An automatized XKS-splitting procedure for large data sets: Extension package for SplitRacer and application to the USArray

Frederik Link [*], Miriam Christina Reiss, Georg Rümpker

Institute of Geosciences, Goethe-University Frankfurt, Frankfurt, Germany

A B S T R A C T

Recent technological advances have led to community wide use of large-scale seismic experiments which produce seismic data on previously impossible scales. Standard processing procedures thus require automatization to facilitate a fast and objective analysis of the data. Among these, XKS-splitting is an important tool to derive first insights into the Earth's deformation regimes at depth by studying seismic anisotropy. Most often, shear-wave splitting is interpreted to represent crystallographic preferred orientation (CPO) of mantle minerals like olivine as dominating feature and can thus be used as a proxy of mantle flow processes. Here, we introduce an addition to the MATLAB®-based SplitRacer tool box (Reiss and Rümpker 2017) which automatizes the entire XKS-splitting procedure. This is achieved by the automatization of 1) choosing a time window based on spectral analyses and 2) categorization of results based on three different XKS-splitting methods (energy minimization, rotation correlation and splitting intensity). This provides effective and objective results for splitting as well as null-measurement results. This extension allows to use SplitRacer without a graphical interface and introduces a bootstrapping statistics as error estimate of the single layer joint splitting method. The procedures are designed to allow a fast and more objective analysis of a vast amount of data, as produced by recent seismic deployments (e.g. USArray, AlpArray). We test this automatization by applying the analysis to the USArray data set, which has approximately 1900 stations with between two to fifteen years of data. We can reproduce the general pattern of the results from former studies with the more objective automatic analysis. Based on a joint-splitting approach, we approximate the splitting effect at individual stations by a single anisotropic layer. As we include null-measurements as well as a larger data set as previous studies, we can provide improved statistical evidence for these effective splitting parameters.

1. Introduction

In modern seismology, seismic anisotropy is often used to study the complex structure and internal processes of the Earth, as it is mostly related to deformation processes or structural properties at depth (Savage 1999). Intrinsic anisotropic minerals, e.g., Olivine, can be aligned into direction of strain. This causes large-scale features in the mantle and provides important insight into deformation processes (Karato et al., 2008). The bulk anisotropy produces characteristic features in seismic phases, which results most prominently in a split of shear waves. The split phases show perpendicular polarizations with the fast-travelling phase polarized parallel to the preferred orientation of the (mantle) minerals or shape preferred orientation caused by, e.g., aligned cracks or melt-filled lenses. The lag time of the phases recorded at the receiver (referred to as splitting time in the following) scales with the

strength of anisotropy and the extend of the anisotropic volume. Usually, core-mantle converted S-phases like PKS, PKIKS, SKS, SKKS and others, here summarized as XKS-phases, are used for the analysis of the splitting, as the S-wave leg converted at the core-mantle boundary is considered to be polarized in the radial direction (Silver and Chan 1991).

The analysis is commonly based on the inverse splitting operator introduced by Silver and Chan (1991), which removes the anisotropic effect by testing all possible parameter combinations of the fast polarization and splitting time minimizing the transverse energy or the second eigenvalue of the covariance matrix of the particle motion. Bowman and Ando (1987) introduced a rotation correlation method which is based on searching for the maximum cross-correlation coefficient for the trial fast polarization and derives the splitting time from the time lag of the maximum of the cross correlation. Another possible approach to

characterize the splitting of shear waves is based on the observation that the transverse component is proportional to the time derivative of the radial component if the dominant period is significantly larger than the delay time (Chevrot 2000). Splitting can then be characterized by a single property called the splitting intensity. However, the characteristic azimuthal variations of the conventional splitting parameters due to layered anisotropy are obscured when using splitting intensity (e.g., Kong et al., 2015).

There are a number of software tools to process and analyze shear wave splitting of XKS-phases which differ in their approaches from manual to semi-automatic and automatic processing (Reiss and Rümpker 2017; Silver and Chan 1991; Wüstefeld et al. 2008, 2010; Walther et al., 2014; Teanby et al., 2004; Liu and Gao 2013; Savage et al., 2010). Besides the specific analysis method, key tasks for any splitting code are the selection of appropriate phases and the time window for the analysis as well as the classification of the results. While automatization provides the potential to objectify the processing and categorization of measurements (Savage et al., 2010) and allows for fast processing of huge data sets (Walther et al., 2014), it can lead to an increase of rejected waveforms in comparison to a manual approach (Evans et al., 2006; Wüstefeld et al., 2010). Thus, identifying and characterizing the most appropriate time window and phase as well as categorizing the results retaining a maximum possible number of events are the main challenges for a successful automatization.

In the community, there has been quite an effort to minimize the influence of analysts' bias. SplitLab (Wüstefeld et al., 2008) which is the most widely used teleseismic splitting code, facilitates the splitting analysis with a graphical user interface in the MATLAB environment. While the authors first focused on a manual procedure, they later introduced a comparison of results from two independent analysis techniques to estimate the quality of the measurements (Wüstefeld et al., 2010). A new extension for SplitLab also allows a joint analysis of different events measured at a station by stacking of individual error surfaces (Grund 2017), which stabilizes the inversion for single layer anisotropy.

Furthermore, Walther et al. (2014) provide a window selection procedure by identifying the onset of the core-mantle converted phase with results from an fk-analysis. The optimum window is then chosen from a moving window of different length for a number of filters.

Walker and Wookey (2012) developed a MATLAB based toolbox to model splitting based on the elastic tensors in two anisotropic layers. This allows to introduce additional effects from dipping symmetry axes and realistic anisotropic mineral properties. However, the splitting parameters have to be derived separately and can then be compared with the modelling.

The application of the splitting analysis to local S-waves raises additional challenges (compared to teleseismic events), due to the unknown orientation of the polarization which depends on the focal mechanism and the frequency content. Teanby et al. (2004) propose a cluster search for randomly applied time windows of different length to automatize and improve objectivity of the window selection process. Savage et al. (2010) introduce an automatic procedure for local events, that requires only the manual pick of the S-phase. The window selection is based on the method of Teanby et al. (2004), while the length depends on the dominant frequency of the event. They introduce a quality assignment of the splitting results that is based on different criteria, but mainly on the statistical scatter of the results, either from the window selection or the stability of the results based on the shape of the error surface. Most null-measurements are discarded in the cluster comparison. However, remaining null-measurements are identified with a second criterion, which compares the polarization and fast axis direction. A recent publication by Spingos et al. (2020) combines manual and automatic procedures for analyzing shear wave splitting of local earthquakes in a user-friendly graphical interface for Python.

Here, we focus on the automatization of the SplitRacer code (Reiss and Rümpker 2017), which combines a user-friendly interface with the

possibility to adjust the processing and analysis criteria to the specific needs of an advanced user. We combine a time-frequency analysis with a random variation of the time windows to select the optimum window for the analysis. We also employ the splitting intensity to identify null-measurements, while further quality criteria are determined based on parameters describing the stability and quality of the energy minimization technique (Silver and Chan 1991). To test the applicability of the automatized procedures, we apply our new approach to the USArray data set (Transportable Array, TA) and compare our results with previous studies (Liu 2009; Refayee et al., 2014; Yang et al. 2014, 2017; Liu et al., 2014; Liu and Gao 2013).

2. Method

2.1. Summary SplitRacer

SplitRacer is an open source software for the MATLAB environment (MATLAB 2020) and is available for the Linux, Mac and Windows operating systems (Reiss and Rümpker 2017). It is easily operated using a graphical user interface and allows the user to download data from different sources using the FDSN web services. The pre-processing consists of three stages: 1) check for gaps in the data, 2) calculation of travel times for XKS and other main phases, 3) calculation of the signal to noise ratio (SNR) and elimination of data below a user-set threshold. This is followed by the quality check, in which the user manually checks (and alters) the automatically selected time window for the splitting analysis and discards unusable events. As a major advantage, sensor misalignment is calculated from the long-period particle motion (Rümpker and Silver 1998). Remaining events and phases are then analyzed using the transverse energy minimization method (Silver and Chan 1991), which yields a fast polarization and delay time for each waveform. The analysis in repeated for a predefined number of time windows, which are selected randomly around the preset times picked by the user. The splitting time and fast polarization are derived from the minimum of the stacked transverse-energy grids and the errors are calculated as proposed by Walsh et al. (2013). Following the analysis, the user is required to manually evaluate the waveforms before and after correction for the effect of anisotropy and to assign the quality categories "null", "good", "average" or "poor". The program provides the possibility to present the individual results in an overview and fit one- or two-layer models. The results can further be improved by applying a joint-splitting analysis. For this, the splitting analysis is performed simultaneously for all events by summing up the value of the error surfaces of all events for each possible splitting parameter combination. This suppresses the error of single measurements, incorporates their ambiguities and allows to include "null"-measurements when searching for a one- or two-layer model.

