Computers & Geosciences 134 (2020) 104346

Contents lists available at ScienceDirect

Computers and Geosciences

journal homepage: www.elsevier.com/locate/cageo

Pytheas: An open-source software solution for local shear-wave splitting studies

Ioannis Spingos *, George Kaviris, Christos Millas, Panayotis Papadimitriou, Nicholas Voulgaris

Section of Geophysics–Geothermics, Department of Geology and Geoenvironment, National and Kapodistrian University of Athens, Panepistimiopolis, 157 84, Zografou, Greece

ARTICLE INFO

Keywords:
Data processing
Geophysics
Seismology

ABSTRACT

Shear-wave splitting analysis of local earthquakes is a challenge tackled by a long history of published software. However, there is currently no complete solution that integrates manual and (partially or fully) automatic methods. The Pytheas software aims to achieve this by offering a collection of popular analysis techniques accompanied by an intuitive Graphical User Interface. The software offers its own implementation of the following methods: (a) Visual inspection of polarigrams and hodograms (manual), (b) Rotation-Correlation (semi-automatic), (c) Eigenvalue (semi-automatic) and (d) Minimum Energy (semi-automatic). Moreover, it automates the last three operations by using Cluster Analysis. The software also utilizes databases to effectively manage results storing. It is compatible with standard file formats for event and station metadata (QuakeML and StationXML), as well as waveform data (such as SAC and miniSEED). Pytheas also introduces a novel algorithm for automatically assigning quality grades to measurements. To showcase the utility of the software, we analyzed a set of 263 events recorded in the Western Gulf of Corinth (Greece). The workflow included the establishment of a control group through manual analysis, the fully automatic processing of the catalogue and the verification of the results through visual inspection of a sample. Automatic and manual measurements show good agreement, validating the potential of Cluster Analysis. The presented software aims to provide a friendly experience to the end-user in all stages of its usage (installation, operation and database management) and trivialize the time-consuming part of a local shear-wave splitting study, leaving more time for interpretation and analysis of results.

1. Introduction

Shear-wave Splitting (SwS) is the branch of seismic anisotropy that refers to the generation of two orthogonally polarized shear-waves in an anisotropic medium. The two new components of polarization propagate with different velocities and, as such, the S_{fast} (higher velocity) and S_{slow} (lower velocity) elements can be distinguished. Although SwS has been traditionally studied in teleseismic waves characterizing the upper mantle (e.g. Fukao, 1984; Hatzfeld et al., 2001; Kaviris et al., 2018a; Vinnik et al., 1989), the phenomenon is also present in the upper crust (Crampin, 1986). Crustal anisotropy is generally caused by structural features (e.g. vertical fluid-saturated microcracks) and is often related to the local stress field. According to the Anisotropic Poro-Elasticity (APE) model (Crampin and Zatsepin, 1997; Zatsepin and Crampin, 1997), changes in the properties of fluids and the regional stress affect the characteristics (geometry and density) of the microcracks. Analysis of waveforms from local events yields the polarization direction of the S_{fast} (ϕ), oriented according to the microcracks, and the time-delay between the arrivals of split shear-waves at the station (t_d), which are indicators of the stress state in the propagation medium. Both quantities have been used to identify precursory phenomena of earthquakes (e.g. Aster et al., 1990; Booth and Park, 1988; Crampin et al., 1999, 2008; Peacock et al., 1988) and volcanic eruptions (e.g. Bianco et al., 2006; Miller and Savage, 2001).

Analyzing waveforms of direct local S-wave recordings can be an arduous task. Such processing requires a rather large initial dataset of candidate event-station pairs. Rejection of candidates is usually attributed to poorly identified shear-waves (due to noise and emergent arrivals) and the arrival of secondary phases (converted or scattered) near the expected direct arrival. In an attempt to exclude the latter before the analysis, only rays included in the shear-wave window (a maximum angle of incidence derived from the velocity structure in the area) are considered (Booth and Crampin, 1985; Evans, 1984). Finally, common pitfalls in existing analysis methods lead to further rejection of data

* Corresponding author.
E-mail address: ispingos@geol.uoa.gr (I. Spingos).

https://doi.org/10.1016/j.cageo.2019.104346
Received 16 April 2019; Received in revised form 26 July 2019; Accepted 1 November 2019
Available online 4 November 2019

(Crampin and Gao, 2006). The above factors converge to a single conclusion; to effectively obtain a robust SwS dataset we need to devote a lot of resources to analyzing the largest possible number of seismograms.

Manual methods require the complete involvement of the analyst. The ϕ quantity is initially identified in a particle motion diagram and the analyst seeks the t_d that best linearizes it, after correcting the waveforms for anisotropy. A common particle motion diagram is the hodogram, which exhibits the continuous motion of a ground particle in the horizontal NE plane (Buchbinder, 1989; Chen et al., 1987). Another variant of these diagrams is the polarigram (Bernard and Zollo, 1989), where a vector of motion is displayed for each sample of the recordings. The horizontal components (N-S and E-W) are rotated to the system defined by the determined ϕ (F-S, where F denotes the direction of the S_{fast} polarization and S its perpendicular, which corresponds to S_{slow}) and the t_d is measured. The recordings are then rotated back to the initial axial system, to validate the measurement. The visual inspection of polarigrams and hodograms (hereafter noted as MAN) has been extensively used in crustal SwS studies in Greece (Bouin et al., 1996; Bernard et al., 1997; Kaviris et al., 2015, 2017; 2018b, 2018c; Papadimitriou et al., 1999).

Semi-automatic methods require the specification of a signal window for analysis, before the determination of the optimal ϕ and t_d pair. The technique proposed by Silver and Chan (1991) performs a grid search for the pair of parameters that best minimizes the energy in the transverse component, by either obtaining the signal energy itself (Minimum Energy; ME) or evaluating the second eigenvalue of the covariance matrix of the radial and transverse components (Eigenvalue; EV). This technique is popular amongst upper mantle studies (e.g. Savage and Silver, 1993; Baccheschi et al., 2008), but has also been implemented in crustal research (e.g. Nolte et al., 2017; Teanby et al., 2004a). A second popular semi-automatic method involves the rotation of the horizontal waveforms to various F-S systems and the application of a range of t_d values for each rotation. The pair that leads to the highest correlation coefficient between the two corrected waveforms is the optimal. This method, aptly named Rotation-Correlation (RC), was first introduced in teleseismic studies (Bowman and Ando, 1987), but has also been used in investigating the upper crust (e.g. Chang et al., 2009; Giannopoulos et al., 2015). To fully automate the SwS analysis, Teanby et al. (2004b) integrated the EV technique with Cluster Analysis (CA). Their processing scheme first obtains parameter pairs for a range of pre-determined signal windows around the S_{fast} arrival with EV, forms clusters based on the proximity of the measurements (in a space defined by ϕ and t_d) and then retrieves the most stable solution (as characterized by the most constrained cluster). CA has been applied in various case studies (e.g. Roman et al., 2011; Walpole et al., 2014).

The described methodologies have been included in a plethora of published software (indicatively we mention Gao and Liu, 2012; Grund, 2017; Hao et al., 2008; Piccinini et al., 2013; Reiss and Rümpker, 2017; Savage et al., 2010; Wüstefeld et al., 2008; Zaccarelli et al., 2012), with the exception of the manual method. However, there is no currently released package that combines all of the above. In the Pytheas software, we have incorporated the following features:

(a) An intuitive Graphical User Interface (GUI)
(b) The seamless integration of manual and automatic (partially or fully) techniques
(c) The incorporation of databases
(d) A results management system that includes publication-ready figures

Pytheas was developed in the Python programming language (van Rossum, 1995), which ensures the open-source character of the code and doesn't require the use of fee-only software. Moreover, the operation of the software can be achieved solely through the GUI (from data analysis to figure production), without requiring the knowledge of terminals and other cumbersome operating system features. Following, we describe the main features of Pytheas in detail.

