Geophysical Journal International

Geophys. J. Int. (2010) **181,** 1159–1170 doi: 10.1111/j.1365-246X.2010.04570.x

Automated determination of *P*-phase arrival times at regional and local distances using higher order statistics

L. Küperkoch,[1] T. Meier,[2] J. Lee,[1] W. Friederich[1] and EGELADOS Working Group[3]

[1] *Ruhr-University Bochum, Institute of Geology, Mineralogy and Geophysics, Ruhr-University Bochum, Universitätsstr.* 150, 44780, D-44801 *Bochum, Germany. E-mail: ludger.kueperkoch@rub.de*

[2] *Institute of Geophysics, Christian-Albrechts-University Kiel,* D-24118 *Kiel*

[3] *Aristotle University Thessaloniki, National Observatory Athens, Technical University of Crete, Istanbul Technical University, University of Hamburg, GeoForschungszentrum Potsdam, Ruhr University, Bochum, Germany*

Accepted 2010 February 17. Received 2010 February 16; in original form 2009 July 8

SUMMARY

We present an algorithm for automatic *P*-phase arrival time determination for local and regional seismic events based on higher order statistics (HOS). Using skewness or kurtosis a characteristic function is determined to which a new iterative picking algorithm is applied. For *P*-phase identification we apply the Akaike Information Criterion to the characteristic function, while for a precise determination of the *P*-phase arrival time a pragmatic picking algorithm is applied to a recalculated characteristic function. In addition, an automatic quality estimate is obtained, based on the slope and the signal-to-noise ratio, both calculated from the characteristic function. To get rid of erroneous picks, a Jackknife procedure and an envelope function analysis is used. The algorithm is applied to a large data set with very heterogeneous qualities of *P*-onsets acquired by a temporary, regional seismic network of the EGELADOS-project in the southern Aegean. The reliability and robustness of the proposed algorithm is tested by comparing more than 3000 manually derived *P* readings, serving as reference picks, with the corresponding automatically estimated *P*-wave arrival times. We find an average deviation from the reference picks of 0.26 ± 0.64 s when using kurtosis and 0.38 ± 0.75 s when using skewness. If automatically as excellent classified picks are considered only, the average difference from the reference picks is 0.07 ± 0.31 s and 0.07 ± 0.41 s, respectively. However, substantially more *P*-arrival times are determined when using kurtosis, indicating that the characteristic function derived from kurtosis estimation is to be preferred. Since the characteristic function is calculated recursively, the algorithm is very fast and hence suited for earthquake early warning purposes. Furthermore, a comparative study with automatically derived *P*-readings using Allen's and Baer & Kradolfer's picking algorithms applied to the same data set demonstrates better quantitative and qualitative performance of the HOS approach. This study shows, that precise automatic *P*-onset determination is feasible, even when using data sets with very heterogeneous signal-to-noise ratio.

Key words: Time series analysis; Body waves; Early warning.

1 INTRODUCTION

One of the most fundamental tasks in seismology is the precise location of earthquakes. Aside from the obvious requirement to use an accurate velocity model, reliable earthquake location depends strongly upon the accuracy of phase onset determination, which is done traditionally by visual human analyst inspection. With the deployment of digital seismological stations and due to the increasing amount of continuous data, much effort has been spent to develop reliable, automated phase picking algorithms. The most important advantages of automated procedures are their consistency and their capability of processing large data sets, such as those produced during reservoir characterization at enhanced geothermal systems (EGS) and hydrocarbon reservoirs, or in tomographic experiments. Furthermore, the precise and fast evaluation of *P*- and *S*-phase onsets is the basis for any earthquake early warning system (EEWS), which has to operate continuously and as reliably as possible. Nevertheless, the accuracy of experienced analysts has not yet been reached by routine automatic picking algorithms, mainly due to phase misidentifications and due to picking errors caused by the high variability of noise and *P*-phase waveforms. Diehl *et al.* (2009) point out, that phase misidentification may result in errors of up to several seconds. On the other hand, it has been shown that even for clear and unambiguous onsets, the error of manual picks can

be greater than 0.5 s (Douglas *et al.* 1997). Therefore, there is still a need for improved algorithms of automated *P*-phase onset-time determination including the consistent estimation of the pick uncertainty.

One of the first signal detectors was proposed by Freiberger (1962), who applied an approximative comparison of spectral densities for the detection of Gaussian signals in Gaussian noise. This method is suitable for recognizing signals rather than detecting signal onsets. Stewart (1977) developed an automated procedure for *P*-phase detection, *P*-phase processing and coda processing for local seismic event analysis in central California. By computing three 'moving-time noise averages' of the incoming signal it is tested, whether the seismic station is operating within acceptable limits of noise. A *P*-phase is detected and confirmed, if the waveform meets certain requirements based on amplitudes and zero-crossings during the post-detection process.

A fundamental step towards automatic phase-onset determination was the algorithm proposed by Allen (1978, 1982). He introduced the concept of the characteristic function (CF), resulting from a nonlinear transformation of the seismic trace from which the *P*-wave arrival time is estimated. Allen's CF is based on short-term-average to long-term-average ratios (STA/LTA). His picking algorithm is still frequently applied, for example, by the USGS Earthworm system (Johnson *et al.* 1995).

Due to lack of computer power in the early days of digital seismology, Goforth & Herrin (1981) developed an automatic seismic signal detector based on the Walsh transform. It is similar to the Fourier transform, but the CF is generated by a series of rectangular waveforms that take only the values $+1$ and -1, thus requiring less computation time than the Fast Fourier Transform. A signal is detected when the current sum of absolute values of Walsh coefficients exceeds a detection threshold, which is calculated from the previous 512 evaluation of this sum. Michael *et al.* (1982) used this approach to develop a real-time event detection and recording system for the MIT Seismic Network.

Joswig (1990) proposed a pattern recognition technique using characteristic event features in spectrograms to detect events rather than precisely picking onset times.

Baer & Kradolfer (1987) developed an automatic phase picker by modifying Allen's envelope function and incorporated a dynamic signal threshold. This algorithm is still frequently applied, for example, in Programmable Interactive Toolbox for Seismological Analysis (PITSA, Scherbaum & Johnson (1992) and in the automatic repicking system MannekenPix (Aldersons 2004). Earle & Shearer (1994) calculate STA/LTA-ratios of a smoothed envelope function (termed a 'smoothed ratio function' or SRF), determined by a Hilbert transform of the seismogram.

Besides these time and frequency domain approaches, model-oriented algorithms have also proliferated. Most widespread is the use of autoregressive (AR) models. Based on the Akaike Information Criterion (AIC, Akaike 1971; Akaike 1974), which can be considered as a measure of the badness of an estimated model, Takanami & Kitagawa (1988) developed a procedure for the fitting of a locally stationary autoregressive model to seismograms. They implemented this procedure as an on-line system called Fast Univariate Case of Minimum AIC Method of AR (FUNIMAR) model fitting. Leonard & Kennett (1997) propose an autoregressive method that detects increases in the AR-model order caused by the higher complexity of signals compared to preceding noise. The standard autoregressive two-model Akaike Information Criterium (AR-AIC, e.g. Sleeman & van Eck 1998) estimates the AR-coefficients from a fixed noise and a fixed signal window. The AIC is calculated from the prediction error in two adjacent moving time segments as a function of their merging point. Takanami & Kitagawa (1991) extended their FUNIMAR method by using a multivariate locally stationary autoregressive model (MLSAR), where they sum up the three AICs of the univariate autoregressive models of each seismic trace. Gentili & Michelini (2006) propose an artificial neural network approach for *P*- and *S*-phase onset time determination, called innovative model of neural network (IUANT2).

Withers *et al.* (1998) investigated several algorithms in a waveform correlation event detection and location system (WCEDS). They compared an STA/LTA-algorithm with a detector based on *Z*-statistics, which estimates the distance of the data from the mean in units of the standard deviation, an algorithm based on the estimation of power spectral densities, and with an algorithm based on polarization analysis. They concluded, that 'no specific algorithm and set of user-defined parameters is optimal for all scenarios of source, path, receiver, and background noise'. Hence, many new automatic phase-detection algorithms attempt to benefit from the strengths of the different approaches by combining them (e.g. Bai & Kennett 2000; Zhang *et al.* 2003).

Though many location routines require the input of uncertainty estimates, only few automatic picking algorithms determine the quality of the picks. Di Stefano *et al.* (2006) used a hand-picked subset of the available data for calibrating their proposed repicking system MannekenPix (Aldersons 2004), which provides the capability for high-precision automatic earthquake locations by incorporating automatic quality assessment.

Algorithms have been proposed that estimate relative traveltimes instead of absolute ones. Examples are multi-station and array approaches using multichannel cross-correlation methods (VanDecari & Crosson 1990) or adaptive stacking techniques (e.g. Rawlinson & Kennett 2004).

In their proposed *P*-pick extractor module, Gentili & Bragato (2006) and Gentili & Michelini (2006) calculate skewness, kurtosis, a combination of skewness and kurtosis and their time derivatives in a 2.048 s long moving window as an input for a neural network.

Here we modify and apply the scheme proposed by Saragiotis *et al.* (2002), who calculate skewness and kurtosis in sliding windows for *P*-onset determination, to a large data set of a temporary broad-band seismic network. Its performance is compared to manual *P*-phase arrival time readings and established automatic picking routines. A quality estimation scheme is introduced. The proposed iterative algorithm can briefly be described as follows: (1) identification of the *P*-phase by applying AIC to a CF, that has been calculated using higher order statistics (HOS), (2) recalculation of the CF and precise determination of the arrival time using a pragmatic picking algorithm, (3) estimation of the quality of the *P*-onset and (4) identification of erroneous picks.

In our procedure the CF is calculated recursively, making processing much faster. Maeda (1985) derived a formula, from which the AIC can be derived directly from the raw seismogram. We adopt his approach to calculate the AIC from the CF. The quality of the *P*-onset is estimated from the slope of the CF after the derived *P*-onset and from the signal-to-noise ratio (SNR) of the CF. Erroneous picks are recognized by checking the signal length, the time interval of all determined *P*-readings for one event and their individual effect on the onset time variance using a Jackknife procedure. To avoid picking of later phases (e.g. *S*), our algorithm compares the amplitudes of the vertical component around the derived *P*-arrival with the amplitudes on the horizontal components. The proposed algorithm is applied to a large data set resulting from a temporary, regional seismic network of the EGELADOS-project, which covered the entire

Hellenic subduction zone and continuously monitored the seismicity for 18 months. The automatically derived *P*-onsets are compared with the corresponding manual picks. Configuration parameters of the proposed *P*-phase detection algorithm are adjusted by comparison with selected manually derived picks which are divided into three quality classes. Furthermore, in order to test the robustness and performance of our picking procedure, we compare the results of our picker with the results of Allen's and Baer's & Kradolfer's picking algorithms.

1.1 Higher Order Statistics

In the following we give at first a brief introduction to HOS, describe its applications to *P*-phase picking and point out the similarity of the Baer- & Kradolfer-algorithm to HOS estimates.

The expectation of a continuous distribution is given by (e.g. Hartung 1991)

$$E[X] = \int_{-\infty}^{\infty} x p(x)\,\mathrm{d}x \tag{1}$$

with the distribution function $p(x)$ of the random variable X.

Using the expectation the statistic moment α of order k of the random variable X is defined as

$$\alpha_k = E[X^k] = \int_{-\infty}^{\infty} x^k p(x)\,\mathrm{d}x. \tag{2}$$

By analogy the central statistic moment m of order k is defined as

$$m_k = E[(X - E[X])^k], \quad k > 1. \tag{3}$$

The second central moment is the variance, the lowest moment yielding informations about the variability of a random variable

$$\mathrm{Var}[X] := E[(X - E[X])^2] = m_2. \tag{4}$$

The variance defines the mean power of the alternating part of an ergodic process.

The skewness is defined in terms of the third central moment

$$S = \frac{E[(X - E[X]]^3}{E[X - E[X]]^{3/2}} = \frac{m_3}{m_2^{3/2}}. \tag{5}$$

S becomes zero if the distribution is symmetrical. It becomes negative (positive) if the distribution contains outlayers to the left (right). The skewness gives informations about positive or negative deviations of the distribution density function from the mean value.

The kurtosis is defined using the fourth central moment

$$K = \frac{E[(X - E[X]]^4}{E[X - E[X]]^{4/2}} = \frac{m_4}{m_2^2}. \tag{6}$$

K becomes 3 for normally distributed random variables. Positive (negative) deviations of K from 3 indicate widening (narrowing) of the distribution.

An estimation of the central moment from the spot check is

$$\hat{m}_k = \frac{1}{n}\sum_{j=1}^{n}(x_j - \bar{x})^k \tag{7}$$

with $\bar{x}$ being the arithmetic mean.

Estimates of the second central moment, skewness and kurtosis are hence given by

$$\hat{\sigma}^2 = \hat{m}_2 = \frac{1}{n}\sum_{j=1}^{n}(x_j - \bar{x})^2 \tag{8}$$

Table 1. Examples of spot checks and corresponding values for variance, skewness and kurtosis.

Spot check	$\hat{\sigma}^2$	$\hat{S}$	$\hat{K}$
$[1, -1, 1, -1]$	1	0	1
$[1, -1, 2, -2]$	2.5	0	1.36
$[1, -1, 1, -\sqrt{7}]$	2.5	−1.11	2.08

$$\hat{S} = \frac{\hat{m}_3}{\hat{m}_2^{3/2}} \tag{9}$$

$$\hat{K} = \frac{\hat{m}_4}{\hat{m}_2^2}. \tag{10}$$

Table 1 shows some example spot checks which demonstrate the superiority of skewness and kurtosis over variance to detect even small outliers. The first two examples show symmetrically distributed samples with zero skewness. In the second example the variance increases, though no outlier distorts the distribution. The skewness remains zero and kurtosis increases only slightly. In the third example the outlier is detected by skewness and kurtosis, while the estimate of the variance remains the same as for the second example.

Figs 1(a) and (b) show the influence of a signal onset on the statistical properties of a noisy time-series. Shown are the time-series, the distribution density function of the samples and the variance, skewness and kurtosis, determined for a moving window of 10 s length. When the signal onset is reached (Fig. 1b), variance, kurtosis and skewness increase strongly and the shape of the distribution density function is no longer Gaussian.

2 APPLICATION OF HIGHER ORDER STATISTICS FOR DETERMINATION OF THE CF

Throughout this paper, we calculate several CFs by evaluating higher order statistical moments, particularly skewness and kurtosis, in moving time windows. The length of the window depends on the sampling interval and the dominant period. In order to make the calculation fast, a recursive procedure is applied, which can be described as follows:

Let $\{x(j)\}$, $j = 1, \ldots, M$ be a zero-mean, stationary process, T the length of the moving window, $\mathrm{d}\,t$ the sampling interval and

$$N = T/\mathrm{d}t + 1, \tag{11}$$

the number of samples of the window. The current value of central moment of order k of the moving window ending at sample j is

$$\hat{m}_k(j) = \frac{1}{N}\sum_{l=0}^{N-1} x_{j-l}^k. \tag{12}$$

Its estimate at sample j may be calculated from the previous one at sample $j - 1$

$$\hat{m}_k(j) = \hat{m}_k(j-1) - x^k(j-N) + x^k(j). \tag{13}$$

The recursive calculation of the central moment with eq. (13) reduces the computation times by a factor of about 10.

Note, that in case of mean and variance the CF is determined as the ratio STA/LTA, while in case of kurtosis or skewness the calculation is done by using just one long term moving window. Fig. 2 shows a comparison of CFs from mean and variance. The lengths of the

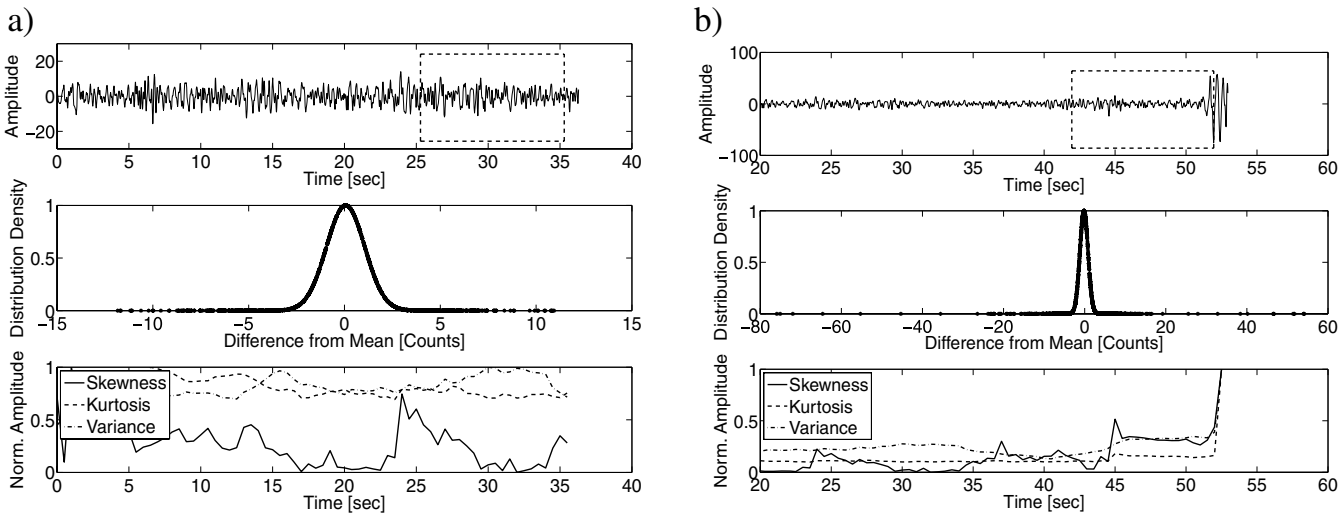

Figure 1. (a) Seismic trace (top panel) showing almost Gaussian distributed noise, as indicated by the statistical parameters variance, skewness and kurtosis (bottom panel) and the corresponding distribution function (middle panel), calculated using a 10 s length moving window (top, indicated by the box). (b) As soon as the moving window reaches the *P*-onset, the distribution is no longer Gaussian (middle panel) and variance, skewness and kurtosis increase strongly (bottom panel).

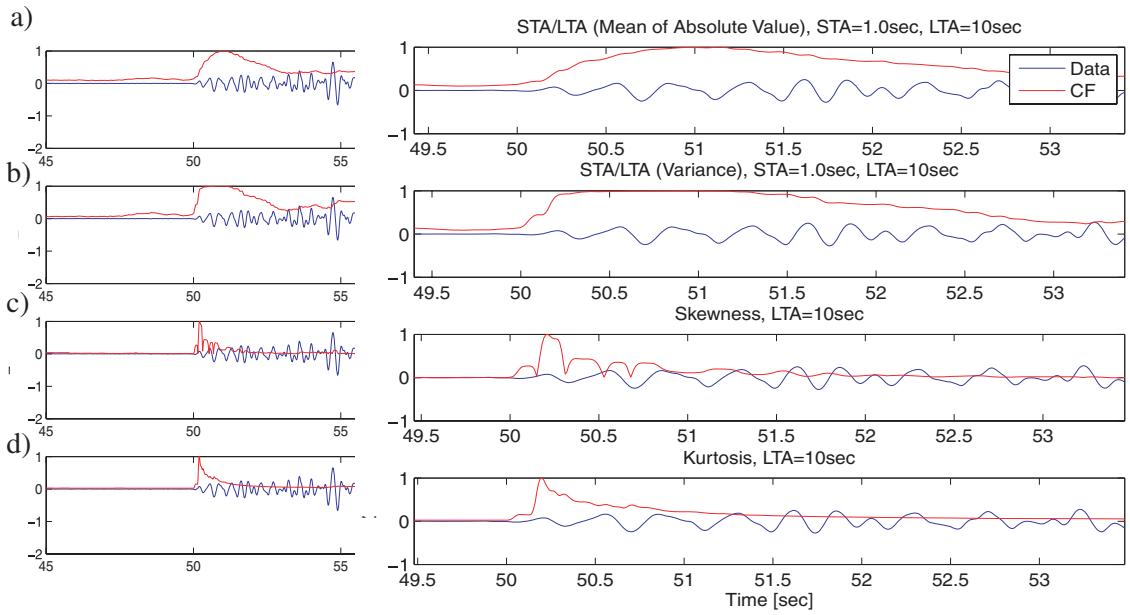

Figure 2. CF for recordings of a local event calculated from STA-/LTA-ratios of the mean (a) and the variance (b) using a 1 s length short-term window and a 10 s length long-term window, as well as skewness and kurtosis (c and d) calculated using a 10 s length moving window. The onset is recognized by each CF, while the CF calculated using skewness and kurtosis show the steepest gradient.

short and long term moving windows are 0.1 and 10 s, respectively. The length of the moving window for the determination of the CFs from skewness and kurtosis is 10 s. Each CF is normalized to its maximum. The onset is recognized by each CF, but the CFs calculated from skewness and kurtosis exhibit steeper gradients, which make these CFs very useful for picking procedures.

The length of the moving window must be adopted to the sampling frequency and the applied filtering. Too short time windows may result in biased estimates of HOS, while too long time windows may result in a smooth CF, which hides little, but important changes in amplitude of the original seismic trace. Gentili & Michelini (2006) use a 2.048 s length moving window for 2–50 and 2–8 Hz bandpass filtered data. Savvaidis *et al.* (2002) applied different time window lengths, ranging from 0.125 to 2 s for 200 Hz data, but found better results when applying a 2 s length moving window for kurtosis calculations. Fig. 3 shows the influence on the shape of the CF when using different time window lengths for kurtosis calculations.

3 ALLEN'S AND BAER AND KRADOLFER'S CF

Allen (1978) introduced the concept of the characteristic function CF, by which the 'character' of the seismic trace is specified.

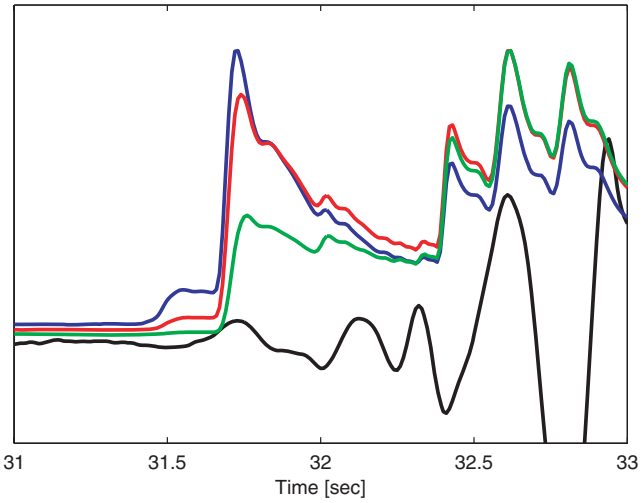

Figure 3. CF determined using kurtosis calculated for moving windows of 7 s (blue), 10 s (red) and 20 s (green), respectively.

However, it should be noted that Allen does not use his CF to determine the phase arrival times, but instead STA/LTA ratios of the CF. In the context of this paper, we follow Baer & Kradolfer (1987), who defined the CF as the time-series, from which the phase arrival time is estimated. Allen defined his CF as the envelope function

$$e_i = x_i^2 + C_i \cdot \dot{x}_i^2 \,. \tag{14}$$

x_i is the time-series under investigation, $\dot{x}_i$ its first difference and C_i a weighting constant with

$$C_i = \frac{\sum_{j=1}^{i} |x_j|}{\sum_{j=1}^{i} |x_j - x_{j-1}|}.$$

to control the relative importance of amplitude and derivative.

Baer & Kradolfer (1987) modified Allen's envelope function e_i to

$$E_i = x_i^2 + \dot{x}_i^2 \frac{\sum_{j=1}^{i} x_j^2}{\sum_{j=1}^{i} \dot{x}_j^2}. \tag{15}$$

By squaring this envelope function and implementing the variance of E_i, they get the following CF:

$$CF_i = \frac{E_i^2 - \overline{E_i^2}}{\sigma^2\left(E_i^2\right)}, \tag{16}$$

where $\overline{E_i^2}$ is the mean of E_i^2 from 1 to i and $\sigma^2(E_i^2)$ is the variance of E_i^2 from 1 to i. Note, that similar to kurtosis this CF also involves fourth powers of the data values. However, in eq. (16) the variance of E_i^2 instead of E_i is taken. This means, that Baer & Kradolfer's CF is pragmatically motivated in contrast to the CF using HOS, that has a specific statistical meaning.

4 TESTS ON SYNTHETIC DATA

In order to verify the potentials of the different CF, tests were carried out on synthetic data. *P*-onsets may be associated with changes in amplitude, frequency and phase, respectively. Therefore, we compared the sensitivities of CF calculated using STA/LTA-ratios, skewness and kurtosis (Fig. 4a), the envelope function proposed by Allen and the CF proposed by Baer & Kradolfer (Fig. 4b) to synthetic data including changes in amplitude, frequency or phase.

Fig. 4 indicates that a change in amplitude only is recognized by all CFs. The CF based on kurtosis shows the most distinct onset and a simple shape, making it best suitable for picking algorithms.

A change from higher frequencies to lower ones is best recognized by skewness, kurtosis and Baer's & Krodolfer's CF (Fig. 4, both middle). The way, in which the CFs using STA/LTA and Allen's envelope function recognize a change in frequency is not suitable for threshold based picking algorithms.

A change in phase is best recognized by skewness and Baer's & Kradolfer's CF (Fig. 4, both bottom). Kurtosis and Allen's envelope function are only weakly affected by a change in phase.

These simple tests indicate that skewness, kurtosis and the CF proposed by Baer & Kradolfer are to be preferred for picking algorithms.

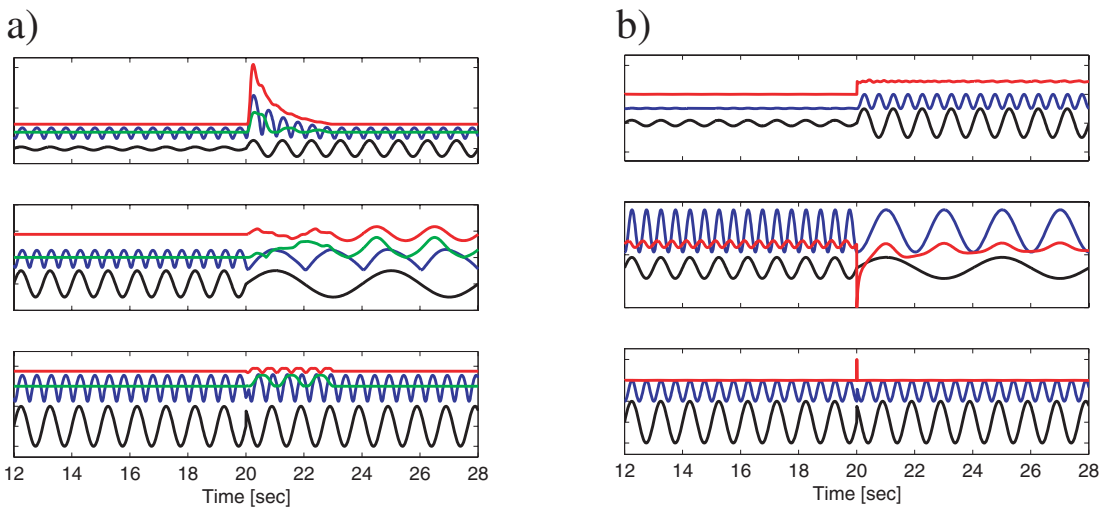

Figure 4. Tests on synthetic data (black) with change in amplitude only (top panel), change in frequency only (middle panel) and change in phase only (bottom panel). Amplitudes are normalized. (a) CF determined using STA/LTA-ratios (blue), skewness (green) and kurtosis (red). (b) CF determined using Allen's envelope function (blue) and the one proposed by Baer & Kradolfer (red).