2.2. Extension SplitRacerAUTO – automatization

The new extension for SplitRacer follows the same procedure as the main code, maintaining a clear structure, while giving access to a full automatization of the whole process and optimizing it for the application to a vast amount of data.

The main differences to the previous version are: 1) automatic selection of the time window for the XKS phase, 2) automatic quality check, 3) automatic categorization of the results, 4) command-line approach. We will discuss the major changes in detail below. Ultimately, no manual input of the user is required during each step of the analysis. However, this new version is fully compatible with the current GUI SplitRacer version and can be used interchangeably at any point of the analysis (see also Fig. 1).

2.3. Automatized selection of time windows

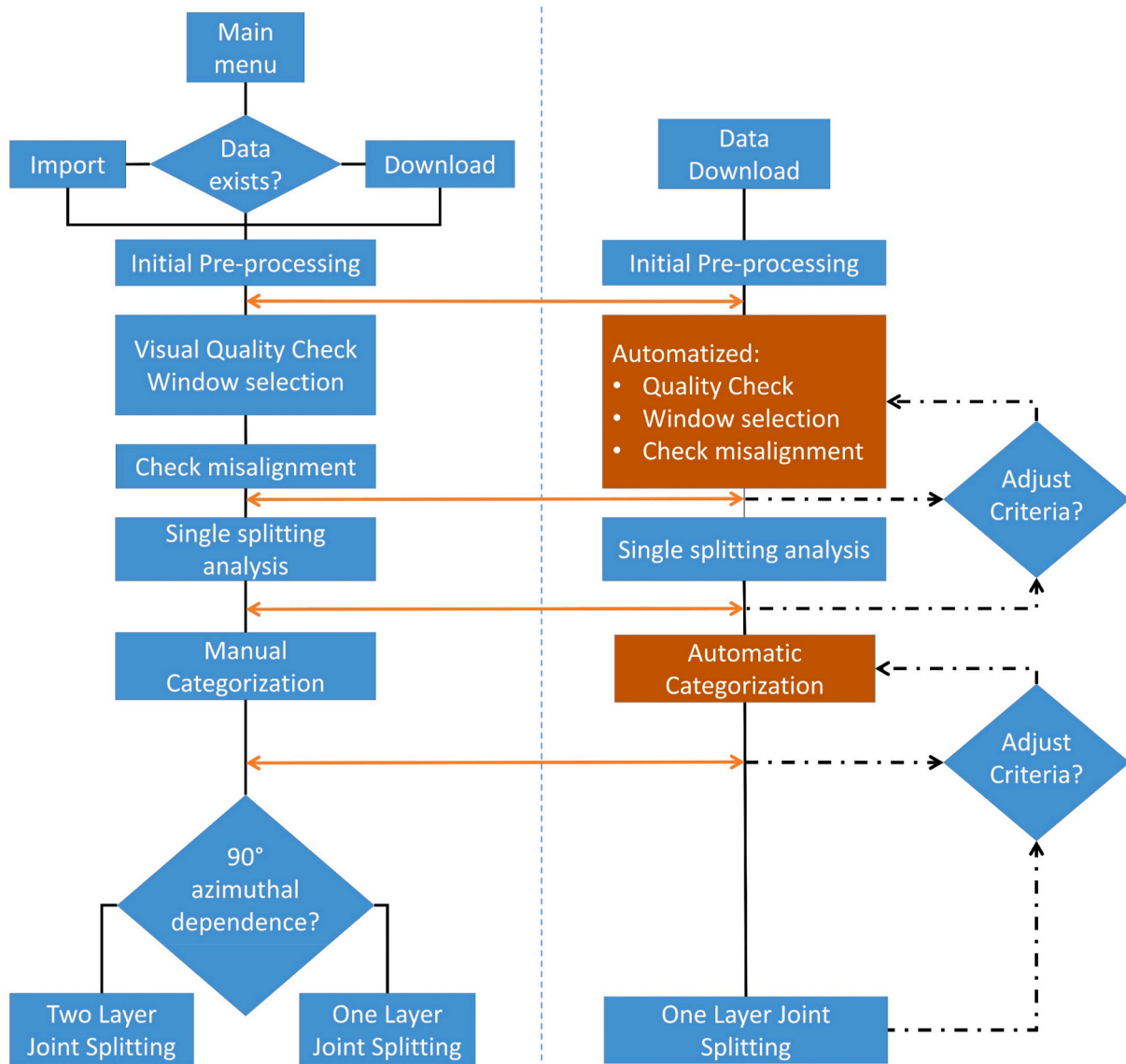To identify the ideal time-window containing an XKS-phase, we

Fig. 1. *Workflow of SplitRacer (left) and the extension for automatization (right). Main operations are presented as blocks; diamonds and polygons indicate two or more possible options in the workflow. The orange-coloured blocks highlight the new automatized steps in the extension. Black solid lines show the regular flow path between the main processes, while dotted lines show optional steps and additional circles and red solid lines mark sequences where intermediate products are interchangeable between the GUI-based SplitRacer program and the extension for automatization. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)*

perform a time dependent spectral analysis in an expected time range. The initial time window for the spectral analysis is chosen ± 50 s around the expected arrival time of a core-mantle phase, which is derived from the IASP91 velocity model (Kennett and Engdahl 1991). To ensure that the selection procedure is not contaminated by different phases (e.g., S, ScS), the time window is shortened accordingly, if those would arrive within the initial time frame. We assume that the appearance of a clear and analyzable XKS-phase produces a significant change of the spectral energy on the radial component, which persists for the duration of the wavelet. A spectrogram is produced using a discrete short-time Fourier transform (STFT; e.g., Quatieri 2006).

The selection process of the waveform used in the later analysis is shown in Fig. 2. We identify the frequency content of the dominant wavelet in the combined spectrum of the radial and transverse STFT and produce a trace of stacked energy from this frequency range. The start time, t_{start}, and end time, t_{end}, of the resulting window is defined by the crossing of its 50% peak level. We extend the window for a time interval depending on the dominant period, $T + 2.5\ s$, to recover the true length

of the wavelet and to account for the effect of the taper window.

The traces are discarded if the start or end times of the determined window are lying outside the original time frame as this indicates that there is no significant wavelet to be analyzed for the expected phase or that there is contamination of different phases (e.g. S, ScS). The trace is also discarded if the dominant frequency range reaches one of the edges of the frequency band as it indicates a superimposed dominant noise source. A stronger criterion, which is always applied, to identify noise dominated signals is the short-time average over long-time average ratio (e.g., Allen 1978; Trnkoczy 2009). A short-time average over 20 s (STA) and long-time average over 50 s (LTA) is computed to allow a signal quality estimation (see Fig. 2e). If the resulting time window shows a maximum STA/LTA-ratio less than 2.1, it is discarded as a lower ratio indicates that no significant signal is present, or that noise is considerably disturbing the signal. This empirical threshold has proven to consistently identify acceptable signals in our application.
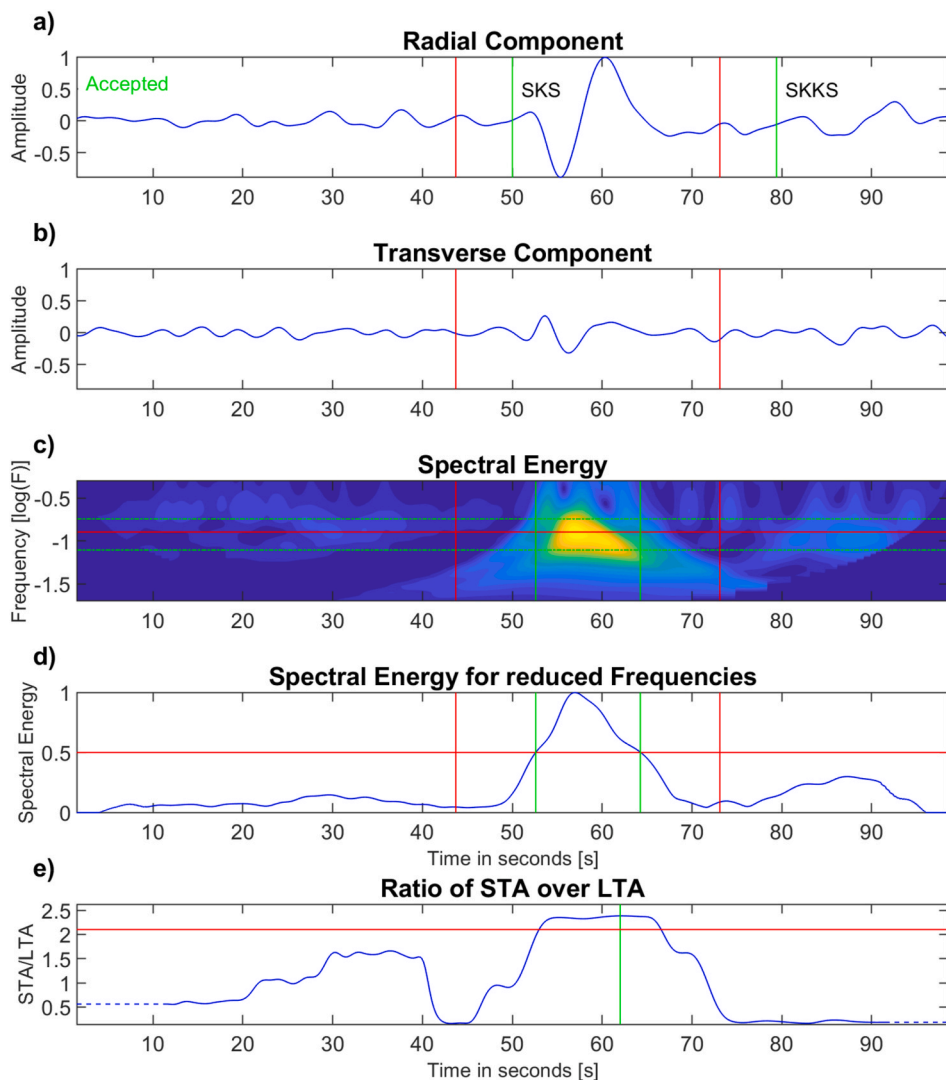
Fig. 2. Time-Spectral analysis for an exemplary event at station T25A in the network TA (as provided to the user). a) and b) radial and transverse component of the recorded waveforms. The green lines mark the theoretical arrival times of seismic phases corresponding to the event of interest. Red vertical lines mark the start and end time of the final window containing the waveform. "Accepted" indicates an accepted waveform. c) summed spectral energy of the time-spectral analysis of the radial and transverse components. The red horizontal line marks the frequency with maximum spectral energy. The green dashed lines show the maximum and minimum frequencies for an appearance of at least 80% of the maximum spectral energy. d) stacked time series over all frequencies within the 80% energy level. (c-d) The green solid lines mark the minimum and maximum time, where the summed spectral energy exceed 50% of the maximum energy. e) quality estimate for the signal with the STA/LTA-ratio. The green vertical line marks the maximum STA/LTA-ratio within the found time-window. The red horizontal line marks the threshold of 2.1 for which the event is accepted or discarded. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

2.4. Modifications in the splitting procedure

The splitting analysis in SplitRacer is based on the energy minimization after applying the inverse splitting operator in a grid search for the delay time δt and fast polarization direction φ (Silver and Chan 1991). Previous studies have shown that a comparison of splitting results derived with different approaches is useful to determine the quality of the analysis (Wüstefeld et al., 2010). However, the main advantage of introducing independent approaches to derive splitting parameters is an improved quality estimate in the automated categorization process.

Accordingly, we introduce the rotation correlation method searching for the largest correlation coefficient in a trial rotation from 0° to 180° fast orientation. The thus found maximum relates to the fast/slow axis orientation, while the time lag of the correlation coefficient provides the splitting time which is positive for a found fast axis orientation and negative for the slow case (Bowman and Ando 1987). The third approach provides a measure for the splitting present in a particular radial and transverse waveform. The so-called splitting intensity (Chevrot 2000; was added to SplitRacer in a 2019 release, and first used in Reiss et al., 2019) only retains information about the fast axis orientation in its sign and can be recovered from the phase of the overall cosine move-out for multiple measurements of different backazimuths. The main result is still derived from the energy minimization method, while the user can get an estimate for a one-layer model using the splitting intensity for a visual comparison using the GUI-based version of

SplitRacer.

2.5. Automatized categorization

The quality of the splitting measurements is usually defined by four categories which are assigned manually by a visual check of the waveforms before and after correction of the apparent splitting:

- "null" for linear particle motion and a clearly visible waveform on the original components
- "good" for a clear split of the waveform with visible energy on the transverse component of the original seismogram and a linearized particle motion after correction
- "average" for a clear split of the waveform, where a higher noise level is allowed, with energy on the transverse component of the original seismogram and fairly linearized particle motion after correction
- "poor" for XKS-waveforms with insufficient correction and/or no possible linearization after correction

For a full automatization, these categories have to be assigned using objective criteria. We divide the problem of the categorization into two steps. First, the null-measurements are identified, then, the quality of the remaining splitting measurements are assigned according to characteristics derived from the three different methods mentioned above.

2.5.1. Assigning null-measurements

The energy minimization method cannot be used directly to identify negligible splitting or null-measurements. The error surface (transverse energy plot) leads to highly ambiguous range of possible splitting parameters. On the contrary, the absolute value of the splitting intensity preserves the information of minor/no splitting for the same set of waveforms. Therefore, when taking the (ideally minimal) ellipticity of the original waveform into account to exclude traces with high noise contamination, the splitting intensity is a strong indicator for waveforms with negligible splitting, with no visual inspection required.

To determine suitable parameter combinations and thresholds for the categorization we examine a manually categorized data set of an arbitrarily chosen exemplary station (T25A) in the network TA. As

described, we rely on the absolute splitting intensity SI and the ellipticity of the original particle motion as sufficient parameters to assign a waveform as null-measurement. Following Vidale (1986), the ellipticity e is determined from the ratio of eigenvalues of the covariance matrix of the original waveform

$$e = \frac{\lambda_1}{\lambda_2}. \qquad\qquad 1$$

We are searching for limits of ellipticity and absolute splitting intensity, which isolate the null-measurements from the non-null measurements. For this purpose, we calculate two fractions. The first fraction R_{null} is defined by the number of null-measurements n_{null} within a trial set of the parameters e and SI incorporating possible limits over
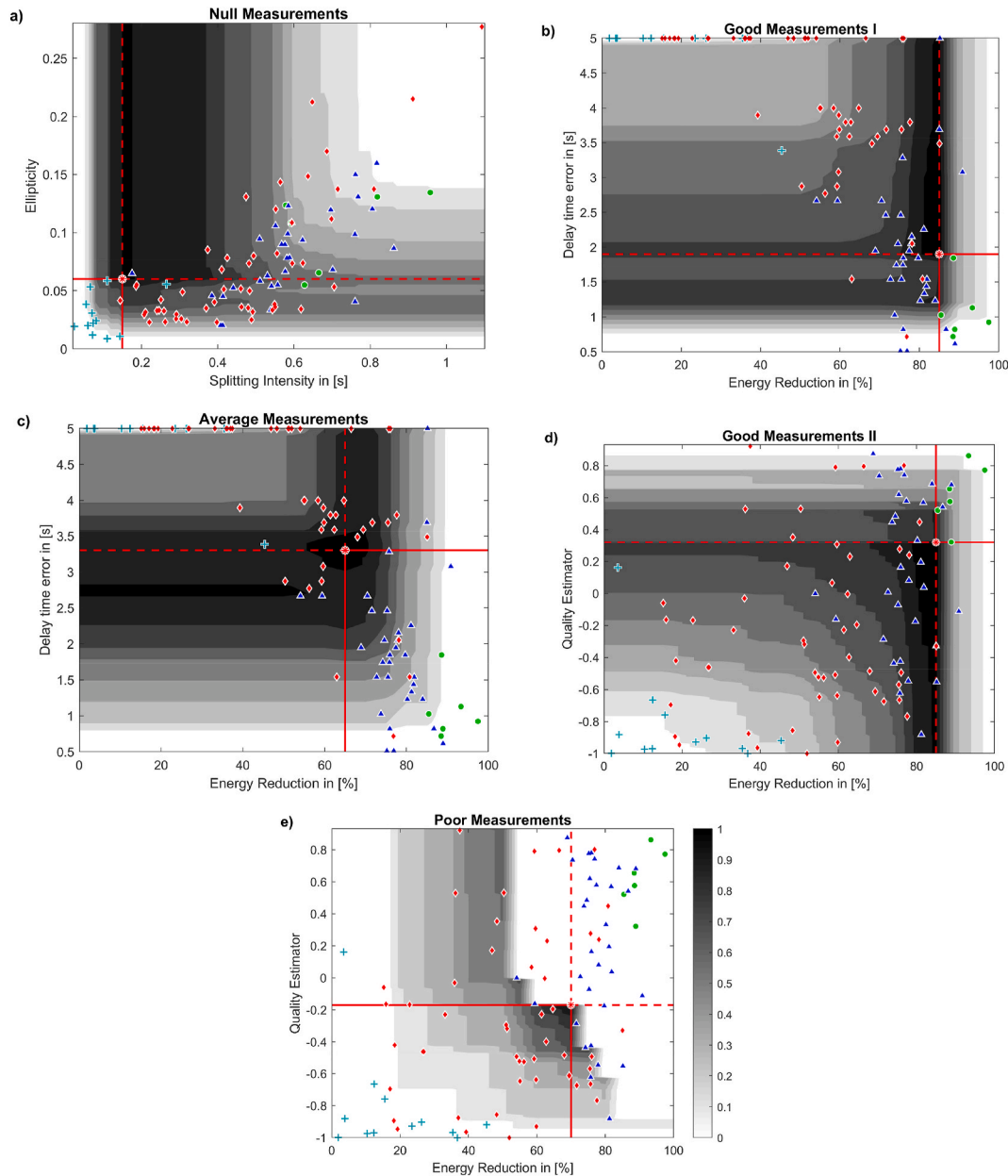

