GFZ
POTSDAM

Originally published as:

Galiana-Merino, J. J., Rosa-Herranz, J. L., Parolai, S. (2008): Seismic P Phase Picking Using a Kurtosis-Based Criterion in the Stationary Wavelet Domain. - IEEE Transactions on Geoscience and Remote Sensing, 46, 11, 3815-3826

DOI: 10.1109/TGRS.2008.2002647

Seismic *P* Phase Picking Using A Kurtosis Based Criterion In the Stationary Wavelet Domain

J.J. Galiana-Merino[(1)], J. Rosa-Herranz[(1)], S. Parolai[(2)]

[(1)]Dept. Physics, Systems Engineering and Signal Theory, University of Alicante,
P.O.Box 99, E-03080 Alicante, Spain
[(2)]GeoForschungsZentrum (GFZ) Potsdam,
Telegrafenberg, 14473 Potsdam, Germany

E-mail: juanjo@dfists.ua.es, julio@dfists.ua.es, parolai@gfz-potsdam.de

Submitted for publication at the IEEE Transactions on Geoscience and Remote Sensing

Submitted: June 12th, 2007

Abstract

The *P* seismic phase first arrival identification is a fundamental problem in seismology. The accurate identification of the *P*-wave first arrival is not a trivial process, especially when the seismograms present a very low signal-to noise ratio (SNR) or are contaminated with artificial transients that could produce false alarms. In this paper, a new approach based on higher-order statistics and the stationary wavelet transform is presented. The *P* onset is obtained under a statistical criterion applied in the time-frequency domain. The results have been compared to those estimated by other *P* phase picking algorithm and *P* onsets picked by expert analysts. The comparison shows that our proposed method efficiently provides a good estimate of the P onset picks that are consistent with analyst picks, especially in the cases of very low SNR.

Index Terms - Kurtosis, *P* phase identification, seismic signal processing, stationary wavelet transform (SWT).

1. Introduction

The detection and accurate picking of the first *P*-wave arrival is a crucial point in seismology as it provides important information about event detection, identification, source mechanism analysis, etc. Traditionally, this task has been carried out by human analysts in a visual way. Visual analysis, however is a very time consuming and subjective task, that cannot manage the huge volume of digital and real-time seismic data recorded today by a seismological observatory. Therefore, there is a need to provide a more efficient, rapid, objective and automatic detector of the *P* onset.

Various techniques can be found in the literature for detecting and picking the arrivals of different seismic waves from single-component as well as three-component recordings. Over the last two decades, numerous algorithms have been developed for *P* arrival identification based on energy analysis [1]-[5], polarization analysis [6]-[10], artificial neural networks [11], [12], maximum likelihood methods [13], [14], fuzzy logic theory [15], autoregressive techniques [16]-[19], higher-order statistics [20]-[24], or the wavelet transform [25]-[27].

Most of these methods satisfactorily determine the *P* phase arrival when the analyzed seismograms are contaminated with relatively low noise. When the SNR is low and the first arrival is not clear, however, most of them do not perform well, particularly when tested on real seismograms containing seismic noise that is not exactly zero-mean Gaussian and exhibits variable time-frequency features.

In this paper, we propose a new algorithm for estimating a reliable *P* onset on seismograms for cases when the SNR is very low (i.e. approximately less than 2dB).

Our scheme applies a kurtosis-based criterion in a time-frequency domain through the stationary wavelet transform (SWT) [28,29]. In this way, the non-Gaussian nature of the *P*-waves [30] and the time-frequency features of the seismic signal and noise are accounted for obtaining the first arrival.

In the following sections a brief theoretical introduction to the stationary wavelet transform and the kurtosis based criterion is presented, as well as details of the proposed method. The algorithm has been tested on synthetic seismograms, contaminated with real noise, to study the robustness of the proposed method under different SNR's. It has also been applied to local seismograms recorded by the seismic network of Alicante province, southeastern Spain, and to seismograms recorded at Bonn, Germany. Results are compared to those estimated by an algorithm based on maximum kurtosis and statistical criteria [22] and the *P* onset picked by expert analysts. The comparison shows that our proposed method provides the results closer to the manual procedure, especially in cases of very low SNR.