5 AUTOMATIC PICKING

A new iterative picking algorithm is applied to the CF in order to detect the *P*-phase onset. The algorithm is organized into four stages. In the first stage, the data are 2–10 Hz bandpass filtered, using a causal third-order Butterworth filter. Then a CF is calculated using skewness or kurtosis from a 3 min time window of the vertical component seismogram. In analogy to Maeda (1985) the AIC is applied to the CF in the following way:

$$AIC(k) = (k-1)\lg\left(\frac{1}{k}\sum_{j=1}^{k} CF_j^2\right) + (L-k+1)\lg\left(\frac{1}{L-k+1}\sum_{j=k}^{L} CF_j^2\right), \quad (17)$$

where is L the length of the CF and k ranges from 0 to L. The applied formula differs from the original version by Maeda (1985) by taking cumulative sums of the squared CF instead of the variance of the data. The *P*-arrival is ascribed to the local minimum of the AIC-function before the global maximum. The minimum is found by searching for a common minimum of the original and a smoothed AIC-function.

In stage 2 the precise *P*-onset is determined. To this purpose the waveform is bandpass filtered between 2 and 15 Hz and the CF is recalculated. In addition, a smooth CF is calculated by taking the averages over the samples in a time interval of length $C_1 \approx 1.5$ s (Table 2). The length of the smoothing interval should be large enough to ensure that only the most prominent extrema remain within the smoothed CF. The picker searches then for a common local minimum of the smoothed and the unsmoothed CF to the right of the initial onset within a certain pick window of length $C_2 \approx 2$ s. C_2 should be large enough to contain at least two full cycles of the dominating period. Looking to the right avoids picking short-term increase of noise. Then the algorithm searches for a common local minimum of the smoothed and unsmoothed CF to the left of the initial onset within the pick window of length C_2 s. The condition for a local minimum is defined as follows:

$$CF_i < (1 + C_3) \cdot CF_{i-1}\,.$$

C_3 is set to 0.005 or to C_4, the smallest non-zero value of $CF' \cdot (1 + C_3)$, if C_4 is less than C_3. CF' denotes the first derivative of CF. If local minima are found to the left as well as to the right of the initial *P*-onset, the lower common minimum is assumed to coincide with the true *P*-arrival time. An example of a calculated CF and the corresponding automatically and manually derived *P*-onset is shown in Fig. 5(c).

Table 2. Constants needed for automatic picking.

Constant	Value/unit	Remark
C_1	1.5 [s]	Window length for moving average calculation for smoothing of CF
C_2	2 [s]	Window length of picking window
C_3	0.005	Factor for artificial up-lifting of smoothed CF-values
C_4	$<C_3$	Calculated from C3 and CF'
C_5	0.3 [s]	Window length for fitting slope to CF
C_6	30 [per cent]	Minimum adjusted standard deviation for slope fitting as percentage of $CF_{\max}$
C_7	2 [s]	Window length for noise level estimation
C_8	0.5 [s]	Window length for signal level estimation

Note: For details see text (Section 5).

The next stage is the automatic quality estimation. Two quality criteria serve as input for the quality evaluation scheme. The first quality criterion is defined as the logarithm of the slope of the CF normalized to the rms of the noise. Therefore, the unit of the first quality criterion becomes dB s^{-1}. The slope is calculated within a time window of length $C_5 \approx 0.3$ s immediately after the determined onset (Fig. 5c). An estimate of C_5 may be obtained by taking the average of the corner periods of the applied band pass filter, here 0.07–0.5 s. The slope is derived by a polynominal fit of order one. In order to evaluate the local properties of the CF from the onset to its first prominent local maximum, the window length for the slope estimation is decreased until the misfit of the slope reduces to $C_6 \approx$ 30 per cent of the global maximum of the CF.

The second quality criterion is the SNR of the CF. The noise window starts $C_7 \approx 2$ s before the determined *P*-onset and ends at the onset time. The signal window starts at the onset time and ends $C_8 \approx 0.5$ s later (Fig. 5c). The SNR in dB is calculated according to

$$\mathrm{SNR} = 20 \cdot \lg \frac{A}{A_0},$$

where A_0 is the rms-value of the noise amplitudes and A the rms-value of the signal amplitudes. While the slope of the CF serves as a

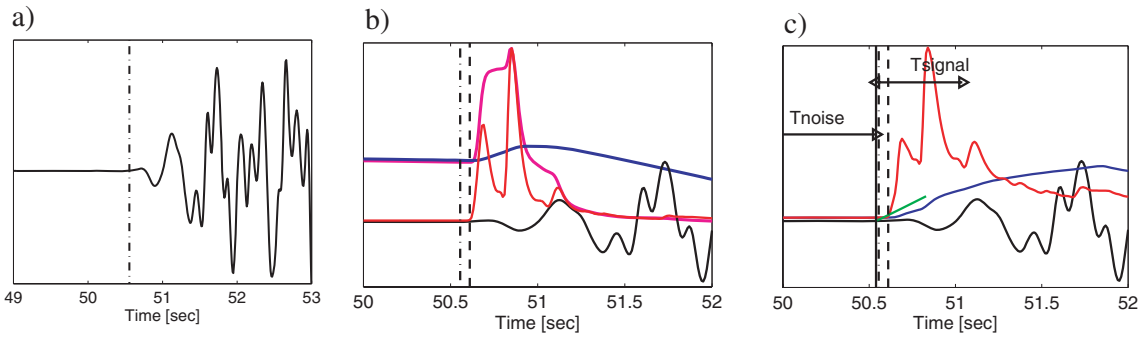

Figure 5. Automatic determination of a *P*-onset in a 2–10 Hz bandpass filtered, local event waveform (a, black) using the iterative picking algorithm. The manual *P*-reading is indicated by the dash–dotted vertical line. (b) Zoomed in portion of the above waveform (black), also showing the CF calculated using kurtosis (red), the unsmoothed AIC-function (cyan) and the smoothed AIC-function (blue). The initial *P*-onset (dashed vertical line) is determined from the two AIC-functions. (c) Zoomed in waveform (black), recalculated, unsmoothed CF (red) using kurtosis, calculated from 2 to 15 Hz bandpass filtered data, and recalculated, smoothed CF (blue). The final automatically derived *P*-onset (black vertical line) is close to the derived manual *P*-reading. The green line indicates the slope fitted to the unsmoothed CF. The noise window T_{noise} and the signal window T_{signal} are used for SNR estimation. T_{noise} is 2 s.

very local quality estimate, the SNR provides a more global quality estimate of the *P*-onset.

In order to decide if the derived picks belong to a seismic event and not noise, it is crucial for every automatic picking algorithm to check the reliability of the determined onset. In stage 4 of our algorithm we implemented three tools to get rid of erroneous picks. At first, the signal length is checked using an envelope function of the filtered trace. Thus, noise peaks are rejected. In order to get rid of *S*-onsets, which were spuriously declared to be *P*-onsets, the amplitudes of the envelope function of the vertical trace are compared with the amplitudes of the envelope functions of the horizontal traces.

The third tool is applied after all available traces for one event have been picked and verifies the following conditions to be met: (1) All onset times must lie within a certain time interval, depending on the dimensions of the seismic network and (2) the single *P*-readings should not distort the estimate of the variance of all determined *P*-onsets, which is tested by applying a Jackknife-procedure. Picks, which lie outside the proposed time interval or distort the variance of all determined *P*-readings, are skipped or down weighted to 4 implying their exclusion from later event location. This procedure contributes to the reliability of the algorithm. Furthermore, the applied location routines HYPOINVERSE (Klein 2002) and HYPOSAT (Schweitzer 2001) recognize outliers due to large rms-traveltime residuals.

6 APPLICATION TO EGELADOS-DATA

The large temporary, regional seismic network of the EGELADOS project (a Greek word for earthquake, here an abbreviation for Exploring the Geodynamics of subducted Lithosphere using an Amphibian Deployment Of Seismographs) monitored the seismicity of the entire Hellenic subduction zone for 18 months (Friederich & Meier 2008). More than 1000 seismic events were detected per month. Fig. 6 shows the locations of the earthquakes used for this study and the available stations of the EGELADOS-network. Using three months of data (2005.12–2006.02), manually derived *P*-onsets, serving as reference picks, are compared with our automatically derived *P*-onsets.

For manual uncertainty/quality estimation a weighting scheme for manual *P*-phase determination is used, which is summarized in Table 3. The manual *P*-weights are determined from errors of onset times provided by the analysis software. The relation between weighting classes and picking uncertainty is given in Table 3, column 2. The applied weighting scheme is adopted from the location routine HYPOINVERSE (Klein 2002), where weight-0 onsets denote excellent or impulsive (100 per cent weight), weight-1 very good (75 per cent weight), weight-2 good (50 per cent weight) and weight-3 intermediate onsets (25 per cent weight). Weight-4 picks are not used for location. In contrast to the manual quality estimation, the automated picking algorithm uses the CF for uncertainty and quality assessment, resulting in different parameters, which are the slope of the CF after the determined *P*-onset and the SNR estimated from the CF (Fig. 5). The thresholds for slope and SNR assigned to the weighting classes 0-4 are summarized in Table 3.

In order to restrict the analysis to well-constrained events, we require more than five automatically and manually derived *P*-readings, a maximum azimuthal gap of 180° and a maximum rms-value of traveltime and single station residuals of two seconds,

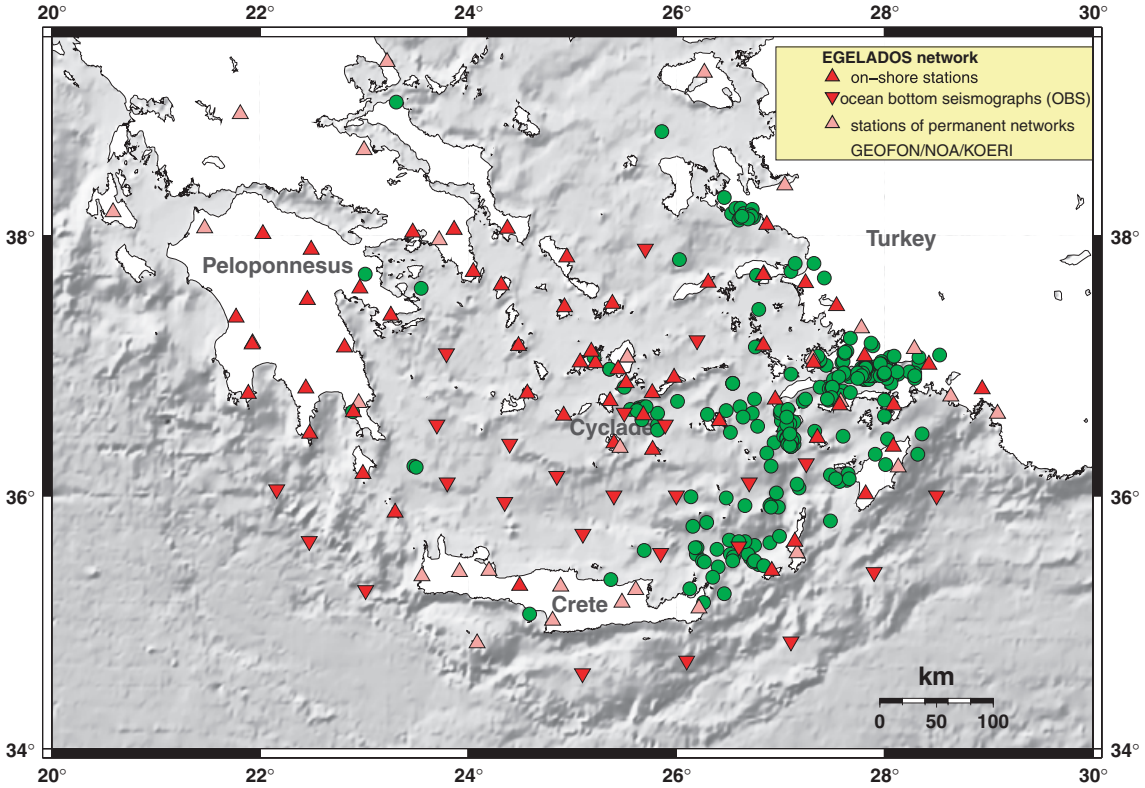

Figure 6. The seismic network of the EGELADOS project, covering the entire Hellenic subduction zone. Green circles indicate the locations of events used for this comparative study, ranging in magnitudes from $M_L = 0.7$–4.3. Hypocentral depths extend from near surface to 144 km. OBS-data are not used in this study.

Table 3. Weighting scheme for manual and automatic picking (kurtosis).

Weight	Man. uncertainty ε (s)	Slope m (dB)	SNR (dB)	Remark
0	$0 < \varepsilon \leq 0.04$	$m \geq 35$	$SNR > 8$	Excellent onset
1	$0.04 < \varepsilon \leq 0.08$	$m \geq 25$	$SNR \geq 5$	Very good onset
2	$0.08 < \varepsilon \leq 0.16$	$m \geq 11$	$SNR \geq 3$	Good onset
3	$0.16 < \varepsilon \leq 0.32$	$m \geq 3$	$SNR \geq 1$	Intermediate onset
4	$\epsilon > 0.32$	$m < 3$	$SNR < 1$	Bad onset, not used

Note: Slope and SNR are automatically derived by the proposed algorithm.

respectively. Using these restrictions, 475 manually located events with 3383 manually determined *P*-readings, serving as reference picks, are available for this study. For the comparison between manual and automatic picks it is required, that the difference between manually and automatically derived source times does not exceed 10 s, otherwise it is assumed that automatic and manual picks belong to different events. Though the available manual *P*-readings do not meet the formal, very restrictive conditions for reference picks as defined by Di Stefano *et al.* (2006) or Diehl *et al.* (2009) (such as equal choice of filtering and equal phase identification), we nevertheless use these manual picks as a reference, as they are appropriate for testing the robustness of the proposed picking algorithm and comparing different automatic algorithms applied to the same data set.

Fig. 7(a) shows the histogram of the differences ($T_{\text{auto}} - T_{\text{manu}}$) between automatically and manually derived *P*-arrival times using skewness for automatically derived quality classes 0-3 and 0, respectively. 132 automatically as well as manually located events meet the introduced conditions. 1461 automatically determined *P*-phase arrival times fall into weighting classes 0-3 and are further used for location. 472 out of them correspond to automatic weighting class 0. For automatic weighting classes 0-3 the average deviation from the manually derived reference picks is 0.380 ± 0.75 s, while for automatic weighting class 0 the average difference is 0.070 ± 0.41 s.

As an upper bound for the acceptable automatic picking error, we adopt the 68 per cent interval (equals the standard deviation in the case of a Gaussian distribution) of the rms traveltime residuals of the manually derived locations. For our reference data the width of this interval is 0.5 s. Thus, the differences between automatically and manually derived *P*-onset times within the 68 per cent interval should be smaller than 0.5 s. This ensures, that the location errors are not dominated by picking errors.

When the CF is calculated from skewness, the width of the 68 per cent interval of the deviations from the reference picks is 0.4 s and hence below this error bound. However, the tail of the histogram in Fig. 7(a) indicates that the automatically derived *P*-onsets are somewhat late compared to manually derived *P*-onsets. This is confirmed by the width of the 80 per cent interval which is 0.75 s. Visual inspections of corresponding waveforms show that the tail is mainly due to different filtering. While the automatic algorithm uses 2–10 and 2–15 Hz bandpass filtering, the analysts very often pick the first break on unfiltered data. Fig. 8 shows an example of a differently filtered waveform and the corresponding manually and automatically derived *P*-onsets, respectively. This example indicates, that a comparison of differently derived picks may be misleading when different filters are applied. This fact also applies to the comparison of manual picks derived from differently filtered seismograms. Furthermore, while the automatic usually picks the first *P*-arrival (Pn), the analysts sometimes chose to pick the later and much clearer Pg-phase, in case that the corresponding Sg-phase

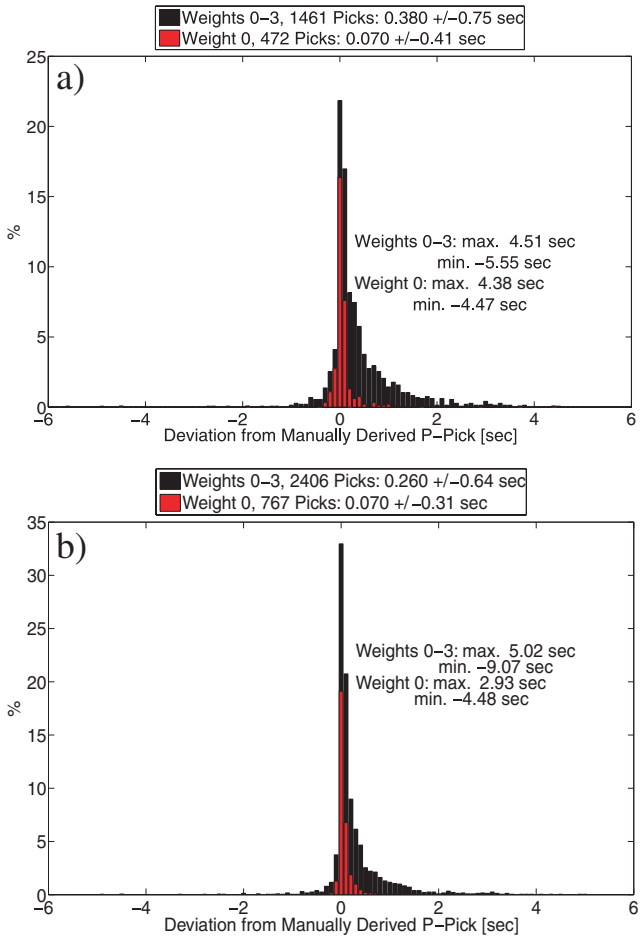

Figure 7. (a) Histogram of differences between manually and automatically derived *P*-onsets using skewness. (b) Same as (a), but *P*-onsets determined using kurtosis. Weighting classes are defined in Table 3.

is also clearer than the preceding Sn-phase. This problem also has been recognized by other authors (e.g. Sleeman & van Eck 1998; Diehl *et al.* 2009). Usually, the number of phase misidentifications is especially large at distances $\Delta > 100$ km, where emergent Pn phases are followed by impulsive *Pg* or *PmP* phases (e.g. Diehl *et al.* 2009). However, such a distance dependence of differences between manually and automatically derived *P*-arrivals is not present in our data set, a fact which might be due to the complex tectonic settings of the Aegean region.

Fig. 7(b) shows the histogram of differences between automatically derived *P*-arrival times using kurtosis and corresponding manually derived *P*-onsets. The automatic algorithm produced 2406 *P*-readings in weighting classes 0-3 leading to 177 comparable events. The average differences from the manual picks are ranging from

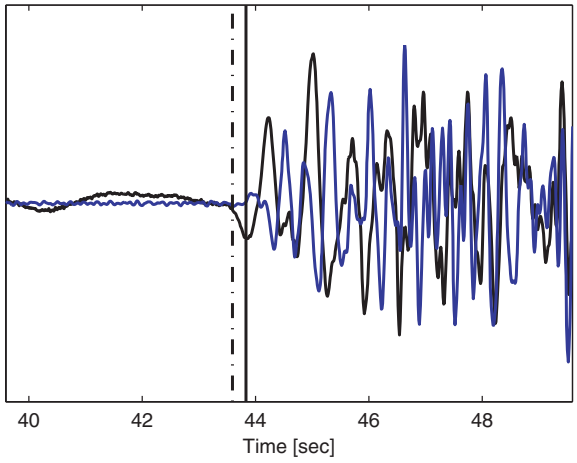

Figure 8. Unfiltered (black) and filtered (blue, third order butterworth, 2–10 Hz bandpass) waveform and corresponding manual (dash–dotted line) and automatic (dashed line) picks demonstrating the effect of different filtering on *P*-onset determination.

0.260 ± 0.64 s for automatic weights 0-3 to 0.070 ± 0.31 s for automatic weight 0. The width of the 68 per cent interval of the differences between manually and automatically derived *P*-picks is 0.2 s and hence below the picking error bound of 0.5 s. This automatic picks also tend to be somewhat late compared to manual *P*-readings due to different filtering and *Pg/Pn* mismatch. However, the width of the 80 per cent interval is 0.45 s compared to 0.75 s for skewness. Since the comparison reveals better qualitative and quantitative performance for the kurtosis application, we will focus on kurtosis for the following comparative studies.

Fig. 9(a) displays the number of picks plotted against the slope of the CF and the deviation from the corresponding manual pick for the kurtosis-picker, hereafter referred to as residuals. The larger the slope, the less the number of outliers. Furthermore, even for automatically low weighted picks, that is, for emergent onsets, the number of *P*-readings with small residuals is nevertheless high. The automatic weighting classes, indicated by red horizontal lines, are defined by visual comparisons of determined slopes and manually determined weighting classes. Fig. 9(b) shows the number of picks versus SNR [dB] determined from the CF and the residuals. The figure indicates a clear correlation between SNR and number of picks with large residuals. All *P*-onsets with SNR smaller than 1 dB are skipped or downweighted to 4. Fig. 10(a) compares the automatically estimated weights of the automatically determined onset times compared with the manually determined weights. This figure demonstrates a strong correlation between manually and automatically assigned weights. However, the automatic weighting scheme seems to be too conservative for manual weight-0 picks, but not strict enough for manual weight-1 picks. The manual weight-2 *P*-readings are well resolved by the automatic. Manual weight-4 picks are upgraded to automatic weight-0 picks in only 0.6 per cent of the cases. The majority of manual weight-4 *P*-readings are upgraded to automatic weight-2 and weight-3 picks. About 44 per cent of the automatic weight-4 picks get the same manual weight. Visual inspections of manual weight-4 picks, upgraded to weight-0 *P*-onsets by the automatic, and their corresponding waveforms indicate, that most of them are not false picks, but either very noisy onsets or good onsets of a different seismic event, occurring nearly at the same time.

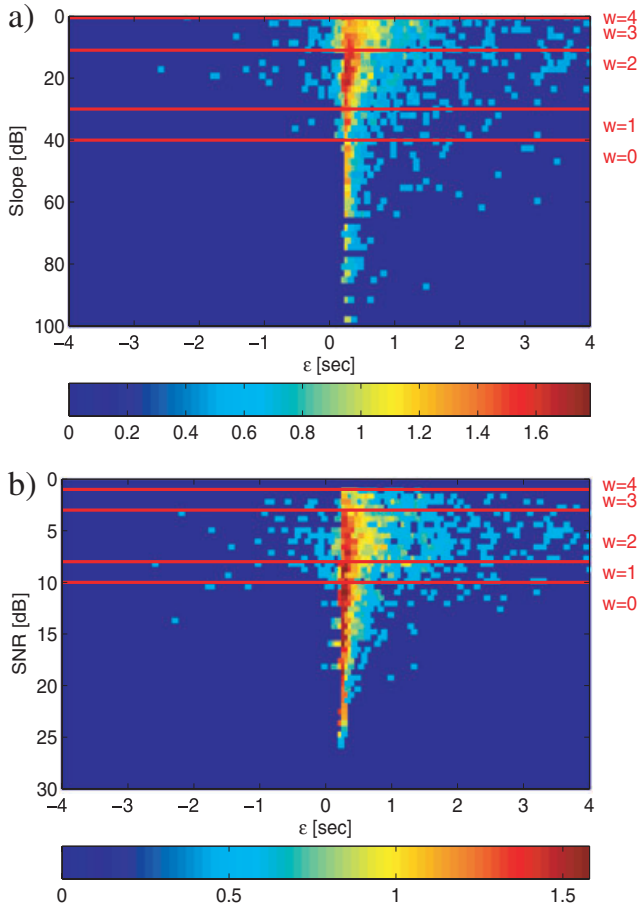

Figure 9. Frequency distribution of automatic picks against slope of CF (a) and SNR (b) and corresponding residuals to manual picks. The width of the frequency distributions increases with decreasing slope (SNR) and thus justifies our use of slope and SNR as quality criteria. The applied weighting classes are indicated by the red horizontal lines. For every weighting class the automatically derived *P*-onset has to meet both criteria. Note the large number of low-weight onsets with small residuals.

7 COMPARISON WITH THE ALLEN- AND THE BAER- AND KRADOLFER-PICKER

In order to test the robustness and reliability of the proposed picking algorithm, we applied the Rex–Allen-picker (RA hereafter) and the Baer- & Kradolfer-picker (BK hereafter) to the same data set. As for the kurtosis-picker, the required parameters are optimized on categorized *P*-onsets. For quality estimation we use the proposed weighting scheme by Allen (1978). However, Allen only gives certain criteria to be met by impulsive *P*-onsets. They are successively relaxed for lower weights 1, 2 and 3. In this study, we tried to adjust these conditions so that the resulting weights match the manually determined weights. Baer & Kradolfer (1987) did not implement an automatic quality assessment. In order to make the results more comparable, we applied Allen's weighting scheme also to the BK picks. As Baer & Kradolfer squared Allen's envelope function, some values adopted from Allen's weighting scheme are squared, too. The conditions to be met by the automatically derived *P*-onsets are summarized in Table 4.

124 events are automatically as well as manually located and meet the introduced restrictions when using RA. Fig. 11(a) shows

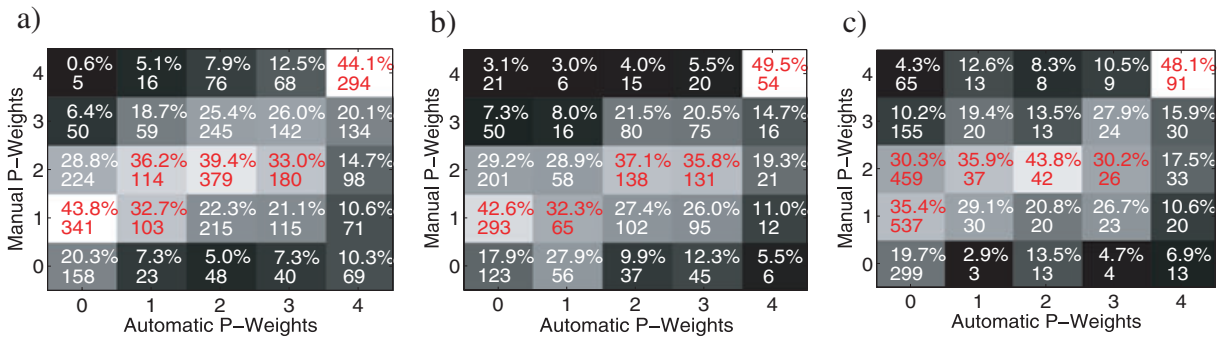

Figure 10. Comparison between frequencies of automatically and manually determined quality classes using a matrix representation. (a) kurtosis-picker, (b) RA (Allen-) picker and (c) BK- (Baer- & Kradolfer-) picker. Each column shows the frequency distribution of weights assigned to those manual picks, which are associated to automatic picks of one fixed weighting class. For example, in panel (a), second column, 23 automatic picks or 7.3 per cent of all available automatic picks of quality class 1 are assigned to manual weight-0 picks, but 103 automatic picks or 32.7 per cent of all available automatic picks of quality class 1 belong to the same manual weighting class. Percentages in one column add up to 100 corresponding to all available automatic picks with manual complements of the certain weighting class.

Table 4. Applied weighting scheme for RA and BK.

Weight	Criterium I	Criterium II	Criterium III	Criterium IV
0	$D > \sqrt{B}$	$A_1 > 450$	$A_1/\sqrt{B} > 4$	$A_2 > 6 \cdot \sqrt{B}$ or $A_3 > 6$
1	$D > 0.5 \cdot \sqrt{B}$	$A_1 > 100$	$A_1/2 \cdot \sqrt{B} > 4$	$A_2 > 3 \cdot \sqrt{B}$ or $A_3 > 5$
2	$D > 0.3 \cdot \sqrt{B}$	$A_1 > 50$	$A_1/\sqrt{B} > 3$	$A_2 > 2 \cdot \sqrt{B}$ or $A_3 > 4$
3	$D > 0.1 \cdot \sqrt{B}$	$A_1 > 20$	$A_1/\sqrt{B} > 2$	$A_2 > \sqrt{B}$ or $A_3 > 3$

Notes: D is the trace first difference and B a measure of the noise level derived from the CF at the onset time, respectively. A_1, A_2 and A_3 are the first three amplitude peaks of the seismic trace. For more details, see Allen (1978). For the BK all values depending on B are squared. For each weight all four criteria must be fulfilled.

the histogram of differences between RA-picks and manually derived *P*-onsets, indicating that this algorithm tends to be somewhat early compared to what an analyst would pick. The average residual is -0.12 ± 1.44 s for automatic weight 0-3 picks and -0.179 ± 0.47 s for automatic weight-0 onsets. 1546 *P*-readings were automatically determined and further used for location, compared to 1419 when using skewness and 2406 *P*-readings when using kurtosis. The 68 per cent interval of all differences between manual and automatic picks is 0.3 s wide and thus within the upper picking error bound. The width of the 80 per cent interval is 0.5 s compared to 0.45 s for kurtosis.

The comparison with the results of the RA confirms the results of Saragiotis *et al.* (2002), who found a better performance of HOS applications. In their study they used 44 seismograms of local and regional seismic events in Northern Greece. Skewness and kurtosis applications yielded similar results and performed better in 75 per cent of the cases. However, automatic quality assessment was not considered in their work.

Fig. 10(b) shows the comparison between manually and automatically derived weights of available *P*-readings when using RA. Allen's weighting scheme is quite conservative as most of the automatic weight-0 picks are downgraded compared to manual weights. The weighting classes 1 and 2 are in good agreement to the manual weights. However, in ≈3 per cent of the cases, manual weight-4 picks are upgraded to automatic weight-0 picks. Manual weight-4 picks are well resolved by RA.

Another test for robustness and reliability of the proposed picking algorithm is the comparison with the BK. The histogram of deviations of the automatically derived picks from manual *P*-readings is shown in Fig. 11(b). 124 events are automatically as well as manually located and meet the introduced restrictions. 1683 *P*-readings were determined by this algorithm and classified as weight 0-3 picks. The overall pattern of the distribution of differences from the manual picks indicates that this algorithm tends to be somewhat late compared to the reference picks. This has also been shown by Sleeman & van Eck (1998). For weight 0-3 *P*-picks the average residual is 0.341 ± 2.43 s. The large number of excellently weighted *P*-onsets indicates that the weighting scheme adopted from the RA needs to be modified for the BK CF. However, the standard deviation slightly decreases for weight-0 *P*-picks. Though this picking algorithm produces some large outliers, the width of the 68 per cent interval is 0.2 s, which is below the introduced acceptable picking error bound and equal to the corresponding width of the 68 per cent interval of differences from the manual picks when using kurtosis. The width of the 80 per cent interval is 0.6 s compared to 0.45 s for kurtosis.

Fig. 10(c) shows the comparison between manually and automatically derived weights when using the BK. About 20 per cent of automatic weight-0 picks are in agreement with manual weight-0 *P*-readings, but ≈4 per cent show an upgrading of manual weight-4 picks to automatic weight-0. The comparison between automatic weight-0 classes derived with BK reveals that the weighting scheme needs to be improved, as to many *P*-readings are automatically classified as weight-0 *P*-onsets. However, weight-2 class picks are well resolved by the applied automatic weighting scheme. In ≈48 per cent of the cases manual weight-4 picks are resolved by the automatic.

For a direct comparison of the RA, the BK and the kurtosis-picker, we selected *P*-readings determined both manually and by all three automatic picking algorithms. Fig. 12 shows the results

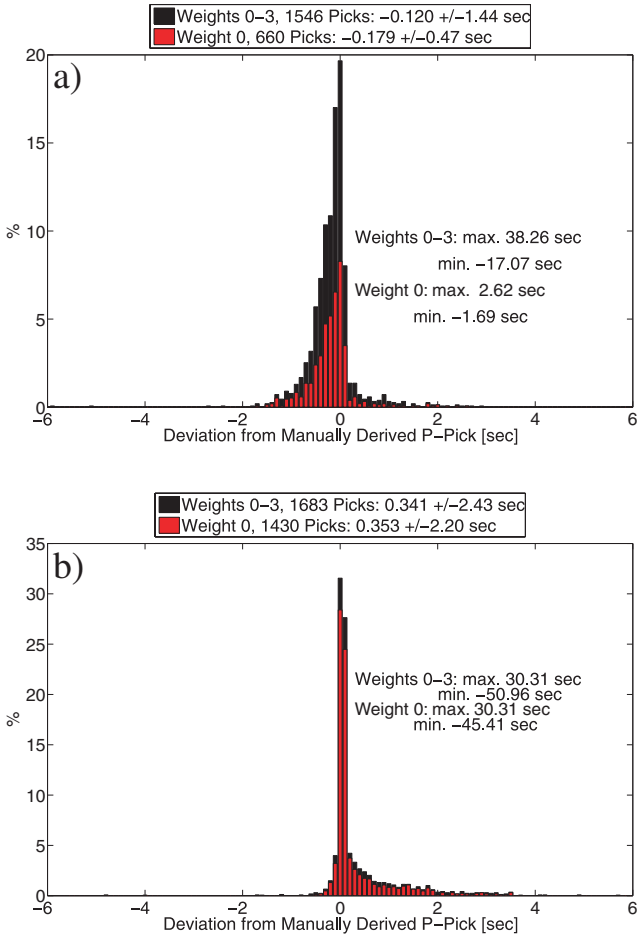

Figure 11. (a) Histogram of differences between manually and automatically derived *P*-onsets using RA. (b) Same as (a), but *P*-onsets determined using BK. Weighting classes are defined in Table 4.

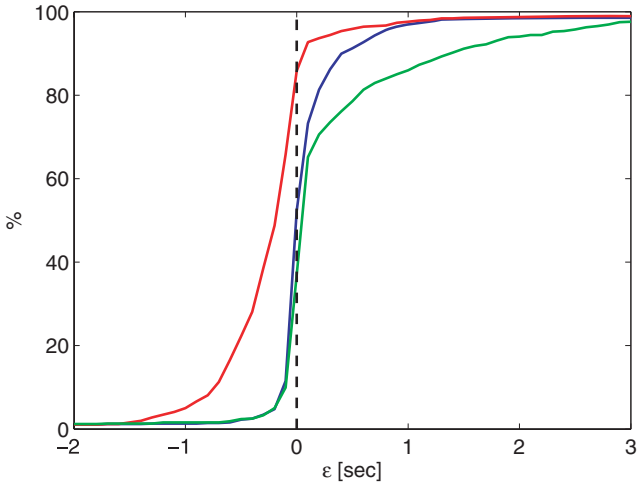

Figure 12. Cumulative frequencies of automatic picks obtained by using kurtosis (blue), RA (red) and BK (green), plotted versus their residuals to associated manual picks.

for automatic weighting classes 0-3 for 755 available *P*-picks as cumulative frequencies of residuals, which reveals the tendency of a picking algorithm whether it picks to late or to early. The figure indicates nearly similar and excellent results of the BK and the kurtosis-picker for negative residuals, that is, for earlier picked onsets compared to the manual ones. The BK shows more outliers for later picked *P*-onsets compared to the kurtosis- and RA, which tends to be somewhat late compared to the manual *P*-readings. The best results are achieved when using the kurtosis-picker, where the residuals are more or less equally distributed around zero.

The quantitative performance is best for the kurtosis picker. 71.1 per cent of 3383 available *P*-onsets are determined and classified as useful picks for location purposes by this automatic, 49.7 per cent by the BK, 45.7 per cent by RA and 43.2 per cent by the skewness-picker. The overall low number of automatically determined *P*-readings is due to low SNR, noise bursts at island stations, and due to the introduced restrictions on the selection of seismic events.

