -\delta t/2 & i = 2 \\ 0 & i = 3 \end{cases}. \quad (13)$$

[33] The time shifts introduced by  $D$  are independent of  $\omega$ , so the net effect of the splitting on the total displacement in ray planar coordinates is

$$\begin{aligned}\mathbf{u}_{s\mathbb{R}}(t; \phi, \delta t) &= \int_0^\infty \mathbf{u}_{s\mathbb{R}}(t, \omega; \delta t) d\omega \\ &= \mathbf{U}_{\mathbb{R}\mathbb{S}}(\phi) \int_0^\infty \mathbf{u}_{\mathbb{S}}(t + \mathbf{\Delta}(\delta t), \omega) d\omega \\ &= \mathbf{U}_{\mathbb{R}\mathbb{S}}(\phi) \mathbf{u}_{\mathbb{S}}(t + \mathbf{\Delta}(\delta t)),\end{aligned} \quad (14)$$

where we have used the shorthand notation  $\mathbf{u}_{\mathbb{S}}(t + \mathbf{\Delta}, \omega) = (\mathbf{u}_{\mathbb{S}1}(t + \Delta_1, \omega), \mathbf{u}_{\mathbb{S}2}(t + \Delta_2, \omega), \mathbf{u}_{\mathbb{S}3}(t + \Delta_3, \omega))^T$ .

### 2.3. Splitting Parameter Estimation

[34] The observed particle motion at the seismometer is a combination of the split wave and observational noise  $\varepsilon(t)$ . In ray planar coordinates, the particle motion is

$$\begin{aligned}\mathbf{u}_{\text{obs}\mathbb{R}}(t) &= \mathbf{u}_{s\mathbb{R}}(t; \phi, \delta t) + \varepsilon(t) \\ &= \mathbf{U}_{\mathbb{R}\mathbb{S}}(\phi) \mathbf{u}_{\mathbb{S}}[t + \mathbf{\Delta}(\delta t)] + \varepsilon(t).\end{aligned}$$

The observed displacement vector  $\mathbf{u}_{\text{obs}}(t)$  is recorded in geographic or bearing coordinates. Given the back azimuth  $\beta$ , it can be converted to its representation in ray planar coordinates  $\mathbf{u}_{\text{obs}\mathbb{R}}(t) = \mathbf{U}_{\mathbb{R}\mathbb{G}}(\beta) \mathbf{u}_{\text{obs}\mathbb{G}}(t)$  or splitting coordinates  $\mathbf{u}_{\text{obs}\mathbb{S}}(t) = \mathbf{U}_{\mathbb{S}\mathbb{R}}(\phi, \theta, i) \mathbf{U}_{\mathbb{R}\mathbb{G}}(\beta) \mathbf{u}_{\text{obs}\mathbb{G}}(t)$ . The goal of a shear wave splitting analysis is to take the observed particle motion and find an optimal pair  $(\hat{\phi}, \hat{\delta t})$  that returns  $\mathbf{u}_{\text{obs}}(t)$  to

a rectified wave  $\tilde{\mathbf{u}}(t; \hat{\phi}, \hat{\delta t})$  that is as close as possible to linear motion along the polarization vector  $\hat{\mathbf{p}}$ :

$$\tilde{\mathbf{u}}(t; \hat{\phi}, \hat{\delta t}) = \mathbf{U}_{\mathbb{R}\mathbb{S}}^T(\hat{\phi}) \mathbf{u}_{\text{obs}\mathbb{S}}(t - \mathbf{\Delta}(\hat{\delta t})). \quad (15)$$

[35] In the most general formulation of the problem, the three angles ( $\theta$ ,  $i$ , and  $\beta$ ) may need to be estimated as well.

[36] The variance-covariance matrix of the displacement of the unsplit noise-free wave is

$$\text{Cov}[\mathbf{u}(t)] \propto \hat{\mathbf{p}} \hat{\mathbf{p}}^T, \quad (16)$$

which is a rank 1 matrix having one nonzero eigenvalue with corresponding eigenvector  $\hat{\mathbf{p}}$ . The best rectification of the split wave occurs when the variance-covariance matrix has a dominant eigenvalue  $\lambda_1$  that is as large as possible:

$$(\hat{\phi}, \hat{\delta t}) = \arg \max_{\phi, \delta t} \lambda_1 \{ \text{Cov}[\tilde{\mathbf{u}}(t; \phi, \delta t)] \}, \quad (17)$$

while at the same time having the other two eigenvalues ( $\lambda_2, \lambda_3$ ) both as close to zero as possible.

[37] In splitting coordinates, the rectified wave is

$$\tilde{\mathbf{u}}_{\mathbb{S}}(t; \hat{\delta t}) = \mathbf{u}_{\text{obs}\mathbb{S}}(t - \mathbf{\Delta}(\hat{\delta t})), \quad (18)$$

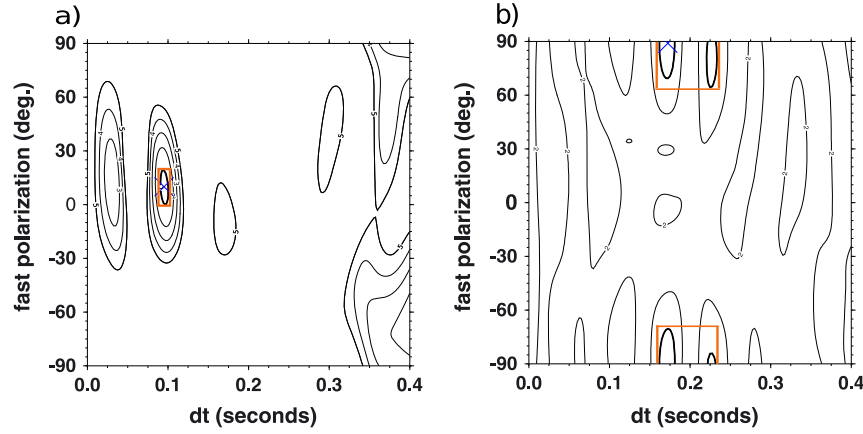
and the variance-covariance matrix of the rectified wave  $\tilde{\mathbf{u}}_{\mathbb{S}}(t; \hat{\delta t})$  has entries (Compare SC (9))

$$\begin{aligned}C_{\mathbb{S}}(\phi, \delta t)_{ij} &= \text{Cov}[u_{\text{obs}\mathbb{S}i}(t - \Delta_i), u_{\text{obs}\mathbb{S}j}(t - \Delta_j)] \\ &= \int_{-\infty}^{\infty} u_{\text{obs}\mathbb{S}i}(t - \Delta_i) u_{\text{obs}\mathbb{S}j}(t - \Delta_j) dt \\ &\quad - \int_{-\infty}^{\infty} u_{\text{obs}\mathbb{S}i}(t - \Delta_i) dt \int_{-\infty}^{\infty} u_{\text{obs}\mathbb{S}j}(t' - \Delta_j) dt' \\ &\text{for } i, j = 1, 2, 3.\end{aligned} \quad (19)$$

This expression differs from SC (9) in three important respects. First, it is a full three-dimensional expression, allowing for nonvertical rays. Second, it is a symmetric variance-covariance matrix, rather than the asymmetric cross covariance defined in SC (9) [Brockwell and Davis, 2006]. However, we believe their intention was in fact to use the covariance matrix because their original computer code implements the covariance matrix rather than cross covariance, and so do the Teanby codes, Mfast, and Splitlab. Third, SC (9) omits the second term in our equation (19): the product of expectations. This is appropriate if the wave train has had its mean subtracted, when this term is zero. However, in software implementations, finite segments of waves are shifted by varying values of  $\delta t$  across analysis subwindows of fixed length before  $C_{\mathbb{S}}$  is calculated, and it cannot be guaranteed that the mean is zero within the analysis window for all shifts. We have found, for example, that Mfast versions (which are based on the Teanby codes) prior to 1.5 did not subtract off the product of expectations. As a result of our work, this has been rectified in Mfast version 1.5, and after. This change has led to minor changes in the estimated splitting parameters for lower quality (weak signal/high noise) observations.

[38]  $\mathbf{C}$  can be converted to its representation in ray planar coordinates  $\mathbb{R}$  by

$$\mathbf{C}_{\mathbb{R}}(\phi, \delta t) = \mathbf{U}_{\mathbb{R}\mathbb{S}}(\phi) \mathbf{C}_{\mathbb{S}}(\phi, \delta t) \mathbf{U}_{\mathbb{S}\mathbb{R}}(\phi). \quad (20)$$



**Figure 2.** Illustration of two different types of confidence regions produced by *Teanby et al.* [2004b] codes used in Mfast for two different events recorded at the BOR station on Reunion Island. The minimum eigenvalue is denoted by a cross, and the 95% confidence region by the bolded lines. The orange rectangle denotes the smallest box bounding all of the 95% confidence region(s). Assuming normal distributions, the corresponding standard errors of  $\phi$  and  $\delta t$  are a quarter of the width and length of this box, respectively. This box represents a visual aid showing how the standard errors are determined by SC. (a) Single confidence region and (b) two multiple confidence regions, which could be mistaken for four if the  $180^\circ$  ambiguity is not taken into account. In practice, results such as Figure 2b would be given a lower quality grade from the Mfast method and may not be relied upon to make inferences.