2. Mathematical background

2.1. The Wavelet and the Stationary Wavelet Transforms

For the case of non-stationary signals, a time-frequency analysis is more appropriate than a separate analysis in the time or in the frequency domain alone [31]. Some examples of time-frequency analysis are the Discrete Wavelet Transform (DWT) [32, 33] and the Discrete Stationary Wavelet Transform (DSWT) [28,29].

The DWT is based on H and G filters, that represent low-pass and high-pass filters, respectively, and on a downsampling operator D, that simply chooses every even

sample of a sequence.

A sub-band coding scheme [34] is used to calculate recursively the wavelet coefficients, $\lambda_{j,1}(n)$, of a discrete time series $s(n) \equiv \lambda_{0,0}(n)$, as follows.

$$\lambda_{j+1,0}(n) = D\,H\,\lambda_{j,0}(n)$$

$$\lambda_{j+1,1}(n) = D\,G\,\lambda_{j,0}(n) \qquad (1)$$

$$j = 0,..., L-1$$

where n is the sample index, j is the scale parameter and L is the maximum decomposition level. Sub-index 0 indicates coefficients from low-pass filters, while sub-index 1 indicates coefficients from high-pass filters. The wavelet coefficients $\{\lambda_{1,1}(n), \lambda_{2,1}(n),..., \lambda_{L,1}(n)\}$ characterize the details of the signal at different scales or resolutions, while the coefficients $\{\lambda_{1,0}(n), \lambda_{2,0}(n),..., \lambda_{L,0}(n)\}$ represent the approximation of the signal at different scales.

For a wavelet decomposition of L scales, the original signal, $\lambda_{0,0}(n)$ can be represented as $\{\lambda_{L,0}(n), \lambda_{L,1}(n),..., \lambda_{1,1}(n)\}$ from which the original signal can be reconstructed perfectly [32], [33].

The DWT is a shift-variant transform due to the downsampling operation. This means that the DWT of a translated version of a signal $s(n)$ is not, in general, the translated version of the DWT of $s(n)$. For some applications, this may not be an important issue as the signal is exactly recovered after applying the inverse transform. For other applications, however, such as the present one, where the sample corresponding to the first arrival has to be determined in the wavelet domain, shift-

invariance is a serious drawback.

To overcome this problem, the basic DWT algorithm can be modified to provide the DSWT that no longer depends on translations of the signal [28].

The DSWT is implemented using the same sub-band coding scheme. In this case, the appropriate low and high pass filters are applied to the data at each scale, but no downsampling is performed. Instead, the filters are modified at each scale by applying an operator *I*, which inserts a zero between every adjacent pair of elements, as follows:

$$H_{j+1} = I\,H_j \;\Rightarrow\; \begin{cases} h_{j+1}(2k) = h_j(k) \\ h_{j+1}(2k+1) = 0 \end{cases} \qquad (2)$$

$$j = 0,\ldots,\, L-2$$

where *k* is the sample index of the filter.

The stationary wavelet and approximation coefficients can be calculated recursively in the following way:

$$\lambda^{sta}_{j+1,0}(n) = H_j\,\lambda^{sta}_{j,0}(n)$$

$$\lambda^{sta}_{j+1,1}(n) = G_j\,\lambda^{sta}_{j,0}(n) \qquad (3)$$

$$j = 0,\ldots,\, L-1$$

where *n* is the sample index, *j* is the scale parameter and *L* is the maximum decomposition level. For a stationary wavelet decomposition of *L* scales, the original signal, $\lambda_{0,0}(n)$ can be represented as $\{\lambda^{sta}_{L,0}(n), \lambda^{sta}_{L,1}(n),\ldots, \lambda^{sta}_{1,1}(n)\}$, where all coefficients are of the same length as the original signal, rather than becoming shorter as the scale increases as in the standard DWT.

