Automatic measurement of shear wave splitting and applications to time varying anisotropy at Mt. Ruapehu volcano, New Zealand

Submitted for publication in Journal of Geophysical Research May 2010, revised August 2010. The citation for the edited paper is:

Savage MK., Wessel A, Teanby NA, Hurst AW, Automatic measurement of shear wave splitting and applications to time varying anisotropy at Mount Ruapehu volcano, New Zealand, *J. Geophys. Res.*, 115, B12321, doi:10.1029/2010JB007722, 2010.

Authors: M. K. Savage[1], A. Wessel[1], N. A. Teanby[2], T. Hurst[3]

1. Institute of Geophysics, Victoria University of Wellington, New Zealand
2. Atmospheric, Oceanic and Planetary Physics, University of Oxford, OX1 3PU, U.K.
3. GNS Science, 1 Fairway Drive, Avalon, Lower Hutt 5010, PO Box 30368, Lower Hutt 5040, New Zealand

Up to 5 Index Terms: 7203 Body Waves; 7223 Earthquake interaction, forecasting, and prediction; 7280 Volcano Seismology; 7230 Seismicity and tectonics; 8419 Volcano Monitoring

Up to 6 Key words: shear wave splitting; anisotropy; automatic; stress; birefringence; temporal variation

Abstract

We present an automatic shear wave splitting measurement tool for local earthquakes, with the sole manual step of choosing an *S* arrival time. We apply the technique to three data sets recorded on Mt. Ruapehu Volcano in New Zealand that have previously been determined to have fast polarizations that vary in time and with earthquake depth. The technique uses an eigenvalue minimisation technique, applied over multiple measurement windows. The dominant period of each waveform sets minimum and maximum window lengths. Cluster analysis determines the best solution among all the windows, and quality grading criteria assess the results automatically. When the same filters are used for events determined to be high quality from manual studies, the automatic technique returns virtually identical results. Scatter increases when the automatic technique chooses the best filter, but the average automatic results remain consistent with the manual results. When the automatic technique is used on sets that include data previously judged as poor quality, some stations yield distributions of fast polarizations that include peaks that were not present in previously published results. The difference may stem from two factors: automatic grading lets through some measurements that independent analysts consider poor quality, but also unconscious bias in the manual selection process may downgrade measurements that do not fit expectations. Nonetheless, the new objective analysis confirms changes in the average fast polarizations between 1994 and 2002, and between shallow and deep events. Therefore, this new technique is valuable for objective assessment of anisotropy and its variation in time.

1. Introduction

Seismic anisotropy, the directional dependence of seismic wave speed, is one of the few means by which directionality can be measured beneath the surface of the earth. It has been used widely to study strain and deformation due to aligned minerals in the mantle [*Fouch and Rondenay*, 2006; *Savage*, 1999] and also to study structural properties and stress due to aligned cracks in the crust [e.g., *Balfour et al.*, 2005; *Boness and Zoback*, 2006; *Crampin*, 1994; *Zinke and Zoback*, 2000]. It has been proposed as a stress monitoring device that could potentially have predictive powers for earthquake occurrence or volcanic eruptions [e.g., *Crampin*, 1994; *Gerst and Savage*, 2004; *Miller and Savage*, 2001]. To reach its full potential for rapid evaluation in the case of monitoring, automatic methods are becoming imperative. Manual measurements are time consuming and tedious, and they are less objective than fully automatic methods.

The most common method to study anisotropy is to examine its birefringent effect on shear waves, which is often called "shear wave splitting". This splits the wave into perpendicular fast and slow components. The first arriving wave has a polarization (ϕ) parallel to the fast orientation of the anisotropic material and the delay time (*dt*) between the two waves depends upon the integrated effect of anisotropy along the travel path. If anisotropy is caused by alignment of near-vertical cracks or microcracks, then ϕ is a measure of the average crack orientation, assumed to be parallel to the maximum horizontal stress (S_{Hmax}) and *dt* is proportional to the crack density [*Hudson*, 1981].

Several techniques have been suggested for automatic analysis of shear wave splitting. Many of them use as their basis the grid search technique pioneered by Silver and Chan [1991] (SC91), which requires the user to select a time window over which to examine a waveform. Their "eigenvalue" method is most relevant here; it is used when the incoming polarization of the waveform is unknown. In this case, the pair of (ϕ, dt) that best returns the waveform to linear particle motion as measured by the smallest eigenvalue of the corrected covariance matrix is determined to be the best measurement.

The major difficulty of the SC91 algorithm is that the measurements can depend upon both the filter used for analysis and also on the measurement window chosen. Several approaches have been suggested for dealing with this. Walker et al. [2004] used the SC91 algorithm on SKS phases for 30 different time windows and averaged the results. A fully automated method for SKS data uses at its base the SC91 algorithm [*Evans et al.*, 2006]. It uses a short term /long term average algorithm to pick the start of the SKS phase, and then it measures splitting parameters on numerous windows to determine the set of windows with the longest stable window length. However, 98% of the data are rejected with this method.

Other researchers use methods that are similar to the SC91 method, or incorporate parts of it. For crustal events, Gao et al. [2006] and Bianco and Zacarelli [2009] published semi-automatic shear wave splitting methods that use the covariance matrix to determine the fast polarization and cross correlation to determine the delay times. But they both require extensive tuning of the parameters before applying the method to any particular dataset, and use a fixed window length once the tuning is complete. Bianco and Zacarelli [2009] used their technique extensively on Italian volcanoes and document changes with time in the polarizations and time delays. Peng and Ben-Zion [2004] also start with a method that is equivalent to SC91. They choose a single frequency band (less than 15 Hz), and use a sliding 0.6 s long time window over the *S* arrival to determine which window yields the most linear particle motion, which is used as their final determination. They also have a series of quality criteria that they apply to weed out bad measurements. They tested their method with local earthquake data recorded near the North Anatolian Fault, Turkey and report a rejection of 70% of the data. Liu and others [2007] use the SC91 method and also a cross correlation method to determine splitting parameters for a large dataset. They also use a single filter (1-10 Hz) and use a fixed start of 0.02 s before the *S* arrival. They use a variable window length and choose the window that gives the maximum cross correlation coefficient as their best window. Again, they have a number of criteria for grading the final measurements.

Teanby et al. [2004a] use a slightly different approach to find the best measurement window. They use the SC91 method on multiple analysis windows and use cluster analysis to determine the best measurement. The best measurement is determined to be within the cluster that has the most measurements with the smallest error bars for the individual measurements. This method has been systematically tested on a dataset in the North Sea and proved useful for analysis of time variation of splitting using local events [*Teanby et al.*, 2004b]; and also for static properties of teleseismic events [*Greve et al.*, 2008; *Savage. et al.*, 2010], but most studies using this method include manual grading to weed out poor measurements.

We present here a method for full automatic measurement of shear wave splitting that allows the quick processing of thousands of events with minimal manual interaction. Our method expands on the Teanby method by first finding the best filter and then using the frequency of the seismogram to determine the lengths of the analysis windows. Using multiple filters allows a larger set of data to be investigated than the previous techniques, which mainly used single filters. We also developed an automatic grading technique that evaluates the measurement based on the clustering characteristics instead of simply choosing the best cluster. We have previously applied the method on earthquakes recorded at volcanoes in Japan [*Savage et al.*, 2008; *Savage. et al.*, 2010], Alaska [*Johnson et al.*, 2010] and Montserrat [*Roman et al.*, 2010], and we found time variations that we interpret to be caused by deformation caused by magmatic movement. Here we present a detailed analysis of the method and we apply it to data analyzed previously around Mt. Ruapehu Volcano, from 1994, 1998 and 2002, in which temporal changes in anisotropy have been reported in association with volcanic activity [*Gerst and Savage*, 2004; *Miller and Savage*, 2001].

2. Method

The method is based on the SC91 algorithm and the cluster analysis method of Teanby et al. [2004a]. An overview of the processing steps is illustrated in Figure 1. Detailed explanations are provided in the manual discussed in the appendix, and here we present a summary. The SC91 analysis is carried out on multiple measurement windows and cluster analysis determines the best window. The cluster that has the minimum variance is chosen as the best cluster, and a final SC91 measurement is made based on the window that gives the minimum error bars on the splitting parameters within the cluster. The main advance presented here is in automatically choosing the range of the measurement windows to use for the cluster analysis based on the period of the event. As part of this advance, we use multiple filters to find the frequency bands with the best signals as measured by the product of the signal-to-noise ratio and the filter bandwidth. Furthermore, automatic grading is carried out to determine the best measurements. These extensions to the SC91 and Teanby et al [2004a] techniques greatly simplify the processing of large datasets, with the only manual step being in picking the *S* arrival time. However, the measurements require several parameters, which could be checked and modified in other studies. These parameters can be easily changed in the code and their definitions and choice of value are fully discussed in Table 2 of the manual.

2.1 Filtering

Since local broadband seismic data are often contaminated by noise, raw data can rarely be processed. The application of a bandpass filter is important for shear wave splitting measurements, because a sufficiently high signal-to-noise ratio is necessary for a high quality measurement (Figure 2). However, narrow band filters should be avoided because they can lead to cycle skipping. Therefore we favor broader band filters when possible. Instead of applying a broad filter to all data or manually selecting a filter for each measurement, a bandpass filter is determined for each measurement on the basis of a signal-to-noise ratio criterion and the width of the filter, as discussed later.

A predefined set of 14 bandpass filters (Table 1), with typical corner frequencies found in previous studies [*Gerst*, 2003] is tested for each measurement and the best filter is selected. Using different filters would make this method applicable to broad classes of data beyond the application to local *S* waves described herein. For example, the Teanby codes can be used on teleseismic data to study mantle anisotropy [*Greve et al.*, 2008; *Savage et al.*, 2007] and straightforward modifications of the codes described herein could be used to create automatic *SKS* measurements. If short period data are used, the long period filters should be modified since they will effectively be cut off at higher frequencies than expected, so some of the resulting filters will give nearly identical responses. For particular datasets different sets of filters may be more appropriate, although those listed in Table 1 provide reasonable starting values for most microseismic studies.

2.2 Signal-to-noise ratio calculation

A new signal-to-noise ratio (SNR) for the filtered data is calculated using the same three second window length for both noise and signal. The noise window precedes the *S* arrival (here defined as time 0 s) and includes an offset to account for possible inaccuracies in the *S* arrivals (here we use -3.05 to -0.05 s). The signal window follows the *S* arrival (here 0.05 to

3.05 s). The noise window is chosen to precede the *S* arrival directly to include the signal-generated noise of the *P* coda, because such signal-generated noise will affect the *S* wave analysis.

The mean of the ratio of the amplitude of the Fast Fourier Transform of the signal and noise windows is used as the SNR. No measurements with SNR less than a maximum value (3 here) are considered for interpretation. In the applications discussed herein, the waveforms filtered with the three filters giving the highest value of the product of the filter bandwidth in octaves and the SNR (if there are three or more filters fitting the SNR criterion) are analyzed. This allows us to examine the frequency dependence of the results, and in averaging the parameters, it ensures that the measurements that are most stable with frequency contribute most to the final measurement. Earlier versions of the code used only the SNR itself rather than the SNR-bandwidth product, and were more prone to cycle skipping, relying more heavily on later grading to weed out poor measurements [*Johnson et al.*, 2010; *Savage. et al.*, 2010].

2.3 Basic splitting measurement technique

The shear wave splitting parameters are measured by applying an inverse splitting operator, which is determined by a grid search over possible values [*Silver and Chan*, 1991]. The more a certain operator removes the splitting of the investigated seismogram, the smaller the second eigenvalue λ_2 of the covariance matrix of particle motion $c(\phi, dt)$ becomes. This is equivalent to maximizing the linearity of the particle motion [*Silver and Chan*, 1991]. The inverse operator that removes the shear wave splitting best gives the resultant shear wave splitting parameters. Contours of λ_2 for all the operators considered give a measure of the confidence region by using an F-test (SC91) (part f in Figure 3,4). Here we search the parameter space in units of 1° and a time unit depending on the scale of the problem for all fast polarizations and for delay times from 0.0 to a maximum value (*tlagmax)*, 1.0 s in this study. For some studies in which there were many local earthquakes close to a volcano, we used a maximum of 0.4 s for the local events [*Johnson et al.*, 2010; *Savage et al.*, 2008; *Savage. et al.*, 2010]. For *SKS* measurements, 4 to 6 s may be used[*Greve et al.*, 2008; *Savage et al.*, 2007].

The results of the grid search for one pair of shear wave splitting parameters can be dependent on the selected measurement window, i.e., the part of the waveform that is actually considered for the measurement. To address this dependency, the analyst usually performs a measurement several times with different measurement windows to confirm the stability of the resulting parameters.

The method published by Teanby et al. [2004a] automatically performs measurements for a large number of window configurations and then determines the most stable solution with cluster analysis. The original method allows one to choose one set of configuration parameters for all measurements or to interactively choose measurement window times. We extended the method to automatically generate a customized configuration file for each single measurement. Measurement window times relative to the *S* arrival are calculated based on the dominant frequency of the signal.

The dominant frequency f_d is calculated from a window (3 s here) which follows the *S* arrival. The maximum and minimum possible frequencies of f_d are limited such that at least one full cycle is included in the calculation window and frequencies do not become unfeasibly large. We used $0.3 \le f_d \le 8$ Hz.

Following suggested guidelines [*Teanby et al.*, 2004a], the minimum window (2-3 in Figure 3b) to use in the splitting cluster analysis is chosen to be one period long (calculated from $1/f_d$), while the maximum window (1-4) is 2.5 periods long. The number of different measurement window end times depends on the minimum and maximum window length (see the manual described in the Appendix for details). Here the number of end windows is between 15 and 25.