This transformed matrix has the same eigenvalues as  $C_{\mathbb{R}}$ , and its dominant eigenvector is the polarization vector  $\hat{\mathbf{p}}$  expressed in ray planar coordinates. The trace of  $\mathbf{C}$  is independent of  $\phi$  and  $\delta t$ :

$$\begin{aligned}
 \text{tr}[\mathbf{C}_{\mathbb{S}}(\phi, \delta t)] &= \text{tr}\{\text{Cov}[\mathbf{D}^*(\delta t)\mathbf{u}_{\text{obsS}}(t)]\} \\
 &= \sum_{i=1}^3 \text{Var}[u_{\text{obsS}i}(t - \Delta_i)] \\
 &= \sum_{i=1}^3 \text{Var}[u_{\text{obsS}i}(t)] \\
 &= \text{tr}\{\text{Cov}[\mathbf{u}_{\text{obsS}}(t)]\} \\
 &= \text{tr}\{\text{Cov}[\mathbf{U}_{\mathbb{S}\mathbb{R}}(\phi)\mathbf{u}_{\text{obsR}}(t)]\} \\
 &= \text{tr}\{\mathbf{U}_{\mathbb{S}\mathbb{R}}(\phi)\text{Cov}[\mathbf{u}_{\text{obsR}}(t)]\mathbf{U}_{\mathbb{R}\mathbb{S}}(\phi)\} \\
 &= \text{tr}\{\text{Cov}[\mathbf{u}_{\text{obsR}}(t)]\}.
 \end{aligned}$$

This trace is the sum of the eigenvalues of  $\mathbf{C}$ . In two dimensions, maximizing the dominant eigenvalue  $\lambda_1$  is therefore equivalent to minimizing the only other eigenvalue  $\lambda_2$ , and this is what is done in software implementations of the SC method (Teanby codes, Mfast, and Splitlab). We note that other equivalent functions for the 2-D case mentioned in *Silver and Chan* [1991] are available in Splitlab (maximization of  $\lambda_1$ , minimization of  $\lambda_1 \times \lambda_2$ , and maximization of  $\lambda_1/\lambda_2$ ). In three dimensions, both of the nondominant eigenvalues must be minimized, or more simply, the dominant eigenvalue should be maximized. For a noisy signal, the values of the nondominant eigenvalues for the optimal solution quantify the noise power: they are the variances of the displacement along the propagation (**b**) and null (**a**) directions. The dominant eigenvalue combines power from both signal and noise.

## 2.4. Confidence Region

[39] From here on, following SC, we restrict ourselves to the two-dimensional problem with a near-vertical seismic ray. Once the pair  $(\hat{\phi}, \hat{\delta t})$  that lead to the smallest second eigenvalue have been identified, error estimates can be calculated. *Silver and Chan* [1991] do this by calculating a quarter of the width and length of a confidence region derived from an assumption that the signal is combined with Gaussian white noise.

[40] If the rectification of the wave has been successful, then the minimum value of the nondominant eigenvalue  $\lambda_2^{\min} = \lambda_2(\hat{\phi}, \hat{\delta t})$  is the variance of displacement on the rectified null component, i.e., the variance of the noise, and is a draw from a chi-squared random variable  $\chi_v^2$  scaled by the noise variance. The value of the degrees of freedom  $v$  is discussed further below. In the absence of a signal, then at any  $(\phi, \delta t)$  pair, the ratio

$$R(\phi, \delta t) = \frac{(\lambda_2(\phi, \delta t) - \lambda_2^{\min})/k}{\lambda_2^{\min}/(v - k)} \sim F_{k, v-k} \quad (21)$$

is a draw from an  $F_{k, v-k}$  distribution, where  $k = 2$ , the number of estimated parameters  $(\phi, \delta t)$ . This ratio can be used to define a  $100(1 - \alpha)\%$  confidence region for  $(\phi, \delta t)$  as the set

$$\{(\phi, \delta t) : R(\phi, \delta t) \leq F_{k, v-k}^{1-\alpha}\}, \quad (22)$$

where  $F_{k, v-k}^{1-\alpha}$  marks the beginning of the  $100\alpha\%$  upper tail of the  $F_{k, v-k}$  distribution.

[41] Figure 2 shows examples of confidence regions for two different events. For each pair of  $\phi$  and  $\delta t$  on a grid, the second eigenvalue  $\lambda_2(\phi, \delta t)$  of the covariance matrix is determined and its values are drawn on a contour map. The smallest  $\lambda_2$  value is marked by a cross. The 95% confidence region, defined by  $\lambda_2^{0.95}(\phi, \delta t) = \lambda_2^{\min}\{1 + [k/(v - k)]F_{k, v-k}^{0.05}\}$ ,

is indicated by a thick black line. This critical value is often used to normalize the contour plot so that the contour for the value 1 corresponds to the boundary of the 95% confidence region and any number less than 1 is inside the confidence region.

[42] If we assume normally distributed parameter errors, then we can obtain approximate standard errors for  $\phi$  and  $\delta t$  as a quarter of the length and width of the ellipse enclosing the minimum. Determining the length and width for a contiguous region like Figure 2a is simple, but for more complex regions like the one in Figure 2b, assigning a sensible value for the standard error is not straightforward. The fast direction  $\phi$  has a  $180^\circ$  ambiguity, so 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 code when estimating the fast polarization error, and find the shortest interval in  $\phi$  that contains all points on the boundary of the confidence region.

## 2.5. Degrees of Freedom

[43] In order to construct the confidence region, the degrees of freedom  $\nu$  must be calculated. We have rederived expressions to estimate  $\nu$  following the approach laid out by *Silver and Chan* [1991], but find equations that differ slightly from theirs.

[44] The smaller eigenvalue  $\lambda_2$  of the covariance matrix  $\mathbf{C}$  above is equal to the power of the rectified wave in the null direction  $\hat{\mathbf{a}}$ . We assume that this rectified displacement  $\eta(t) = \tilde{u}_{S2}(t)$  is the convolution of a Gaussian white noise process  $g(t; \sigma)$ , with noise amplitude  $\sigma$ , and a smoothing filter  $f(t; h)$ , with smoothing width  $h$ . By Parseval's theorem and the convolution theorem, the power in a sample of  $n + 1$  observations of  $\eta(t)$  over time interval  $[0, T]$  at spacing  $\Delta t = T/n$  may be approximated by

$$E = \sum_{i=0}^n a_i |\tilde{f}_i|^2 |\tilde{g}_i|^2$$

with  $a_0 = a_n = 1/2$ , and  $a_i = 1$  for  $i = 2, \dots, n-1$ ,

where  $\tilde{f}_i = \tilde{f}(\omega_i; h)$ ,  $\tilde{g}_i = \tilde{g}(\omega_i; \sigma)$  are the discrete Fourier transforms of  $f(t; h)$  and  $g(t; \sigma)$  evaluated at frequencies  $\omega_i$  spaced at  $2\pi/(n\Delta t) = 2\pi/T$  and  $\tilde{f}_n = \tilde{f}_0$ ,  $\tilde{g}_n = \tilde{g}_0$ .

[45] The first and second moments of  $E$ , averaging over the Gaussian noise process, (which is also Gaussian in the frequency domain) are

$$\langle E \rangle = \sigma^2 \sum_{i=0}^n a_i |\tilde{f}_i|^2 \equiv \sigma^2 F_2 \quad (23)$$

$$\langle E^2 \rangle = 2\sigma^4 \sum_{i=0}^n a_i^2 |\tilde{f}_i|^4 + \langle E \rangle^2 \equiv \sigma^4 (F_4 + F_2^2) . \quad (24)$$

These formulas define the scale-free filter moments  $F_2$  and  $F_4$ , and we have used the fact that for Gaussian noise  $\langle \tilde{g}_i^2 \rangle = \sigma^2$  and  $\langle \tilde{g}_i^4 \rangle = 3\sigma^4$  [Vaseghi, 2008]. We note that equation SC (A8) incorrectly omits the factor of 2 in the definition of  $F_4$ .

[46] For the degrees of freedom calculation below, we need to construct estimators of the filter moments  $F_2$ ,  $F_2^2$ ,

and  $F_4$  from the observed data. We define the frequency domain sample moments

$$E_2 = \sum_{i=0}^n b_i |\tilde{\eta}_i|^2 = \sum_{i=0}^n b_i |\tilde{f}_i|^2 |\tilde{g}_i|^2 \quad (25)$$

$$E_4 = \sum_{i=0}^n c_i |\tilde{\eta}_i|^4 = \sum_{i=0}^n c_i |\tilde{f}_i|^4 |\tilde{g}_i|^4 , \quad (26)$$

with coefficients  $\{b_i\}$  and  $\{c_i\}$  to be determined such that

$$\langle E_2 \rangle = \sigma^2 F_2 = \sigma^2 \sum_{i=0}^n a_i |\tilde{f}_i|^2$$

$$\langle E_4 \rangle = 2\sigma^4 F_4 = 4\sigma^4 \sum_{i=0}^n a_i^2 |\tilde{f}_i|^4 .$$

Averaging over the Gaussian noise process as before, it follows that  $b_i = a_i$  and  $c_i = 4a_i^2/3$  for all  $i$ . This means that  $E_2 = E$  and hence that  $\langle E_2^2 \rangle = \langle E^2 \rangle = \sigma^4 (F_4 + F_2^2)$ . These results also imply that

$$E_4 = \sum_{i=1}^{n-1} \frac{4}{3} |\tilde{f}_i|^4 |\tilde{g}_i|^4 + \frac{1}{3} (|\tilde{f}_0|^4 |\tilde{g}_0|^4 + |\tilde{f}_n|^4 |\tilde{g}_n|^4) , \quad (27)$$

which we have written out explicitly to show that it differs from the equivalent expression given by SC (A10):

$$E_4^{(\text{SC})} = \sum_{i=1}^{n-1} |\tilde{f}_i|^4 |\tilde{g}_i|^4 + \frac{1}{3} (|\tilde{f}_0|^4 |\tilde{g}_0|^4 + |\tilde{f}_n|^4 |\tilde{g}_n|^4) . \quad (28)$$

[47] For large numbers of data points in the analysis window (large  $n$ ), the incorrect  $E_4$  given by SC is dominated by the sum from  $i = 2$  to  $n-1$  and therefore underestimates  $E_4$  by a factor of  $3/4$ .