2. The Pytheas software

The main problem that we aimed to solve with the development of Pytheas was the inclusion of all popular SwS analysis methods under one GUI. The Python programming language was selected because it offers a wide variety of packages under the Python Standard Library (PSL), as well as through independent developers. However, to avoid a complicated installation process, we limited the use of dependencies to a minimum. Thus, we utilized the Obspy toolbox (Beyreuther et al., 2010; Krischer et al., 2015) for handling basic seismological-related operations (such as waveform management and metadata files interaction) and the Matplotlib package (Hunter, 2007) for producing figures. Certain CA functions are executed through the Scikit-learn module (Pedregosa et al., 2011). Finally, the GUI is implemented under the Qt5 framework[1] and the relevant Python bindings included in PyQT5.[2]

2.1. Software design

Pytheas is comprised by a main script and six individual modules. The main script (*pytheas.py*) is responsible for building the GUI and connecting the relevant functionality to it. The GUI has been designed with simplicity in mind, as to not overwhelm the user with overcomplicated features. It encompasses a toolbar, which hosts all the functionality included in the software, the information widget, which displays metadata relevant to the active event-station pair, and the figure widget, which showcases the active waveforms and particle motion diagrams (Fig. 1).

The Pytheas library consists of six modules (*parsers, db_handler, eigenvalue, rotationcorrelation, clustering* and *tools;* Fig. 2). The *parsers* module controls the reading and writing of settings from the configuration files. *db_handler* contains all the functions responsible for the interaction with the database. We have used the SQLite[3] implementation of the PSL to manage the results of the analysis. Utilizing a database permits the storage of large datasets and enables fast access to them by decreasing reading, writing and querying time, compared to archiving in simple ASCII files. SQLite stores the data locally, without the need of a separate server process. Moreover, the flexibility in terms of data types, that characterizes SQLite, permits the archiving of Python objects, such as arrays. The database used contains three tables; (a) *"event"* (containing event metadata), (b) *"station"* (consisting of station metadata) and (c) *"method"*. The *"method"* layer includes all the information associated with a SwS measurement (such as the values of ϕ and t_d). Each unique entry in the database is characterized by three identifying properties, i.e. the event, station and method (used for the analysis) code. These properties are used to relate each measurement to the *"event"* and *"station"* layers.

The *eigenvalue* module contains the functions required for implementing the EV and ME methods (Silver and Chan, 1991). The errors of ϕ and t_d are acquired with the formulations of Walsh et al. (2013). Respectively, *rotationcorrelation* is a collection of all the functions used in the RC method (Bowman and Ando, 1987). The errors of the SwS parameters are obtained by estimating the 95% confidence interval for each, assuming a normal distribution.

The *clustering* module is the implementation of the CA method proposed by Teanby et al. (2004b). It hosts the class object that manages all the individual processes (candidate window generation, cluster formation and selection). For the definition of the window range, we followed an approach similar to that of Savage et al. (2010). The extreme bounds

[1] https://www.qt.io/.
[2] https://www.riverbankcomputing.com/software/pyqt/.
[3] https://www.sqlite.org/.

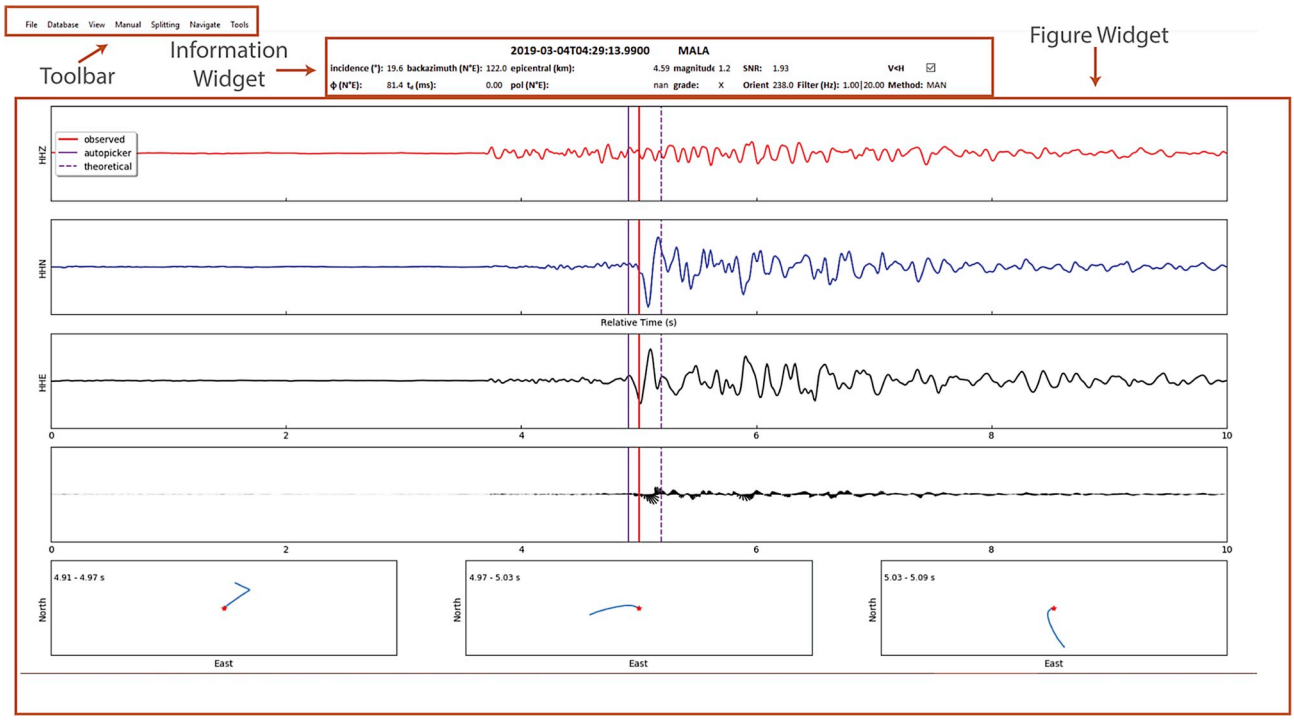

Fig. 1. The main GUI of the Pytheas software. The three separate compartments (toolbar, information widget and figure widget) are highlighted. The information widget displays the following: (upper row) the origin time and the station code, (middle row) the angle of incidence, the backazimuth, the epicentral distance, the magnitude, the Signal-to-Noise Ratio near the S-arrival and a checkbox ("V < H") showcasing whether the amplitude in the vertical component (V) is larger than the ones in the horizontal components (H) close to the S-arrival, (bottom row) the polarization direction of the S_{fast}, the time-delay, the polarization direction of the (corrected for anisotropy) shear-wave, the qualitative grade, the instrument orientation correction, the filter bounds and the method's result displayed.