Fig. 3. Class isolation function and distribution of the categories for parameter combinations of a) Splitting Intensity and ellipticity to isolate null-measurements, b) energy reduction and splitting time error to isolate good measurements, c) again energy reduction and splitting time error to isolate average measurements, d) energy reduction and quality estimator to isolate good measurements and e) energy reduction and quality estimator to estimate a criterion for discarding poor measurements. The categories are represented with green dots for good, blue triangles for average, red diamonds for poor and pluses in light blue for null-measurements. The background gray shading scales with the value of the class isolation function and is normalized to range from 1 to 0. The ideal parameter combination is marked by a red star with crosslines and are gathered in Table 1. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

the total number of null-measurements N_{null}. The second fraction R_{res} is defined similarly based on the number of non-null-measurments outside of these limits n_{res} over the total number of non-null-measurments N_{res}. The optimum set of parameters can be found performing a grid search for the maximum of both fractions multiplied

$$X = R_{null} \cdot R_{res}. \tag{2}$$

The object function X (hereafter class isolation function) becomes small for any region one of the ratios is small and reaches its maximum for both ratios similarly high, marking the set of parameters for which the maximum fraction of null-measurements is inside the corresponding limits while simultaneously excluding the maximum fraction of non-null-measurements. Fig. 3 (a) shows the good isolation of the null-measurements by the set of parameters marking the maximum with single outliers of both groups. The found thresholds are gathered in Table 1.

2.5.2. Assigning good, average, and poor measurements

For the assignment of good, average and poor quality classes to the splitting results, we include the method of Wüstefeld et al. (2010) to express the simultaneous fit to the results from the energy minimization technique and the rotation correlation (XC). For a null-measurement the ratio of splitting times $\Delta = \frac{\delta t_{XC}}{\delta t_{SC}}$ becomes zero, while the fast polarizations differ by $45°$, so $\Omega = \frac{\Phi_{SC} - \Phi_{XC}}{45°}$ equals one. For an exact match, the ratio of splitting times Δ becomes one and the fast polarizations agree such that Ω becomes zero. With this we define the quality Q following Wüstefeld et al. (2010), as

$$Q = 1 - \sqrt{2 \cdot \left((\Delta - 1)^2 + \Omega^2 \right)}. \tag{3}$$

The splitting intensity as well as the rotation correlation depend on the similarity between the fast and the slow waveform, which is not valid generally for real waveforms (Crampin and Gao 2006). Therefore, we introduce additional parameters to assess the categorization in view of the energy minimization. We suggest to prefer waveforms with a high energy reduction on the transverse component and with a considerable nonlinear initial particle motion (considerably high ellipticity). The ratio of corrected over original transverse energy quantifies the quality of a measurement. However, the reliability of this parameter depends on the amount of transverse energy on the original waveform due to the split and on the noise level. The errors for the splitting parameters derived from the energy minimization technique provide a measure of stability of the individual (non-null) measurement and the noise contamination.

For determining the good, average and poor measurements, we calculate similar class isolation functions as used previously for null-measurements but based on the ratio of the category to be isolated and the ratio of the remaining categories to be neglected with a set of parameters used as limits. The good (I) and average measurements can be isolated reliably using the energy reduction and splitting time error (see Fig. 3 b-c). As the good (I) measurements are also within the limits of the average category, we separate average and good solutions using the boundaries found previously for the good (I) category. The quality estimator provides only limited potential to isolate categories compared

to the energy reduction or the splitting time error (see Fig. 3 d). However, we use it in combination with the energy reduction as second criterion for assigning the good (II) category and for neglecting poor measurements (see Fig. 3 d and e, respectively).

2.5.3. Assembling limits for the categorization

We combine the findings from the manual class assignments to propose values for an automated categorization. Measurements are classified as null-measurement, poor, average or good according to the thresholds in Table 1. The measurements are assigned in order according to the priority of the given criteria. First null-measurements are assigned using the splitting intensity and ellipticity. The remaining measurements are checked for the energy reduction (E_{red}), splitting time error ($e_{\delta t}$) and quality estimator (Q). The following steps all assign the categories for the remaining measurements of the preceding step according to the given criteria of the investigated category. Measurements, which remain unclassified are set to poor after the categorization sequence. Thus, we suggest using two consecutive criteria to identify good measurements. Since the energy reduction is the most robust estimator for this selection in combination with the splitting time error, we use this parameter set with higher priority (good I) than the less robust combination of energy reduction and quality identifier, which is applied as last criterion (good II) to additionally test for measurements with similar results of the cross-correlation analysis and the energy reduction technique.

We will show in the application of the automatized software package that the proposed thresholds are viable for a large variety of data. Nevertheless, there might be cases, for which the given values are not suitable for a particular data set. In such a case, we provide a function within the automatized SplitRacer toolbox to determine new thresholds based on manually categorized data. If the results appear to not fit to the user's criteria, we propose that the user analyzes a representative station of the data set manually and applies the provided function to obtain the new thresholds.

2.6. Modifications in the joint-splitting procedure

SplitRacer provides the possibility to invert for a single or two-layer anisotropy for all good, average and null-measurements simultaneously in a multi-phase joint splitting analysis. The inversion procedure follows the same scheme as for the analysis of the individual phases and, in case of single-layer anisotropy, is similar to the method proposed by Wolfe and Silver (1998). First, the inverse splitting operator is calculated for a set (two in case of a single layer inversion and four in case of the two-layer inversion) of splitting parameters and applied for each event individually. Then, the remaining energy on the transverse component after correction is summed up for all events included in the analysis before repeating the calculation for the next set of splitting parameters. This results in an energy grid that represents the overall misfit of the model defined by the set of splitting parameters to the full data set. It has been shown that joint splitting reduces the influence of noise significantly and yields more robust measurement compared with single-phase splitting results (Wolfe and Silver 1998; Wölbern et al., 2014; Homuth et al., 2016; Reiss et al., 2016). Here, we accelerate this calculation by directly stacking the previously compiled energy grids from the single splitting analysis and by applying bootstrap statistics (Efron 1979). Therefore, we introduce an additional estimator for the error of the parameters derived from the joint inversion. For that purpose, subsets of the waveforms are created, discarding a random number of waveforms and replacing them by randomly chosen remaining waveforms. For the new data sets the grid search is repeated. The best model parameters and the error are estimated from a fitted normal distribution with the mean value and the standard deviation as estimators. The bootstrap statistics is only valid for a significant number of waveforms and a large number of iterations. We create 1000 random subsets to fulfil the latter. Naturally, in many applications only a limited number of waveforms is available due to a temporary deployment of seismic stations and the

Table 1
Thresholds for the assignment of splitting measurements to quality classes.

	null	poor	good I	average	good II
priority	I	II	III	IV	V
SI	<0.15s	x	x	x	x
λ_1 / λ_2	<0.06	x	>0.02	>0.02	>0.02
E_{red}	x	<70%	>85%	>65%	>85%
$e_{\delta t}$	x	x	<1.9s	<3.3 s	x
Q	x	< −0.17	x	x	>0.32

noise conditions at a site. While there is no definitive lower limit of sample size to get a reliable distribution using the bootstrapping method, we advise to rely on this error estimate only for data sets that consist of at least 10 waveforms.