[48] We assume that the power  $E$  is a scaled chi-square random variable with  $\nu$  degrees of freedom  $E = \kappa \chi_\nu^2$  for some constant  $\kappa$ . This means that  $\langle E \rangle = \kappa \nu$  and  $\text{Var}(E) = 2\kappa^2 \nu$ , and it follows that

$$\nu = \frac{2\langle E \rangle^2}{\text{Var}(E)} = \frac{2F_2^2}{F_4} \quad (29)$$

(Compare SC (A9)) and

$$\kappa = \frac{\text{Var}(E)}{2\langle E \rangle} = \sigma^2 \frac{F_4}{2F_2} . \quad (30)$$

[49] Using the estimators  $E_2$  and  $E_4$  defined above, suitable (first-order unbiased) estimators for  $\nu$  and  $\kappa$  are

$$\hat{\nu} = 2 \left( \frac{2E_2^2}{E_4} - 1 \right) \quad (31)$$

(Compare SC (A12)) and

$$\hat{\kappa} = \frac{E_4}{4E_2} . \quad (32)$$

[50] The underestimation of  $E_4$  in the SC expression (28) means that implementations of the SC method using the incorrect formula (28) will overestimate the number of degrees of freedom by a factor of approximately  $4/3$ .

[51] We checked some of the original SC software code to see which definition of  $E_4$  was actually used. We found that although equation (28) had been implemented in early versions of the code, it had been replaced later by a version



**Table 2.** Differences Between the Formulas

This Article		Silver and Chan [1991] Article	
(5)	$\mathbf{\Gamma}(\phi, \delta t) \equiv \exp(-i\omega\delta t/2)\widehat{\mathbf{f}}\widehat{\mathbf{f}}^T + \exp(i\omega\delta t/2)\widehat{\mathbf{s}}\widehat{\mathbf{s}}^T + \widehat{\mathbf{b}}\widehat{\mathbf{b}}^T$	SC (5)	$\mathbf{\Gamma}(\phi, \delta t) \equiv \exp(i\omega\delta t/2)\widehat{\mathbf{f}}\widehat{\mathbf{f}}^T + \exp(-i\omega\delta t/2)\widehat{\mathbf{s}}\widehat{\mathbf{s}}^T$
(5), (6)	$\mathbf{\Gamma}(\phi, \delta t) = \exp(-i\omega\delta\mathbf{T})$	SC (8)	$\mathbf{\Gamma}(\phi, \delta t) = \exp(i\omega\delta\mathbf{T})$
(11)	$u_{sr}(t, \omega; \phi, \delta t) = [\cos^2 \phi \exp[-i\omega(t + \delta t/2)] + \sin^2 \phi \exp[-i\omega(t - \delta t/2)]]w(\omega)$	SC (14)	$u_{sr}(t, \omega; \phi, \delta t) = \cos^2 \phi w(t + \delta t/2) + \sin^2 \phi w(t - \delta t/2)$
(11)	$u_{sr}(t, \omega; \phi, \delta t) = \frac{1}{2} \sin(2\phi) [\exp[-i\omega(t + \delta t/2)] - \exp[-i\omega(t - \delta t/2)]]w(\omega)$	SC (15)	$u_{sr}(t, \omega; \phi, \delta t) = -\frac{1}{2} \sin(2\phi) [w(t + \delta t/2) - w(t - \delta t/2)]$
(19) <sup>a</sup>	$\mathbf{C} = \begin{bmatrix} \int_{-\infty}^{\infty} u_1^2(t)dt & \int_{-\infty}^{\infty} u_1(t)u_2(t - \delta t)dt \\ \int_{-\infty}^{\infty} u_1(t)u_2(t - \delta t)dt & \int_{-\infty}^{\infty} u_2^2(t)dt \end{bmatrix}$	SC (9)	$\mathbf{C} = \begin{bmatrix} \int_{-\infty}^{\infty} u_1(t)u_1(t - \delta t)dt & \int_{-\infty}^{\infty} u_1(t)u_2(t - \delta t)dt \\ \int_{-\infty}^{\infty} u_2(t)u_1(t - \delta t)dt & \int_{-\infty}^{\infty} u_2(t)u_2(t - \delta t)dt \end{bmatrix}$
(24)	$F_4 = \sum_{n=1}^{N-1} 2 \tilde{f}_n ^4 + \frac{1}{2}( \tilde{f}_0 ^4 +  \tilde{f}_N ^4)$	SC (A8)	$F_4 = \sum_{n=1}^{N-1}  \tilde{f}_n ^4 + \frac{1}{2}( \tilde{f}_0 ^4 +  \tilde{f}_N ^4)$
(27)	$E_4 = \sum_{n=1}^{N-1} \frac{4}{3} \tilde{f}_n ^4 + \frac{1}{3}( \tilde{f}_0 ^4 +  \tilde{f}_N ^4)$	SC (A10)	$E_4 = \sum_{n=1}^{N-1}  \tilde{f}_n ^4 + \frac{1}{3}( \tilde{f}_0 ^4 +  \tilde{f}_N ^4)$

<sup>a</sup>We have written (19) in this form to highlight the differences between our formula and SC (9). SC (9) is equivalent to the 2-D version of (19) assuming that  $u_1(t)$  and  $u_2(t)$  have zero means.

of (28) where the  $(1/3)$  factor for  $i = 0, n$  was replaced by  $(1/2)$ :

$$E_4^{(\text{SC-Code})} = \sum_{i=1}^{n-1} |\tilde{f}_i|^4 |\tilde{g}_i|^4 + \frac{1}{2} (|\tilde{f}_0|^4 |\tilde{g}_0|^4 + |\tilde{f}_n|^4 |\tilde{g}_n|^4). \quad (33)$$

[52] A comment included in the code indicated that better performance was obtained using this version of  $E_4$ , which is also used in the Teanby codes, Mfast versions below 2.0, and Splitlab. All these differences are summarized in Table 2.

### 3. Testing

[53] 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. In the first, we evaluate differences in the degrees of freedom estimate using the three formulas for  $E_4$  in section 2.5. The second test determines whether the integrated power of the noise component has a chi-squared distribution. The final test determines the reliability of the standard errors calculated by *Silver and Chan* [1991] using a set of similar earthquakes observed by a single seismometer.

#### 3.1. Test 1: Degrees of Freedom Estimate

[54] This test compares the degrees of freedom estimates found when using our formula (27) for  $E_4$ , compared to the formula (28) presented by SC, and the modified formula (33), and determines the extent to which  $\hat{\nu}$  from equation (31) is a biased estimate of  $\nu$  from equation (29).

[55] Recall that we have digital data comprising a sample of  $n + 1$  data points over the interval  $[0, T]$  to create a vector  $t_i$  with a sampling frequency of  $n/T$ . We assume the observed noise can be written as a convolution of Gaussian noise  $g$  with variance  $\sigma^2$  and a filter  $f(s; h)$  which we define to be a Gaussian

$$f(s; h) \propto \exp \left[ -\frac{1}{2} \left( \frac{s}{hT} \right)^2 \right], \quad (34)$$

where the parameter  $h$ , the proportion of the observation window  $T$ , controls the smoothing width  $hT$ .

[56] Next, we sample the Fourier transform of this filter  $\{f_i\}$  at frequencies  $\{\omega_i = 2\pi i/T, i = 0, \dots, n\}$  and evaluate the filter moments  $F_2$  and  $F_4$  by the use of (23) and (24). We then calculate the number of degrees of freedom  $\nu$  of the

noise using equation (29) under the approximation that the integrated filtered noise follows a chi-square distribution.

[57] For a chosen value of the smoothing width  $h$ , we generate a realization of Gaussian noise  $\{\tilde{g}_i\}$  at the set of  $n + 1$  frequencies  $\{\omega_i\}$  and multiply them by  $\{\tilde{f}_i\}$  to form a simulated power spectrum  $|\tilde{\eta}_i|^2 = |\tilde{f}_i|^2 |\tilde{g}_i|^2$ . We take  $n = 52$  to match the size of our test data sets used later in this section: our data are sampled at 100 Hz, and the observation window is typically 0.52 s wide, containing 52 observations.