8 CONCLUSIONS

We present an iterative algorithm for automated *P*-onset determination for local and regional seismic events based on HOS and Akaike Information Criterion. The algorithm also accounts for quality assessment, based on the slope of the characteristic function after the derived *P*-onset and its SNR. Using a large data set of waveforms recorded at the temporary, regional seismic network of the EGELADOS-project in the southern Aegean, we compare the performance of the proposed algorithm with Allen's (RA) and Baer- & Kradolfer's (BK) picking algorithms. 475 well constrained events with 3383 manually derived *P*-readings, serving as reference picks, are available for this study. The comparison shows similar and very good results for kurtosis and skewness applications with a better quantitative performance of the kurtosis-picker. Adopting the weighting scheme applied in the location routine HYPOINVERSE (Klein 2002), the average deviation from the reference picks for the kurtosis application is 0.070 ± 0.31 s to 0.260 ± 0.64 s, depending on the automatically derived weighting class. For the skewness application the corresponding average deviation is 0.070 ± 0.41 to 0.380 ± 0.75 s, respectively. Both, kurtosis and skewness applications, performed well below the picking error bound of 0.5 s, which is estimated from the width of the 68 per cent interval of the rms traveltime residuals obtained from the manual locations. In this way it is ensured, that the location error is not dominated by picking errors. For the kurtosis application 80 per cent of the differences between automatic picks and the corresponding manual picks remain below 0.5 s. Furthermore, even emergent or noisy *P*-onsets are determined reliably by the proposed algorithm.

Erroneous picks are found by checking the signal length determined from the envelope function of the seismic trace, by application of a Jackknife procedure applied to the variance of all available automatic *P*-readings, and a comparative envelope function analysis to get rid of noise peaks and spuriously picked *S*-phases, which enhances the reliability of the presented picking algorithm significantly. The majority of the rejected automatic picks (≈75 per cent) do not pass the requirement of a minimum signal length. Random inspections on these rejected automatic picks revealed that this quality control reliably detects noise peaks or short-term increase of noise. The low number of large outliers confirms the robustness of the proposed weighting scheme. However, the automatic quality assessment based on the slope and SNR of the CF is somewhat conservative for weight-0 picks compared to manual weighting. Only 0.6 per cent of automatic weight-0 picks got weight 4 by the operator. This is an indication for the reliable performance of the proposed weighting scheme.

Comparison of the proposed algorithm using kurtosis with the RA and the BK shows better results for the kurtosis application

regarding both the quantitative and the qualitative performance. When considering the 80 per cent interval the differences between automatic picks and the corresponding reference picks exceed 0.5 s for both the RA and the BK.

Allen's automatic quality estimation scheme assigns weight 0 to ≈3 per cent of the manually declared weight-4 *P*-readings. For the BK Allen's weighting scheme is slightly modified to account for the squared envelope function but needs further improvement when implemented into the BK, as too many picks are upgraded to higher quality classes.

On the basis of this comparative study, we conclude that the proposed iterative algorithm provides an efficient way to pick reliable *P*-wave arrivals. The speed of the presented algorithm makes it well suitable for implementation into near-real time processing schemes. The proposed automatic weighting scheme works reliably and assigns similar weights to the *P*-onsets as the analyst does.

ACKNOWLEDGMENTS

We thank Tobias Diehl and an anonymous reviewer for their valuable comments that helped improve the original manuscript. This study was funded by the German Research Foundation (DFG) within the Collaborative Research Centre 526 'Rheology of the Earth – From the Upper Crust to the Subduction Zone'.

REFERENCES

Akaike, H., 1971. Autoregressive model fitting for control, *Ann. Inst. Statist. Math.,* **23,** 163–180.

Akaike, H., 1974. Markovian representation of stochastic processs and its application to the analyses of autoregressive moving average processes, *Ann. Inst. Statist. Math.,* **26,** 363–387.

Allen, R.V., 1978. Automatic earthquake recognition and timing from single traces, *Bull. seism. Soc. Am.* **68,** 1521–1532.

Allen, R.V., 1982. Automatic phase pickers: their present use and future prospects, *Bull. seism. Soc. Am.* **72,** 225–242.

Aldersons, F., 2004. Toward three-dimensional crustal structure of the Dead Sea region from local earthquake tomography, *PhD thesis*, Tel Aviv University, Israel.

Bai, C.Y. & Kennett, B.L.N., 2000. Automatic phase-detection and identification by full use of single three-component broadband seismogram, *Bull. seism. Soc. Am.* **90,** 187–198.

Baer, M. & Kradolfer, U., 1987. An automatic phase picker for local and teleseismic events, *Bull. seism. Soc. Am.* **77,** 1437–1445.

Diehl, T., Kissling, E., Husen, S. & Alderson, F., 2009. Consistent phase picking for regional tomography models: application to the Greater Alpine region, *Geophys. J. Int.* **176,** 542–554.

Di Stefano, R., Aldersons, F., Kissling, E., Baccheschi, P. & Chirabba, C., 2006. Automatic seismic phase picking and consistent observation error assessment: application to the Italian seismicity, *Geophys. J. Int.* **165,** 121–134.

Douglas, A., Bowers, D. & Young, J.B., 1997. On the onset of P seismograms, *Geophys. J. Int.* **129,** 681–690.

Earle, P.S. & Shearer, P.M., 1994. Characterization of global seismograms using an automatic-picking algorithm, *Bull. seism. Soc. Am.* **84,** 366–376.

Freiberger, W.F., 1962. An approximate method in signal detection, Technical Report No. 12, Department of the Army, Division of Applied Mathematics, Brown University.

Friederich, W. & Meier, T., 2008. Temporary seismic broadband network acquired data on Hellenic subduction zone, *EOS, Trans Am. geophys. Un.,* **89**(40), 378.

Gentili, S. & Bragato, P., 2006. A neural-tree-based system for automatic location of earthquakes in Northeastern Italy, *J. Seismol.* **10,** 73–89.

Gentili, S. & Michelini, A., 2006. Automatic picking of P and S phases using a neural tree, *J. Seismol.* **10,** 39–63.

Goforth, T. & Herrin, E., 1981. An automatic seismic signal detection algorithm based on the Walsh transform, *Bull. seism. Soc. Am.* **71,** 1351–1360.

Hartung, J., 1991. Statistik - Lehr- und Handbuch der angewandten Statistik, Oldenbourg publishing company, Vienna, Munich.

Joswig, M., 1990. Pattern recognition for earthquake detection, *Bull. seism. Soc. Am.* **80**(1), 170–186.

Johnson, C.E., Bittenbinder, A., Bogaert, B., Dietz, L. & Kohler, W., 1995. Earthworm: a flexible approach to seismic network processing, *Incorp. Res. Instit. Seismol. Newslett.,* **14**(2), 1–4.

Klein, F.W., 2002. User's guide to HYPOINVERSE-2000, a Fortran program to solve for earthquake locations and magnitudes, Open File Report 02-171, U.S. Geological Survey.

Leonard, M. & Kennett, B.L.N., 1999. Multi-component autoregressive techniques for the analysis of seismograms, *Phys. Earth planet. Int.* **113,** 247–263.

Maeda, N., 1985. A method for reading and checking phase times in auto-processing system of seismic data, *Zisin=Jishin* **38,** 365–379.

Michael, A.J., Gildea, S.P. & Pulli Jay, J., 1982. A real-time digital seismic event detection and recording system for network applications, *Bull. seism. Soc. Am.* **72**(6), 2339–2348.

Rawlinson, N. & Kennett, B.L.N., 2004. Rapid estimation of relative and absolute delay times across network by adaptive stacking, *Geophys. J. Int.* **157,** 332–340.

Saragiotis, C.D., Hadjileontiadis, L.J. & Panas, S.M., 2002. PAI-S/K: a robust automatic seismic P phase arrival identification scheme, *IEEE Trans. Geosci. Remote Sens.* **40,** 1395–1404.

Savvaidis, A., Papazachos, C., Soupios, P., Galanis, O., Grammalidis, N., Saragiotis, Ch., Hadjileontiadis, L. & Panas, S., 2002. Implementation of additional seismological software for the determination of earthquake parameters based on MatSeis and an automatic phase-detector algorithm, *Seism. Res. Lett.,* **73**(1), 57–69.

Scherbaum, F. & Johnson, J., 1992. Programmable Interactive Toolbox for Seismological Analysis (PITSA). *IASPEI Software Library,* **5,** Seismological Society of America, El Cerrito.

Schweitzer, J., 2001. HYPOSAT - An enhanced routine to locate seismic events, *Pure appl. geophys.,* **158,** 277–289.

Sleeman, R. & van Eck, T., 1998. Robust automatic P-phase picking: an on-line implementation in the analysis of broadband seismogram recordings, *Phys. Earth planet. Int.* **113,** 265–275.

Stewart, S.W., 1977. Real-time detection and location of local seismic events in Central California, *Bull. seism. Soc. Am.* **67**(2), 433–452.

Takanami, T. & Kitagawa, G., 1988. A new efficient procedure for estimation of onset times of seismic waves, *J. Phys. Earth* **36,** 267–290.

Takanami, T. & Kitagawa, G., 1991. Estimation of the arrival times of seismic waves by multivariate time series model, *Ann. Inst. Stat. Math.* **43,** 407–433.

VanDecari, J.C. & Crosson, R.S., 1990. Determination of teleseismic relative phase arrival times using multi-channel cross-correlation and least-squares, *Bull. seism. Soc. Am.* **80,** 150–169.

Withers, M., Aster, R., Young, C., Beiriger, J., Harris, M., Moore, S. & Trujillo, J., 1998. A comparison of select trigger algorithms for automated global seismic phase and event detection, *Bull. seism. Soc. Am.* **88**(1), 95–106.

Zhang, H., Thurber, C. & Rowe, C., 2003. Automatic P-wave arrival detection and picking with multiscale wavelet analysis for single-component recordings, *Bull. seism. Soc. Am.* **93,** 1904–1912.