The minimum and maximum times of the measurement window start are less critical than the window end times [*Teanby et al.*, 2004a]. We consider five window start times in our application of the cluster analysis, which are usually 0.2 s apart, so that t=-0.3, -0.5, -0.7, -0.9 and -1.1 s, relative to the *S* arrival. However, to minimize interference of the *P* wave for close earthquakes, if the time between the *S* and *P* arrivals (*Ts-Tp*) is less than 2.2 s, we make the shortest time window begin at t=-0.1 s and the longest time window begins at *-(Ts-Tp)*/2; the other three time window start times are scaled accordingly. The total number of measurement window *Ntot* is thus between 75 and 125, and is directly proportional to the processing time.

The cluster analysis searches the parameter space of the *Ntot* pairs of measurements ($\phi(i)$, $dt(i)$), where i is an index to indicate the measurement number, to determine clusters of measurements with similar values, and is described more fully elsewhere [*Teanby et al.*, 2004a]. Several sets of statistics are used to describe the clusters, the most important of which is the total variance of each cluster. It depends on both the average variance of the individual measurements within each cluster and the variance of measurements within the cluster. The cluster with the minimum total variance is chosen as the best cluster, and within that cluster, the measurement with the minimum variance is chosen as the best measurement, and is used as the final measurement for that phase at that filter. Measurements from different filters are compared as discussed below.

2.4 Grading criteria

One problem plaguing shear wave splitting measurements is that of cycle skipping, in which the splitting program may mismatch waveforms by an integer number of half-cycles. If the waveform is mismatched by one half cycle, then the fast and slow waves may be interchanged, and *dt* differs by one half period [e.g., *Matcham et al.*, 2000]. This is particularly a problem with narrow band filters, and is a cause of concern at volcanic areas, because the stress field near a dyke is proposed to reorient by nearly 90º after the intrusion [*Gerst and Savage*, 2004]. Even when cycle skipping is not present, sometimes a cluster of windows will include scattered phases that result in multiple solutions that differ from each other by values that are other than an integer half cycle or 90°. For example, see Figure 4(c) and (d), which has three groups of clusters with different ϕ and *dt*. Most studies use manual checks to alleviate the problem, but it can be time consuming and also difficult to be objective during manual checks. So we introduce an automatic technique instead to eliminate measurements with multiple solutions.

We automatically grade the measurements based on the cluster analysis. Instead of simply using the results from the best cluster, all clusters with measurement numbers above the minimum threshold are compared to the chosen "best cluster". We try to reject measurements in which there are secondary clusters with similar quality to the best cluster, but with very different shear wave splitting parameters. Let *var(k)*, $\phi(k)$, *dt(k)*, and *Nmeas(k)* be the average variance, fast polarization, delay time, and number of measurements in cluster *k*, respectively. Let *kbest* be the cluster number of the measurement with *var(kbest) = min(var(k))*. This is the cluster chosen by the original Teanby et al. [2004a] program to be the best measurement. *Nmeas(kbest)* is thus the number of measurements in the best cluster. To consider clusters of similar quality, the cluster grading considers all clusters with *var(k) <5*var(kbest)*. Within these "OK clusters" we consider the differences between the fast polarizations and delay times of each cluster compared to the best cluster. Therefore, we define *ϕdiff = abs(ϕ(k)-ϕ(kbest))* and *tdiff = (dt(k)-dt(kbest))* to assess the reliability of the best cluster. Table 2 includes a description of the cluster grading methods.

Another concern is for "null" measurements, which can occur when there is no anisotropy in the plane of the *S* wave particle motion, or when the initial shear wave is polarized along the fast or slow orientation of the medium, so that no orthogonal wave exists to split (Figure 5; see also [*Silver and Chan*, 1991; *Wüstefeld and Bokelmann*, 2007]). Often they will be given a poor grade in the cluster grading since different windows may return ϕ or *dt* that are far apart. But in case the cluster grading has not removed them, these null measurements are treated separately from regular splitting measurements. We use a geometrical criterion [*Peng and Ben-Zion*, 2005; *Savage et al.*, 1996]; we compare the initial polarization α determined from the SC91 inversion program to the fast polarization ϕ*.* Measurements are considered as null, if they do not fulfill the criterion $20^{\circ} \leq |\phi - \alpha| \leq 70^{\circ}$. For a uniform distribution of incoming polarizations, we expect 4/9 of the measurements to be rejected by this criterion.

Measurements that result in a delay time close to the maximum indicate cycle skipping or noisy data [*Evans et al.*, 2006], therefore measurements close to the maximum *dt* are also rejected. The mean delay time obtained by Gerst and Savage [2004] for local earthquakes at Mt. Ruapehu in 2002 is 0.11 s for shallow events (z<35 km) and 0.27 s for deep events (z>55 km). For local events the delay times are generally expected to be between 0.1 and 0.6 s (Table 1 of Savage [1999]). Therefore we choose the maximum delay time for the grid search (*tlagmax* described above) to be 1.0 s and subsequently rejected all measurements with a delay time greater than 0.8 times this value. A final grade of A or B is made based on whether the measurement has a cluster grade of *Acl* or *Bcl*, and values of the SNR and 95% confidence interval of the ϕ measurement (Table 2).

Finally, we developed a criterion based on our manual grading techniques, which involve examining the plots of the contours of the eigenvalues of the covariance matrix of the final best measurement (e.g., Figures 3, 4 (f)). A small range of contours indicates that the best result is not much better than the worst result. Within the SC91 code, the eigenvalues are rescaled so that the 95% confidence interval has a value of 1 and contours are multiples of this interval. We use the maximum value of the error contours as a further grading parameter. This quantity is correlated with the signal-to-noise ratio, and on the formal error bars of the final best splitting measurement, which we also include as grading criteria (errors in ϕ must be less than 25º for quality A and B measurements), and which is also based on the contours. But the energy criterion is distinct, as it applies to the whole error surface, not just the region around the minimum. Below we use a value of 8 or greater of the maximum error contour to define a high quality measurement (Eng8). The value of 8 was chosen because it left roughly the same number of measurements in the high quality manual (A and B) and the automatic technique in an early implementation of the SNR criteria.

The manual results against which we compare our data were hand graded with published results ranging from C to A, including half-grades [*Gerst*, 2003]. It is difficult to exactly match hand graded versus automatic grades. However, the measurements graded A and AB in Gerst [2003] have similar results to those graded A and B in our analysis (see below); we show below that when they are used on the same waveforms, they give closely matched results. Therefore we consider that the automatic A and B analysis is similar to the A and AB manual grades.

2.5 Averages

To calculate average parameters over multiple events sampling the same anisotropy, we use Gaussian statistics for the delay times, and the Von Mises criterion [*Mardia*, 1972], a circular analogue to the normal distribution, for the polarizations. Along with the calculation of a mean fast polarization, a test for non-randomness must be conducted [*Davis*, 1986]. The calculation for the mean of the polarizations involves adding unit vectors with orientations given by the measured values and dividing by the number of measurements. The ambiguity of 180° in polarization is taken care of by doubling the angles before the vector addition, and halving the resultant angle. The length of the resultant vector, R, gives a measure of the misfit between 0 and 1. R=1 when all the polarizations are exactly lined up. Errors presented are the standard error, which is valid if the distributions are approximately normal. However, many of the distributions are bimodal and thus are not well described by normal distributions, so we recommend caution in using the averages and standard errors.

2.6 Use of multiple filters

Several of our earlier studies presented results from the same event-station pair using multiple filters in the rose diagrams and used these multiple filters in the averages. The rationale was that if an event had the same results with different filters, including it several times would effectively weight the results more heavily than if it had different results with different filters. However, such results should not have been treated with Gaussian statistics because the results from different filters will correlate with each other. It is also difficult to decide which filter results to compare with each other. Therefore, unless otherwise noted, for the measurements shown here we include a further step, which is effectively another grading step: for a given event-station pair, if more than one filter produced a result that has passed the grading criteria, we compare the results and remove the entire measurement if they are too different from each other, using the criteria described above for the cluster grading, i.e., if the time difference is greater than one eighth the maximum allowed value or if the angular distance is greater than $\pi/8$. If the results are similar, we choose the one with the smallest error bars, as calculated by the sum of the relative errors.

3. Data

The method presented here can generally be used to perform shear wave splitting measurements on any three-component data with *S* phases that are discernable above the noise. Because different measurement windows are processed by the cluster analysis, and because the windows begin a set time before the *S* phase, small errors in the *S* arrivals are mitigated.

The cluster analysis method was extensively tested and compared with manual measurements [*Teanby et al.*, 2004a]. In the initial development stage of the modified method [*A Wessel*, 2008], we used the manual results from sixteen high quality events recorded at four stations of the 2002 CHARM deployment on Ruapehu [*Gerst*, 2003; *Gerst and Savage*, 2004] to guide our choice of the best parameters to try to match the manual results. The method was further developed to work at Asama volcano in Japan. Filters were changed slightly from the original study. Because many of the seismograms were emergent at Asama, we used a longer window before the *S* arrivals, which we carry over into the analysis here. We also developed the grading technique by examining data recorded at Asama to find the most stable results [*Savage. et al.*, 2010].

To test the newly developed method, we used data that were recorded by temporary deployments in 1994, 1998 and 2002 on Mt. Ruapehu Volcano, New Zealand, and which were already analyzed manually [*Gerst*, 2003]. The data sets from 1994 and 1998 were sparse (5 and 3 stations, respectively), while 8 stations contributed data in 2002. All *S* phases used in the automatic processing for comparison to the manual methods were re-picked manually, because the previously processed data [*Gerst*, 2003; *Gerst and Savage*, 2004] did not indicate the *S* arrivals, but only the window about the *S* wave used in the final analysis. Events are separated into shallow ($z \leq 35$ km) and deep ($z \geq 55$ km) events. Only the 1994 deep and 1998 shallow events were reprocessed manually to compare to the 2002 dataset [*Gerst and Savage*, 2004], and so we do not include the 1994 shallow or 1998 deep events processed in an earlier study [*Miller and Savage*, 2001].

4. Results

4.1 Comparison of individual measurements

Figure 6 presents the distribution of the frequencies used in the analysis, by plotting histograms of the numbers of measurements as a function of the dominant frequency in the measurement window. Before filtering, the distribution of the dominant frequencies is relatively uniform between 2 and 7 Hz. The measurements with the best signal-to-noise ratio tend to come from filters with lower frequencies, so that the dominant period of the seismograms for most measurements is 3 Hz. The dominant frequencies of the data filtered with the best SNR-bandwidth product also have a peak at 3 Hz, but there are more measurements at higher frequencies than when the SNR alone is used.

We performed several tests of the method. First, we selected the measurements for the 2002 data determined by Gerst [2003], that had A and AB ratings, and compared these manual results to the automatic results using the same filters on the same waveforms (Figure 7; Table 3). The difference between the two methods is that the automatic program chooses the best window by cluster analysis, while the manual method relies on a human to test several windows and filters. If two or three filters gave different results they are all included in the analysis.

The results are shown for increasing levels of grading (Table 2; section Methods). The scatter in the relation increases when poorer quality data are used, but the number of data points is increased (Figures 7, 8, Table 3). However, the majority of all the measurements are within 25 degrees (minimum 75%) and 0.2 s (minimum 85%) of each other, even when no quality criteria are applied (Table 3). The "Energy contour ratio > 8" (Eng8) criterion seems to get rid of almost all outliers once all the other criteria have also been applied (Figure 7). But it does this by leaving out many more measurements. The “all” or “unculled” group (no quality criteria applied beyond the original Teanby et al. methodology) has 212 measurements, reducing to 160 when the null criteria and the D quality measurements from clustering are left out (noNnoD). The AB group, which are the quality that were used in previous publications [*Johnson et al.*, 2010; *Savage. et al.*, 2010], have 145 measurements and when only the top “A” quality measurements and the "Eng8" criteria are applied it reduces to 75, leaving out almost 2/3 of the measurements. Interpreting this must be tempered by the fact that we chose the “Eng8” criterion in order to get roughly the same number of measurements as were in the A and AB manual measurements in an earlier version of the codes.

Another test is to see whether the results will be the same if the automatic code chooses the filter. We compare the same A and AB measurements from Gerst [2003], but allow the automatic program to choose the filters based on the signal-to-noise ratio-bandwidth criterion (Figure 9, Table 3). Both the automatic and manual analyses sometimes give different results for different filters. We have applied the criterion described above separately to the manual measurements and the automatic measurements to weed out measurements that have different results at different filters, so that we compare only one measurement for each event-station pair. Surprisingly, there are fewer outliers for this case compared to when we used the exact same filters, and there is less difference between the various grading criteria; we discuss the reasons below in the discussion section.

Using three filters requires more computer processing time, and in monitoring situations it would be preferable to have a rapid evaluation. Therefore we test the results for the best single filter compared to the manual results (Figure 10). As in the case when the exact filters were compared (Figure 7), increasingly higher quality criteria lead to smaller scatter at the expense of leaving out more measurements. In this case the “ABeng8” criterion reduces the results to 33 from 116 measurements, or 28% of the total.

Finally, we apply the most rigorous test, which is to start with all quality data rather than the selected A and AB quality measurements to see if our technique leaves out the same events and gives the same results (Figure 11; 12, Table 4). This test most closely matches the envisioned applications, when investigators will be using the automatic techniques to evaluate data that has not yet been examined. We started with all the data that had splitting measurements made by Gerst [2003], including the null measurements (431 measurements). We ran the automatic method on these phases, allowing the automatic technique to choose the best filter and keeping the results from the best three filters if they passed the other quality criteria. Rose diagrams made with measurements fitting the ABeng8 criterion (162 events) are compared to the A and AB quality non-null measurements determined by Gerst (168 measurements) (Figure 11, 12). Table 4 compares the results separated into years and deep and shallow events for several methods of measurements compared to the manual results. The rose diagrams use results from all three filters in all studies because they most closely match the conditions used by Gerst and Savage [2004]. The tables present more results using our selection technique to only choose one filter in each measurement set, for both the original AB files and those with eng8.

4.2 Comparison of averages

In all cases, the automatic results are less coherent than the manual ones, as measured by smaller values of R (Table 4). The mean for all delay time measurements in each depth group differs by less than 0.04 s for manual vs. automatic, although in a few cases they are slightly beyond the 95% confidence interval for the hypothesis that they are taken from the same population. The delay times in the deep groups are significantly larger (around 0.25 s) than the shallow group (around 0.14 s). The most scattered results, with the smallest R, are for the 1998 shallow events in all groups. 1998 and 2002 automatic measurements are within the 95% confidence intervals of the manual ones, but the automatic measurements for 1994 deep are more northerly (-25 +- 6) than the manual measurements (-43+-4). For the most numerous 2002 dataset, the biggest differences between the manual and automatic results are for deep events at stations LHUT2 and FWVZ (Figure 11).

5. Discussion

5.1 Comparison of Individual Measurements-Scatter Plots

As discussed in Section 4.1, when the automatic technique chooses the best filters, the measurements are slightly more scattered, but the average values are close and surprisingly, there are fewer outliers than when the exact same filters are used (compare Figs. 7 and 9; Table 3). The results for different quality criteria are also somewhat unexpected, because there are fewer measurements in the group without quality control applied (“all or unculled”), than some of the others. To understand this, recall that the quality criteria are applied before the program that determines whether measurements at different filters should be discarded. So for the “unculled” measurements, if a poor quality result at one frequency gives a different answer than a higher quality result, all the results for that event/station pair are thrown out.

On the other hand, in the "culled" plots, we first get rid of measurements with poor quality criteria, e.g., if an individual measurement at one frequency has too high error bars, it will be thrown out and the other two measurements will be left to compare to each other to see which one is kept for the final figure. This suggests that an effective way of removing poor measurements may be just to measure results on the three best filters and compare the results, without doing any of other quality checking.

Delay times from shear wave splitting measurements from local events generally yield a high scatter of approx. 80%, and the large range of delay times in our data confirms this [*Crampin*, 1999; *Gerst and Savage*, 2004]. Nonetheless, delay times are consistent between the manual and automatic method (Fig. 7-10). Therefore, the scatter in delay times is not a function of processing errors, but must result from some physical process. That process could be rapid stress changes occurring between each measurement [*Crampin et al.*, 2004], but we think that is unlikely, especially since there is a scatter for measurements made at a single waveform but at different frequencies. Instead, phases converted or scattered from interfaces with strong velocity contrasts, such as magma pockets, could be arriving before or after the initial wave and interfering with some of the measurements. In this case, slight changes in paths could make a change in the splitting measurement [e.g., *Liu et al.*, 2004].

The histograms of differences between manual and automatic results (Figs. 8,9) do not show a strong signal from peaks at 90 degrees, suggesting that cycle skipping is not a problem with this data.

5.2 Comparison of Individual Stations-Rose Diagrams

Despite the close relationship when comparing waveforms already judged as high quality by the manual method, if all data are included in the processing, histograms of fast polarizations of the top quality automatic measurements vary at some stations compared to the manual measurements (Figure 11; Figure 12). Results and result qualities vary from station to station in both data sets. In particular, at most stations the automatic measurements are more scattered than the manual measurements, although some stations (e.g., LHOR; FWVZ shallow) have less scatter with the automatic than with the manual methods (Figure 11).

Hypotheses for differences that we have explored are (a) the different techniques reject null measurements at a different rates (b) manual analysis is better than automatic at leaving out nulls and mismatched phases (c) the manual analysis is unintentionally biased. The proportions of nulls rejected are similar for the automatic (15%) versus manual (20%), so (a) is not the answer. Hypotheses (b) and (c) are examined by detailed analysis of waveforms and their grading at two key stations, using plots such as those in Figure 3 and 4 (Table 5). The biggest differences between manual and automatic measurements occurred for station LHUT2 for the 2002 deep measurements (Figure 11). LHUT2 has a single NE/SW peak in the manual measurements, but an additional second peak in the rose diagrams for the automatic methods makes them bipolar. The new automatic technique gives three measurements for two events with values between -81 and -85 ("E-W" measurements), with the rest of the results spread between 22 and 65 (Table 5). The manual measurements gave similar values for the two "E-W" measurements reported, of -71 and -88. But only one filter instead of two was judged high quality for one of the measurements, and the other was graded B, so it was left out of the A/AB plot. More measurements were graded high that gave results between 10 and 70 degrees. Some but not all of these measurements had *dt* values that were less than 0.1 s (Tables 3, 4), which could be indicative of small anisotropy that could easily have polarizations change. The filters chosen by the automatic method had overlapping frequency bands, which could bias them by allowing multiple measurements with the same values. However the filters used in the manual technique also had overlapping frequency. Finally, earlier versions of the automatic routine that used the SNR alone instead of the SNR-bandwidth product had stronger bimodality in the solutions than the SNR-bandwidth product (Figure 11, 12), which suggests that the bimodality could be caused by cycle skipping. So we conclude that both factors (b) and (c) may be playing a role in the differences: there may be some unintentional bias in the manual determination that rejected more measurements that did not fit the expected polarization, but the manual grading may also be better able to reject poor or marginal measurements. We think that it is natural for manual graders to tend to favor measurements that are consistent with previous measurements at the same station, leading to a narrower scatter for manual measurements than for automatic ones. Table 5 shows that two independent observers do not always grade the same measurement the same way.

5.2 Average results and time variations

The average results across all stations for the 2002 data are similar for the manual and automatic measurements (Figure 11; Table 4). Although the 2002 shallow events have average results that differ between manual and automatic measurements by more than the standard error, the absolute difference between the manual and automatic techniques is less than 10˚. The biggest difference in the average results over all stations for the manual vs. automatic techniques occurs for the shallow measurements in 1998, which had almost twice as many automatic as manual measurements, and the averages differ by 52°. The differences are not too surprising, however, if we consider that the main conclusion about the 1998 shallow events was that they were more scattered than the 1998 deep events or than the 1994 measurements from earthquakes at all depths [*Miller and Savage*, 2001].

For the automatic technique, differences in the average fast polarizations between the 1994 deep events and the 1998 shallow events are not statistically significant, but the difference between 1994 deep events and 2002 deep events (59±19) is robust (Table 4). Differences in average ϕ in 2002 between shallow and deep events that were pointed out Gerst and Savage [2004] are not as strong with the automatic technique, but they are statistically significant (differences are 37±12). Therefore, spatial and time variations of seismic anisotropy are confirmed for Mt. Ruapehu when a fully objective, automated method is used.

The stations that had the most scatter in 2002 are close to each other (Figure 11), near the summit of Mt. Ruapehu. We are presently operating another network to study spatial variation of anisotropy on Mt. Ruapehu, which might help to clarify whether this is a trend.

5.3 Recommendations:

The method can in principal be used with any three-component local earthquake data set. So far it has been tested on other data sets on Mt. Asama volcano [*Savage. et al.*, 2010] and three other volcanoes in Japan [*Savage et al.*, 2008], Okmok, Alaska [*Johnson et al.*, 2010] and at Soufriere Hills volcano in Montserrat [*Roman et al.*, 2010]. But we have also made some preliminary studies in non-volcanic areas of New Zealand and have found it to work equally well. With changes in filters we expect that it could readily be modified for teleseismic events, because the original Teanby code has already been modified for teleseismic analysis [*Savage et al.*, 2007].

If there is plenty of time for analysis such as when real-time applications are not being used, then since the shapes of the histograms for different quality criteria don't change too much with respect to outliers (Figure 9; 10), it would be best to keep the results that give the most total measurements, i.e., in this case the three filters with the "AB" criteria (86 measurements). However, using only the best “fb1” filter yields almost as many (80) measurements, and will take only 1/3 as much processor time. Therefore it is certainly acceptable in real-time analysis, and may also be reasonable for other applications. Using the "energy >8" criterion reduces outliers almost to zero, but it also reduces the measurements considerably and is probably only acceptable in situations where there are large quantities of measurements to choose from, such as when time variations are not observed. Using a smaller level for the energy contour grading criteria could also be explored.

We expect that in different areas, different filters might be more suitable. It should be simple to customize the filters for different regions (see details in the manual discussed in the Appendix).

Visual analysis of waveform fits will be necessary in any study in a new area to check that the parameters used in the automatic techniques are suitable. We recommend that in new studies, manual results for a subset of events first be checked against the automatic results and that several sets of parameters be tried and adjusted before the automatic technique is used on the rest of the data. The method may need to be customized for some studies to allow more or less scatter, and special studies may wish to use hand grading. But this new technique should be valuable for rapid and objective overall assessment of anisotropy and its variation in time.

6. Conclusions

The automatic technique returns results that are nearly identical to manual techniques when used on equivalent datasets. Scatter increases when the automatic technique chooses the best filter and increases more when it is used on data previously judged as poor quality. This scatter is caused mainly by the manual and automatic grading assigning different qualities to the same measurements. Some of the difference in quality may stem from a better ability to reject unsuitable waveform fits by a visual analysis, but some may stem from an unconscious bias in manual grading, which downgrades measurements that do not fit expectations. The new objective analysis confirms changes in the average fast polarization between deep events at Mt. Ruapehu in 1994 and 2002, and between shallow and deep events in 2002.

Acknowledgements

The New Zealand Earthquake Commission provided the bulk of the support for this project. Additional funding came from the Earthquake Research Institute, University of Tokyo, the New Zealand Marsden Fund, the New Zealand Foundation for Research, Science and Technology, and the UK Science and Technology Facilities Council. We acknowledge the New Zealand GeoNet project and its sponsors EQC, GNS Science and FRST, for providing the earthquake locations used in this study. Most figures were prepared with GMT [*P Wessel and Smith*, 1991]. Data were processed in SAC format [*Goldstein and Snoke*, 2005]. The TauP toolkit [*Crotwell et al.*, 1999] was used to calculate the incidence angles.We thank T. Ohminato, E. Smith, J. Townend and R. Arnold for helpful discussions and K. Unglert, J. Johnson, and S. Karalliyadda and K. Araragi for help debugging the codes used for distribution. We also thank Francesca Bianco, an anonymous reviewer and the associate editor for helpful reviews.

References

Balfour, N. J., M. K. Savage, and J. Townend (2005), Stress and crustal anisotropy in Marlborough, New Zealand: Evidence for low fault strength and structure-controlled anisotropy., *Geophys. J. Int.*, *163*, 1073-1086, doi: 1010.1111/j.1365-1246X.2005.02783.x

Bianco, F., and L. Zaccarelli (2009), A reappraisal of shear wave splitting parameters from Italian active volcanic areas through a semiautomatic algorithm, *Journal of Seismology*,doi: 10.1007/s10950-008-9125-z.

Boness, N. L., and M. D. Zoback (2006), A multiscale study of the mechanisms controlling shear velocity anisotropy in the San Andreas Fault Observatory at Depth, *Geophysics*, *71*(5), F131-F146,10.1190/1.2231107.

Crampin, S. (1994), The fracture criticality of crustal rocks, *Geophys. J. Int.*, *118*, 428 - 438

Crampin, S. (1999), Calculable fluid-rock interactions, *J. Geol. Soc.*, *156*, 501–514

Crampin, S., S. Peacock, Y. Gao, and S. Chastin (2004), The scatter of time-delays in shear-wave splitting above small earthquakes, *Geophys. J. Int.*, *156*, 39-44

Crotwell, H. P., T. J. Owens, and J. Ritsema (1999), The TauP Toolkit: flexible seismic travel-time and ray-path utilities, *Seism. Res. Lett.*, *70*, 154–160

Davis, J. C. (1986), Analysis of directional data, in *Statistics and Data Analysis in Geology - Second Edition*, edited, pp. 314-330, John Wiley & Sons, Inc.

Evans, M. S., J.-M. Kendall, and R. J. Willemann (2006), Automated SKS splitting and upper-mantle anisotropy beneath Canadian seismic stations, *Geophys. J. Int.*, *165*(3), 931-942

Fouch, M. J., and S. Rondenay (2006), Seismic anisotropy beneath stable continental interiors, *Phys. Earth Planet. Inter.*, *158*, 292–320

Gao, Y., P. Hao, and S. Crampin (2006), SWAS: a shear-wave analysis system for semi-automatic measurement of shear-wave splitting above small earthquakes, *Phys. Earth Planet. Inter.*, *159*, 71-89

Gerst, A. (2003), Temporal changes in seismic anisotropy as a new eruption forecasting tool?, MSc thesis, 184 pp, Victoria University of Wellington, Wellington.

Gerst, A., and M. K. Savage (2004), Seismic anisotropy beneath Ruapehu Volcano: A possible eruption forecasting tool, *Science*, *306*, 1543-1547

Goldstein, P., and A. W. Snoke (2005), SAC availability for the IRIS community, *Incorporated Research Institutions for Seismology Newsletter*, *7*(1)

Greve, S. M., M. K. Savage, and S. D. Hofmann (2008), Strong variations in seismic anisotropy across the Hikurangi subduction zone, North Island, New Zealand., *Tectonophysics*, *462*(1-4), 7-21,10.1016/j.tecto.2007.07.011.

Hudson, J. A. (1981), Wave speeds and attenuation of elastic waves in material containing cracks, *Geophysical J. R. Aston. Soc*, *64*, 133-150

Johnson, J., S. Prejean, J. Townend, and M. Savage (2010), Anisotropy, repeating earthquakes, and seismicity associated with the 2008 eruption of Okmok Volcano, Alaska, *Journal of Geophysical Research*, *in press*,10.1029/2009JB006991.