3. Test of the automatic approach

In the following, we compare the analysis of three reference stations in the USArray Transportable Array (121A, MDND and TIGA) by combining the semi-automatic version of SplitRacer (Reiss and Rümpker 2017) with the fully automatized procedures presented in the previous section. We downloaded data from June 2008 to January 2020 for events at distances between 88° and 140° with magnitudes above 5.5, which results in 8690 event files with almost equal fractions for each station. The first pre-processing step is the same for the semi-as for the fully automatized version. The time series are filtered using a second-order Butterworth filter with cut-off frequencies at 0.02 Hz and 0.25 Hz and cut to sequences of ±50 s centered around the theoretical arrival time of a core-mantle converted phase. We discard events with a signal to noise ratio below 2 or general issues in the time sequence (e.g., data gaps or dead channels). For the window selection process 1093 events and 1258 phases remain. In the semi-automatic version, the windows for the XKS-phases are picked manually and poor waveforms can be discarded directly before analysis. Here, we keep 638 events with a total of 693 XKS phases for the splitting analysis. In the automated version, events are only discarded in this step when no window in the time range ±50 s can be found based on the criteria described above. This yields 754 events and 784 phases. In the manual and automatic approaches, phases are discarded based on different criteria. In the manual selection, we handled noise contaminated signals more strictly, while the automated selection discarded mainly phases, due to inferences with other phases (e.g., S and ScS). For the comparison, 595 concurring phases remain in both data sets.

We compare the manual and automatic time picks and calculate the deviation, normalized to the window length of the manual selection. The start time picks agree quite well for both procedures with most of the picks not differing more than 20% (see Fig. 4 a). If the deviation of the time picks had an impact on the splitting analysis, this should also affect the assignment of the categories. Such a relation is not supported by the automatic assignment, as shown in the following comparison.

The assignment of the categories based on the automatic procedure is mostly in good agreement with the manual classification (see Fig. 5). A few poor and average assignments are exchanged, which indicates an increased noise contamination of the particular waveforms with fairly

acceptable results. The same conclusion holds for the measurements assigned to categories poor versus null and good versus average. This is indicative for events which cannot clearly be assigned to one of the two categories. The automatic procedure provides a decision based on objective criteria.

Finally, we compare the results of the joint splitting of all measurements assigned to be good, average or null to invert for a single anisotropic layer. For the automatic analysis, the accelerated inversion based on the stack of the energy grids from the single splitting analysis and the following bootstrap statistics with 1000 random sample sets is used, while for the semi-automatic analysis, the joint splitting is calculated independently from the single splitting analysis and the error is calculated from the 50 randomly selected time windows. Despite the difference of the procedures the results are in good agreement for the splitting parameters as well as for the errors (see Fig. 6), which supports the reliability of the automatic procedures. Differences in the width of confidence levels stem from different methods for error estimation.

4. Application to the USArray data set

We apply the automatic SplitRacer toolbox to the full data set of the USArray (Transportable Array) to test and present its usability when analyzing large data sets. In total, there are 1900 stations with high quality data between 2 and 15 years recording time starting in 2007, including reference stations continuously running throughout the whole recording period. Shear wave splitting was analyzed for stations set up before 2015. The results are available from data archives by Liu et al. (2014, based on measurements from Liu 2009; Refayee et al., 2014; Yang et al., 2014; Liu et al., 2014) and Yang et al. (2017). While they also include results of preliminary projects and permanent stations of the regional networks, we focus in our comparison on data collected in the USArray project only. The total number of measurements for the stations of the Transportable Array in the previous studies amounts to 12845 pairs of shear wave splitting parameters from 1589 stations (the shift to Alaska and Canada is not included). In our study, we analyzed events for a distance range between 88° and 140° from each individual station and considered only magnitudes above 5.8. As first quality measure, we discard events with a signal to noise ratio below 2. In total, we receive 37852 measurements for 1886 stations, classified as good, average and null according to the processes described above. Due to different selection criteria in our study and the fact that in the previous studies null-measurements are not included, only 7648 phases at 1398 stations remain which can be compared directly. About half of the events, which are not considered in our analysis, but are found in the
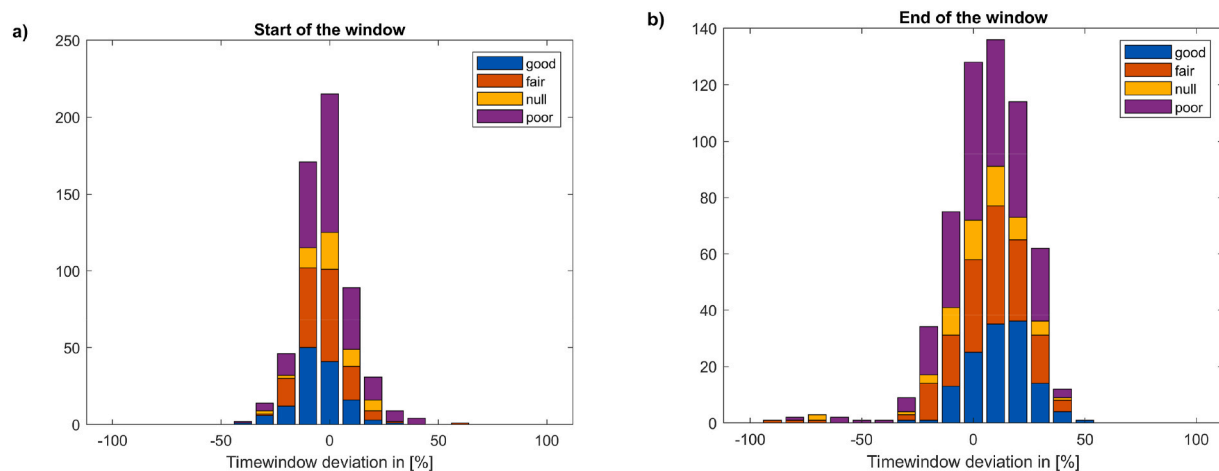

Fig. 4. *Deviation (in percent of the total time window length) of the picked start (a) and end time (b) from the automatic compared to the manual selection in percentage. The assigned categories in the following processing steps are shown by color indicating no clear correlation with deviation of the time picks.* (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)
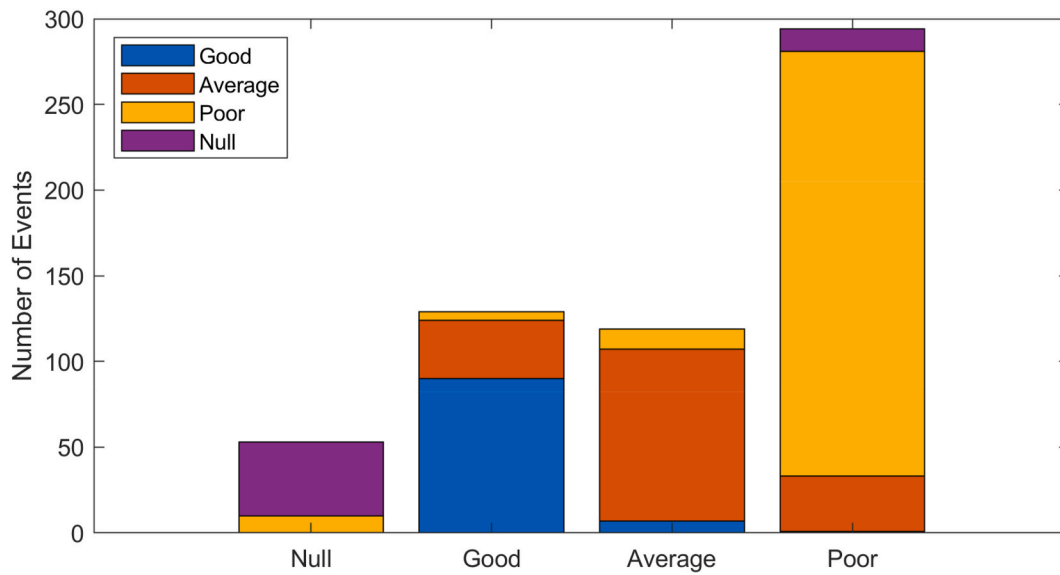
Fig. 5. Comparison of the assigned categories of the manual (colors) and automatic procedure (bars). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)
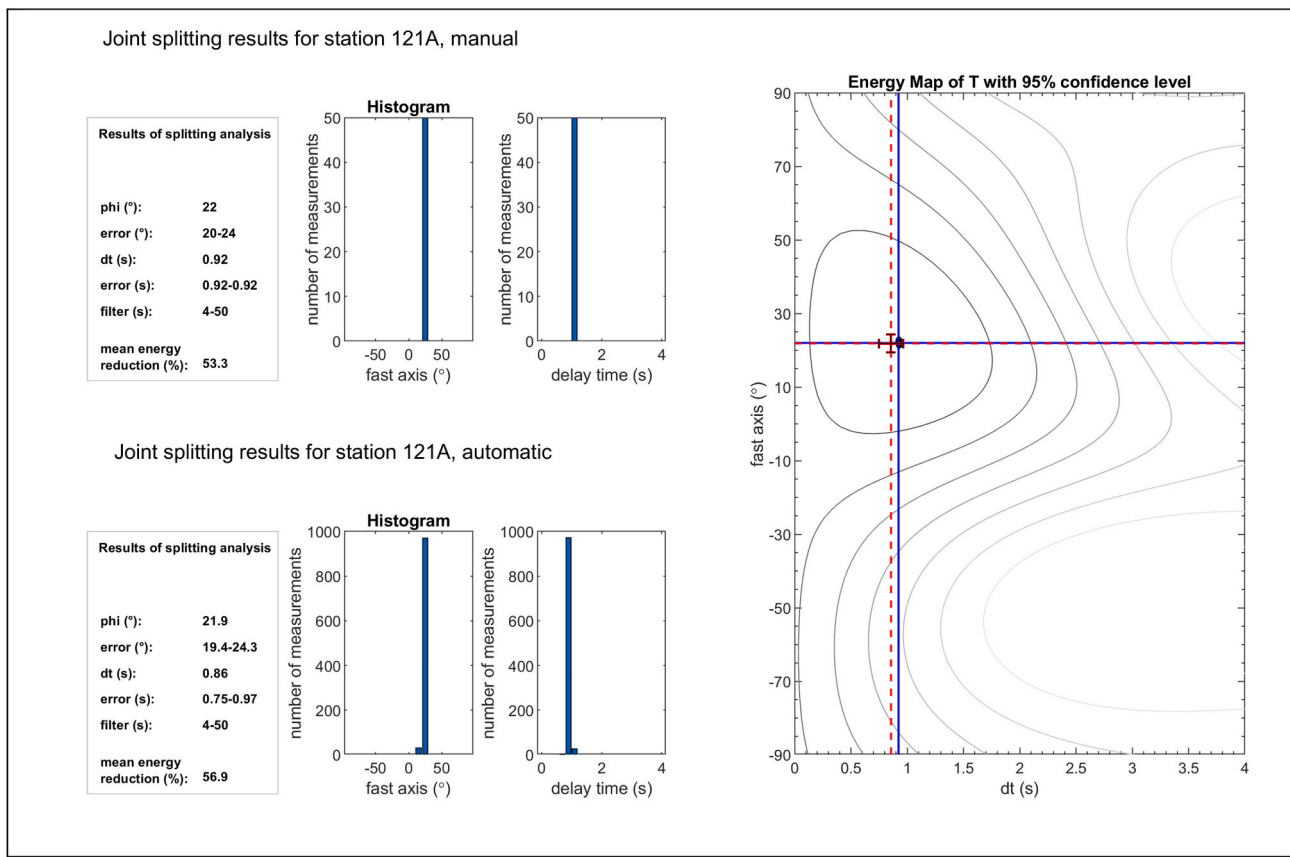

Fig. 6. Results from the joint splitting inversion for one anisotropic layer based on the semi-automatic (denoted "manual") and automatic analysis procedure for stations 121A. The bar graphs for the automatic analysis is based on a bootstrap analysis for the stack of 1000 subsets of the individual energy grids from the single splitting analysis. In the semi-automatic procedure, the bar plots show the distribution of the results from the analysis of 50 random time windows. The blue crosslines in the energy map mark the results from the semi-automatic approach with the black area representing the error surface. The red crosslines and corresponding error bars show the results from the automatic analysis. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

databases, have been discarded due to low data quality based either on the low signal to noise ratio or the low STA/LTA-ratio. Most of the remaining phases not used in our analysis are discarded in the automatic window selection, as they interfere with large amplitude S-phases (e.g., S and ScS).

In Fig. 7, we show the deviation of the fast axis direction Δφ and splitting time Δt. The deviation within one standard deviation, corresponding to 68.27% of the measurements, is smaller than 0.15 s for the splitting time and 4° fast axis orientation. The distribution does not perfectly follow a normal distribution. We therefore also highlight the interval for 95.45% of the measurements corresponding to the 2σ-interval of the normal distribution. From this, the good agreement becomes evident, while larger deviations are likely due to the influence of noise.

While the previous studies derive effective splitting parameters per station using simple averaging, we perform a joint splitting analysis for one anisotropic layer to derive the apparent splitting parameters. This allows to consider null-measurements without including their associated (usually) large errors. We use the advantage of the automatization to analyze the full data set of the Transportable Array and compare the results from the joint splitting with station averages of the databases, where available (see Fig. 8) based on a normalized RMS. We combine the deviation of the results for the fast axes direction and the splitting time.

$$RMS = \sqrt{\frac{1}{2}\left(\frac{(\delta t_1 - \delta t_2)^2}{4} + \frac{(\varphi_1 - \varphi_2)^2}{90}\right)}, \qquad 4$$

with δt the splitting time and φ the fast axis direction at a certain station. The indices 1 and 2 represent the joint splitting results from our automatic and the station averages of previous analyses, respectively.

We only compare stations with at least one "good" or "average" measurement and an error less than 45° for the fast axis. Larger errors are either indicative for a multilayer anisotropy or an insufficient data quality to allow a representative one-layer anisotropic approximation. We find 86 stations with an RMS larger than 0.2. The general increase of the RMS-level and the increased number of stations exceeding the 0.2 level can be traced back to a lower splitting time found in the joint splitting compared to the averaging of the results from the database. This is to be expected, as single splitting measurements inferred from the energy minimization method tend to overestimate the anisotropic strength of the medium. The simultaneous minimization of the energy of all events enables here a more realistic result. Also, the automatization

of the analysis allows us to include a larger data set compared to the collection in the databases (in average by a factor of 2), which strengthens the statistical analysis.

In Fig. 9 we present the results from the joint inversion. The main patterns coincide very well with the previous studies. Here, we omit a detailed discussion and interpretation of the results, as this has been done previously (see Liu et al., 2014; Yang et al., 2014; Refayee et al., 2014; Hongsresawat et al., 2015; Venereau et al., 2019; Wang and Becker 2019; McPherson et al., 2020).

5. Discussion and conclusion

In this study we present an extension for the MATLAB-based SplitRacer toolbox (Reiss and Rümpker 2017) to fully automatize the analysis of teleseismic XKS shear wave splitting. The automatization aims at a more rapid analysis of large data sets, as provided by recent large seismological networks, and at an improvement in the objectivity of processing and analysis procedures. The software is tested by applying it to the large data set of the USArray (Transportable Array) with more than 1900 seismic broadband stations. The new version provides an automated window selection for the waveforms, which is based on a time-spectral analysis to identify a dominant wavelet in a pre-selected section of the traces. The STA/LTA-ratio within the found window provides a first quality estimate and noise dominated traces are discarded. The calculation of the splitting intensity (Chevrot 2000) and the rotation correlation (Bowman and Ando 1987) is included, which provides additional information about data quality and stability. The comparison of the different techniques with the energy minimization (Silver and Chan 1991) allows the formulation of thresholds to objectively classify the measurements. The splitting intensity proves to be an ideal identifier for waveforms with only small or insignificant splitting. The remaining categories are best assigned using the splitting time error from the 95% confidence region provided by the single splitting analysis of the energy minimization and the energy reduction for the found splitting parameters. A comparison with a manual categorization shows, that the comparison of the rotation correlation and energy minimization by using the quality estimator (Wüstefeld et al., 2010) has only weak sensitivity for the true quality of the present splitting in a waveform. The comparison with the manual classification allows to define thresholds that match the criteria of the user. The thresholds given here might not suit all data sets and an adjustment might be necessary to improve the particular analysis. For this purpose, we provide a function that adopts the mechanisms presented here. Nevertheless, the application to the
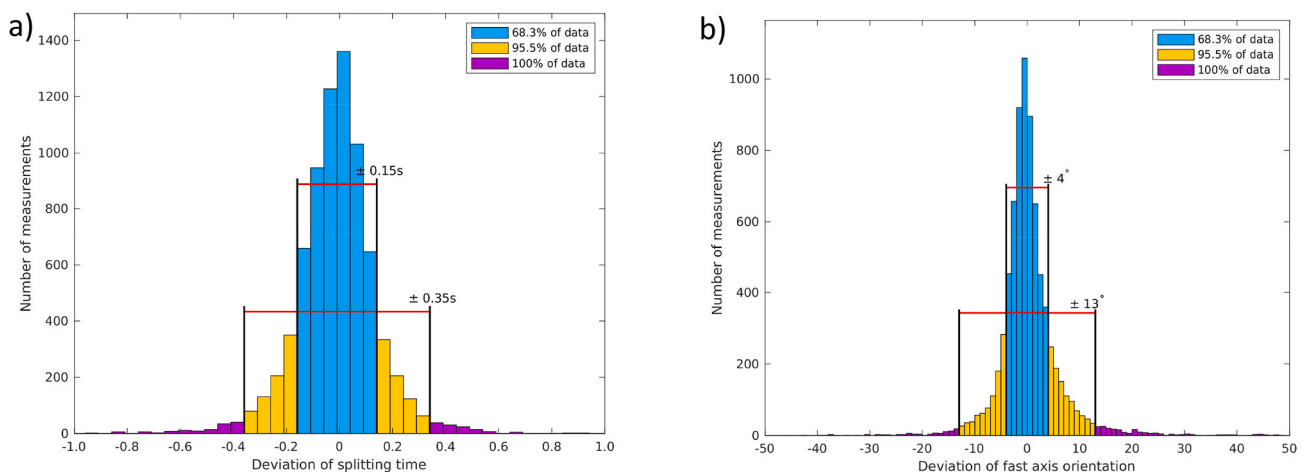

Fig. 7. Histograms for the deviation of splitting time (a) and fast axis orientation (b) from previous studies and the automatic analysis with SplitRacer. The blue color assembles 68.27% of all measurements within the indicated error. The yellow color adds up to 95.45% of all measurements. The purple color mark the remaining measurements, which are less than 5% of all compared pairs of splitting parameters. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)
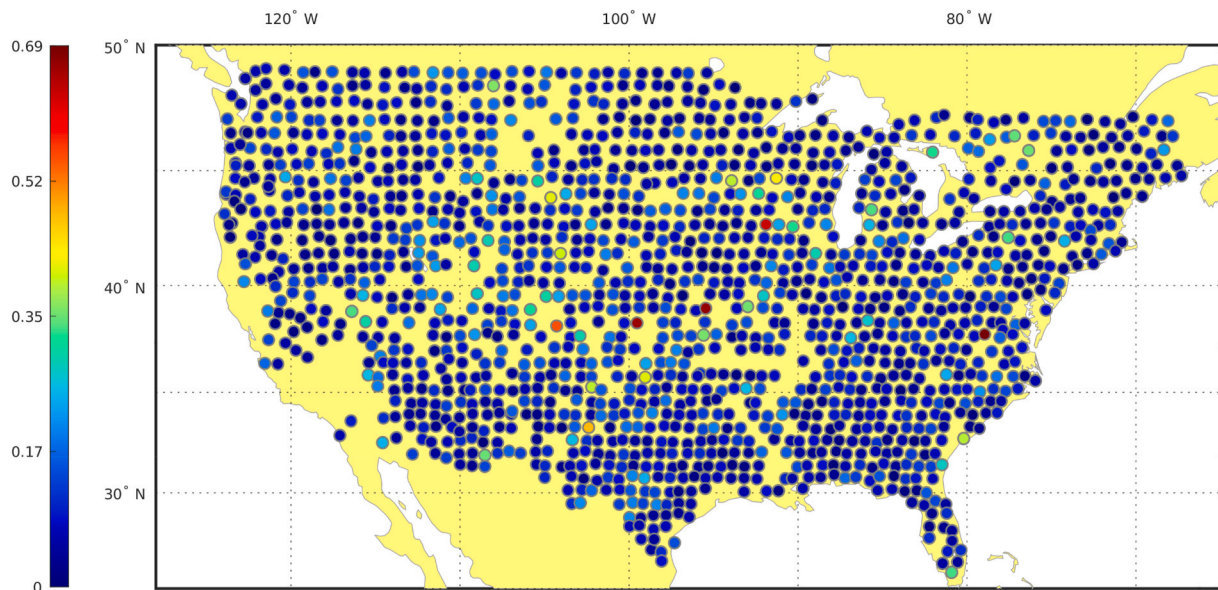
Fig. 8. Deviation of the results derived from the joint splitting in SplitRacer from the station averages of the splitting parameters of previous studies represented as normalized rms at each station location (see eq. (4)). Large rms are characterized by yellow to red colors while low rms is characterized by green to blue colors. 86 locations show rms-values above 0.2. See text for discussion. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)
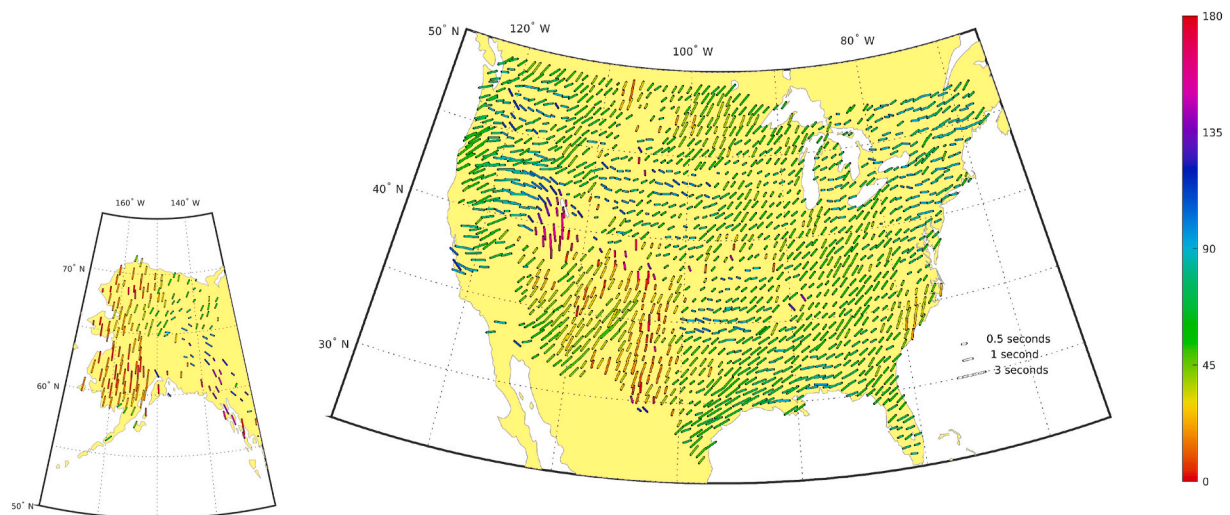

Fig. 9. Splitting results for the 1-layer joint splitting for the full data set of the USArray Transportable Array including the extension to Western Canada and Alaska. The length of the bars is scaled with the splitting time and are oriented parallel to the fast axis direction. Additionally, the colors of the bars indicate the fast axis direction. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

diverse data set of the USArray shows the viability of the proposed thresholds in a general context. The extension for SplitRacer is also designed to allow an accelerated analysis procedure. Therefore, the extension is provided as command line-based code, giving access to the familiar tools in the SplitRacer software without losing resources to the graphical interface. The joint splitting analysis for one layer is accelerated and a bootstrapping statistics (Efron 1979) allows a fast analysis for the best one layer splitting parameters considering all measurements including null-measurements.

To test the stability and reliability of the automatic procedure, we present the application to the complete USArray (Transportable Array) and compare our results with previously measured shear-wave splitting (Liu 2009; Refayee et al., 2014; Yang et al. 2014, 2017; Liu et al., 2014; Liu and Gao 2013). The results of the single measurements attest to the very good agreement of the automatic procedure in comparison to the

previous studies. The automatic procedure yields a data set of 37852 (23242 only good and average) splitting parameter pairs including null measurements. For the main experiment (without the expansion to Alaska) this represents 19256 good and average measurements, compared to the previously reported 13436 splitting measurements for the same data set. Thus, we can clearly improve the statistical basis for the subsequent joint splitting analysis. The inclusion of null-measurements in our joint procedure further stabilizes the effective splitting at the individual stations compared to the station-averaged splitting parameters from previous studies.

Authorship statement

FL developed and performed the data analysis and wrote the first draft of the manuscript. MCR performed the analysis of a subset of the

presented data for independent data classification. FL, MCR and GR jointly discussed the results and contributed to the manuscript.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cageo.2021.104961.

Computer code availability

The original SplitRacer toolbox was developed 2017 by Miriam Christina Reiss (Institute of Geosciences, Goethe-University Frankfurt, Germany, phone: +49 69 798 40139, email: reiss@geophysik.uni-frankfurt.de) and Georg Rümpker (Institute of Geosciences, Goethe-University Frankfurt, Germany, email: rumpker@geophysik.uni-frankfurt.de). It is available for Linux and Windows from the repository: https://www.geophysik.uni-frankfurt.de/64002762/Software. The Code can be downloaded as compressed .tar.gz archive (Linux) or .zip archive (Windows). The required storage space is 5.5 MB. The Code is written for the MATLAB-environment and compatible with the version 2015b or newer. The expansion package SplitRacerAuto was developed by Frederik Link (Institute of Geosciences, Goethe-University Frankfurt, Germany, phone +49 69 798 40117, email: link@geophysik.uni-frankfurt.de), Miriam Christina Reiss (Institute of Geosciences, Goethe-University Frankfurt, Germany, phone: +49 69 798 40139, email: reiss@geophysik.uni-frankfurt.de) and Georg Rümpker (Institute of Geosciences, Goethe-University Frankfurt, Germany, email: rumpker@geophysik.uni-frankfurt.de). The software package is available at the same repository as the original SplitRacer Code: https://www.geophysik.uni-frankfurt.de/64002762/Software. The Code can be downloaded as compressed .tar.gz archive (Linux) or .zip archive (Windows). The required storage space is 0.5 MB. The Code is written for the MATLAB-environment and compatible with the version 2015b or newer. Supporting information for the installation of the Software package can be found in the User Guide included in the archive.