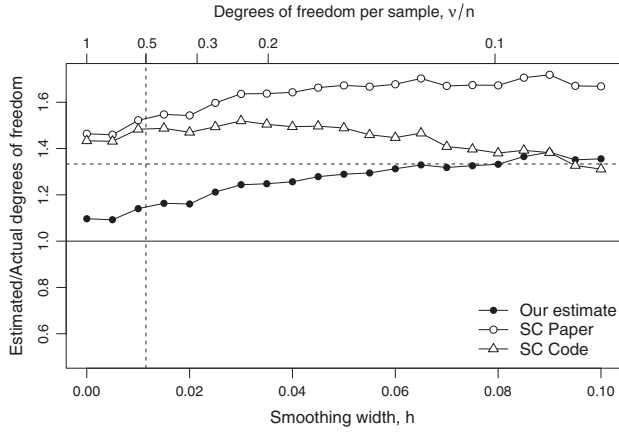
[58] For each realization, we evaluate the estimators  $E_2$  and  $E_4$  using equations (25) and (26), and use them to estimate the degrees of freedom  $\nu$  from equation (31). We calculate  $E_4$  three ways: using our equation (27), the SC formula (28), and the modified SC formula implemented in code equation (33). This gives us three estimates  $\hat{\nu}$ ,  $\hat{\nu}^{(\text{SC})}$ , and  $\hat{\nu}^{(\text{SC-Code})}$ , respectively. We repeat this experiment 1000 times and compute the mean of the degrees of freedom estimates  $\bar{\nu}$ ,  $\bar{\nu}^{(\text{SC})}$  and  $\bar{\nu}^{(\text{SC-Code})}$  across the 1000 replicates. Dividing these means by the degrees of freedom  $\nu$  from equation (29), we obtain a ratio which is unity if the chi-square approximation holds and the estimates are unbiased. The results are shown in Figure 3 where the ratios are plotted against smoothing width  $h$ , and also shown is the number of degrees of freedom per observation  $\nu/n$  corresponding to each smoothing width  $h$ .

[59] All three estimates are positively biased. At the lowest smoothing widths ( $h \simeq 0$ ), our estimate is the closest to  $\nu$ , and as predicted, the two SC estimates are both approximately  $4/3$  larger than  $\nu$  (for very large  $n$ , e.g.,  $n = 1000$ , we find our estimates are exactly unbiased at  $h \simeq 0$ , and the SC estimates exactly  $4/3$  larger). For larger smoothing widths, the bias increases, although the SC code estimate (33) mitigates this increase somewhat.

[60] SC found approximately 0.3 degrees of freedom per observation, and in our test data set, we find a similar value of 0.5: a vertical dashed line marks this value in Figure 3. Our estimates are biased upward by a factor of approximately 1.1 at this degree of smoothing, whereas both SC estimates are biased upward by a factor of approximately 1.5. The SC code estimate (33), introduced for its empirically superior performance to the SC paper estimate, does show slightly lower bias.

[61] There are two conclusions to be made from this test. The first is that for the number of points in typical data





**Figure 3.** Comparison of the mean estimated degrees of freedom for different methods as a function of smoothing widths  $h$  using three different estimation methods. Each point is the mean of 1000 estimates of the degrees of freedom from 1000 simulated data sets of  $n = 52$  observations at the given smoothing width  $h$ , scaled by the “true” degrees of freedom  $\nu$  from equation (29). Filled circles are estimates from our equation (27), open circles are from the SC paper, equation (28), and the open triangles are what is implemented in a version of the SC code, equation (33). If the estimates are unbiased, they would all lie on the horizontal line at 1. The horizontal dashed line shows the expected 4/3 bias of the SC estimate for low smoothing widths, and the vertical dashed line shows the typical 0.5 degrees of freedom per observation seen in our example data sets.

sets ( $n = 52$  here), the degrees of freedom formula (31) is a positively biased estimate of  $\nu$  from equation (29). The second conclusion is that the calculation of the moment  $E_4$  should be carried out using the revised equation (27) rather than the formula (28) presented by SC, or the version

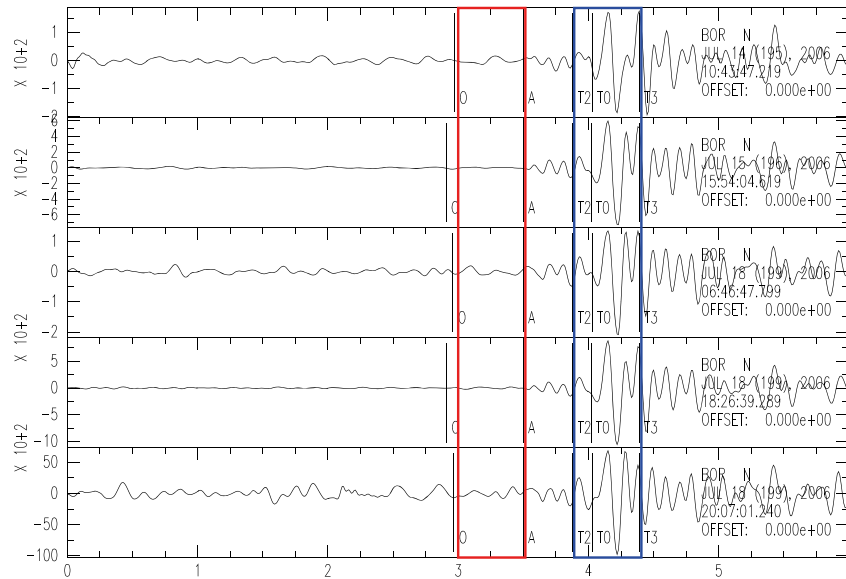
implemented in software (33). A new version of Mfast (v2.0) was released in 2013 incorporating equation (27) as its key change.

### 3.2. Test 2: Gaussian Noise Assumption

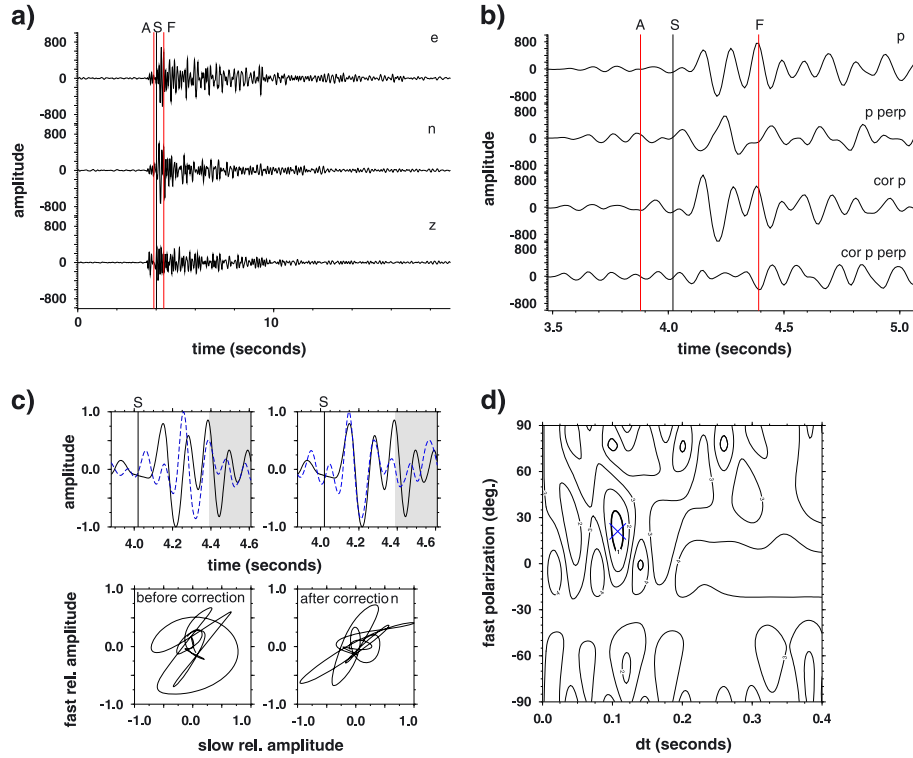
[62] We next test whether the Gaussian noise assumption is correct in a real data set. To test this, we have identified a data set of real observations from a cluster of earthquakes observed on the station BOR (courtesy of Rivemale *et al.* [2010]) on Reunion Island off the East coast of Africa. The events are replicates in the sense that they are all from BOR, they are in close proximity temporally (end of 2004 to mid-2006), all events have similar waveforms and were all sampled at 100 samples per second. Our expectation is that each event has a similar underlying  $S$  waveform, similar fast polarization ( $\phi$ ) and similar delay time ( $\delta t$ ). In that case, we consider the main difference between the observations to be noise, and this provides us with an opportunity to observe the characteristics of that noise.

[63] As this is a real data set, there is a possibility that it violates the assumptions of the methodology. In particular, there could be multiple layers of anisotropy, dipping layers, and heterogeneous anisotropy as well as scattered arrivals interfering with the  $S$  arrivals used. The similarity of the waveforms suggests that if such effects are present, they are having the same effect on all measurements. However, as we show below, it is quite likely that some of the variation we see is caused by non-Gaussian effects such as scattered arrivals coming in from slightly different paths on each waveform. Therefore, the measurement noise that we consider will include both natural noise variation such as wind and wave action as well as unmodeled path effects for waves that take slightly different paths.

[64] Each event records activity for 19 s and has an  $S$  wave arrival at around 4 s. Figure 4 shows an example of the North component for five different events with esti-



**Figure 4.** Example of replicates of the North component from BOR station from 0 to 6 s. The 0 label represents the event origin; A, the  $P$  wave arrival; T2, the window beginning; T3, the window end; and T0, the  $S$  wave arrival. Red box shows the Pre  $P$  noise, blue box shows the window used to calculate the splitting noise after the wave has been desplit.



**Figure 5.** Example of an Mfast output plot for the second waveform in Figure 4. (a) filtered East (e), North (n), and vertical (z) components. Red vertical bars denote the start (A) and end (F) times. Solid black line denotes the  $S$  wave arrival. (b) Waveforms rotated into the incoming polarization direction ( $p$ ) and its perpendicular value ( $p$  perp) for both the original (top two) and corrected (bottom two) waveforms. Solid red and black lines are the same as those in Figure 5a. (c) The fast and slow waveforms (top) and particle motion (bottom) for the original (left) and corrected (right) waveforms. (d) Contour plot of the second eigenvalue of the covariance matrix. Blue cross denotes the best set of parameters. Other output plots are fairly similar and have not been displayed.

mated fast polarization between  $0^\circ$  and  $45^\circ$ . An example of the Mfast output for the second waveform is shown in Figure 5.

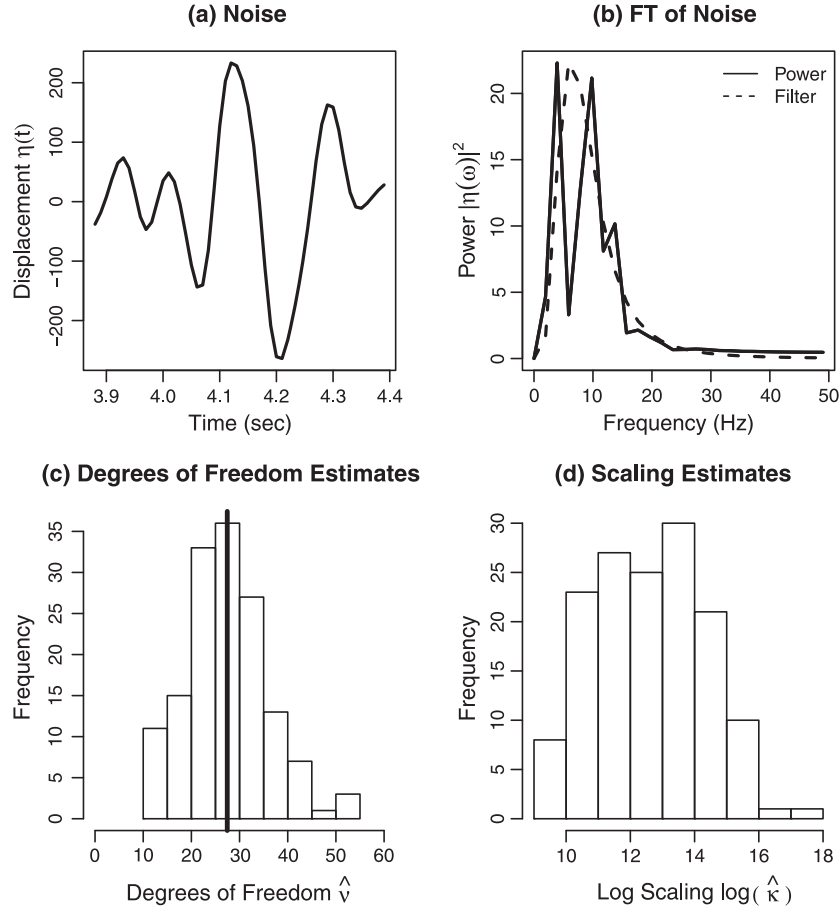
[65] The data set was provided to us by Rivemale *et al.* [2010] and is the largest single family (Family 1) taken from a larger data set that comprises 25 families of earthquakes. To group the earthquakes into families, the cross correlation between vertical components at each frequency (i.e., the coherency) was calculated. Events in each family had a coherency greater than 90%. The initial absolute locations of the earthquakes in Family 1 were from 0.1 to 0.8 km from the station, with depths of 1.2 to 2.5 km beneath the station (elevation 2540 m). Using the velocity model used in the location, this led to angles of incidence at the surface all less than  $33^\circ$ , which should be unaffected by  $S$ - $P$  conversions at the surface [Nuttli, 1961]. Events of the multiplet were selected, cross-correlated, and relocated using cross-spectral analysis and earthquake multiplet double-difference relocation [Got *et al.*, 1994], which led to an average distance of 100 m between events.

[66] The data set provided originally contained 513 events that were recorded between 2004 and 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 [Savage *et al.*, 2010; Teanby *et al.*, 2004b; Wessel, 2010].

[67] Prior to processing, the data are filtered to remove unwanted frequencies that interfere with the signal. A set of filters are applied, and Mfast selects the one that results in the largest signal-to-noise ratio. The standard Mfast algorithm first determines the best filter for each event using the signal-to-noise bandwidth product around the  $S$  arrival. Then the best filter is applied to the data. However, instead of using a different filter for each event, we chose a single filter for all of the events to ensure consistency and comparability of the noise. The filter that resulted in the most high quality results (A and B grades) was a two-pole Butterworth band-pass filter that had corner frequencies of 4 and 10 Hz. A total of 268 high quality events were retained in the data set after filtering.

[68] Examining the fast polarization estimates for these 268 events, we found three groups of observations. Group 1 consisted of events with fast polarizations between  $0^\circ$  and  $45^\circ$  (146 events), Group 2 consisted of events with fast polarizations between  $90^\circ$  and  $120^\circ$  (89 events), and Group 3 contained a further 33 that did not belong to either Group 1 or 2. Only events in Group 1 were retained for analysis.

[69] Finally, a fixed time window to be analyzed during shear wave splitting (the area between  $T2$  and  $T3$  in Figure 4 and A and F in Figure 5) was chosen and used for all of the events, starting 3.88 s after the  $P$  wave arrival and ending at

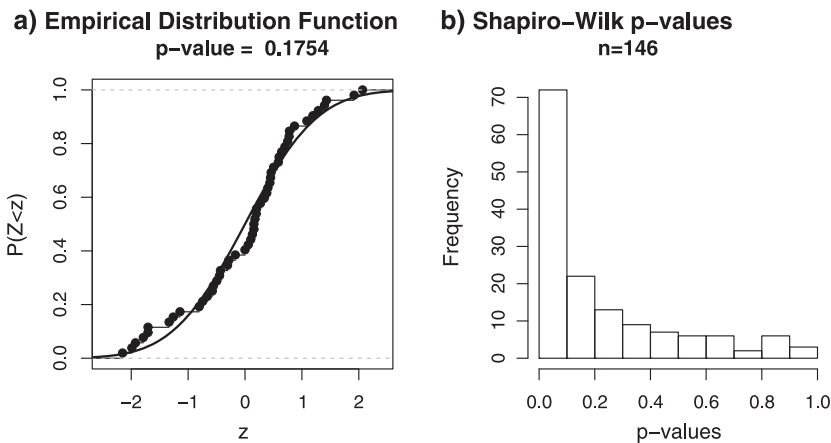


**Figure 6.** An example of the noise component of a high quality event. (a) The normalized, filtered displacement in the time domain, and (b) the power spectrum (up to the Nyquist frequency). The two-pole Butterworth filter used in the analysis is shown as a dashed line in Figure 6b. (c) The set of 146 estimates of the degrees of freedom  $\hat{v}$  and (d) estimates of the log scaling parameter  $\log(\hat{\kappa})$ .

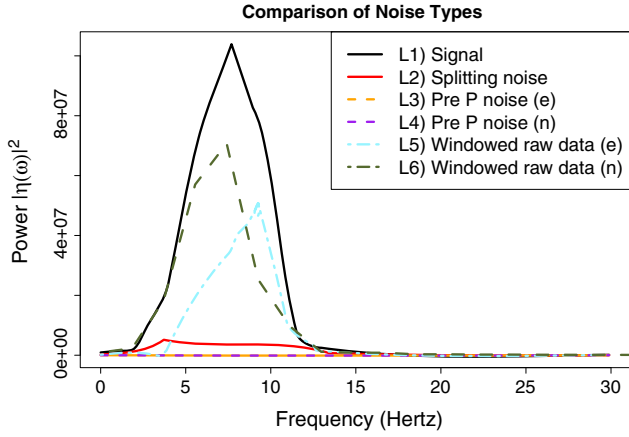
4.39 s. Note that when both the window and filters are fixed, Mfast is equivalent to SC, and so Figures 5 and 2 should be identical to a figure based exactly on the SC method.

[70] After the wave has been rotated and time shifted by the desplitting process, the null component should contain only noise. For each event  $r$ , we have computed the power

spectrum of the noise component  $|\tilde{\eta}_{(r)}(\omega)|^2$  and an estimate of the noise variance  $\hat{\sigma}_r^2$ . An example of the noise component and its scaled power spectrum  $|\tilde{\eta}_{(r)}(\omega)|^2/\sigma_r^2$  is shown in Figures 6a and 6b. Superimposed on the Fourier transform is the two-pole Butterworth filter used to remove the low and high frequencies during data processing. This filter



**Figure 7.** (a) The cumulative distribution function of the noise component of the same high quality event shown in Figure 6. (b) Histogram of  $p$  values from the Shapiro-Wilk test of normality for all 146 events.



**Figure 8.** Average power of  $n = 107$  high quality events for: L1 shows the power of the signal along the polarization direction, L2 shows the power of the noise along the null direction, L3 and L4 show the power of the noise prior to the  $P$  wave arrival (Figure 4), red box) on the East and North components, and L5 and L6 show the power of the windowed East and North components used during the splitting analysis (Figure 4), blue box).

contributes additional correlation to that already present in the observed noise.

[71] For each of the 146 events, we have estimated the degrees of freedom  $\nu$  and the scaling  $\kappa$  using the estimators in equations (31) and (32). Histograms of these estimates are shown in Figures 6c and 6d. The mean number of degrees of freedom is 27.4, which corresponds to 0.53 degrees of freedom per observation. The (squared) noise scaling  $\kappa$  varies strongly between observations due to the large variations in the ambient and signal generated noise between events, and is plotted here on a logarithmic scale to show this large range.

[72] By assumption, the noise component should follow a Gaussian (normal) distribution, and we have tested for this by performing a Shapiro-Wilk test for normality [Shapiro and Wilk, 1965] on each of the 146 data sets. Figure 7a shows the empirical distribution function of the (standardized) noise from the event plotted in Figure 6a, with the standard normal distribution function shown as a solid line. The Shapiro-Wilk test does not reject normality ( $p = 0.1754$ ). A histogram of the  $p$  values of all 146 tests shown in Figure 7b should show a uniform distribution if the noise is normal across all of the events. The histogram shows a clear excess of small  $p$  values, rejecting normality, with 37% of the  $p$  values below the nominal 0.05 value

where 5% of  $p$  values would be expected. This confirms some, although not extreme, departures from normality in the noise.

[73] If we assume that all 146 events have the same number of degrees of freedom  $\nu$ , estimated by the mean 27.4, and differ only in their scaling values  $\kappa_r$ , then we can make a further test. The set of scaled observations  $E_{2(r)}/\sqrt{\kappa_r}$  (for event  $r$ ) should come from the same chi-squared distribution  $\chi^2_\nu$ . However, a Kolmogorov-Smirnov test convincingly rejects this with  $p$  value 0.0042.

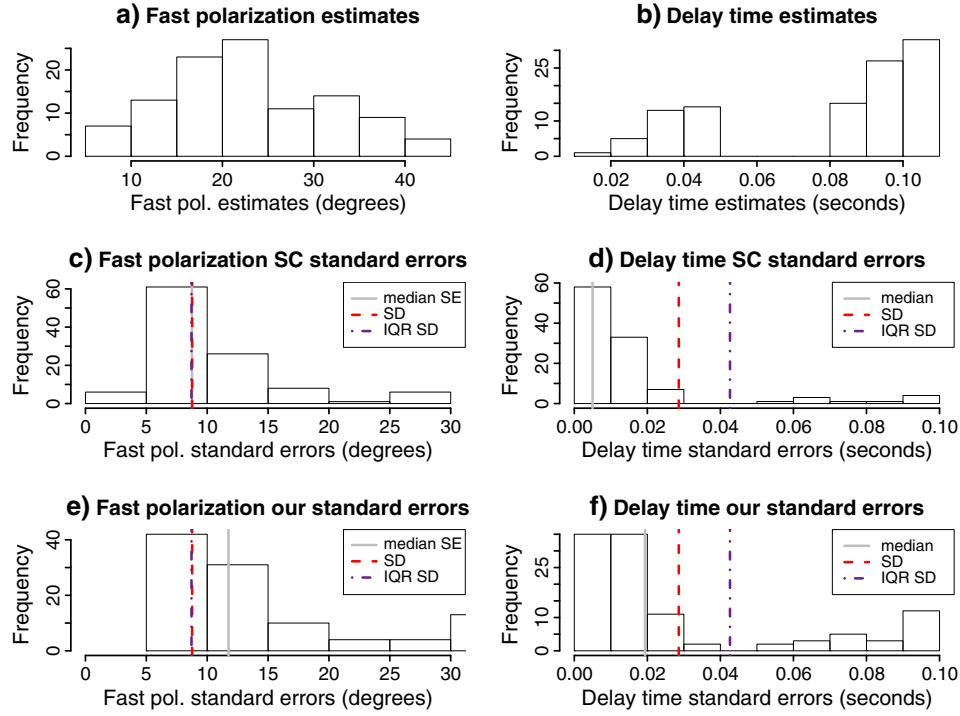
### 3.2.1. Comparison to Background Noise

[74] The Gaussian noise assumption does not always hold so the errors produced by the SC method may not be reliable. Our findings above suggested that there was too much structure in the windowed noise for it to be Gaussian noise. Here we compare the splitting noise from within the measurement window to noise on the East component before the  $P$  wave arrival (background noise) to see if the two have similar power. Our background noise consists of a sample of the data prior to the  $P$  wave (see Figure 4, red box). We compare the windowed noise to this to see if they have a similar frequency content and amplitudes. To do this, we look at the power plots and create summary statistics for all events. Both types of noise are constructed with 27 points of data up to the Nyquist frequency. Figure 8 shows the average power of 107 events (Mfast measurements with A or B grades) for the signal (L1), the splitting noise (L2), background noise from the East and North components (L3 and L4) as well as the power of the windowed East and North components (Figure 4, blue box) used for the splitting analysis (L5 and L6). As expected, after the wave has been desplit, most of the power is contained in the signal along the polarization direction ( $\hat{\mathbf{p}}$ ) with a small amount of splitting noise along the null ( $\hat{\mathbf{a}}$ ) component. Most of the power is concentrated between 4 and 10 Hz as a consequence of our Butterworth filter and as a result of noise properties at the BOR station. The background noise on both the East and North components (Figure 8: L3 and L4) have considerably less power. This difference could be caused by some of the signal still remaining on the null ( $\hat{\mathbf{a}}$ ) component even after the desplitting process. The final two lines (L5 and L6) were used to confirm the process of obtaining the power had been implemented correctly. As anticipated, their power is smaller than the signal because it is mixed across both the East and North components. Summary statistics of the maximum amplitude of the power are displayed in Table 3. 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 much larger for windowed noise (Splitting noise) than the background noise (pre- $p$  noise) with the medians differing by a factor of more than 100.

**Table 3.** Summary Statistics ( $\times 10^6$ )<sup>a</sup>

	Minimum	Median	Mean	Maximum
Signal (L1)	0.0637	0.182	10.8	104
Splitting noise (L2)	0.0270	0.0745	0.863	5.58
Pre $p$ noise - East (L3)	0.0000362	0.0000882	0.00335	0.0237
Pre $p$ noise - North (L4)	0.0000230	0.0000631	0.00257	0.0281
Windowed - East (L5)	0.0358	0.0977	4.38	50.8
Windowed - North (L6)	0.0675	0.207	7.01	70.7

<sup>a</sup>Summary statistics are given for the maximum amplitude for the two types of noise.



**Figure 9.** Histograms of the high quality events (108 events) showing: (a) fast polarization estimates, (b) delay time estimates, (c) *Silver and Chan* [1991] standard errors of the fast polarization using (33), (d) *Silver and Chan* [1991] standard errors of the delay time using (33), (e) our standard errors of the fast polarization using (27), and (f) our standard errors of the delay time using (27).

### 3.3. Test 3: Reliability of the Standard Errors

[75] Lastly, we test how reliable the errors obtained from the *Silver and Chan* [1991] method are. This is done by creating plots of these errors (Figure 9) and marking on these plots three different measures of spread: (1) a solid gray line representing the median of the standard error estimates: median  $(\widehat{SE}(\phi_1), \dots, \widehat{SE}(\phi_n))$  and median  $(\widehat{SE}(\delta t_1), \dots, \widehat{SE}(\delta t_n))$ , (2) a dashed red line representing the standard deviation of the parameter estimates: SD  $(\hat{\phi}_1, \dots, \hat{\phi}_n)$  and SD  $(\hat{\delta t}_1, \dots, \hat{\delta t}_n)$ , and (3) a dash-dotted purple line representing the interquartile range (IQR) based estimates of the standard deviation:  $SD_{IQR} = (UQ_{\hat{\phi}} - LQ_{\hat{\phi}})/1.349$  and  $SD_{IQR} = (UQ_{\hat{\delta t}} - LQ_{\hat{\delta t}})/1.349$ . (For a normal distribution,

the standard deviation is equal to  $IQR/1.349$ , see *Suykens et al.* [2002]). If the standard deviation (dashed red line) and the standard error (solid gray) are equal, then the *Silver and Chan* [1991] method is producing unbiased estimates of the errors. If the standard deviation exceeds the standard error, then the method is underestimating the errors, and if the standard deviation is smaller than the standard error, then the *Silver and Chan* [1991] errors are being overestimated. We also check that the IQR-based estimate of the standard deviation (dash-dotted purple line) is not significantly larger than the regular standard deviation, indicating that the distribution is skewed and that the regular standard deviation may be unfairly large.

**Table 4.** Comparison of Errors<sup>a</sup>

	Median Error	Standard Deviation	IQR Standard Deviation
SC $\phi$ (33)	8.75		
Our $\phi$ (27)	11.75	8.76	8.71
SC $\delta t$ (33)	0.0050		
Our $\delta t$ (27)	0.0190	0.029	0.043
SC small $\delta t$ , one 95% contour(33)	0.0034		
Our small $\delta t$ , one 95% contour(27)	0.0047	0.0061	0.0051
SC large $\delta t$ , one 95% contour(33)	0.0034		
Our large $\delta t$ , one 95% contour(27)	0.0056	0.0061	0.0056

<sup>a</sup>Median standard errors, a regular standard deviation and a standard deviation based on the interquartile range (IQR) for different groups of events. Note that the standard deviations are based on the parameter estimates  $(\hat{\phi}, \hat{\delta t})$ , so they remain the same regardless of whether (33) or (27) are used to calculate the standard errors  $(\widehat{SE}(\phi), \widehat{SE}(\delta t))$ . First two rows contain  $n = 108$  events, third row  $n=16$ , and the final row  $n=48$ . Small delay times correspond to the left mode of Figure 9b ( $\delta t \leq 0.06$ ), while large delay times correspond to the right mode of Figure 9b ( $\delta t > 0.06$ ).

[76] Figure 9 shows the standard errors for high quality events (A and B). Figures 9a and 9b show the fast polarization estimates and the delay time estimates, respectively. The fast polarization error estimates from the *Silver and Chan* [1991] method using equation (33) for  $E_4$  (Figure 9c) are unbiased. On the other hand, our  $E_4$  equation (27) (Figure 9e) produces fast polarization standard errors that are slightly overestimated. The delay time estimates using the SC formula for  $E_4$  (Figure 9d) appear to be underestimated by a factor of 2 while the standard errors obtained from our  $E_4$  formula (Figure 9f) appear to be underestimated by a factor of roughly 1.5. However, the IQR-based standard deviation is large in Figure 9b because the distribution is bimodal. This is caused by cycle skipping, where the waveforms are mismatched by an integer number of half cycles. This bimodal distribution has a standard deviation much larger than the spread within each of the modes. To adjust for this, we remove events with ambiguous results (multiple 95% confidence regions in either the SC (33) based contour plot or our (27) based contour plot) and split the delay times into two groups with small ( $\leq 0.06$ , 16 events) and large ( $> 0.06$ , 48 events) delay times (last two rows of Table 4). Both SC and our errors are consistent with the spread of the delay time estimates.

#### 4. Discussion and Conclusion

[77] We have re-presented the method given in *Silver and Chan* [1991] in detail. In section 2, we have identified certain differences between the formulas in our derivations and those given by SC; these are listed in Table 2.

[78] Discrepancies between what was written in SC and what is implemented in the codes have also been identified. While the mistake in calculating the covariance matrix is relatively minor, using the wrong coefficients for the degrees of freedom (cf. 27, 28, and 33) and inconsistent definitions of the standard errors are a concern. We have demonstrated that for uncorrelated Gaussian noise, the formula for the frequency domain sample moment  $E_4$  (28, 33) used by *Silver and Chan* [1991] produces a degrees of freedom estimate that is an overestimate by a factor of about 4/3 (33). Consequently, the standard errors should be too small. This is of concern because commonly used programs (Teauby codes, Mfast (prior to v2.0 which was released July 2013), and Splitalab) use these expressions. Our rederived formula (27) produces an estimate of the degrees of freedom that is unbiased for data with a high sampling rate, when the noise is Gaussian. However, equation (27) is still positively biased by a factor of 1.1 for our particular data set. Our results apply to individual measurements. Averaging of separate results from SC may be less sensible and could introduce a bias [Montellier and Chevrot, 2010]. The *Wolfe and Silver* [1998] technique effectively uses these errors to normalize the contributions of each measurement in its stacking. Therefore, using these new error estimates and contour plots as inputs should also improve the error estimate outputs of the *Wolfe and Silver* [1998] codes. We note that the *Chevrot* [2000] and *Wolfe and Silver* [1998] averages are best for the single-layer, homogeneous, and time-invariant anisotropy case for which they were designed. However, the weighting of the contours in the *Wolfe and Silver* [1998] method means that the measurements with biggest contrast in the errors

(usually the highest signal-to-noise ratio and therefore highest magnitude events) are most highly weighted in the averages, which means that a single measurement from a particular location and time could unfairly bias the estimates. This is an area that we think should be investigated more fully in further research.

[79] In order to obtain a confidence region for the splitting parameters ( $\phi, \delta t$ ), it is assumed that the noise is Gaussian, and consequently, the integrated noise comes from a chi-square distribution. Our results show that the noise cannot be guaranteed to be Gaussian most likely because of contamination by other coherent (nonrandom) signals from scattering and from other events. Further, the power of the splitting noise is significantly larger than the power of the background noise, which is most likely caused by scattering or some other coherent signal remaining in the splitting noise. This raises questions regarding the suitability of theoretical standard errors that require assumptions. However, most methods require either geological or statistical assumptions [e.g., *Silver and Chan*, 1991; *Wolfe and Silver*, 1998; *Menke and Levin*, 2003] and those that do not are only applicable to a specific type of data [e.g., *Chevrot*, 2000]. Investigating such assumptions would be a good area for further research.

[80] The assumptions used to construct the errors do not hold and simulations indicate that the degrees of freedom are positively biased, so the standard errors should be too small. Despite this, tests on very similar earthquakes suggest that the fast polarization errors match the spread of the fast polarization estimates. On the other hand, the delay time errors appear to be underestimated as a result of cycle skipping rather than a flaw in the method itself. When accounting for cycle skipping, the delay time error estimates using equations (33) and (27) also match the empirical spread of delay time estimates from our set of similar events. This could be caused by scattering.

[81] We note that these results were obtained using  $S$  phases from a data set of shallow earthquakes. The frequency content (4–10 Hz) in our test data set is significantly higher than the 0.03–0.1 Hz band often used in SKS studies. Microseismic noise is often large at SKS frequency bands and may also be contaminated by secondary arrivals from phases such as the SpdKS phase. Local  $S$  phases tend to have frequency contents that are more distinct from background noise, but have more contamination by scattered phases. However, since there are no data sets with sufficient number of identical teleseisms, we were unable to verify whether SKS noise biases exist. One way around both these problems could be to test synthetic seismograms with simulated noise and is another area that would benefit from future research.

#### Appendix A: Euler Rotation Matrices

[82] The standard Euler matrices  $\mathbf{E}_i(\theta)$  rotate a vector by an angle  $\theta$  anticlockwise around axis  $i$ , for  $i = 1, 2, 3$ . In this paper, we use the matrices  $\mathbf{E}_2(\theta)$  and  $\mathbf{E}_3(\theta)$  which are defined

$$\mathbf{E}_2(\theta) = \begin{bmatrix} \cos \theta & 0 & \sin \theta \\ 0 & 1 & 0 \\ -\sin \theta & 0 & \cos \theta \end{bmatrix}; \quad \mathbf{E}_3(\theta) = \begin{bmatrix} \cos \theta & -\sin \theta & 0 \\ \sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (\text{A1})$$



## Notation

$\phi$	fast polarization;
$\delta t$	delay time;
$i$	angle of incidence;
$\theta$	SV angle;
$\beta$	back azimuth;
$\mathbf{u}(t, \omega)$	shear wave displacement;
$w(\omega)$	wavelet function;
$\Gamma(\phi, \delta t)$	splitting operator;
$\delta \mathbf{T}$	time difference tensor;
$\mathbf{U}_{\mathbb{A}\mathbb{B}}$	transformation matrix from coordinate system $\mathbb{B}$ to $\mathbb{A}$ ;
$\mathbf{u}_s(t, \omega; \phi, \delta t)$	split shear wave;
$\tilde{\mathbf{u}}(\omega, t)$	rectified (desplit) shear wave;
$W_{\pm}(\delta t; t, \omega)$	shorthand for $w(\omega)e^{-i\omega(t \pm \delta t/2)}$ ;
$\hat{\mathbf{e}}_i$	$i$ th column of the identity matrix;
$\mathbf{E}_i(\theta)$	Euler rotation matrix for angle $\theta$ anticlockwise about axis $i$ ;
$\mathbf{C}(\phi, \delta t; \omega)$	covariance matrix of the rectified wave $\tilde{\mathbf{u}}(t, \omega; \phi, \delta t)$ ;
$\lambda_i[\mathbf{C}]$	$i$ th eigenvalue of the covariance matrix $\mathbf{C}$ , in descending order of magnitude ( $\lambda_1 \geq \lambda_2 \geq \lambda_3$ );
$D_{ii}^*(\delta t, \omega)$	time shift operator;
$\langle \cdot \rangle$	mean value, averaged over a Gaussian noise process;
$k$	number of splitting parameters;
$n$	number of observations;
$\nu$	residual degrees of freedom after $n$ point Gaussian noise is filtered;
$F_{k, \nu-k}^{1-\alpha}$	$1 - \alpha$ quantile of the $F_{k, \nu-k}$ distribution;
$100(1 - \alpha)\text{CR}$	$(1 - \alpha)$ confidence region;
$\lambda_2^{(1-\alpha)}$	eigenvalue corresponding to the boundary of the $(1 - \alpha)$ confidence region;
$E$	signal power;
$T$	time interval of the analysis window;
$f(t; h), \tilde{f}(\omega; h)$	noise filter in the time and frequency domains;
$h$	smoothing width of noise filter;
$g(t), \tilde{g}(\omega)$	Gaussian noise with zero mean and variance $\sigma^2$ ;
$\sigma^2$	variance of the Gaussian noise distribution;
$\eta(t), \tilde{\eta}(\omega)$	filtered noise in the time and frequency domains.