Liu, Y., T.-L. Teng, and Y. Ben-Zion (2004), Systematic analysis of shear-wave splitting in the aftershock zone of the 1999 Chi-Chi, Taiwan earthquake: shallow crustal anisotropy and lack of precursory variations, *Bulletin - Seismological Society of America*, *94*(6), 2330-2347

Liu, Y., H. Zhang, C. Thurber, and S. Roecker (2007), Shear wave anisotropy in the crust around the San Andreas fault near Parkfield: spatial and temporal analysis, *Geophys. J. Int.*,doi: 10.1111/j.1365-246X.2007.03618.x.

Mardia, K. V. (1972), *Statistics of Directional Data*, Academic Press, New York and London.

Matcham, I., M. K. Savage, and K. R. Gledhill (2000), Distribution of seismic anisotropy in the subduction zone beneath the Wellington region, New Zealand, *Geophys. J. Int.*, *140*, 1-10

Miller, V. L., and M. K. Savage (2001), Changes in seismic anisotropy after a volcanic eruption: Evidence from Mount Ruapehu, *Science*, *293*, 2231-2233

Peng, Z., and Y. Ben-Zion (2004), Systematic analysis of crustal anisotropy along the Karadere-Duzce branch of the North Anatolian fault, *Geophys. J. Int.*, *159*(1), 253-274

Peng, Z., and Y. Ben-Zion (2005), Spatiotemporal variations of crustal anisotropy from similar events in aftershocks of the 1999 M7. 4 Izmit and M7. 1 Düzce, Turkey, earthquake sequences, *Geophys. J. Int.*, *160*(3), 1027

Roman, D. C., M. K. Savage, S. DeAngelis, and J. L. Latchman (2010), Forward modeling of seismic anisotropy images volcanic conduit inflation months prior to eruption, *J. Geophys. Res.*, *submitted April 2010*

Savage, M. K. (1999), Seismic anisotropy and mantle deformation: what have we learned from shear wave splitting?, *Rev. Geophys.*, *37*(1), 65-106

Savage, M. K., A. F. Sheehan, and A. Lerner-Lam (1996), Shear-wave splitting across the Rocky Mountain Front, *Geophys. Res. Lett.*, *23*(17), 2267-2270

Savage, M. K., M. Duclos, and K. Marson-Pidgeon (2007), Seismic Anisotropy in the South Island, New Zealand, in *A Continental Plate Boundary: Tectonics at South Island, New Zealand*,175 edited by D. Okaya, T. Stern and F. Davey, pp. 95-116, Am. Geophys. Union Geophys. Monograph Series.

Savage, M. K., T. Ohkura, K. Umakoshi, H. Shimizu, Y. Kohno, M. Iguchi, A. Wessel, and J. V. Mori (2008), Variations in Seismic Anisotropy with time on Volcanoes in Kyushu Island, Southern Japan, *Eos Trans. AGU, 89* (53), S53A-1813

Savage., M. K., T. Ohminato, Y. Aoki, H. Tsuji, and S. Greve (2010), Absolute stress and its temporal variation at Mt. Asama Volcano, Japan, from seismic anisotropy and GPS, *Earth Planet. Sci. Lett.*, *290*(3-4), 403-414,10.1016/j.epsl.2009.12.037.

Silver, P. G., and W. W. Chan (1991), Shear wave splitting and subcontinental mantle deformation, *J. Geophys. Res.*, *96*, 16429–16454

Teanby, N., J.-M. Kendall, and M. van der Baan (2004a), Automation of shear-wave splitting measurements using cluster analysis, *Bull. Seism. Soc. Amer*, *94*, 453-463

Teanby, N., J. M. Kendall, R. H. Jones, and O. Barkved (2004b), Stress-induced temporal variations in seismic anisotropy observed in microseismic data, *Geophys. J. Int.*, *156*, 459-466

Walker, K., G. Bokelmann, and S. Klemperer (2004), Shear-wave splitting beneath the Snake River Plainsuggests a mantle upwelling beneath eastern Nevada, USA, *Earth Planet. Sci. Lett.*, *222*, 529-542

Wessel, A. (2008), Automatic shear wave splitting measurements at Mt. Ruapehu volcano, New Zealand, MSc thesis, 206 pp, Victoria Univ. Wellington, Wellington.

Wessel, P., and W. H. F. Smith (1991), Free software helps map and display data, *EOS, Trans. Am. Geophys. Union*, *72*(441), 445-446

Wüstefeld, A., and G. Bokelmann (2007), Null Detection in Shear-Wave Splitting Measurements, *Bull. Seismol. Soc. Am.*, *97*(4), 1204-1211

Zinke, J. C., and M. D. Zoback (2000), Structure-related and stress-induced shear-wave velocity anisotropy: observations from microearthquakes near the Calaveras Fault in central California, *Bull. Seismol. Soc. Am.*, *90*(5), 1305-1312

Figure Captions.

Figure 1. Flowchart of data processing steps. Dashed lines are optional steps.

Figure 2. Comparison of waveforms after the application of different bandpass filters. Although a change in frequency content is visible on the top trace (raw data), the *S*-wave is masked by long period noise. The application of a bandpass filter (BP) emphasizes the signal, but the narrow 2-3 Hz filter appears "ringy" and is susceptible to cycle skipping. (a) small event (M=3.8) recorded at station LHOR. The best filters as measured by the maximum of the product of the SNR and the bandwidth (fb1 through fb3) are at high frequencies, but the best frequency still has a narrow 1-Hz bandwidth. (b) larger event (M=4.2) , which has better response at long periods.

Figure 3. High quality, A grade (Table 2) measurement recorded at station LHOR for a regional event. The grey boxes in panels (a), (b) and (e) delineate the time window used for the final measurement. (a) filtered East (E) North (N) and vertical (Z) waveforms. The solid line is the *S* arrival. The dashed lines are the minimum start (1) and maximum end (4) times for windows used in the processing, as in (b) . (b) the waveforms rotated into the SC91-determined incoming polarization direction (p) and its perpendicular value ($p^{\perp}$), for the original filtered waveform (top) and the waveforms corrected for the SC91-determined *dt* (bottom) for the window shown in grey. The straight black line is the *S* arrival. The two sets of dashed lines on either side of the straight line show the range of allowed starting (1 and 2) and ending (3 and 4) windows for the SC91 measurements. (c) ϕ and *dt* determined for each measurement window as a function of window number. (d) all the clusters of 5 or more measurements, with the large cross being the chosen cluster. (e) waveforms (top) and particle motion (bottom) for the original (left) and corrected (right) waveform according to the final

chosen SC91 window. (f) contours of the smallest eigenvalue of the covariance matrix for the final chosen SC91 measurement.

Figure 4. Sample C quality measurement, as in Figure 3. This sample presents good waveform fits and has a high signal-to-noise ratio for the best measurements, but other windows with qualities that are not much different have ϕ that vary by tens of degrees, so the measurement may exhibit cycle skipping. We do not analyse this grade of measurement.

Figure 5. Null measurement: an unsplit shear wave shows a lack of energy on the component perpendicular to the initial polarization.

Figure 6. Histograms of dominant frequencies in the measurement windows used in the analyses. (a) defining the best filters by using only the maximum SNR, (b) defining the best filters by comparing the product of the SNR and the filter bandwidth. (c) The raw, unfiltered data.

Figure 7. Comparison of manual vs. automatic measurements of fast polarizations (a) and delay times (b) for all stations, using the same filters in both studies (the filters used are those determined by Gerst & Savage (2004) qualitatively to yield the best measurements), and showing results using different quality criteria in each box (Table 2; 3). Note that wrap-around effects make the measurement at (80, -85) appear to be far apart, while in fact they differ by less than 15º. The headings are defined in Table 2.

Figure 8. Histograms showing the distribution of differences between ϕ for the manual and automatic measurements in Figure 7. The histograms all have the same scale.

Figure 9. Comparison of 2002 measurements from automatic and manual methods, allowing the program to choose the filters based on the SNR and the set of 14 filters that are

shown in Table 1. Only the “AB” and “AB eng8” sets (Table 2; 3) results are plotted. Again, most of the measurements lie close to the 1-1 line. (top) comparison of ϕ values. (middle) comparison of *dt* values. (bottom) histograms of the difference between the automatic and manual measurements of *dt*, all using the same scale.

Figure 10. Scatter plots of measurements from the single best filter compared to the manual measurements, for no culling, using the “AB’ criterion, and the “AB eng8” criterion. (top) comparison of ϕ values. The number of measurements in each plot is given in the upper left hand corner. (bottom) comparison of *dt* values.

Figure 11. Rose diagrams of manual and automatic results compared at each station. Plots on left are manual, middle are automatic based on using filters fitting the SNR criteria alone and plots on right are automatic using the SNR-bandwidth product to determine which filters are best. Top is for events with depth < 35 km. Bottom is for events with depth > 55 km. Up to three different filters per event are included. This is to account for the relatively sparse data set.

Figure 12. Same as Figure 11, but for 1994 deep and 1998 shallow results.

Table 1. Filters tested

Filter number	Low freq (Hz)	High freq (Hz)	Bandwidth (octave)
1	0.4	4	5
2	0.5	5	5
3	0.2	3	7.5
4	0.3	3	5
5	0.5	4	4
6	0.6	3	2.5
7	0.8	6	3.75
8	1	3	1.5
9	1	5	2.5
10	1	8	4
11	2	3	0.75
12	2	6	1.5
13	3	8	1.3
14	4	10	1.25

Table2. Quality Criteria

Grade name	Criterion
N (null)	If the fast polarisation f is between -20 to 20 or 70 to 110 degrees of the incoming polarization
Dcl	Cluster D grade: If there is any cluster k for which the following holds: *nmeas(k)*>*Nmeas(kbest)*/2 and *var(k)* < 5 var(kbest) and also: (*tdiff(k)*> *tlagmax*/4 or (π/4 < *phidiff(k)* < 3π/4))
Ccl	Cluster C grade: If the cluster is not D grade and there is any cluster i for which the following holds: nmeas(k)>Nmeas(kbest)/2 and var(k) < 5 var(kbest) and also: *tdiff(k)* > *tlagmax*/8 or π/8 < *phidiff(k)* < 7π/8.
Bcl	Cluster B grade: If the cluster is not grade D or C and there is any cluster *k* for which the following holds: *var(k)* < 5**var(kbest)* and *nmeas(k)* > *Ncmin* (5 here) and also: *tdiff(k)* > *tlagmax*/8 or π/8 < *phidiff(k)* < 7π/8
Acl	Cluster A grade: If the cluster is not grade D or C or B
ABPAR	based on parameters alone: not null, *dt* < 0.8**tlagmax*, SNR>3, $d\phi$<25
APAR	based on parameters alone: *dt* < 0.8**tlagmax, SNR*> 4, $d\phi$ < 10, where $d\phi$ = standard deviation of ϕ
B	Cluster B, not null, *dt* < 0.8**tlagmax*, *SNR*>3, $d\phi$<25
A	Cluster A, *dt* < 0.8**tlagmax*, *SNR* > 4, $d\phi$ < 10
AB	A or B
Eng8	As described in text, maximum value of contour energy plots is greater than 8

Table 3. Comparison of Measurements for different quality criteria.

Method of culling	Total number of events	Proportion Of total (%)	Number within 25 deg	Proportion within 25 deg(%)	Number within 0.2 s	Proportion within 0.2 s (%)
Compare same filters						
All meas	212	100	158	75	181	85
NoN (No Null meas.)	179	84	157	88	166	93
NoD (No meas. with DCl)	178	84	151	85	163	92
NoN,NoD	160	75	142	89	151	94
NoN, dt<0.8	175	83	153	87	164	94
NoN, dt<0.8,SNR>3	169	80	146	86	157	93
NoN, dt<0.8,SNR>3, $d\phi$<25	169	80	146	86	157	93
NoN, dt<0.8,SNR>4, $d\phi$<10	137	65	123	90	134	98
AB	145	68	131	90	138	95
A	110	52	101	92	107	97
ABeng8	88	42	87	99	87	99
Aeng8	75	35	74	98	75	100
Compare different filters						
All filts (3 best)	77	100	67	87	72	94
AB (3 best)	86	112	73	85	80	93
AB eng 8	56	73	52	93	54	96
fb1	116	100	93	80	96	83
fb1 AB	80	69	74	93	76	95
fb1 AB eng8	33	28	32	97	33	100

Table 4. Average results from the 1994 and 1998 and 2002 deployments (all stations)

Data set	Method	ϕ [°]	± ϕ [°]	δt [s]	±δt [s]	|R|	#
1994 deep	Automatic-up to 3 filters ABeng8	-29	4	0.24	0.02	0.65	44
	Automatic best meas of 3 filters	-25	6	0.26	0.03	0.46	43
	Auto eng8 then best of 3	-25	7	0.26	0.04	0.54	26
	Manual	-42.8	3.6	0.23	0.021	0.707	37
1998 shallow	Automatic up to 3 filters ABeng8	-13	11	0.13	0.009	0.2	79
	Automatic AB best meas of 3 filters	-18	14	0.16	0.01	0.14	101
	Auto eng8 then best of 3	-5.2	16	0.14	0.01	0.17	49
	Manual	13.4	5.8	0.11	0.0093	0.516	39
2002 deep	Automatic up to 3 filters ABeng8	10	6	0.24	0.02	0.4	62
	Auto AB best of 3 filters	26	10	0.29	0.02	0.23	72
	Auto eng8 then best of 3	10	7	0.24	0.025	0.43	37
	Manual	19.2	2.7	0.272	0.016	0.63	117
2002 shallow	Automatic up to 3 filters ABeng8	-17	5	0.12	0.008	0.41	100
	Auto AB best of 3 filters	-16	6	0.14	0.009	0.35	100
	Auto eng8 then best of 3	-16	7	0.13	0.01	0.37	58
	Manual	-30	2.4	0.107	0.0048	0.69	123

Table 5. Comparison of manual and automatic measurements at station LHUT2 [i].

EQ ID	filter (Hz)	*dt* (s)	ϕ (°)	Auto grade	pick grade	Gerst *dt*	Gerst ϕ	Gerst grade	new grade
2002.126.20	0.5-5	0.06	-85	BCl	1	0.05	-71	B	1
2002.126.20	0.1-2					0.1	24	A	
2002.125.17	0.1-3	0.1	22	ACl	2			X	1
2002.125.17	2.0.6-3	0.1	26	ACl	2	0.1	31	A	1
2002.132.04	0.2-2					0.1	23	nullA	
2002.132.04	0.4-4	0.31	38	ACl	1	0.5	39	AB	1
2002.132.04	1-3					0.5	33	C	
2002.015.09	0.4-4	0.3	59	ACl	1	0.28	54	AB	1
2002.015.09	0.5-5	0.28	60	ACl	1			X	1
2002.016.12	0.3-3	0.28	65	ACl	1	0.1	39	AB	1
2002.073.02	0.1-8	0.14	-85	ACl	1	0.05	-88	AB	1
2002.073.02	1-5	0.14	-81	ACl	1			x	1

[i]All measurements for deep events with automatic grades A or B and energy contour greater than 8 are included. Manual measurements are reported in the same line as the closest automatic measurement. If there were more manual measurements than automatic for an event or vice versa, an extra line is included with the single measurement. Measurements are sorted by fast polarization. Column Auto Grade gives the grades determined by the automatic grading described in Table 3. Column Pick grade is the grade given during manual picking of the S wave and represents how well the analyst thought the S wave arrival time was determined (1=good, 2=fair). Columns labeled "Gerst" were grades determined by Gerst [2003] (X=not used). Column labeled "new grade" is for grades determined by Teanby before he had read an initial version of the manuscript and was invited to be an author, so we consider him to have been an independent observer.

Appendix

Obtaining and using the programs

The codes for this method have mainly been written in BASH syntax, SAC scripts [*Goldstein and Snoke*, 2005], Fortran and GMT [*P Wessel and Smith*, 1991] and all necessary programs are freely available. The programs and a detailed technical manual can be found online (http://mfast-package.geo.vuw.ac.nz/).

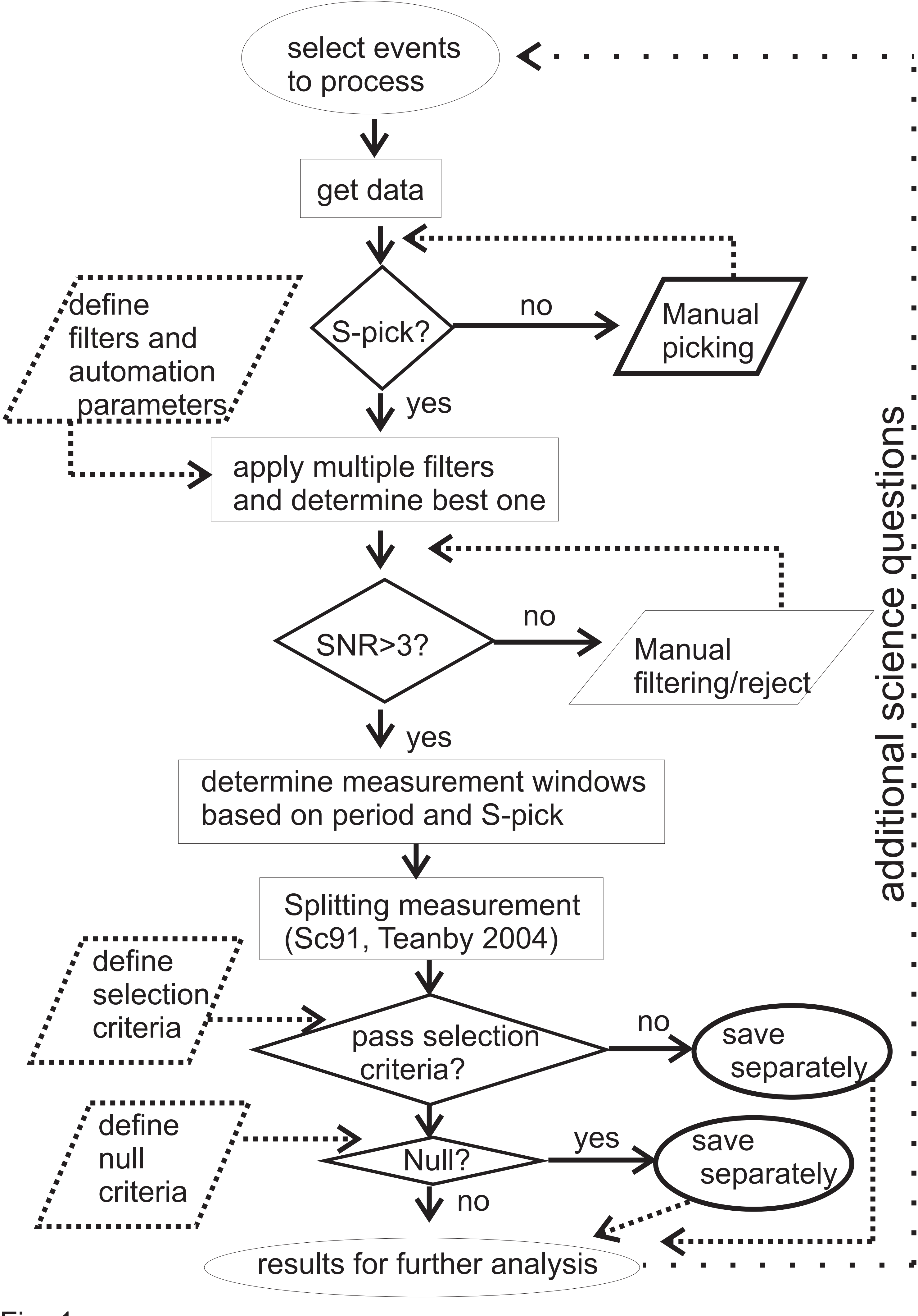

select events
to process
get data
define
filters and
automation
parameters
S-pick?
no
Manual
picking
yes
apply multiple filters
and determine best one
SNR>3?
no
Manual
filtering/reject
yes
determine measurement windows
based on period and S-pick
Splitting measurement
(Sc91, Teanby 2004)
define
selection
criteria
pass selection
criteria?
no
save
separately
define
null
criteria
Null?
yes
save
separately
no
results for further analysis
additional science questions

Fig. 1

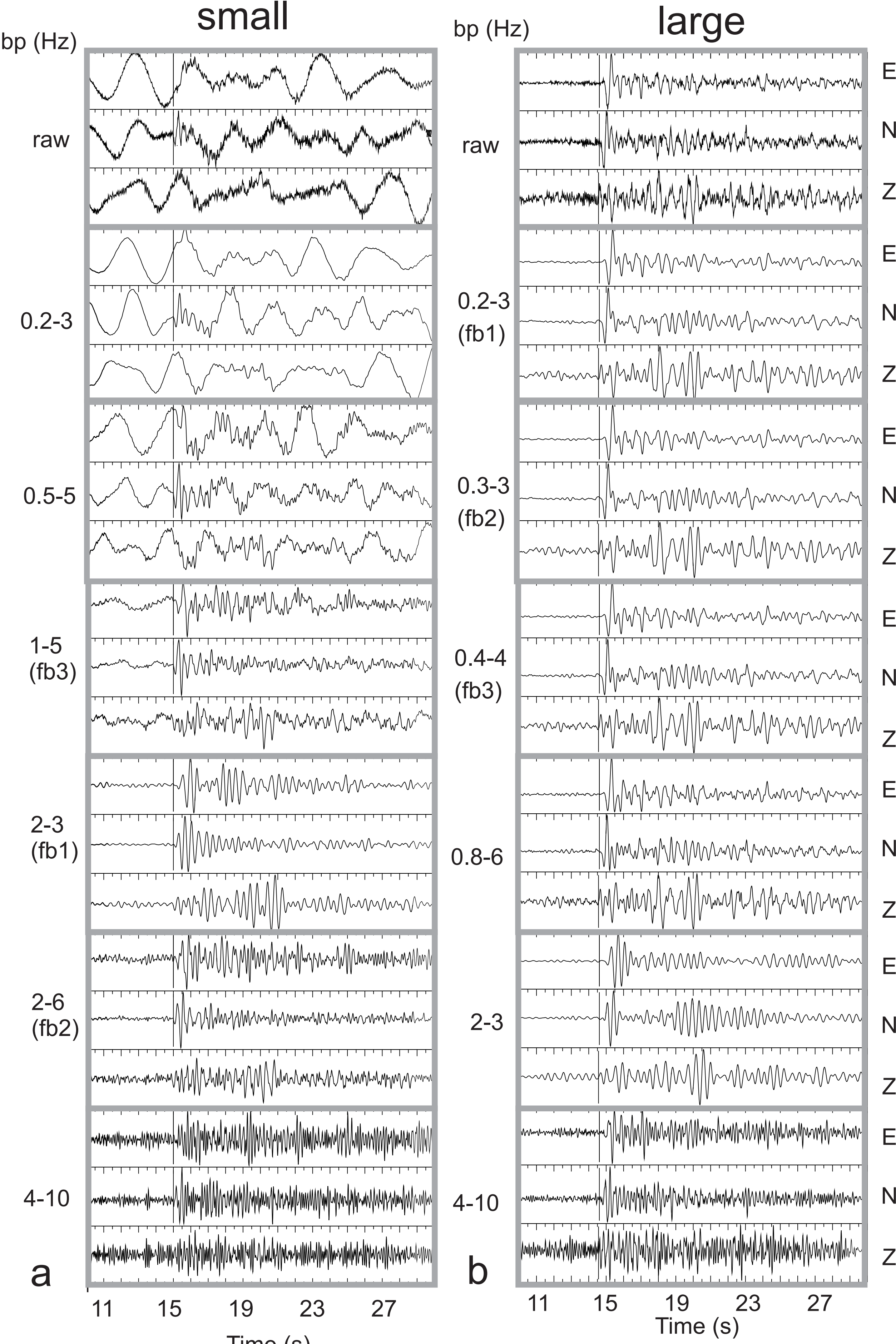

small
large
bp (Hz)
bp (Hz)
raw
0.2-3
0.5-5
1-5
(fb3)
2-3
(fb1)
2-6
(fb2)
4-10
raw
0.2-3
(fb1)
0.3-3
(fb2)
0.4-4
(fb3)
0.8-6
2-3
4-10
E
N
Z
a
b
11
15
19
23
27
Time (s)

Fig. 2

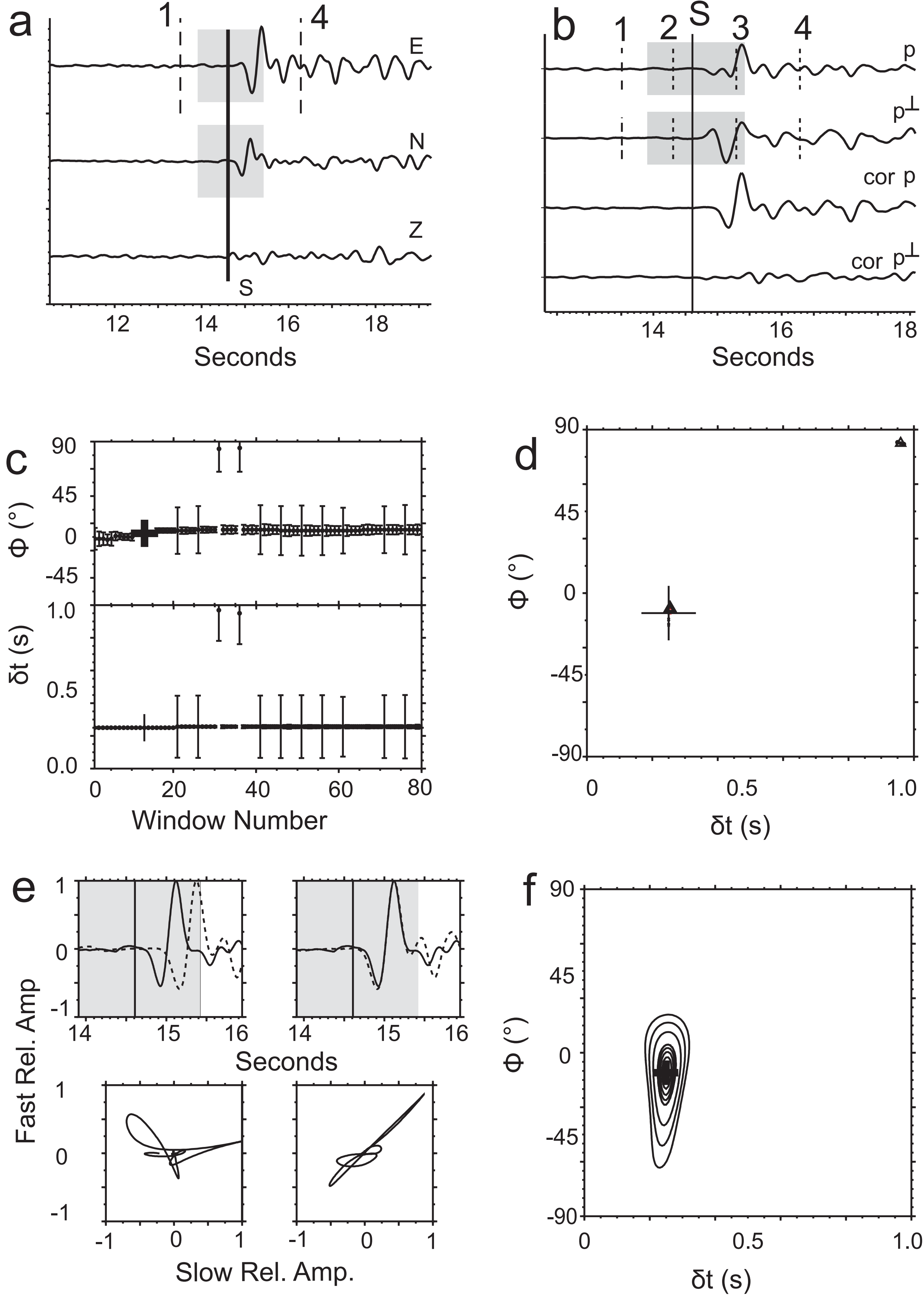
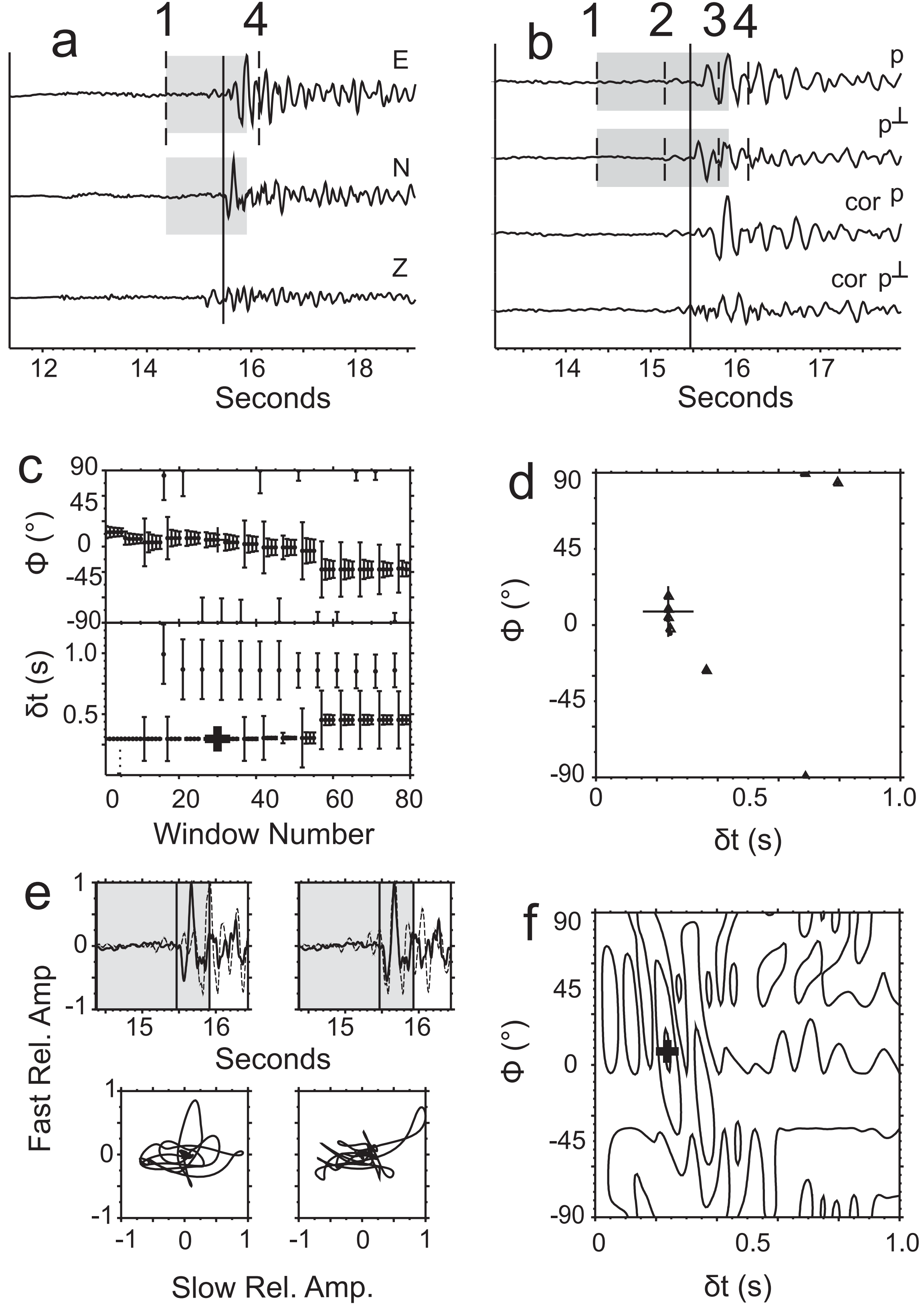
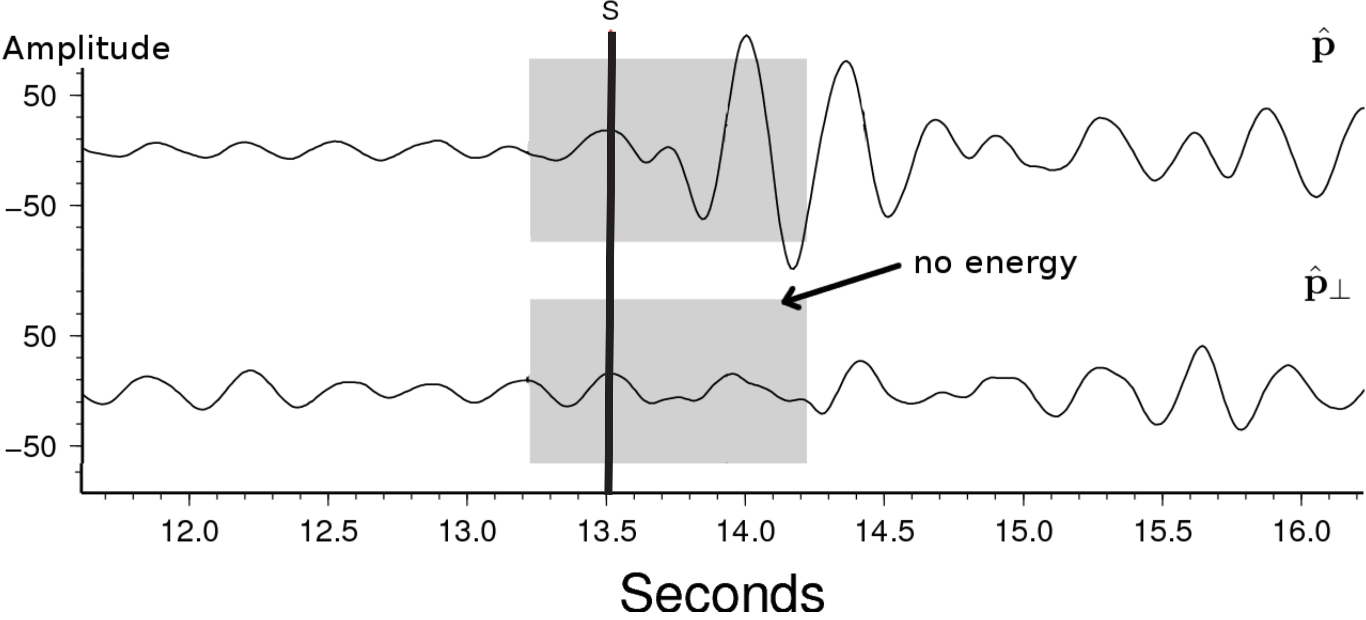

a
1
4
E
N
Z
S
12
14
16
18
Seconds
b
1
2
S
3
4
p
p⊥
cor p
cor p⊥
14
16
18
Seconds
c
90
45
0
-45
Φ (°)
1.0
0.5
0.0
δt (s)
0
20
40
60
80
Window Number
d
90
45
0
-45
-90
Φ (°)
0
0.5
1.0
δt (s)
e
1
0
-1
14
15
16
Seconds
Fast Rel. Amp
Slow Rel. Amp.
f
90
45
0
-45
-90
Φ (°)
0
0.5
1.0
δt (s)

Fig. 3



a
1
4
E
N
Z
12
14
16
18
Seconds
b
1
2
3
4
p
p⊥
cor p
cor p⊥
14
15
16
17
Seconds
c
90
45
0
-45
-90
Φ (°)
1.0
0.5
δt (s)
0
20
40
60
80
Window Number
d
90
45
0
-45
-90
Φ (°)
0
0.5
1.0
δt (s)
e
1
0
-1
15
16
15
16
Seconds
Fast Rel. Amp
1
0
-1
-1
0
1
-1
0
1
Slow Rel. Amp.
f
90
45
0
-45
-90
Φ (°)
0
0.5
1.0
δt (s)

Fig. 4



S
Amplitude
$\hat{\mathbf{p}}$
50
−50
no energy
$\hat{\mathbf{p}}_{\perp}$
50
−50
12.0
12.5
13.0
13.5
14.0
14.5
15.0
15.5
16.0
Seconds

Fig. 5

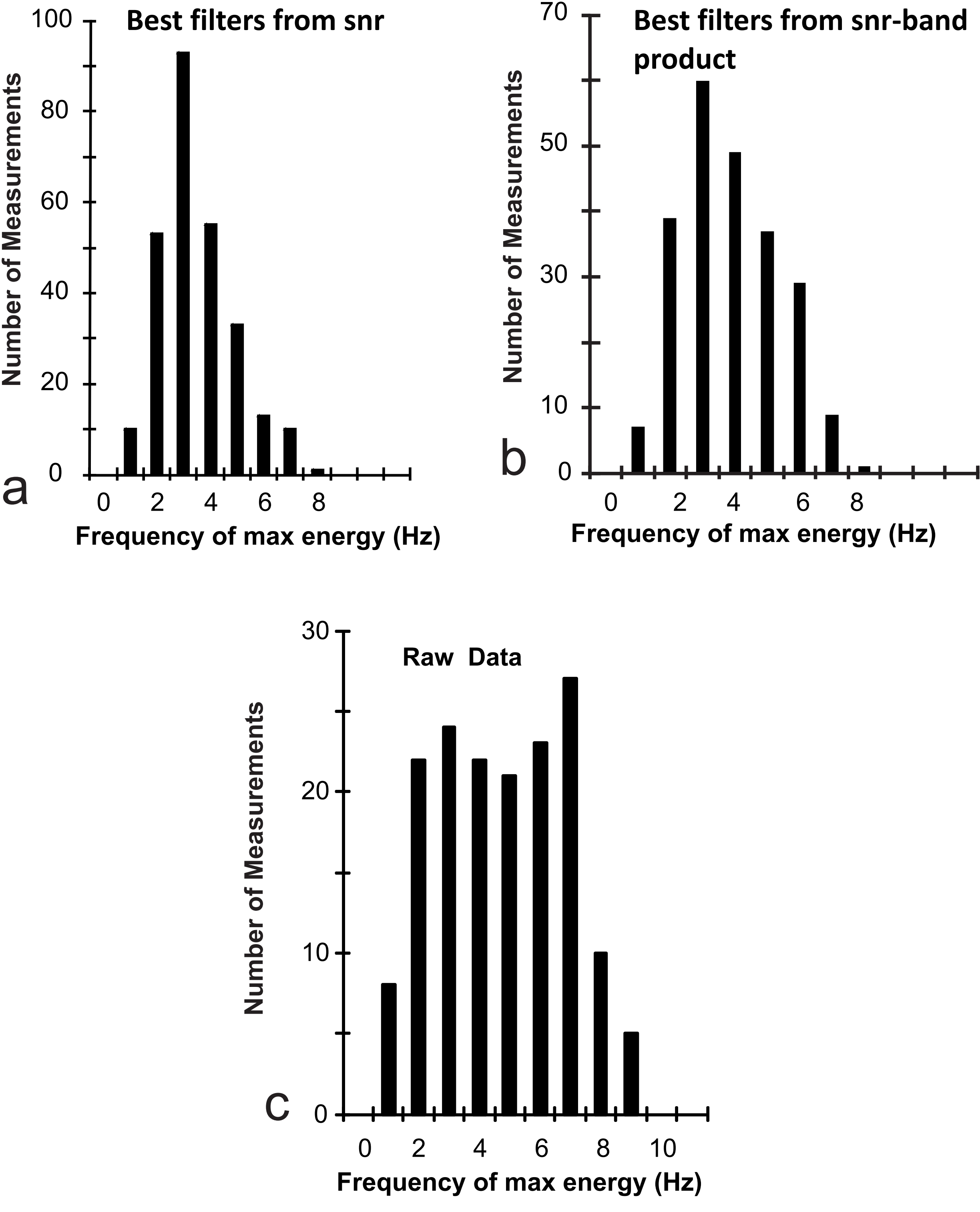

Best filters from snr
Number of Measurements
100
80
60
40
20
0
a
0 2 4 6 8
Frequency of max energy (Hz)
Best filters from snr-band product
Number of Measurements
70
50
30
10
0
b
0 2 4 6 8
Frequency of max energy (Hz)
Raw Data
Number of Measurements
30
20
10
0
c
0 2 4 6 8 10
Frequency of max energy (Hz)

Fig. 6

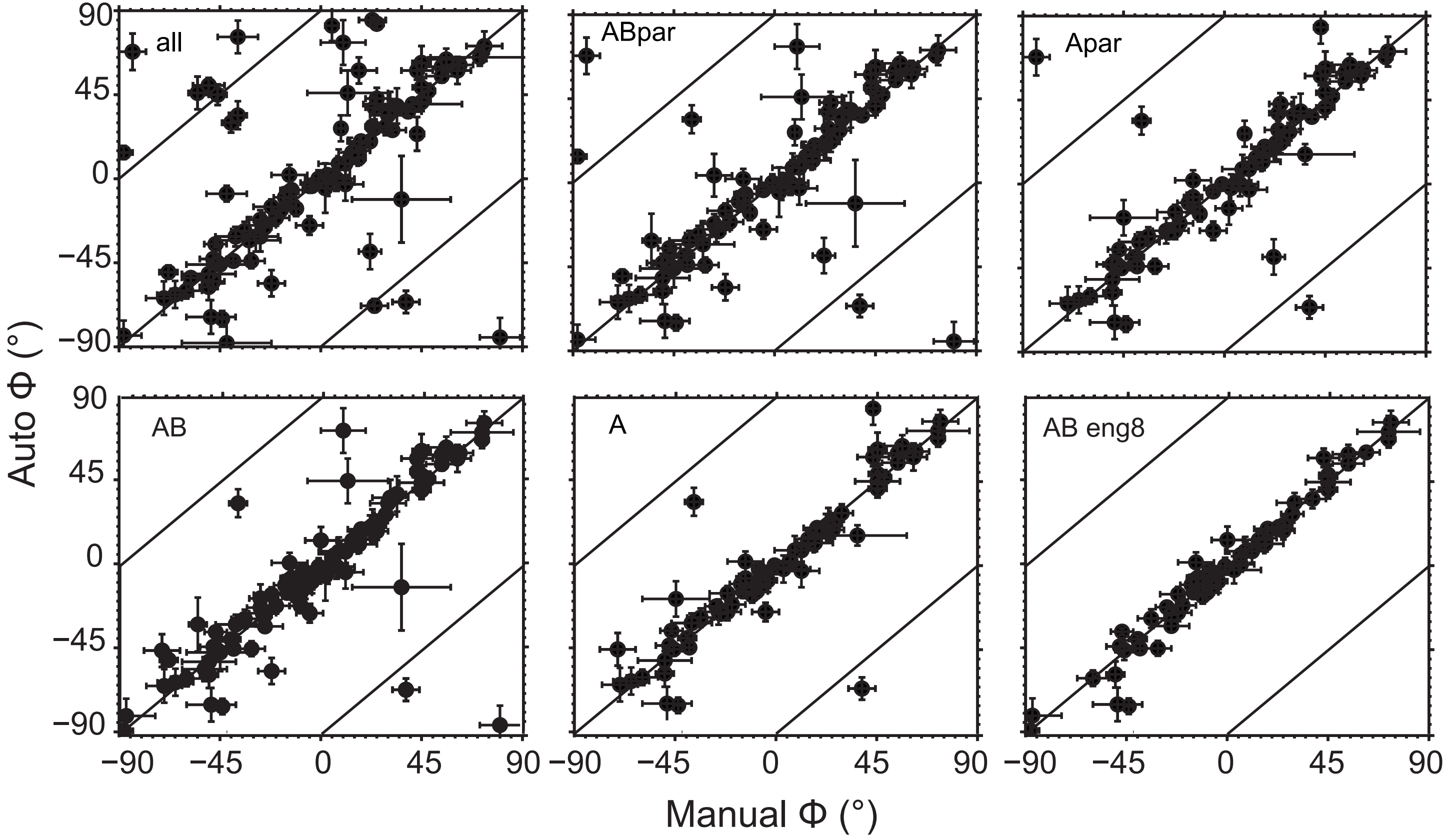

all
ABpar
Apar
AB
A
AB eng8
90
45
0
−45
−90
−90
−45
0
45
90
Auto Φ (°)
Manual Φ (°)

Fig. 7a

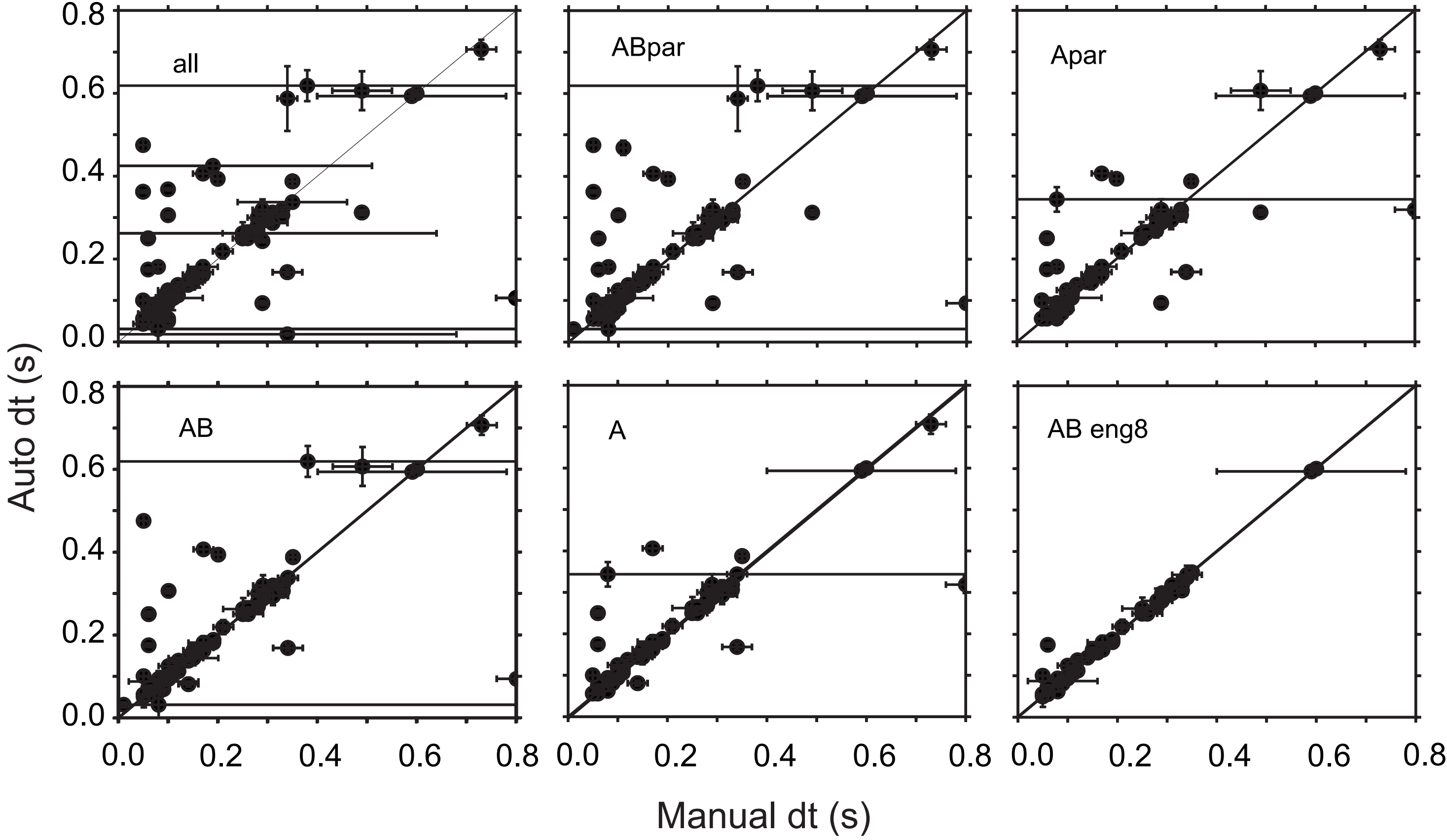


Fig. 7b

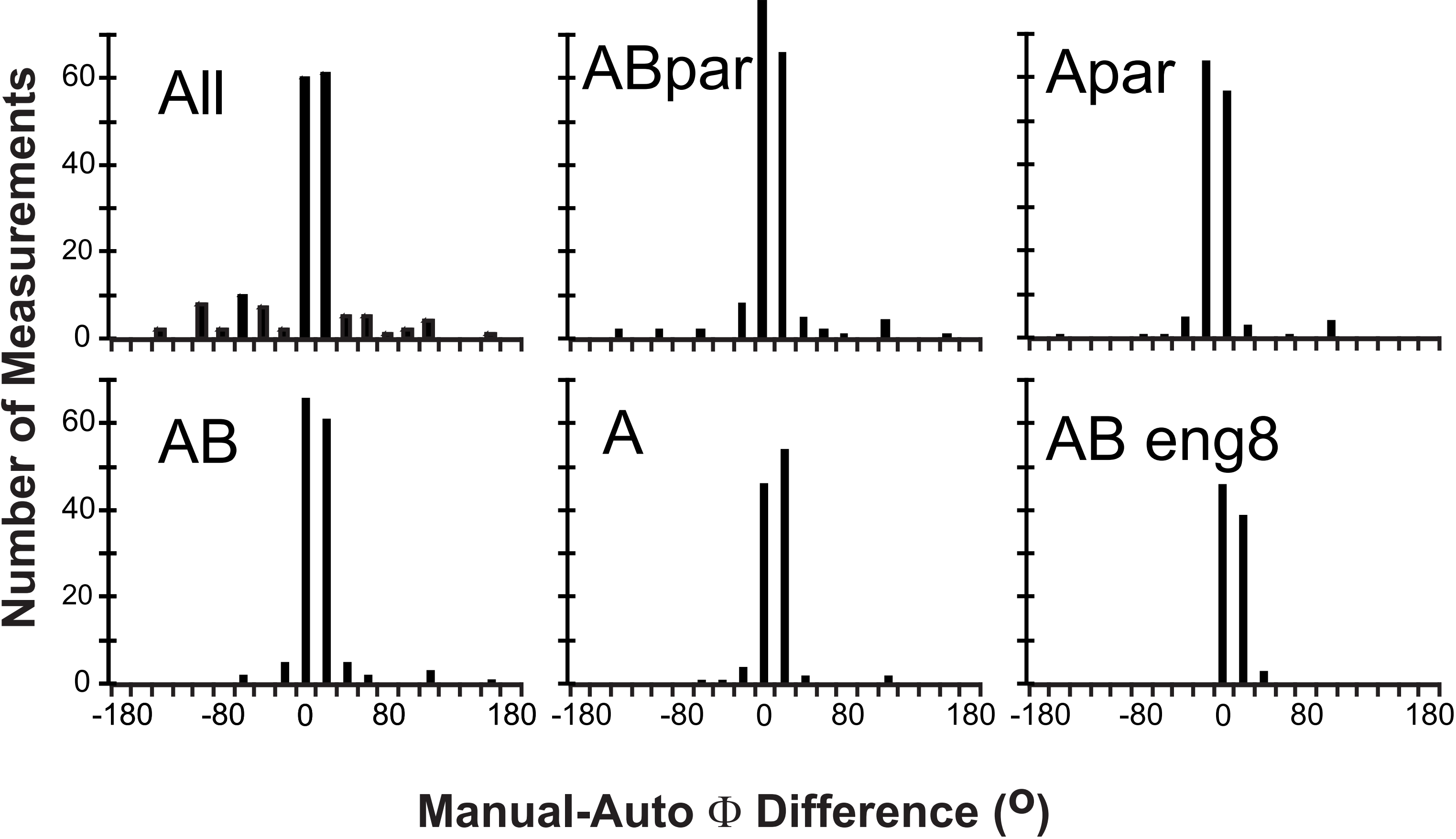
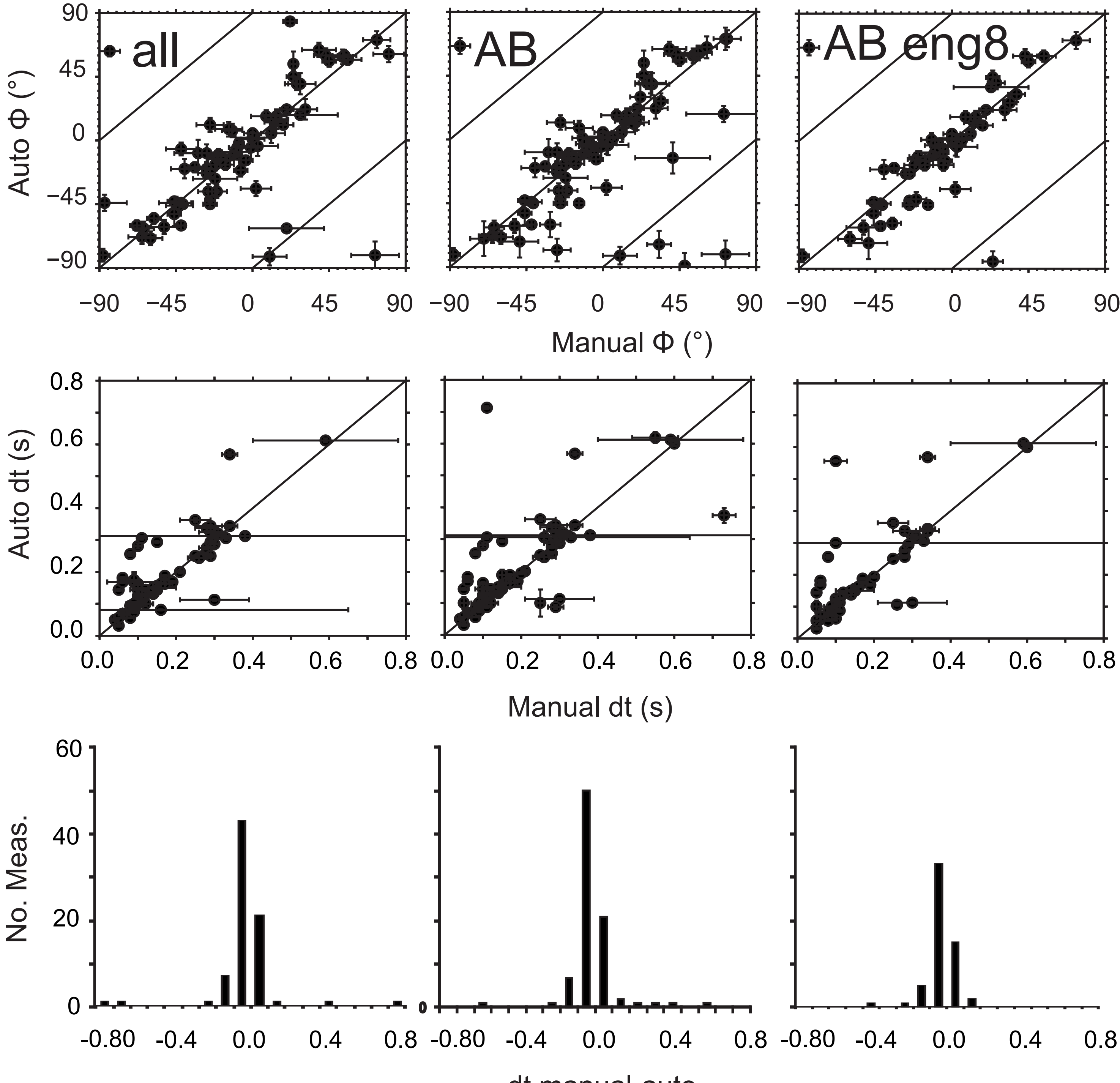
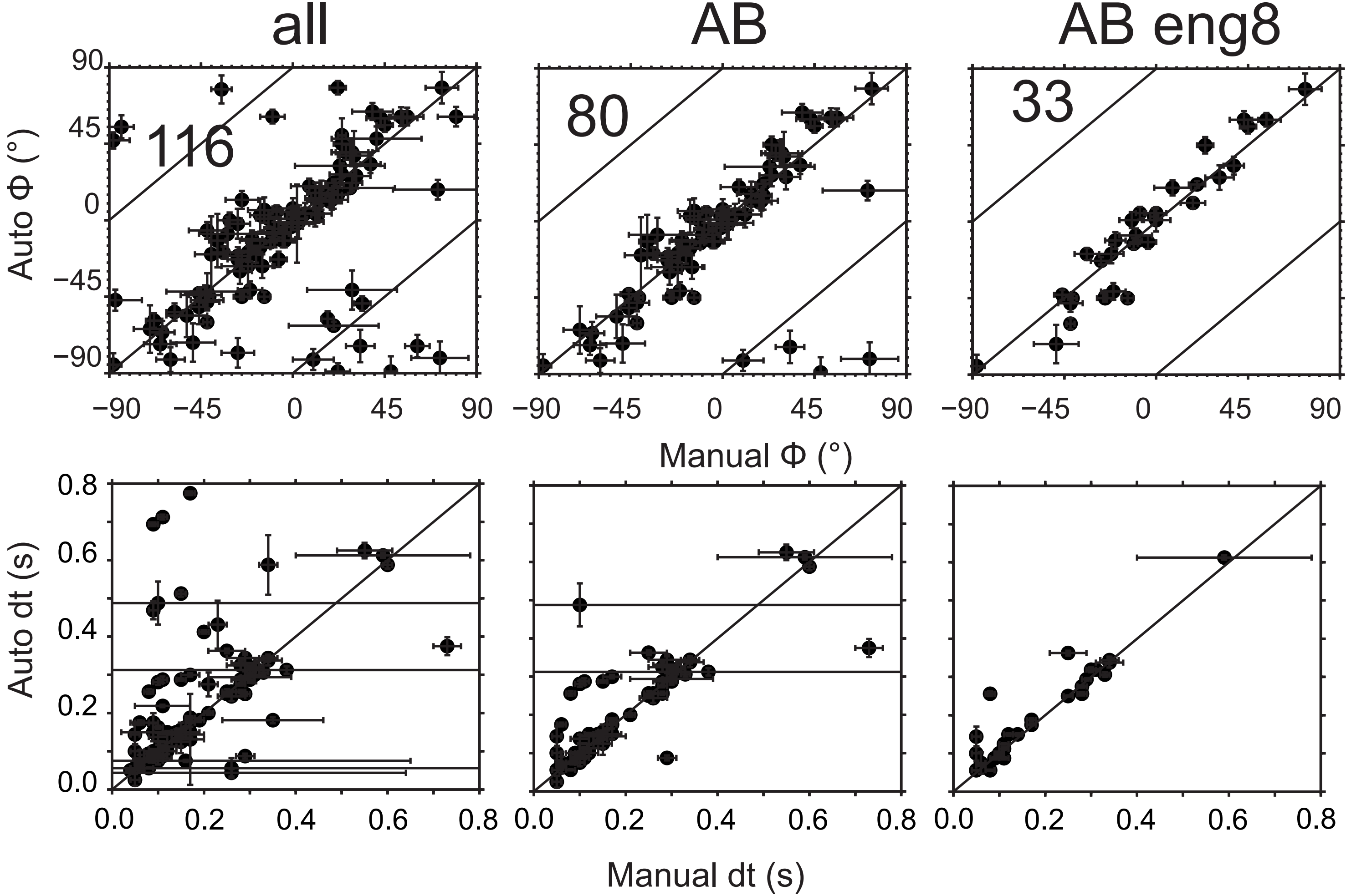