2.2. Kurtosis-Based Criterion

For a finite-length sequence $s(n)$, $n = 1,\ldots,N$, such as the real seismograms, with zero mean and fourth-order stationarity, the kurtosis is defined as [35]

$$K[s] = \frac{\sum_{n=1}^{N} (s(n) - \mu_s)^4}{(N-1)\cdot \sigma_s^4} - 3 \tag{4}$$

where μ_s and σ_s are the estimates of the mean and the standard deviation of the signal $s(n)$.

Kurtosis is a measure of the sharpness of a distribution. A Gaussian distribution has a kurtosis of 0. In contrast, outliers due to non-Gaussian signals (such as seismic events, and especially *P* arrivals) produce high kurtosis values, where the distribution has a sharper peak than normal distribution. Therefore, kurtosis can be an effective statistical tool for identifying signals with non-Gaussian features.

Real seismograms, however, are contaminated with noise, which does not always have a Gaussian distribution and could contain some spurious signals that maybe identified as false arrivals. Moreover, subsequent wave arrivals are non-Gaussian signals and could also have a high kurtosis value, possibly exceeding that of the first *P* onset. In these real cases, a kurtosis analysis in only the time-domain could fail, as the distribution associated with the noise or other arrivals may show high kurtosis values prior to or after the *P* arrival. Our proposed statistical time-frequency analysis circumvents this problem as the variations of the temporal and spectral characteristics of the seismic signal and noise are accommodated. Only the frequency bands related to the possible range of frequencies of the expected *P* arrivals are analyzed from a statistical

point of view. The effect of high frequency noise or artifacts, as well as that of low frequency oscillations or trends is strongly reduced. Moreover, the statistical analysis is also carried out on the rate of variation of the kurtosis to reduce the chances of false detections related to the high values of kurtosis reached when secondary arrivals are recorded (e.g. S-waves, surface waves). In section 4, we demonstrate the advantages of this kind of analysis.

3. Proposed method

3.1. Outline of the Algorithm based on Maximum Kurtosis and κ-Statistic Criteria

In the algorithm [22], the kurtosis is estimated along the seismic trace using a sliding time-window of length M with one-sample shift. For each window, a confidence interval for the kurtosis values is estimated to determine whether the M-sample observation follows a Gaussian distribution. This interval depends on M, and a specific probability factor, q. The P onset is estimated as the last sample preceding the point of maximum kurtosis for which the sequence follows a Gaussian behavior.

This process is repeated for different M and q values, providing a large set of P onset estimates. Initially, the most frequent value (MFV criterion) is selected as the P onset, with a confidence percentage c% (i.e., the fraction of occurrences to the total of estimates). If the corresponding confidence c is not sufficient, i.e., $c \leq 0.6$, then the frequency of occurrence histogram is smoothed iteratively until it exceeds the minimum required percentage (MF-I criterion). This provides an interval (instead of a single value) within which the P onset is contained.

3.2. Proposed scheme

The proposed *P* phase picking algorithm operates as follows:

1) Given the recorded signal $s(n)$, n=1,…,N , we apply the DSWT of L levels to obtain L+1 signals, each one of them associated with a specific frequency band. This new set of signals will be noted as $\lambda_{j,i}^{sta}(n)$, j=1,…L+1; n=1,…,N; i=0,1. In any case, only a subset associated with the frequency band of the analyzed seismic signal will be selected for subsequents steps. In our case, a 5-level decomposition has been performed using a Haar wavelet [32] as the mother wavelet. The Haar wavelet tends to emphasize discontinuities in a signal, such as the one associated with the first seismic arrival, in contrast with other smoother wavelet families more suitable for detecting less abrupt variations [32]. Tests done with different wavelets confirm this point. In general, scales 1 and 2 are very noisy with some transients that may produce false detections (see Fig. 1). Hence, only the scales 3, 4 and 5 ($\lambda_{3,1}^{sta}(n)$, $\lambda_{4,1}^{sta}(n)$ and $\lambda_{5,1}^{sta}(n)$) are selected- