A new automatic S-onset detection technique; application in microseismic data analysis

Athanasios Lois(1), Efthimios Sokos(1), Nikolaos Martakis(2), Paraskevas Paraskevopoulos(1) and G-Akis Tselentis(1)

(1) University of Patras, Seismological Laboratory, Rio 26504, Greece
(2) Landtech Enterprises, Athens Greece

ABSTRACT

Automatic S-wave onset time identification constitutes a difficult problem for seismologists, due to the increased level of seismic energy by P-coda waves, before the S-wave arrival. Most of the algorithms that have been proposed up to now, are mainly based on the polarization features of the seismic waves. In this work we propose a new time domain technique for the automatic determination of the S-phase arrival onsets, and present its implementation on three-component single station data. Over small time windows, the eigenproblem of the covariance matrix is solved and the eigenvalue, corresponding to the maximum eigenvector is retained for further processing. In this way we obtain a time series of maximum eigenvalues, which serves as a characteristic function, whose statistical properties provide a first S-phase arrival time estimation. A multi-window approach combined with an energy-based weighting scheme is also applied, in order to reduce the algorithm’s dependence on the moving window’s length and provide a weighted S phase onset. Automatic picks are compared against manual reference picks, resulting in sufficiently good results, regarding the accuracy as well as noise robustness.
INTRODUCTION

A large number of algorithms have been proposed and used on seismic networks, for phase identification. The accurate determination of both compressional and transversal seismic waves is important for earthquake location and focal mechanism determination, but also essential for other applications such as passive seismic tomography investigations (Tselentis et al., 2011b). Regarding the event detection as well as P-phase identification, a number of techniques exists, involving energy criteria (Allen, 1978; Baer and Kradolfer, 1987), polarization analysis tests (Montalbetti and Kanasewich, 1970; Vidale, 1986; Magotra et al. 1987; Ruud and Husebye, 1992), fuzzy logic tests (Chu and Mendel, 1994), artificial neural networks (Dai and MacBeth, 1995), higher order statistics (Saragiotis et al., 2002), wavelet analysis (Anant and Dowl, 1997) etc.

In contrast to P-onset time estimation, algorithms that deal with S-phase arrival are mainly based on polarization attributes of the seismic signal, and their development is a topic of ongoing research. The identification of phase-arrival times is traditionally done by experienced seismologists, but due to the character of the later arriving shear waves, even manual S-wave picking is often uncertain and inconsistent. Moreover the S-onset time identification problem can be further complicated due to converted waves, which can be misinterpreted as the direct S onset (Sokos et al., 2012).

It is well known that polarization measurements may indicate the arrival of seismic phases, since highly linear particle motion may be associated with body wave arrivals. Using three-component data and linear algebra’s fundamentals, Flinn (1965) designed a time-varying non-linear filter to enhance particle motion which is rectilinear in a particular direction in three dimensional space. The usefulness of such a processor lies in its ability to separate compressional wave motion from the shear or surface wave motion, when the distance and azimuth to the seismic source is specified.
Montalbetti and Kanasewich (1970) modified the time domain polarization filter, originally proposed by Flinn, and used it to increase the signal to noise ratio of teleseismic body wave phases. A complex polarization filter was proposed by Vidale (1986) as an extension of the Montalbetti and Kanasewich (1970) scheme, where the imaginary part of the signal is the Hilbert Transform of the real part. Another time domain technique based on Flinn’s method was proposed by Jurkevics (1988). He extended Flinn’s work by including a frequency decomposition and the application to arrays of three-component sensors.

Cichowich (1993) combined the significant characteristics of an S-wave arrival into one characteristic function that consists of a product of different polarization filters such as rectilinearity, directivity and the ratio between transverse and total energy. Finally hybrid methods, have also been proposed, by Wang and Teng (1997), Bai and Kennett (2000), Saragiotis et al. (2000) Diehl et al. (2009), Gentili and Michelini (2006), Nippres et al. (2010), Kuperkoch et al. (2011), that involve in addition to polarization analysis, time series analysis techniques, autoregressive prediction, STA/LTA detectors, pattern recognition schemes, artificial neural networks, wavelet analysis and higher order statistics. In this paper we propose an S-wave detector scheme, which is comprised by a mixture of polarization analysis and higher order statistics theory in time domain. It is an almost parameter free algorithm, which is straightforward to implement and demands low computational resources.

THEORY

In the following section, we describe the basic principles of polarization analysis and explain how the diagonalization of the data covariance matrix relates to the polarization features of a seismic wave. Moreover a brief introduction to Higher Order Statistics (HOS) is presented, and their usefulness on seismic phase automatic picking is also discussed.
Polarization Analysis

The problem can be formulated as follows. Given a zero-mean (3x1) data vector \( s \) in z, n and e space (z=vertical, n=north-south, e=east-west), we need to identify the direction in which projection \( y \) exhibits the maximum variance (Magotra et al., 1987). This can be written as a dot product:

\[
y = \hat{u}^T \bar{s} = [u_1 \ u_2 \ u_3]^T \begin{bmatrix} s_z \\ s_n \\ s_e \end{bmatrix}
\]

where \( \hat{u} \) is a unitary vector (\( \|\hat{u}\|=1 \)), in the direction of the source’s azimuth. Vector \( \bar{s} \) is assumed to be zero mean, so for the mean value and variance of \( y \) we can write:

\[
E[y] = E[\hat{u}^T \bar{s}] = 0
\]

\[
\sigma_y^2 = E[y^2] = \hat{u}^T C_s \hat{u}
\]

where \( E[.\] denotes the statistical expectation and \( C_s \) the covariance matrix of \( \bar{s} \):

\[
C_s = E[\bar{s}\bar{s}^T] = \begin{bmatrix} s_z^2 & s_z s_n & s_z s_e \\ s_n s_z & s_n^2 & s_n s_e \\ s_e s_z & s_e s_n & s_e^2 \end{bmatrix}
\]

where \( s_z^2 = E[z^2], s_n^2 = E[n^2], s_e^2 = E[e^2] \) and \( s_z n, s_n e, s_e z \) are the covariances between \( z, n, e \) components. The projection’s variance \( \sigma_y^2 \) can be maximized, using Lagrange multipliers, given the constraint \( \|\hat{u}\|=1 \). The above procedure is equal to the maximization of the following expression:

\[
\Phi(\hat{u}; \lambda) = \sigma_y^2 - \lambda (\|\hat{u}\| - 1) = \hat{u}^T C_s \hat{u} - \lambda (\hat{u}^T \hat{u} - 1)
\]
By differentiating this expression with respect to $\vec{u}$ and $\lambda$, the following partial differential equations are obtained:

$$\frac{\partial \Phi}{\partial \vec{u}} = (\vec{u}^T \vec{u} - 1), \text{ and}$$

$$\frac{\partial \Phi}{\partial \vec{u}} = C_{ij} \vec{u} - \lambda \vec{u} = (C - \lambda I) \vec{u} = 0$$

The first one defines the initial constraint and the second one suggests that $\lambda$ is eigenvalue of the covariance matrix, while $\vec{u}$ is the corresponding eigenvector. Moreover the eigenvectors, obtained by the diagonalization of the covariance matrix, represent an orthonormal base of the 3-dimensional space and form an ellipsoid (polarization ellipsoid), that best fits to the data in a least-squares sense. From equation (6) it is clear that the polarization characteristics of a signal are obtained by solving the eigenproblem of the data covariance matrix since, once the principal axes of the polarization ellipsoid are estimated, the particle motion is determined. Using attributes computed from the principal axes, information describing the degree of linear polarization, the directivity of the particle motion, the azimuth of P-wave propagation as well as the apparent incidence angle of rectilinear motion, is extracted. For example, according to Jurkevics, (1988) if $\lambda_1 > \lambda_2 > \lambda_3$, the rectilinearity is given by the following relation:

$$Rect = 1 - \frac{\lambda_2 + \lambda_3}{2\lambda_1}$$

which is expected to be close to one for both P and S phases, while the planarity of the particle motion is given by:

$$Plan = 1 - \frac{2\lambda_3}{\lambda_1 + \lambda_2}$$

which is expected to be close to zero for the P arrival and close to one for the first S wave arrival.
Additionally if $\vec{u}_{ij}$ are the corresponding eigenvectors, with $i=1, 2, 3$ the three components (vertical, north-south and east-west) and $j=1, 2, 3$ the three direction cosines then the azimuth of P-wave propagation is given by:

$$P_{\text{azimuth}} = \tan^{-1}\left(\frac{\vec{u}_{i2}}{\vec{u}_{i3}}\right)$$

(9)

and the incident angle of rectilinear motion can be obtained by the relation:

$$P_{\text{incidence}} = \cos^{-1}\left(|\vec{u}_{i1}|\right)$$

(10)

**Higher Order Statistics**

The mean value, variance, autocorrelation and power spectrum, constitute the first and second-order statistics, respectively, and are extensively used to describe processes that are linear and Gaussian distributed. However, most of the processes in earth sciences deviate from linearity and Gaussianity. Such processes can be studied through Higher Order Statistics (HOS).

Specifically, let’s assume the N-sample, real and zero–mean process $\{X(k)\}$, that is fourth–order stationary. Its second-, third- and fourth–order moments are defined as (Nikias et al. 1993):

$$R_2(m) = E\{X(k)X(k + m)\}$$
$$R_3(m, n) = E\{X(k)X(k + m)X(k + n)\}$$
$$R_4(m, n, l) = E\{X(k)X(k + m)X(k + n)X(k + l)\}$$

(11)

$E\{\cdot\}$ denotes the expectation, and for a continuous random variable $x$ is given by:

$$E\{x\} = \int_{-\infty}^{\infty} xf(x)dx$$

(12)

where $f(x)$ is the probability density function of $x$. Note that, $R_2(0)$ equals to the variance $\sigma^2(x)$ of the random variable $x$. Another set of statistical parameters that can be used, due to their excellent noise-suppressing properties, are cumulants and can be expressed in terms of the
moments. The following formulas denote the third- and fourth- order cumulant sequences of \(\{X(k)\}\):

\[
C_3(m, n) = R_3(m, n) \\
C_4(m, n, l) = R_4(m, n, l) - 3(R_2(m))^2
\]  \(\text{(13)}\)

For the zero-lag case, that is \(m=n=l=0\), we obtain the skewness \(C_3(0,0)\) and kurtosis \(C_4(0,0,0)\).

Skewness provides a measure of symmetry of the distribution and is expected to become zero if the distribution is symmetrical. Furthermore, it takes negative values if the distribution contains outliers to the left and positive values in the opposite case. The fourth-order zero-lag cumulant, the kurtosis, provides a measure of heaviness of the tails of the distribution, and takes the value three for Gaussian distributed random variables. Kurtosis values larger than three, indicate widening of the distribution, while narrowing is indicated for values smaller than three.

HOS parameters, were first used in seismic phase automatic identification by Saragiotis et al., 2002 who developed the PAI-S/K algorithm in order to identify the P-onset time of a seismic event. According to this algorithm, skewness and kurtosis, as measures of asymmetry and non-Gaussianity respectively, are estimated over a moving time window, and they are expected to present maxima in the neighborhood of the P arrival, due to the changes of the signal statistics. The location of the maximum slope of these curves is assigned as the final P onset time estimation. A number of automatic P-picking experiments on microseismic data (Lois et al., 2010, Tselentis et al., 2011a) have shown that the best results are obtained using the kurtosis criterion, thus only this HOS parameter is used in the proposed method.

**METHODOLOGY**

Given a N-length segment of the record, where a seismic event exists and the P-arrival time has been estimated, a M-sample time moving window is applied which divides this segment into
overlapping parts of the record, with \([(M-1)/M]\) x100% of overlap. On each section, the algebraic eigenproblem of the data covariance matrix is solved, that is, the covariance matrix is diagonalized and three eigenvalues \(\lambda_1 > \lambda_2 > \lambda_3\) with their corresponding eigenvectors are obtained. Since the same procedure takes place for each time window, three different sequences \(\lambda_1(t), \lambda_2(t), \lambda_3(t)\), \(t=1..N\), are formed, giving a measure of the energy on the direction of the three principal axes of the polarization ellipsoid (Figure 1).

In this work, we consider only the sequence corresponding to the maximum eigenvalue \(\lambda_1(t)\), since it is more sensitive to energy changes in the direction of signal’s propagation. Moreover, we choose as a characteristic function, the square root of the maximum eigenvalue’s curve \(f(t) = \sqrt{\lambda_1(t)}\) (Figure 2), rather than \(\lambda_1(t)\) itself, due to the property of the square root function to compress the signal, that is, reduce its dynamical range by increasing its lower values and decreasing its higher values. This enables us to observe small changes in the signal’s energy. The next step is to evaluate the square root values of kurtosis \(\sqrt{kur(f(t))}\), on the part of the characteristic function \(f(t)\), that corresponds to the time section starting a few samples after the P-arrival \(t_P\), up to the end of the seismic event \(t_{coda}\), which have been estimated by the detection algorithm. The selection of the specific segment of the record is necessary, for avoiding erroneous picks such as the P-arrival, and for providing also a time window where S-wave exists. Specifically, on this part of the record, kurtosis is evaluated over a sliding M-sample overlapping moving window, using the estimator:

\[
kur(f(t)) = \frac{\sum_{t=1}^{M}(f(t) - \hat{m}_f)^4}{(M-1)\hat{\sigma}_f^4}
\]

(14)

where \(\hat{m}_f\) and \(\hat{\sigma}_f\) are the estimators of the mean value and standard deviation of \(f(t)\) respectively and \(M\) is the length of the moving window.
During the S-wave arrival (Figure 3b), the values of kurtosis of f(t) present a steep increment due to the change of the signal statistics (Figure 3d). Furthermore, with close inspection we can observe that the S-onset time coincides with the point where the values of the sequence begin to increase and not with the maximum value of the curve. Thus, the maximum slope is needed to be assessed through signal’s first derivative (Figure 3e). The location of the initial S-onset time estimation is given by the maximum value of the first derivative of the sequence $\sqrt{\text{kur}(f(t))}$, that is:

$$S_{on} = \max_t \left( \frac{d}{dt} \sqrt{\text{kur}(f(t))} \right), \quad t \in [t_p, t_{\text{coda}}]$$  \hspace{1cm} (15)

An important issue that needs to be addressed is the choice of the sliding window’s length which is the only parameter of the algorithm that has to be set. A too-short window results in early picks, since the algorithm becomes too sensitive to small changes. On the other hand by setting long time window duration, it is possible to obtain picks that are beyond the real S wave arrival.

To overcome the two aforementioned cases of false alarm, a multi-window approach accompanied with a weighting scenario is proposed, as a correction procedure. For each S arrival time estimation an automatically evaluated uncertainty index is introduced for evaluating the probability of a false alarm. This quality measure, similar to Signal to Noise Ratio (SNR), is based on an energy ratio estimated on the two horizontal components, into a predefined time section $[S_{on} - t, S_{on} + t]$ and is given in dB by the relation:

$$q = 20 \log_{10} \left( \frac{\sigma_{S_{on}}}{\sigma_{\text{coda}}} \right)$$  \hspace{1cm} (16)

where $S_{on}$ is the automatically estimated S-arrival time, $\sigma_{S_{on}}$ and $\sigma_{\text{coda}}$ are the standard deviations evaluated on the time windows $[S_{on}, S_{on} + t]$ and $[S_{on} - t, S_{on}]$ respectively and t is time
in seconds that is selected empirically according to the S-P difference of the examined seismic events. For example, during the arrival of the S wave the energy of the signal usually increases, thus if the automatic pick is close to the real one, positive values of $q$ are expected. On the other hand an early pick is expected to have $q$-values close to zero, while a negative $q$ indicates a false pick, since there is no probability of existing P-coda waves with higher energy than the S waves. Nevertheless, it has to be mentioned that a $q$ value close to zero which corresponds to a low quality pick, does not necessarily indicate a false pick, but shows a case where the seismic signal’s energy during the S-wave arrival does not change significantly and careful human analyst’s inspection is needed.

The algorithm is applied on the seismic data using windows with various lengths, providing a set $\{S'_{on}\}_{i=1,w}$ of S-arrival times with the corresponding $q$’s which are used as weights. The final S-onset time estimation is given by the weighted mean of the set $\{S'_{on}\}_{i=1,w}$:

$$
S_{final} = \frac{\sum_{i=1}^{w} S'_{on} q_i}{\sum_{i=1}^{w} q_i}
$$

where $S'_{on}$ are the estimations obtained by the algorithm and form the set of solutions, $q_i$ are the corresponding weights, while the index $i=1,2,..w$ indicates the number of the different time windows used. Applying this weighting scheme, possible outliers (false picks) are eliminated by the algorithm, since they obtain weights with values close to zero. Following the same approach, the overall quality $q_{final}$ of the final estimation $S_{final}$ is evaluated and four classes of uncertainty A, B, C and D are defined as follows:

$$
\text{Uncertainty Class} = \begin{cases} 
A, & \text{if } q_{final} > 10 \\
B, & \text{if } 6 < q_{final} < 10 \\
C, & \text{if } 2 < q_{final} < 6 \\
D, & \text{if } q_{final} < 2 
\end{cases}
$$
Class A corresponds to high quality picks, B and C to moderate while D shows poor quality picks corresponding to high probability of false alarm. A pick which is assigned to a negative quality index, is indicated as a false pick as previously mentioned.

**DATA**

After the occurrence of a $M_w$ 4.7, August 7, 2011, seismic event at Nafpaktos, Greece, a 3-component temporary seismic station was installed in the area, to monitor the aftershock sequence. The station consisted of a 1-Hz LandTech LT100 borehole seismometer and 24-bit LandTech LTSR-24 recorder connected to a Global Positioning System (GPS). The recording took place for a period of 15 days with a sampling frequency of 100 Hz. From the continuous recording a total number of almost 150 microearthquakes, of magnitudes 0.5-1.5 $M_w$, were detected using a chi-squared based test statistic, (Lois et al., 2010) and the P-phase arrival times were identified using the kurtosis criterion (Tselentis et al., 2011, Saragiotis et al., 2002). The validity of the above algorithms’ results was confirmed by an expert analyst who finally chose 110 seismic events as input dataset for the proposed S wave identification algorithm. The criterion for the selection of the specific dataset was the ability of the analyst to pick the S-onset times, with low degree of uncertainty.

**RESULTS**

The proposed algorithm has been applied on the specific set of microseismic data and the results are compared with the S-picks provided by the analyst. Although manual S-picking is a difficult task, due to subjective factor inherent in this procedure, in real data case the picks obtained by a human expert is the only dataset that can be used as an a priori reference, in order to evaluate the algorithm’s efficiency. From the 110 manual picks, 39 picks (35.4%) considered by the analyst as
good quality picks, 67 (60.9%) as average quality picks and 4 picks (3.7%) as poor quality picks indicating high level of uncertainty. Moreover, the dataset consisted of microearthquakes superimposed over various levels of noise, specifically from 3 dB corresponding to low signal to noise ratio (SNR), up to 35 dB corresponding to high quality signals. In Figure 4, examples of seismic events with high and low SNR, are illustrated. The time windows’ lengths were empirically selected to be from 0.4 seconds up to 0.7 seconds, according to the duration of the microseismic events used on the experiment, t selected to be 0.8 seconds for q evaluation and also no filtering procedure took place, in order to have a clear view of the algorithm’s performance.

The implementation of the proposed technique resulted in an average residual time of 0.0517 seconds corresponding to almost 5 samples. We have to specify that the term residual time refers to the mean value of the absolute difference of the manual pick to the automatic pick. Moreover, 32 picks (29.1%) were classified into uncertainty class A, 62 (56.4%) into class B, 16 picks (14.5%) into class C and no picks were classified into class D. From the above results, it is evident that the algorithm is able to perform sufficiently well on this kind of data.

In order to elaborate the evaluation of the proposed technique, a noise robustness test was designed using both artificial and real seismic noise. An example of the effect of this procedure for both cases is depicted in Figure 5. In the first case, Gaussian distributed noise was scaled and added to the initial dataset, in order to achieve SNR’s range from 0 up to 8 dB. The algorithm’s implementation on the new dataset resulted to average accuracy of 0.079 seconds and from the new 110 automatic picks, 6 (5.5%) were classified into class A, 63 (57.3%) into class B, 41 (37.2) into class C and again no picks into class D.

In the second case, real seismic noise was added to the initial dataset and the SNR varied from -1 up to 8 dB. Although both tests took place on almost similar noise levels, the algorithm seems to be affected more by the addition of real seismic noise, since the implementation of the algorithm resulted to an average accuracy of 0.092 sec. The classification of the automatic picks to
the uncertainty classes was the following: 5 picks (4.6%) into class A, 63 picks (57.3%) into class B, 41 picks (37.2%) into class C and one pick (0.9%) into class D. The distribution of the residual times as well as the picks’ classification for all cases, are presented in Figure 6. However, in spite of the wide reduction of the SNR, the algorithm sustained a good performance, as long as the final mean residuals’ time on both cases, did not exceed the acceptable value of 0.1 sec (10 samples).

CONCLUSIONS

In this paper a new approach dealing with the automatic determination of the S-wave onset time, is proposed. Given the detected seismic event and the P arrival, the algebraic eigenproblem of data covariance matrix is solved over small time intervals. From the above analysis a characteristic function based on the maximum eigenvalue is formed, whose statistical attributes provide a first estimation of S-onset time. Furthermore, since the algorithm’s performance depends on the size of the used time window, we follow a multi-window scheme along with a weighting scenario based on energy criteria. This approach concludes to a set of solutions whose weighted mean provides a final, weighted S-onset time. The algorithm’s implementation on real microseismic data provided sufficiently good results in comparison with the manual picks, used as a reference dataset. Moreover the technique was subjected to a noise robustness test, using artificial and real seismic noise, resulting in an average accuracy of less than 0.1 seconds. In general the proposed method is straightforward to implement, demands low computational resources and the only parameters that have to be set are the lengths of the time moving windows the algorithm uses. Furthermore, earthquake location parameters are not necessary for the proposed algorithm in order to work properly. It is also understood that good quality P-pick is a prerequisite to conclude on correct estimations of S-arrival time. Due to its efficiency, the specific technique can be used as a useful tool for processing seismograms obtained by microseismic networks, minimizing the necessity for human intervention. An extension of the proposed method for regional data is possible, since the
characteristic function is quite sensitive to secondary phase arrivals, but this is beyond the scope of this paper.

ACKNOWLEDGMENTS

The authors would like to thank Prof. J. Zahradnik for his constructive criticism and valuable suggestions, as well as “Karatheodori” research program (University of Patras) which financed part of this work. We want also to thank LandTech Enterprises for their field and data processing support.

REFERENCES


CAPTIONS OF FIGURES

Figure 1: (a) Seismic signal and (b) the corresponding time functions obtained by the three eigenvalues.

Figure 2: (a) Seismic signal and (b) the characteristic function based on the maximum eigenvalue.

Figure 3: Example of the algorithm’s performance on a good quality signal, using a 0.6 sec moving window. (a) Seismic signal, (b) The selected part of the signal, as indicated in (a), by a dashed rectangle, (c) Kurtosis calculated for the characteristic function, (e) The maximum slope of kurtosis evaluated through its first derivative. The red dashed line corresponds to the automatic pick as estimated by the algorithm.

Figure 4: Examples of seismic signals with (a) high SNR and (b) low SNR.

Figure 5: The effect of the addition of synthetic and real seismic noise on the initial signal and the corresponding P (dashed line) and automatic S (solid line) picks. (a) An example of seismic signal, selected from the data set, (b) the same signal contaminated by Gaussian noise and (c) the initial signal contaminated by real seismic noise.

Figure 6: Histograms of the residual times and picks’ classification evaluated by the algorithm. Top panel shows the histograms of the residual times for (a) the initial dataset, (b) the dataset resulting by the addition of Gaussian noise and (c) the dataset resulting by the addition of real seismic noise. Bottom panel (d, e, f) shows the classification of the automatic picks into the four classes, for the three aforementioned cases.
Fig. 1
Fig. 2
Fig. 3
Example of High SNR signal

Example of Low SNR signal

Fig. 4
Fig. 5
Fig. 6