Number of Measurements
All
ABpar
Apar
AB
A
AB eng8
0
20
40
60
-180
-80
0
80
180
Manual-Auto Φ Difference (o)

Fig. 8



all
AB
AB eng8
Auto Φ (°)
Manual Φ (°)
90
45
0
−45
−90
Auto dt (s)
Manual dt (s)
0.0
0.2
0.4
0.6
0.8
No. Meas.
60
40
20
0
-0.80
-0.4
0.0
0.4
0.8
dt manual-auto

Figure 9.



all
AB
AB eng8
116
80
33
Auto Φ (°)
Manual Φ (°)
Auto dt (s)
Manual dt (s)
90
45
0
−45
−90
0.8
0.6
0.4
0.2
0.0

Fig. 10

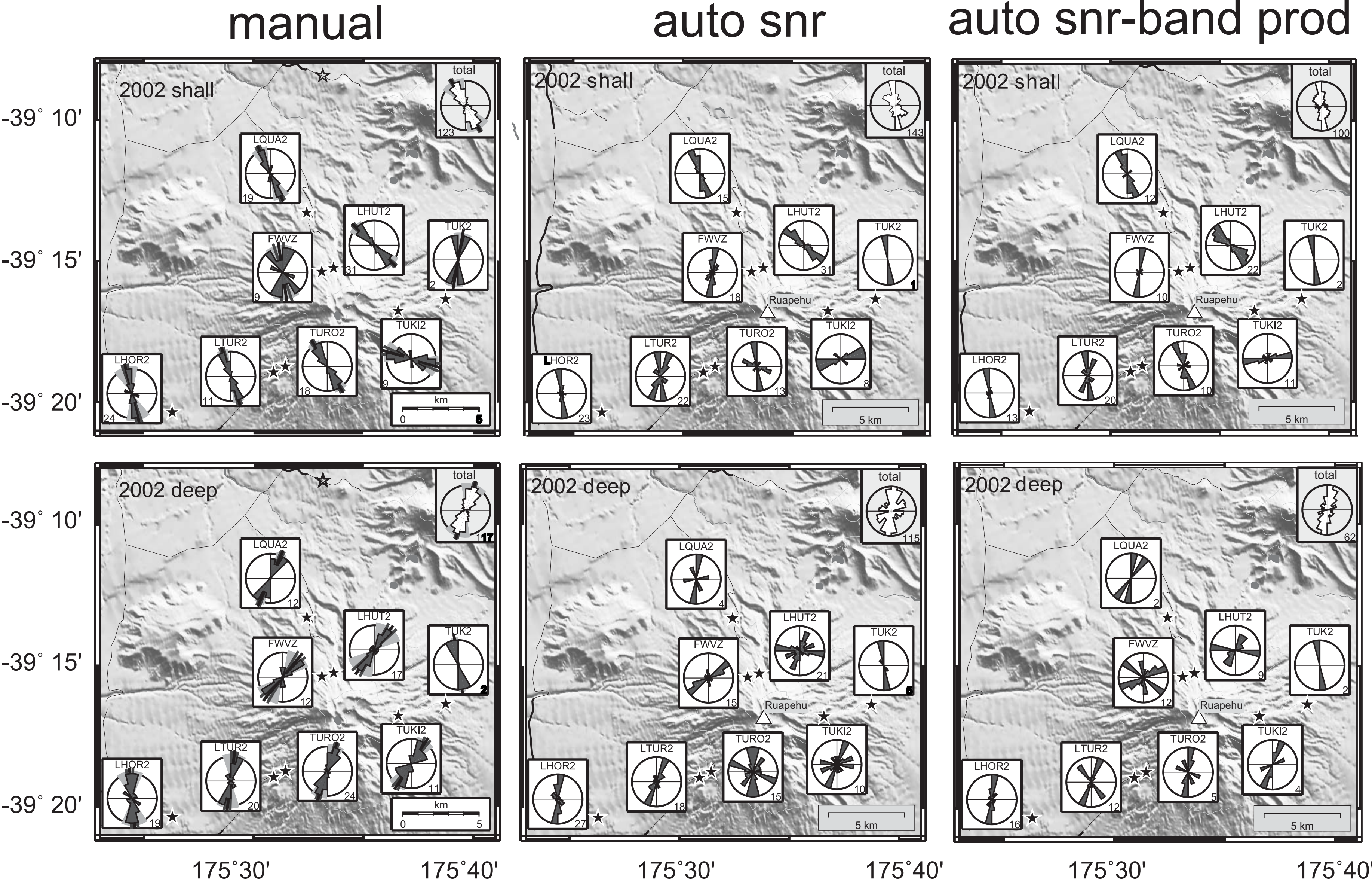
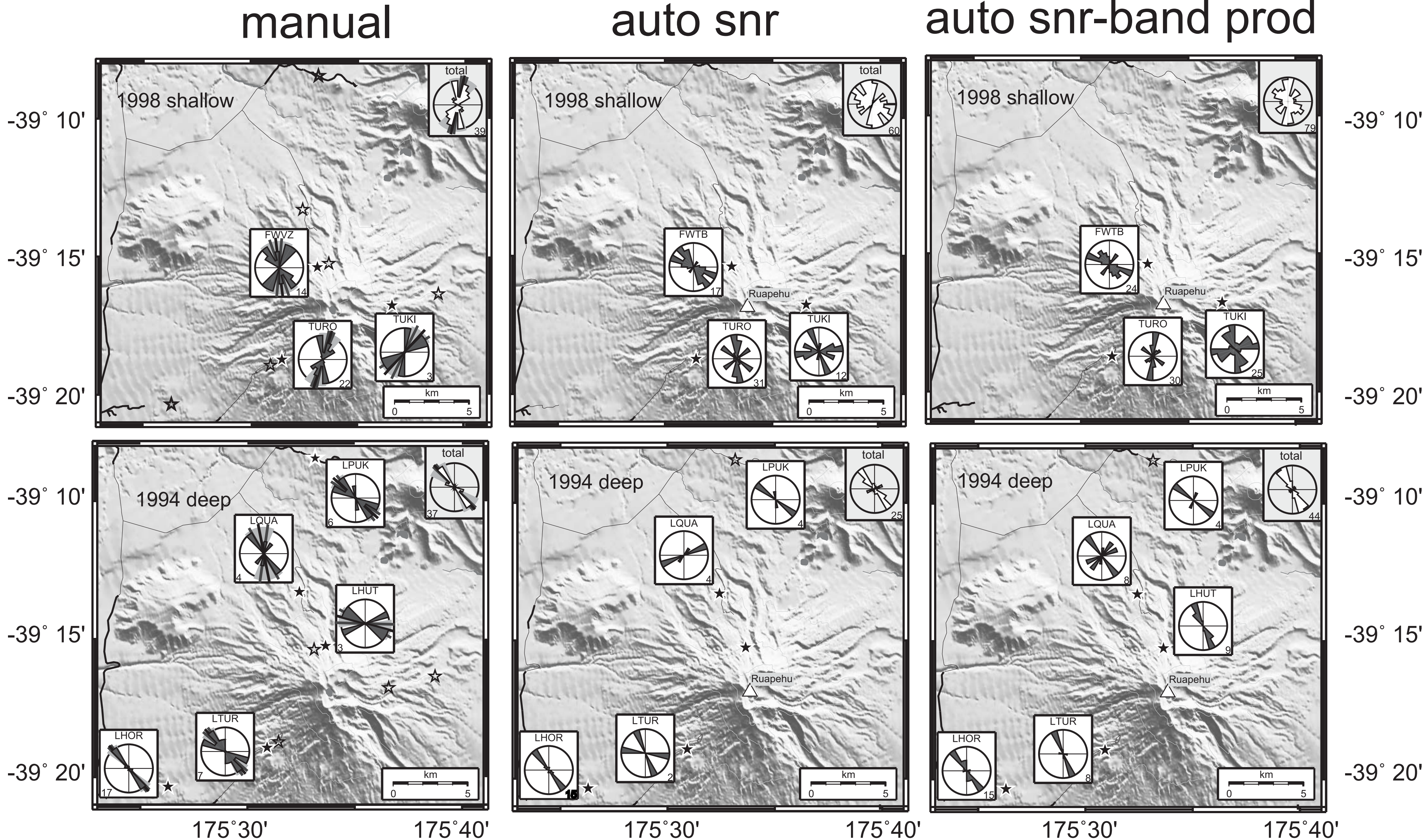

manual
auto snr
auto snr-band prod
2002 shall
2002 deep
total
LQUA2
LHUT2
FWVZ
TUK2
TUKI2
TURO2
LTUR2
LHOR2
Ruapehu
km
5 km
-39˚ 10'
-39˚ 15'
-39˚ 20'
175˚30'
175˚40'

Fig. 11



manual
auto snr
auto snr-band prod
1998 shallow
1994 deep
total
FWVZ
TURO
TUKI
FWTB
Ruapehu
LPUK
LQUA
LHUT
LHOR
LTUR
km
0
5
-39° 10'
-39° 15'
-39° 20'
175°30'
175°40'

Fig 12.