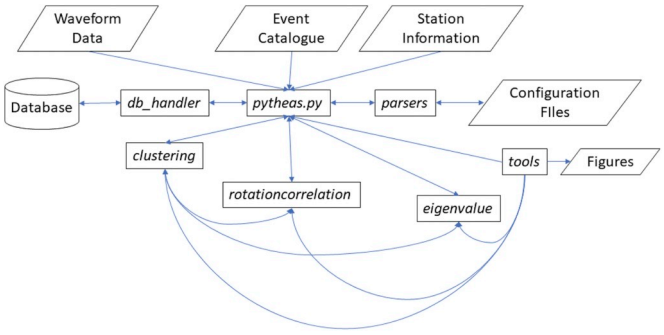

Fig. 2. Chart showcasing the relation between the modules of Pytheas and relevant input/output categories. All the respective functionality is controlled through the GUI in *pytheas.py*.

of the windows range are acquired from the half of the S-P time (for the leftmost window-start point) and the dominant frequency near the S-wave arrival, multiplied by a user-defined factor (for the rightmost window-end point). The use of fixed boundaries for the whole dataset, instead of relative ones, would reduce the effectiveness of the process, since there is no ubiquitous value applicable to all possible event-station combinations. Moreover, Pytheas offers four criteria for linking clusters during the hierarchical clustering, all using the Euclidean distance as a similarity measure. The "ward" criterion (Ward, 1963) seeks the merged clusters with the minimum variance, "average" merges the two clusters with the minimum average distance between observations, "complete" utilizes the minimum of the maximum distances between points and "single" consolidates clusters using the minimum of distances.

Finally, the *tools* module is a collection of various definitions, used in all other functions of the software. For example, it contains the functions that construct the particle motion diagrams. An important feature of *tools* is the automatic grading algorithm. Pytheas accommodates a novel approach to automatically grade results. Initially, it performs a check for null measurements, by verifying that ϕ is not sub-parallel or sub-perpendicular to the polarization direction of the corrected shear-wave. The limits for a measurement to be considered null are specified by the user in the options. In the presented results, we used an angular difference of $\pm 15^{\circ}$ (to consider the two directions sub-parallel) and $90 \pm 15^{\circ}$ (to consider the directions sub-perpendicular). If that condition is met, the measurement is characterized as null ("N" grade). Null measurements refer to either an absence of splitting or a S-wave polarization similar to either the S_{fast} or the S_{slow} wave, rendering anisotropy impossible to distinguish (Wüstefeld and Bokelmann, 2007). A minimum Signal-to-Noise Ratio (SNR) is used afterwards to reject (graded as "E") recordings with significant noise content. The algorithm then calculates four individual scores:

$$s_{\delta\phi} = \frac{\delta\phi}{\delta\phi_{max}} \quad s_{\delta t_d} = \frac{\delta t_d}{\delta t_{d_{\max}}} \quad s_{CC_{FS}} = \frac{1 - |CC_{FS}|}{1 - CC_{FS_{\min}}} \quad s_{CC_{NE}} = \frac{1 - |CC_{NE}|}{1 - CC_{NE_{min}}} \tag{1}$$

where $\delta\phi$ the error of ϕ, δt_d the error of t_d, CC_{FS} the correlation coefficient of the corrected horizontal waveforms in the FS axial system, CC_{NE} the correlation coefficient of the corrected horizontal waveforms in the NE axial system and s the score for each parameter. The *min* and *max* indices denote the lower and upper accepted threshold, respectively. If any of these scores is greater than 1.0, the measurement is rejected ("E" grade"). Otherwise, we obtain the normalized total score (S) from the mean of the individual ones:

$$S = \frac{s_{\delta\phi} + s_{\delta t_d} + s_{CC_{FS}} + s_{CC_{NE}}}{4} \tag{2}$$

The final grade (ranging between "D", i.e. worst, and "A", i.e. best) is determined from S, as each letter corresponds to a score range. The thresholds for each quality parameter and the grading classes are user-configurable. The grading algorithm is not applicable to the manual

method, where SwS parameters' errors cannot be estimated. *tools* also includes a wrapper for using the Obspy implementation of TauP (Crotwell et al., 1999) to calculate theoretical S-arrivals and angles of incidence, as well as the AR-AIC automatic picker (Akazawa, 2004).

2.2. Analysis workflow

Here, we present a synopsis of the suggested workflow for all methods. It is important to note that the user can combine any techniques they prefer. In any case, the software requires three mandatory inputs to operate, i.e. an events catalogue file, a station information file and the master path where waveforms are stored. Regarding the catalogue file, the user can use either a QuakeML (Schorlemmer et al., 2011) or a simple ASCII file. The latter must contain the origin time, location information and magnitude of the events. Nevertheless, the use of QuakeML is preferred, since it can store arrival information as well. Accordingly, station information can be retrieved from either StationXML[4] or plain text (containing the station and network codes, as well as location information) files. Standardized file formats (i.e. QuakeML and StationXML) render the usage of the software easier, as they eliminate the necessity for developing dedicated software for converting between formats, at least for most users. Moreover, both formats allow the inclusion of more information (such as the arrival times in QuakeML and the instrument orientation in StationXML). Waveform data can be in any format readable by Obspy (e.g. SAC and miniSEED). However, each event's data must be stored in its respective folder (named after its origin time) and each station's waveforms must be in a separate file. The software uses the folder code and the station's name to catalog available data faster. Regardless, all of the above constitute an inherent interface between Pytheas and the international Federation of Digital Seismograph Networks (FDSN) services, which provide the necessary tools to obtain event and station metadata, as well as waveform data, in standardized formats.

2.2.1. Manual method

The MAN method is similar in execution to the analysis performed by Kaviris et al. (2015, 2017, 2018b, 2018c). The arrival of the S_{fast} is first picked. The software obtains its polarization direction (i.e. ϕ) from the corresponding vector of motion in the NE plane. The horizontal components are then rotated to the FS axial system, where the t_d is determined by temporally shifting the S_{slow}, in order to match the arrivals of the two split waves. The waveforms are then rotated back to the original NE system. The polarization direction of the corrected shear-wave is determined automatically by Pytheas at the arrival picked at the start. A grade needs to be provided manually - broad guidelines for determining the qualitative weight can be found in Kaviris et al. (2018b). A figure summarizing the results can be produced (Fig. 3). Even though this process requires a short amount of time, it still demands the full involvement of the analyst and is prone to human bias. Furthermore, the uncertainties of both ϕ and t_d cannot be estimated (they are fixed at 1° and 1 sample, respectively).

2.2.2. Semi-automatic methods

The semi-automatic methods (RC, EV and ME) are easily executed by simply selecting the analysis window on the waveforms. The polarigram and hodogram are always displayed, to facilitate the determination of the best window. When the analysis is complete, the software will display a figure for assessing the quality of the measurement (Fig. 4). However, the process still requires the user to adjust the window multiple times and make a decision on the final one.

2.2.3. Automatic method

The CA method can be executed with two different approaches. In the first, the user needs to select the station-event pair, determine the arrival of the S_{fast} (either manually or passively through the event catalogue) and then select one of the combinations (i.e. CA-RC, CA-EV or CA-ME). At the end of the process, a quality assessment figure for the clustering (Fig. 5) will be shown, along with one for the semi-automatic method used (as in Fig. 4). The second approach refers to using Pytheas for automatically analyzing all available event-station pairs in the dataset. In this case, arrivals must be provided in the QuakeML file. It is noted that the user can select the shear-wave window, with all pairs outside it being skipped. In both approaches, the process can be cancelled and resumed at a later time.

The database stores all graded results for each event, station and method. Individual measurements, whether automatic or not, can be loaded at any time through the database interface implemented in Pytheas (albeit, there are freely available applications that can manage SQLite compatible databases, from third parties). A proposed workflow involving all methods would include the following steps. Initially, a control group of measurements in the dataset is established by manual analysis, mainly for determining the more stable features of splitting (i.e. ϕ). Then, the rest of the catalogue is processed with the fully automatic method. Results under a selected grade threshold (e.g. "E" and "D") can be rejected altogether. Finally, a manual validation is performed on a, larger, control group. Outliers and spurious measurements, such as abnormally high t_d (an indicator of cycle skipping), are also checked. The final product is a database of SwS measurements, with the option to be exported in a human comprehensible ASCII format, which can be used in other software to produce the summary plots (such as rose

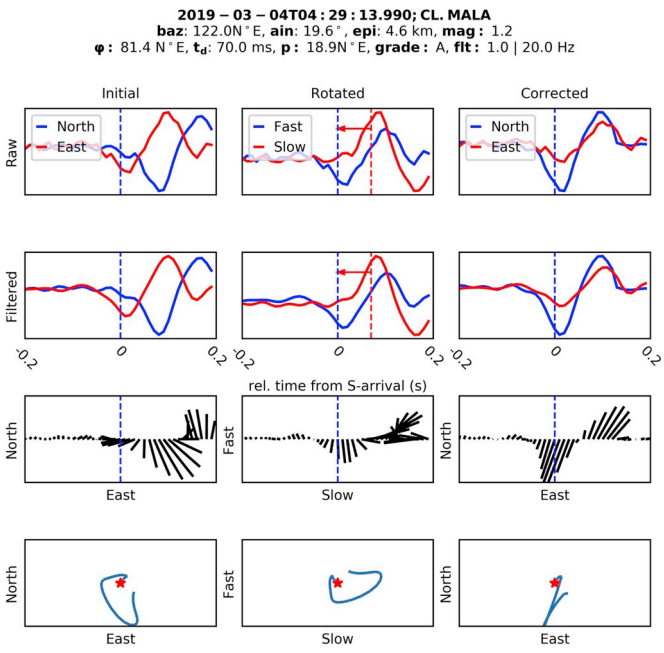

Fig. 3. Summary plot of the analysis with the MAN method. First row (top): Raw horizontal waveforms in the initial NE system (first column), rotated FS system and corrected for anisotropy (second column) and rotated NE system and corrected for anisotropy (third column). Second row: same as first row, but for filtered waveforms. Third row: polarigrams for each stage. Fourth row: hodograms for each stage. The picked S arrivals are denoted by the blue vertical dashed lines in the waveform and polarigram plots. The red vertical dashed lines in the "Rotated" stage indicate the S_{slow}, with the horizontal red arrow showing the temporal shift performed during the analysis. At the top of the figure, the relevant information is displayed: "baz" the backazimuth, "ain" the angle of incidence, "epi" the epicentral distance, "mag" the magnitude, "ϕ" the S_{fast} polarization direction, "t_d" the time-delay, "p" the polarization direction of the corrected S-wave, "grade": the qualitative grade and "flt" the filter used. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

[4] https://www.fdsn.org/xml/station/.

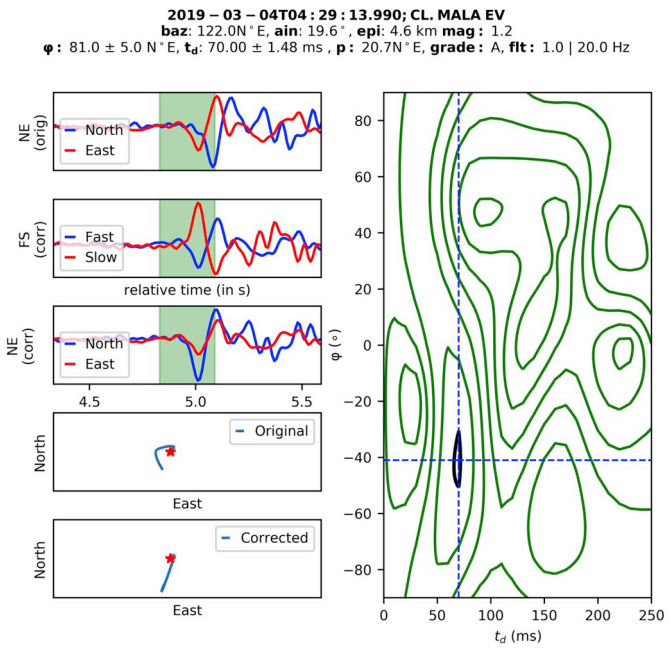

Fig. 4. Summary plot of the analysis with the EV method. Left: Filtered horizontal waveforms in the initial NE axial system (top), the rotated FS system corrected for anisotropy (middle) and the NE axial system corrected for anisotropy (bottom). The two particle motion diagrams for the initial (top) and corrected (bottom) waveforms are shown. Right: Contour plot of the second eigenvalue for each ϕ and t_d pair. The bold black contour highlights the 95% confidence level. Information notation at the top of the figure as in Fig. 3.

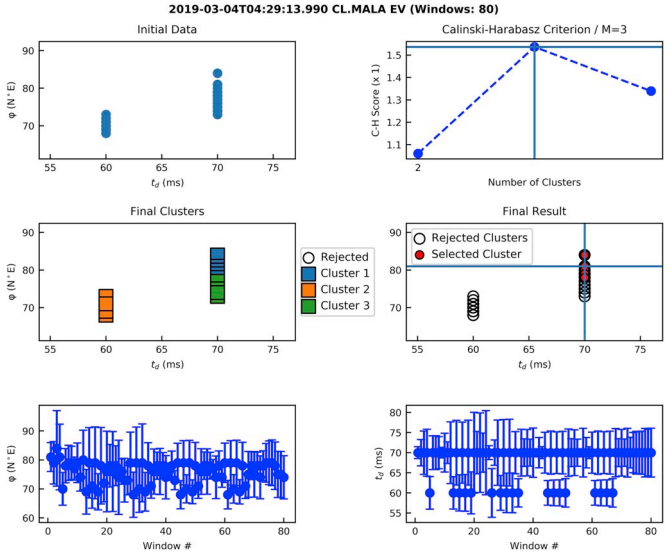

Fig. 5. Summary plot of the analysis with the CA method. Top left: Distribution of initial data, with each point representing a candidate signal window. Top right: Caliński and Harabasz (1974) score for each candidate number of clusters M (for more information see Teanby et al., 2004b). Middle left: Clustering results. Middle right: Cluster and window selection. Bottom left: Variation of ϕ for different windows. Bottom right: Variation of t_d for different windows. The window duration increases with the window number.

diagrams or variations of SwS parameters).

3. Case study: the Western Gulf of Corinth

To examine the performance and applicability of the software we utilized recordings from the Western Gulf of Corinth (WGoC), an area in Central Greece characterized by intense seismicity (Makropoulos et al., 2012) and high tectonic activity (Armijo et al., 1996). The area is monitored by two dense seismological networks. The Hellenic Unified Seismological Network (HUSN) is a joint venture of Greek institutions that has been in operation since 2008. The Corinth Rift Laboratory Network (CRLN) is a local network that covers the majority of the WGoC since the early 2000s (Lyon-Caen et al., 2004). The combination of high seismic activity and dense networks offers the opportunity for building a large initial dataset. We selected 263 events, with magnitudes in the range of 0.6–4.0, between the 1st and 31st of March 2019, a month-long period that includes the occurrence of a $M_w = 5.1$ earthquake on 30/03/2019 10:46:19 UTC. The shear-wave window was set at 45°, according to previous studies (Papadimitriou et al., 1999; Kaviris et al., 2017, 2018b), and the final dataset consisted of 1447 event-station pairs. The angle of incidence for each pair was calculated with TauP, through the software, with the local velocity model of Rigo et al. (1996). A control group of 56 randomly selected events (135 pairs) was analyzed with the MAN method. All events were processed with the CA-EV method, with the preferences showcased in Table 1. Phase arrivals, as well as the event catalogue, were obtained from the Seismological Laboratory of the National and Kapodistrian University of Athens.

The grading algorithm determined 136 null measurements, while the number of results per grade was 930 for "E", 20 for "D", 146 for "C", 177 for "B" and 30 for "A". 7 pairs did not produce a result due to errors during CA (no clusters were produced after applying all relevant rejection criteria). Concerning the null measurements, 61 corresponded to a station whose N-S component did not operate for the whole study period, meaning that only 75 measurements were truly "null". Moreover, "E"-graded results included 158 measurements with a t_d of 200.0 ms or higher, indicative of cycle-skipping. We then compared common results obtained from both methods, accepting only the ones graded between "C" and "A". To investigate the association (assumed to be linear) between the manual and automatic measurements, we employed the Pearson Correlation Coefficient (PCC) which indicates the distance of the data from the best fit model, in our case $y = x$ (Asuero et al., 2006). The PCC ranges between −1 and +1 (negative and positive correlation, respectively), with values close to 0 indicating no association between the data and the model. For ϕ, manual and automatic measurements show a PCC of 0.71 (Fig. 6). It is noted that the circular character of ϕ was considered in estimating the correlation (Berens, 2009). For t_d, the same quantity is 0.53. This indicates that, after enforcing strict criteria, the automatic method can yield satisfyingly reliable results. The grading algorithm managed to successfully identify problematic event-station pairs, such as all recordings in defective stations. Further details about the results are presented in the Appendix.

Table 1
Parameters used in the automated analysis of the WGoC dataset. C_{crit} is the minimum value for the Duda and Hart (1973) criterion (for more information see Teanby et al., 2004b).

Parameter	Value
Maximum trial t_d (ms)	250.0
Minimum window-start point (s)[a]	−0.10
Step for window-start points (s)[a]	0.20
Minimum window-end point (s)[a]	0.10
Step for window-end points (s)[a]	0.02
Maximum S-P (s)	1.50
Signal window for period determination (s)[a]	0.50
Minimum S-wave period (s)	0.10
Maximum S-wave period (s)	2.00
Period Factor	3.0
C_{crit}	3.2
Maximum number of clusters	20
Minimum number of points per cluster	5
Linkage Criterion	"ward"

[a] Relative to the shear-wave arrival.

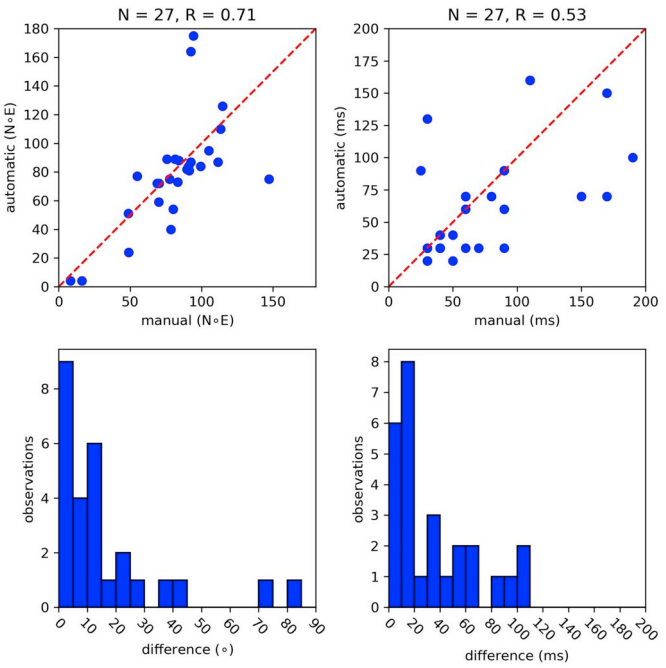

Fig. 6. Top: Comparison of common measurements, with the MAN and CA-EV methods, for ϕ (left) and t_d (right). The points close to the dashed line are indicative of high similarity. Bottom: Histograms of the differences between common measurements for the two parameters. For ϕ, the true angular difference was considered. Only results graded "C", "B" and "A" were used.

4. Conclusions

In the current article we present the first publicly available version of the Pytheas software. Pytheas is an open-access application that aims to combine existing SwS analysis methods with an intuitive GUI. The software was designed and developed with ease of use in mind, as well as compatibility with standard file formats. Although we showcased a suggested workflow, the user can freely adopt any of the offered methods and combine them as they prefer, to maximize productivity and effectiveness. Integration of databases and popular formats are a necessity for trivializing the time-consuming stages of data collection and results retrieval, enabling the analyst to focus more on the purely scientific part of a study.

The application of the software in the WGoC, an area with a vast wealth of available data, exhibited Pytheas' strong points. It took 30.0 h (using a desktop machine with an Intel i5 8400 processor and 8 GB of RAM) to perform a complete analysis on 1447 event-station pairs. The high rejection percentage (65%) was offset by the number of produced results (499 for one month). The manual processing of the 135 event-station pairs required 4.5 h of active user involvement. Thus, it is evident that the automatic analysis (~1.2 min per pair) offers a much more effective way of analyzing data, than manually (~2.0 min per pair). The initial and final manual validation were paramount in obtaining constraints for the automatic process (CA-EV), which, in turn, reduced the analysis time by processing the rest of the dataset. Pairs analyzed by both methods (MAN and CA-EV) provided a moderate correlation of results (0.71 for ϕ and 0.53 for t_d), essentially verifying the robustness of the automatic method. The grading algorithm achieved the improvement of the final database, by rejecting possible artifacts and low-quality measurements.

One of the major planned updates to Pytheas is the addition of a mode for performing splitting studies on teleseismic phases. Additionally, an interface has already been designed for acquiring event and station metadata from FDSN, as well as waveform data, through the software itself. The integrated database will also enable the relatively straightforward implementation of interpretation-related functions, such as the production of rose-diagrams and temporal variations plots of both ϕ and normalized t_d. The latter are essential for identifying earthquake and volcanic precursors through SwS (e.g. Crampin et al., 2015). The refinement of the grading algorithm will also be a major addition to the software, by introducing weighting for individual scores and rendering the process modular. Filter selection and optimization is also an available space for future expansion of Pytheas, since filtering is a parameter that can greatly affect results (e.g. Savage et al., 2010). Finally, to maximize effectiveness and fully exploit the potential of parallel computing in modern multi-core systems, a multiprocessing solution will be implemented for the CA method.

In conclusion, we hope that the presented software will prove useful for the seismological community. By offering an accessible application that reduces the effort required to analyze SwS from local waveforms, the goal is to assist broadening the audience of the field. Pytheas has been developed to be suitable for a wide range of researchers. Students and young scientists can benefit from the GUI-driven approach of the software to start their projects on SwS. At the same time, Pytheas can assist experienced researchers by providing a complete framework for analysis, customizing the workflow to their respective needs and harnessing the potential of directly evaluating automatically acquired results through manual inspection, leading to a higher quality dataset.

Software availability

Pytheas (Version 0.1.0, April 2019) can be accessed through its dedicated repository in GitHub (https://www.github.com/ispingos/pytheas-splitting) under the GNU General Public License v3.0 (or later). The software is written in Python (3.7). Software and hardware requirements are described in the software's documentation.

Authorship statement

Ioannis Spingos designed and developed Pytheas and was responsible for the general supervision of the project (including data acquisition, testing coordination and article writing). George Kaviris and Christos Millas participated in designing and testing Pytheas. Panayotis Papadimitriou and Nicholas Voulgaris aided in implementing the shear-wave splitting analysis methods. All authors were involved in the writing and editing of the article.

Acknowledgements

The authors would like to thank the two anonymous reviewers for their constructive comments, as well as the personnel of the Hellenic Unified Seismological Network and the Corinth Rift Laboratory Network (https://doi.org/10.15778/RESIF.CL) for the installation and operation of the stations used in the current article. Moreover, they would like to express their gratitude to all the developers of packages Pytheas is making use of, for providing their work freely to the scientific community.

Appendix

The dominant polarization direction in the WGoC is WNW-ESE, perpendicular to the extension direction of the rift, with the exception of the Sergoula-Pyrgos area, located near the northern coast (e.g. Giannopoulos et al., 2015; Kaviris et al., 2017). Results from the analysis of the current

dataset showcased similar orientations. Indicatively, we present rose diagrams for five stations, with results acquired through the CA-EV (Fig. A.1) and MAN (Fig. A.2) methods. Four of the stations show a mean ϕ in the general direction of WNW-ESE. This is in accordance with the APE model; the regional stress aligns microcracks according to its orientation. The SERG station (located in the Sergoula-Pyrgos area) is deviating towards a NE-SW orientation, a difference that could be potentially linked to a local stress field related to tectonic structures in the region (Kaviris et al., 2018b).

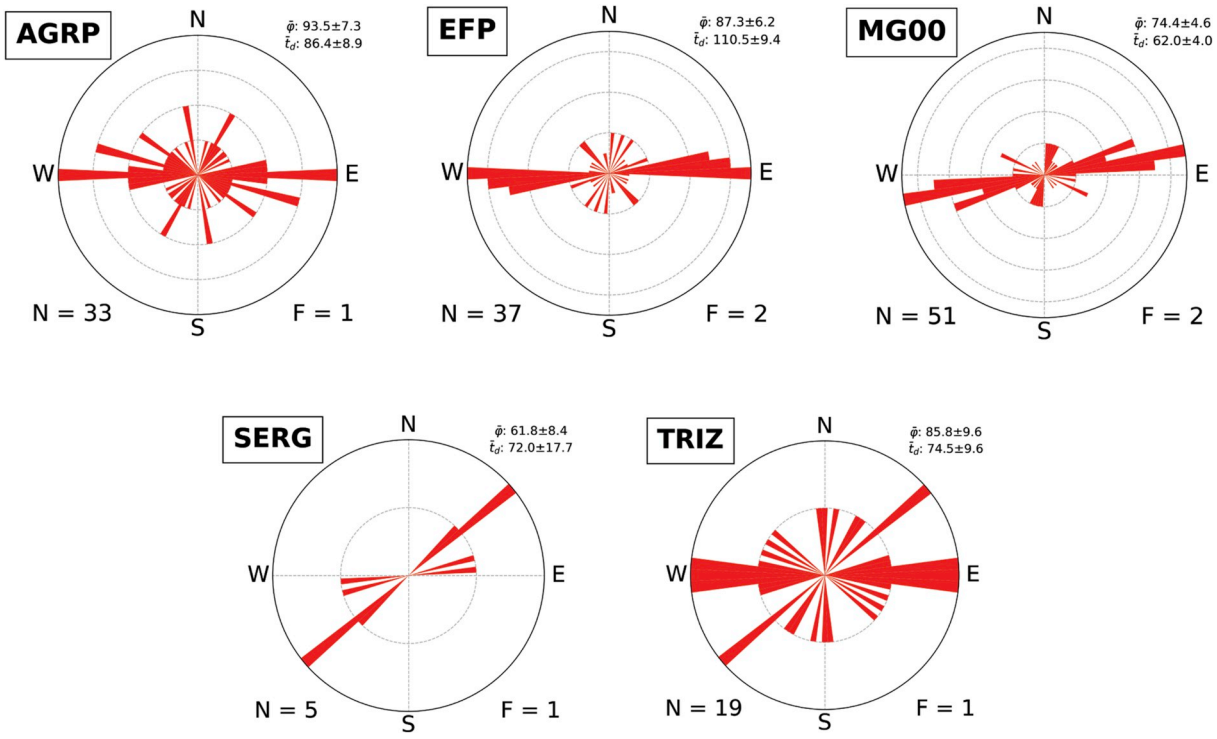

Fig. A.1. Rose diagrams showcasing results from the analysis of the March 2019 WGoC dataset with the CA-EV method. Measurements presented are graded as "C", "B" or "A" (the same grades used for the comparison between methods in Fig. 6). *N* is the number of measurements, *F* the number of measurements per grid interval, $\overline{\phi}$ the mean polarization direction and $\overline{t_d}$ the mean time-delay. The last two parameters are accompanied by their respective standard error of mean.

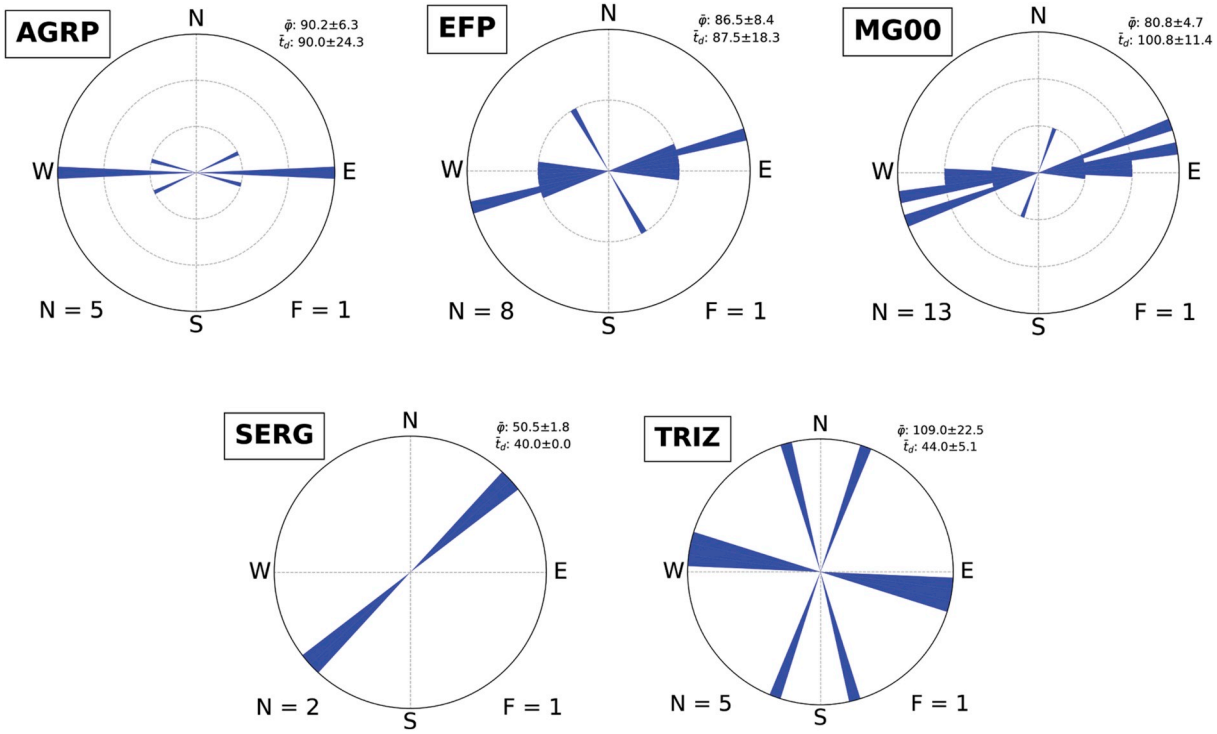

Fig. A.2. Rose diagrams showcasing results from the analysis of the March 2019 WGoC dataset with the MAN method. Measurements presented are graded as "C", "B" or "A" (the same grades used for the comparison between methods in Fig. 6). Notation as in Fig. A1.

References

Akazawa, T., 2004. A technique for automatic detection of onset time of P- and S-phases in strong motion records. In: 13 Th World Conf. Earthq. Eng, vol. 786.

Armijo, R., Meyer, B., King, G., Rigo, A., Papanastassiou, D., 1996. Quaternary evolution of the Corinth Rift and its implications for the late Cenozoic evolution of the Aegean. Geophys. J. Int. 11–53.

Aster, R.C., Shearer, P.M., Berger, J., 1990. Quantitative measurements of shear wave polarizations at the Anza Seismic Network, southern California: implications for shear wave splitting and earthquake prediction. J. Geophys. Res. 95, 12449. https://doi.org/10.1029/JB095iB08p12449.

Asuero, A.G., Sayago, A., González, A.G., 2006. The correlation coefficient: an overview. Crit. Rev. Anal. Chem. 36, 41–59. https://doi.org/10.1080/10408340500526766.

Baccheschi, P., Margheriti, L., Steckler, M.S., 2008. SKS splitting in Southern Italy: anisotropy variations in a fragmented subduction zone. Tectonophysics 462, 49–67. https://doi.org/10.1016/j.tecto.2007.10.014.

Berens, P., 2009. CircStat: a MATLAB toolbox for circular statistics. J. Stat. Softw. 31 https://doi.org/10.18637/jss.v031.i10.

Bernard, P., Chouliaras, G., Tzanis, A., Briole, P., Bouin, M.P., Tellez, J., Stavrakakis, G., Makropoulos, K., 1997. Seismic and electrical anisotropy in the Mornos delta, Gulf of Corinth, Greece, and its relationship with GPS strain measurements. Geophys. Res. Lett. https://doi.org/10.1029/97GL02102.

Bernard, P., Zollo, A., 1989. Inversion of near-source S polarization for parameters of double-couple point sources. Bull. Seismol. Soc. Am. 79, 1779–1809.

Beyreuther, M., Barsch, R., Krischer, L., Megies, T., Behr, Y., Wassermann, J., 2010. ObsPy: a Python toolbox for seismology. Seismol. Res. Lett. 81, 530 LP–533.

Bianco, F., Scarfì, L., Del Pezzo, E., Patanè, D., 2006. Shear wave splitting changes associated with the 2001 volcanic eruption on Mt Etna. Geophys. J. Int. 167, 959–967. https://doi.org/10.1111/j.1365-246X.2006.03152.x.

Booth, D.C., Crampin, S., 1985. Shear-wave polarizations on a curved wavefront at an isotropic free surface. Geophys. J. Int. 31–45.

Booth, D.C., Park, M., 1988. Precursory Shear Wave Splitting Anza Network and Seismic Gap Structure of Hartzell and Brune [1979], vol. 93, pp. 3339–3356.

Bouin, M.-P., Téllez, J., Bernard, P., 1996. Seismic anisotropy around the Gulf of Corinth, Greece, deduced from three-component seismograms of local earthquakes and its relationship with crustal strain. J. Geophys. Res. Solid Earth 101, 5797–5811. https://doi.org/10.1029/95JB03464.

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

Buchbinder, G.G.R., 1989. Azimuthal variations in P-wave travel times and shear-wave splitting in the Charlevoix seismic zone. Tectonophysics 165, 293–302. https://doi.org/10.1016/0040-1951(89)90054-1.

Chang, E.T.Y., Liang, W.-T., Tsai, Y.-B., 2009. Seismic shear wave splitting in upper crust characterized by Taiwan tectonic convergence. Geophys. J. Int. 177, 1256–1264. https://doi.org/10.1111/j.1365-246X.2009.04110.x.

Chen, T.C., Booth, D.C., Crampin, S., 1987. Shear-wave polarizations near the North Anatolian Fault - III. Observations of temporal changes. Geophys. J. R. Astron. Soc. 91, 287–311. https://doi.org/10.1111/j.1365-246X.1987.tb05228.x.

Crampin, S., 1986. Anisotropy and transverse isotropy. Geophys. Prospect. 94–99.

Crampin, S., Gao, Y., 2006. A review of techniques for measuring shear-wave splitting above small earthquakes. Phys. Earth Planet. Inter. 159, 1–14. https://doi.org/10.1016/j.pepi.2006.06.002.

Crampin, S., Gao, Y., Bukits, J., 2015. A review of retrospective stress-forecasts of earthquakes and eruptions. Phys. Earth Planet. Inter. 245, 76–87. https://doi.org/10.1016/j.pepi.2015.05.008.

Crampin, S., Gao, Y., Peacock, S., 2008. Stress-forecasting (not predicting) earthquakes: a paradigm shift. Geology 36, 427. https://doi.org/10.1130/G24643A.1.

Crampin, S., Volti, T., Stefánsson, R., 1999. A successfully stress-forecast earthquake. Geophys. J. Int. 1–5.

Crampin, S., Zatsepin, S., 1997. Modelling the compliance of crustal rock—II. Response to temporal changes before earthquakes. Geophys. J. Int. 129, 495–506. https://doi.org/10.1111/j.1365-246X.1997.tb04489.x.

Caliński, T., Harabasz, J., 1974. A dendrite method for cluster analysis. Commun. Stat. 3, 1–27. https://doi.org/10.1080/03610927408827101.

Crotwell, H.P., Owens, T.J., Ritsema, J., 1999. The TauP toolkit: flexible seismic travel-time and Ray-path utilities. Seismol. Res. Lett. 70, 154–160. https://doi.org/10.1785/gssrl.70.2.154.

Duda, R.O., Hart, P.E., 1973. Pattern Classification and Scene Analysis, first ed. Wiley, New York.

Evans, R., 1984. Effects of the free surface on shear wavetrains. Geophys. J. Int. 76, 165–172. https://doi.org/10.1111/j.1365-246X.1984.tb05032.x.

Fukao, Y., 1984. Evidence from core-reflected shear waves for anisotropy in the Earth's mantle. Nature 309, 695–698. https://doi.org/10.1038/309695a0.

Gao, S.S., Liu, K.H., 2012. AnisDep: a FORTRAN program for the estimation of the depth of anisotropy using spatial coherency of shear-wave splitting parameters. Comput. Geosci. 49, 330–333. https://doi.org/10.1016/j.cageo.2012.01.020.

Grund, M., 2017. StackSplit - a plugin for multi-event shear wave splitting analyses in SplitLab. Comput. Geosci. 105, 43–50. https://doi.org/10.1016/j.cageo.2017.04.015.

Giannopoulos, D., Sokos, E., Konstantinou, K.I., Tselentis, G.-A., 2015. Shear wave splitting and VP/VS variations before and after the Efpalio earthquake sequence, western Gulf of Corinth, Greece. Geophys. J. Int. 200, 1436–1448. https://doi.org/10.1093/gji/ggu467.

Hao, P., Gao, Y., Crampin, S., 2008. An Expert System for measuring shear-wave splitting above small earthquakes. Comput. Geosci. 34, 226–234. https://doi.org/10.1016/J.CAGEO.2007.03.010.

Hatzfeld, D., Karagianni, E., Kassaras, I., Kiratzi, a., Louvari, E., Lyon-Caen, H., Makropoulos, K., Papadimitriou, P., Bock, G., Priestley, K., 2001. Shear wave anisotropy in the upper mantle beneath the Aegean related to internal deformation. J. Geophys. Res. 106, 30737. https://doi.org/10.1029/2001JB000387.

Hunter, J.D., 2007. Matplotlib: a 2D graphics environment. Comput. Sci. Eng. 9, 90–95. https://doi.org/10.1109/MCSE.2007.55.

Kaviris, G., Fountoulakis, I., Spingos, I., Millas, C., Papadimitriou, P., 2018. Mantle dynamics beneath Greece from SKS and PKS seismic anisotropy study. Acta Geophys. https://doi.org/10.1007/s11600-018-0225-z.

Kaviris, G., Millas, C., Spingos, I., Kapetanidis, V., Fountoulakis, I., Papadimitriou, P., Voulgaris, N., Makropoulos, K., 2018. Observations of shear-wave splitting parameters in the Western Gulf of Corinth focusing on the 2014 Mw = 5.0 earthquake. Phys. Earth Planet. Inter. 282, 60–76. https://doi.org/10.1016/j.pepi.2018.07.005.

Kaviris, G., Papadimitriou, P., Kravvariti, P., Kapetanidis, V., Karakonstantis, A., Voulgaris, N., Makropoulos, K., 2015. A detailed seismic anisotropy study during the 2011-2012 unrest period in the Santorini Volcanic Complex. Phys. Earth Planet. Inter. 238, 51–88. https://doi.org/10.1016/j.pepi.2014.11.002.

Kaviris, G., Spingos, I., Kapetanidis, V., Papadimitriou, P., Voulgaris, N., Makropoulos, K., 2017. Upper crust seismic anisotropy study and temporal variations of shear-wave splitting parameters in the western Gulf of Corinth (Greece) during 2013. Phys. Earth Planet. Inter. 269 https://doi.org/10.1016/j.pepi.2017.06.006.

Kaviris, G., Spingos, I., Millas, C., Kapetanidis, V., Fountoulakis, I., Papadimitriou, P., Voulgaris, N., Drakatos, G., 2018. Effects of the January 2018 seismic sequence on shear-wave splitting in the upper crust of Marathon (NE Attica, Greece). Phys. Earth Planet. Inter. 285, 45–58. https://doi.org/10.1016/j.pepi.2018.10.007.

Krischer, L., Megies, T., Barsch, R., Beyreuther, M., Lecocq, T., Caudron, C., Wassermann, J., 2015. ObsPy: a bridge for seismology into the scientific Python ecosystem. Comput. Sci. Discov. 8 https://doi.org/10.1088/1749-4699/8/1/014003.

Lyon-Caen, H., Papadimitriou, P., Deschamps, A., Bernard, P., Makropoulos, K., Pacchiani, F., Patau, G., 2004. First results of the CRLN seismic network in the western Corinth Rift: evidence for old-fault reactivation. Compt. Rendus Geosci. 336, 343–351. https://doi.org/10.1016/j.crte.2003.12.004.

Makropoulos, K., Kaviris, G., Kouskouna, V., 2012. An updated and extended earthquake catalogue for Greece and adjacent areas since 1900. Nat. Hazards Earth Syst. Sci. 12, 1425–1430. https://doi.org/10.5194/nhess-12-1425-2012.

Miller, V., Savage, M., 2001. Changes in seismic anisotropy after volcanic eruptions: evidence from mount Ruapehu. Science 293, 2231–2233. https://doi.org/10.1126/science.1063463.

Nolte, K.A., Tsoflias, G.P., Bidgoli, T.S., Watney, W.L., 2017. Shear-wave anisotropy reveals pore fluid pressure–induced seismicity in the U.S. midcontinent. Sci. Adv. 3, e1700443 https://doi.org/10.1126/sciadv.1700443.

Papadimitriou, P., Kaviris, G., Makropoulos, K., 1999. Evidence of shear-wave splitting in the eastern Corinthian Gulf (Greece). Phys. Earth Planet. Inter. 114, 3–13.

Pedregosa, F., Varoqaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. Scikit-learn: machine learning in Python. J. Mach. Learn. Res. 12, 2825–2830. https://doi.org/10.1007/s13398-014-0173-7.2.

Peacock, S., Crampin, S., Booth, D.C., Fletcher, J.B., 1988. Shear wave splitting in the Anza seismic gap, southern California: temporal variations as possible precursors. J. Geophys. Res. 93, 3339–3356. https://doi.org/10.1029/JB093iB04p03339.

Piccinini, D., Pastori, M., Margheriti, L., 2013. ANISOMAT+: an automatic tool to retrieve seismic anisotropy from local earthquakes. Comput. Geosci. 56, 62–68. https://doi.org/10.1016/j.cageo.2013.01.012.

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, 392–409. https://doi.org/10.1785/0220160191.

Rigo, A., Lyon-Caen, H., Armijo, R., Deschamps, A., Hatzfeld, D., Makropoulos, K., Papadimitriou, E., Kassaras, I., 1996. A microseismic study in the western part of the gulf of Corinth (Greece) implication for large-scale normal faulting mechanisms. Geophys. J. Int. 126, 663–688.

Roman, D.C., Savage, M.K., Arnold, R., Latchman, J.L., De Angelis, S., 2011. Analysis and forward modeling of seismic anisotropy during the ongoing eruption of the Soufrire Hills Volcano, Montserrat, 1996-2007. J. Geophys. Res. Solid Earth 116, 1996–2007. https://doi.org/10.1029/2010JB007667.

Savage, M.K., Silver, P.G., 1993. Mantle deformation and tectonics: constraints from seismic anisotropy in the western United States. Phys. Earth Planet. Inter. 78, 207–227. https://doi.org/10.1016/0031-9201(93)90156-4.

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. https://doi.org/10.1029/2010JB007722.

Schorlemmer, D., Schorlemmer, D., Euchner, F., Kästli, P., Saul, J., 2011. QuakeML: status of the XML-based seismological data exchange format. Ann. Geophys. 54, 59–65. https://doi.org/10.4401/ag-4874.

Silver, P.G., Chan, W.W., 1991. Shear wave splitting and sub continental mantle deformation. J. Geophys. Res. 96, 429–454. https://doi.org/10.1029/91JB00899.

Teanby, N., Kendall, J.-M., Jones, R.H., Barkved, O., 2004. Stress-induced temporal variations in seismic anisotropy observed in microseismic data. Geophys. J. Int. 156, 459–466. https://doi.org/10.1111/j.1365-246X.2004.02212.x.

Teanby, N., Kendall, J.-M., van der Baan, M., 2004. Automation of shear-wave splitting measurements using cluster analysis. Bull. Seismol. Soc. Am. 94, 453–463. https://doi.org/10.1785/0120030123.

van Rossum, G., 1995. Python Tutorial. Technical Report CS-R9526. Amsterdam.
Vinnik, L.P., Farra, V., Romanowicz, B., 1989. Azimuthal anisotropy in the earth from observations of SKS at geoscope and Nars broadband stations. Bull. Seismol. Soc. Am. 79, 1542–1558.
Walpole, J., Wookey, J., Masters, G., Kendall, J.M., 2014. A uniformly processed data set of SKS shear wave splitting measurements: a global investigation of upper mantle anisotropy beneath seismic stations. Geochem. Geophys. Geosyst. 15, 1991–2010. https://doi.org/10.1002/2014GC005278.
Walsh, E., Arnold, R., Savage, M.K., 2013. Silver and chan revisited. J. Geophys. Res. Solid Earth 118, 5500–5515. https://doi.org/10.1002/jgrb.50386.
Ward, J.H., 1963. Hierarchical grouping to optimize an objective function. J. Am. Stat. Assoc. 58, 236–244. https://doi.org/10.1080/01621459.1963.10500845.
Wüstefeld, A., Bokelmann, G., 2007. Null detection in shear-wave splitting measurements. Bull. Seismol. Soc. Am. 97, 1204–1211. https://doi.org/10.1785/0120060190.
Wüstefeld, A., Bokelmann, G., Zaroli, C., Barruol, G., 2008. SplitLab: a shear-wave splitting environment in Matlab. Comput. Geosci. 34, 515–528. https://doi.org/10.1016/j.cageo.2007.08.002.
Zaccarelli, L., Bianco, F., Zaccarelli, R., 2012. Splitting parameter yield (SPY): a program for semiautomatic analysis of shear-wave splitting. Comput. Geosci. 40, 138–145. https://doi.org/10.1016/j.cageo.2011.08.006.
Zatsepin, S., Crampin, S., 1997. Modelling the compliance of crustal rock—I. Response of shear-wave splitting to differential stress. Geophys. J. Int. 129, 477–494. https://doi.org/10.1111/j.1365-246X.1997.tb04488.x.