References

Allen, R., 1978. Automatic earthquake recognition and timing from single trace. Bull. Seismol. Soc. Am. 68 (5), 1521–1532.

Bowman, J.R., Ando, M., 1987. Shear-wave splitting in the upper-mantle wedge above the Tonga subduction zone. Geophys. J. 88, 25–41.

Chevrot, S., 2000. Multichannel analysis of shear wave splitting. J. Geophys. Res.: Solid Earth 105 (B5), 21579–21590.

Crampin, S., Gao, Y., 2006. A review of techniques for measuring shear-wave splitting above small earthquakes. Phys. Earth Planet. In. 159, 1–14.

Efron, B., 1979. Bootstrap methods: another look at the jackknife. Ann. Stat. 7 (1), 1–26.

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

Grund, M., 2017. StackSplit - a plugin for multi-event shear wave splitting analyses in SplitLab. Comput. Geosci. 105, 43–50.

Homuth, B., Löbl, U., Batte, A.G., Link, K., Kasereka, C.M., Rümpker, G., 2016. Seismic anisotropy of the lithosphere/asthenosphere system beneath the Rwenzori region of the Albertine Rift. Int. J. Earth Sci. 105, 1681–1692.

Hongsresawat, S., Panning, M.P., Russo, R.M., Foster, D.A., Monteiller, V., Chevrot, S., 2015. USArray shear wave splitting shows seismic anisotropy from both lithosphere and asthenosphere. Geology 43, 667–670.

Karato, S., Jung, H., Katayama, I., Skemer, P., 2008. Geodynamic significance of seismic anisotropy of the upper mantle: new insights from laboratory studies. Annu. Rev. Earth Planet Sci. 36, 59–95.

Kennett, B.L.N., Engdahl, E.R., 1991. Traveltimes for global earthquake location and phase identification. Geophys. J. Int. 105 (2), 429–465.

Kong, F., Gao, S.S., Liu, K.H., 2015. A systematic comparison of the transverse energy minimization and splitting intensity techniques for measuring shear-wave splitting parameters. Bull. Seismol. Soc. Am. 105, 230–239.

Liu, K.H., 2009. NA-SWS-1.1: a uniform database of teleseismic shear wave splitting measurements for North America. G-cubed 10 (5).

Liu, K.H., Gao, S.S., 2013. Making reliable shear-wave splitting measurements. Bull. Seismol. Soc. Am. 103 (5), 2680–2693.

Liu, K.H., Elsheikh, A., Lemnifi, A., Purevsuren, U., Ray, M., Refayee, H., Yang, B.B., Yu, Y., Gao, S.S., 2014. A uniform database of teleseismic shear wave splitting measurements for the western and central United States. G-cubed 15 (5), 2075–2085.

MATLAB, 2020. Version 9.8.0 (R2020a. The MathWorks Inc., Natick, Massachusetts.

McPherson, A.M., Christensen, D.H., Abers, G.A., Tape, C., 2020. Shear wave splitting and mantle flow beneath Alaska. J. Geophys. Res.: Solid Earth 125 (4), e2019JB018329.

Quatieri, T.F., 2006. Discrete-time Speech Signal Processing: Principles and Practice. Pearson Education India.

Refayee, H.A., Yang, B.B., Liu, K.H., Gao, S.S., 2014. Mantle flow and lithosphere–asthenosphere coupling beneath the southwestern edge of the North American craton: constraints from shear-wave splitting measurements. Earth Planet Sci. Lett. 402, 209–220.

Reiss, M.C., Rümpker, G., 2017. SplitRacer: MATLAB code and GUI for semiautomated analysis and interpretation of teleseismic shear-wave splitting. Seismol Res. Lett. 88 (2A), 392–409.

Reiss, M.C., Rümpker, G., Tilmann, F., Yuan, X., Giese, J., Rindraharisaona, E.J., 2016. Seismic anisotropy of the lithosphere and asthenosphere beneath southern Madagascar from teleseismic shear wave splitting analysis and waveform modeling. J. Geophys. Res.: Solid Earth 121, 6627–6643.

Reiss, M.C., Long, M.D., Creasy, N., 2019. Lowermost mantle anisotropy beneath Africa from differential SKS-SKKS shear-wave splitting. J. Geophys. Res.: Solid Earth 124.

Rümpker, G., Silver, P.G., 1998. Apparent shear-wave splitting parameters in the presence of vertically varying anisotropy. Geophys. J. Int. 135 (3), 790–800.

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

Savage, M.K., Wessel, A., Teanby, N.A., Hurst, A.W., 2010. Automatic measurement of shear wave splitting and applications to time varying anisotropy at Mount Ruapehu volcano, New Zealand. J. Geophys. Res.: Solid Earth 115 (B12321).

Silver, P.G., Chan, W.W., 1991. Shear wave splitting and subcontinental mantle deformation. J. Geophys. Res.: Solid Earth 96 (B10), 16429–16454.

Spingos, I., Kaviris, G., Millas, C., Papadimitriou, P., Voulgaris, N., 2020. Pytheas: an open-source software solution for local shear-wave splitting studies. Comput. Geosci. 134, 104346.

Teanby, N.A., Kendall, J.-M., der Baan, M.V., 2004. Automation of shear-wave splitting measurements using cluster analysis. Bull. Seismol. Soc. Am. 94 (2), 453–463.

Trnkoczy, A., 2009. *New Manual of Seismological Observatory Practice (NMSOP)*, Chap. Understanding and Parameter Setting of STA/LTA Trigger Algorithm. Deutsches GeoForschungsZentrum GFZ, pp. 1–20.

Venereau, C.M.A., Martin-Short, R., Bastow, I.D., Allen, R.M., Kounoudis, R., 2019. The role of variable slab dip in driving mantle flow at the eastern edge of the Alaskan subduction margin: insights from shear-wave splitting. G-cubed 20 (5), 2433–2448.

Vidale, J.E., 1986. Complex polarization analysis of particle motion. Bull. Seismol. Soc. Am. 76 (5), 1393–1405.

Walker, A.M., Wookey, J., 2012. Msat - a new toolkit for the analysis of elastic and seismic anisotropy. Comput. Geosci. 49, 81–90.

Walsh, E., Arnold, R., Savage, M.K., 2013. Silver and chan revisited. J. Geophys. Res.: Solid Earth 118, 5500–5515.

Walther, M., Plenefisch, T., Rümpker, G., 2014. Automated analysis of SKS splitting to infer upper mantle anisotropy beneath Germany using more than 20 yr of GRSN and GRF data. Geophys. J. Int. 196 (2), 1207–1236.

Wang, W., Becker, T.W., 2019. Upper mantle seismic anisotropy as a constraint for mantle flow and continental dynamics of the North American plate. Earth Planet Sci. Lett. 514, 143–155.

Wölbern, I., Löbl, U., Rümpker, G., 2014. Crustal origin of trench-parallel shear-wave fast polarizations in the Central Andes. Earth Planet Sci. Lett. 392, 230–238.

Wolfe, C.J., Silver, P.G., 1998. Seismic anisotropy of oceanic upper mantle: shear wave splitting methodologies and observations. J. Geophys. Res. 103, 749–771.

Wüstefeld, A., Bokelmann, G., Zaroli, C., Barruol, G., 2008. SplitLab: a shear-wave splitting environment in Matlab. Comput. Geosci. 34 (5), 515–528.

Wüstefeld, A., Al-Harrasi, O.H., Verdon, J.P., Wookey, J., Kendall, J.-M., 2010. A strategy for automated analysis of passive microseismic data to image seismic anisotropy and fracture characteristics. Geophys. Prospect. 58, 755–773.

Yang, B.B., Gao, S.S., Liu, K.H., Elsheikh, A.A., Lemnifi, A.A., Refayee, H.A., Yu, Y., 2014. Seismic anisotropy and mantle flow beneath the northern Great Plains of North America. J. Geophys. Res.: Solid Earth 119 (3), 1971–1985.

Yang, B.B., Liu, Y., Dahm, H., Liu, K.H., Gao, S.S., 2017. Seismic azimuthal anisotropy beneath the eastern United States and its geodynamic implications. Geophys. Res. Lett. 44 (6), 2670–2678.