[83] **Acknowledgments.** This work was supported by the New Zealand Earthquake Commission, the New Zealand Marsden Fund, and a Victoria University Wellington Masters by Thesis Scholarship. We thank F. Brenguier and E. Rivemale (Institut de Physique du Globe de Paris, Observatoire volcanologique du Piton de la Fournaise, UnderVolc project) for providing us with the data set. We also thank J. Got for his help with identifying the earthquake locations.

## References

- Angerer, E., S. Crampin, X. Li, and T. Davis (2002), Processing, modelling and predicting time-lapse effects of overpressured fluid-injection in a fractured reservoir, *Geophys. J. Int.*, **149**, 267–280, doi:10.1046/j.1365-246X.2002.01607.x.
- Babuska, V., and M. Cara (1991), *Seismic Anisotropy in the Earth*, Kluwer Academic Publishers, Netherlands.
- Brockwell, P., and R. Davis (2006), *Time Series: Theory and Methods* chap. 11, Springer, New York.
- Chevrot, S. (2000), Multichannel analysis of shear wave splitting, *J. Geophys. Res.*, **105**, 21,579–21,590, doi:10.1029/2000JB900199.
- Crampin, S. (1984), Effective anisotropic elastic-constants for wave propagation through cracked solids, *Geophys. J. R. Astron. Soc.*, **76**, 135–145, doi:10.1111/j.1365-246X.1984.tb05029.x.
- Crampin, S., and S. Chastin (2003), A review of splitting in the crack critical crust, *Geophys. J. Int.*, **155**, 221–240, doi:10.1046/j.1365-246X.2003.02037.x.
- Crampin, S., and Y. Gao (2006), A review of splitting techniques for measuring shear-wave splitting above small earthquakes, *Phys. Earth Planet. Inter.*, **159**, 1–14, doi:10.1016/j.pepi.2006.06.002.
- Fukao, Y. (1984), Evidence from core-reflected shear waves for anisotropy in the Earth's mantle, *Nature*, **309**, 695–698, doi:10.1038/309695a0.
- Gerst, A., and M. Savage (2004), Seismic anisotropy beneath Ruapehu Volcano: A possible eruption forecasting tool, *Science*, **306**, 1543–1547, doi:10.1126/science.1103445.
- Got, J., J. Frechet, and F. Klein (1994), Deep fault plane geometry inferred from multiplet relative relocation beneath the south flank of Kilauea, *J. Geophys. Res.*, **99**, 15,375–15,386, doi:10.1029/94JB00577.
- Joppa, L., G. McNerny, R. Harper, L. Salido, K. Takeda, K. O'Hara, D. Gavaghan, and S. Emmott (2013), Troubling trends in scientific software use, *Science*, **340**, 814–815, doi:10.1126/science.1231535.
- Long, M., and P. Silver (2009), Shear wave splitting and mantle anisotropy: Measurements, interpretations and new directions, *Surv. Geophys.*, **30**, 407–461, doi:10.1007/s10712-009-9075-1.
- Long, M., and R. van der Hilst (2005), Estimating shear-wave splitting parameters from broadband recordings in Japan: A comparison of three methods, *Bull. Seismol. Soc. Am.*, **95**, 1346–1358, doi:10.1785/0120040107.
- Marson-Pidgeon, K., and M. Savage (2004), Modelling shear wave splitting observations from Wellington, New Zealand, *Geophys. J. Int.*, **157**, 853–864, doi:10.1111/j.1365-246X.2004.02274.x.
- Matcham, I. (1997), Seismic anisotropy in the Wellington region from local events recorded at the IRIS station SNZO, Master's thesis, Victoria University of Wellington, New Zealand.
- Menke, W., and V. Levin (2003), The cross-convolution method for interpreting SKS splitting observations, with application to one and two-layer anisotropic Earth models, *Geophys. J. Int.*, **154**, 379–392, doi:10.1046/j.1365-246X.2003.01937.x.
- Montellier, V., and S. Chevrot (2010), How to make robust splitting measurements for single-station analysis and three-dimensional imaging of seismic anisotropy, *Geophys. J. Int.*, **182**, 311–328, doi:10.1111/j.1365-246X.2010.04608.x.
- Nur, A., and G. Simmons (1969), Stress-induced velocity anisotropy in rock: An experimental study, *J. Geophys. Res.*, **74**, 6667–6674, doi:10.1029/JB074i027p06667.
- Nuttli, O. (1961), The effect of the Earth's surface on the S wave particle motion, *Bull. Seismol. Soc. Am.*, **51**, 237–246.
- Ozalaybey, S., and M. Savage (1994), Double-layer anisotropy resolved from S phases, *Geophys. J. Int.*, **117**, 653–664, doi:10.1111/j.1365-246X.1994.tb02460.x.
- Rivemale, E., F. Brenguier, F. Ferrazzini, V. Battaglia, J. Got, J. Kowalski, P. Nercessian, and B. Taisne (2010), Processes of volcanic unrest inferred from 10 years of micro-seismicity at Piton de la Fournaise volcano, oral talk given at AGU, 91(52), Fall Meeting Suppl., Abstract V32B-06.
- Rumpker, G., and P. Silver (1998), Apparent shear-wave splitting parameters in the presence of vertically varying anisotropy, *Geophys. J. Int.*, **135**, 790–800, doi:10.1046/j.1365-246X.1998.00660.x.
- Sandvol, E., and T. Hearn (1994), Bootstrapping shear-wave splitting errors, *Bull. Seismol. Soc. Am.*, **84**, 1971–1977, doi:10.1029/94GL01386.
- Savage, M. (1999), Seismic anisotropy and mantle deformation: What have we learned from shear wave splitting?, *Rev. Geophys.*, **37**, doi:10.1029/98RG02075.
- Savage, M., A. Wessell, N. Teanby, and A. Hurst (2010), Automatic measurement of shear wave splitting and applications to time varying anisotropy at Mount Ruapehu Volcano, New Zealand, *J. Geophys. Res.*, **115**, doi:10.1029/2010JB007722.
- Savage, M., E. Walsh, and R. Arnold (2011), *What Really Causes the Noise in Shear Wave Splitting Measurements?*, International Union of Geology and Geophysics IUGG XXV General Assembly, Melbourne, Australia, 2011, abstract 4282.
- Shapiro, S., and M. Wilk (1965), An analysis of variance test for normality (complete samples), *Biometrika*, **52**, 591–611, doi:10.1093/biomet/52.3-4.591.
- Sileny, J., and J. Plomerova (1996), Inversion of shear-wave splitting parameters to retrieve three-dimensional orientation of anisotropy in continental lithosphere, *Phys. Earth Planet. Inter.*, **95**, 277–292, doi:10.1016/0031-9201(95)03121-9.



- Silver, P., and W. Chan (1991), Shear wave splitting and subcontinental mantle deformation, *J. Geophys. Res.*, *96*(10), 16,429–16,454, doi:10.1029/91JB00899.
- Silver, P., and M. Savage (1994), The interpretation of shear-wave splitting parameters in the presence of two anisotropic layers, *Geophys. J. Int.*, *119*, 949–963, doi:10.1111/j.1365-246X.1994.tb04027.x.
- Suykens, J., T. V. Gestel, J. D. Brabanter, B. D. Moor, and J. Vandewelle (2002), *Least Squares Support Vector Machines*, chap. 5, World Scientific Publishing, Singapore.
- Teanby, N., J. Kendall, and M. van der Baan (2004a), Stress-induced temporal variations in seismic anisotropy observed in microseismic data, *Geophys. J. Int.*, *156*, 459–466, doi:10.1111/j.1365-246X.2004.02212.x.
- Teanby, N., J. Kendall, and M. van der Baan (2004b), Automation of shear-wave splitting measurements using cluster analysis, *Bull. Seismol. Soc. Am.*, *94*, 453–463, doi:10.1785/0120030123.
- Vaseghi, S. (2008), *Advanced Digital Signal Processing and Noise Reduction*, Wiley, West Sussex, section 18.4.7.7.
- Vecsey, L., J. Plomerova, and V. Babuska (2008), Shear-wave splitting measurements—Problems and solutions, *Tectonophysics*, *462*, 178–196, doi:10.1016/j.tecto.2008.01.021.
- Walsh, E. (2013), Measuring shear wave splitting using the Silver and Chan method, Master's thesis, Victoria University of Wellington, New Zealand.
- Walsh, E., M. Savage, R. Arnold, F. Brenguier, and E. Rivemale (2012), Monitoring temporal changes with shear wave splitting: Testing the methodology, poster given at *Eos Trans. AGU*, *93*(52), Fall Meeting Suppl., Abstract S43B-2530.
- Wessel, A. (2010), Automatic shear wave splitting measurements at Mt. Ruapehu Volcano, New Zealand, Master's thesis, Victoria University of Wellington, New Zealand.
- Wolfe, C., and P. Silver (1998), Seismic anisotropy of oceanic upper mantle: Shear wave splitting methodologies and observations, *J. Geophys. Res.*, *103*(B1), 749–771, doi:10.1029/97JB02023.
- Wustefeld, A., and G. Bokelmann (2007), Null detection in shear-wave splitting measurements, *Bull. Seismol. Soc. Am.*, *97*(4), 1204–1211, doi:10.1785/0120060190.
- Wustefeld, A., G. Bokelmann, C. Zaroli, and G. Barruol (2008), Splitlab: A shear-wave splitting environment in Matlab, *Comput. Geosci.*, *34*, 515–528, doi:10.1016/j.cageo.2007.08.002.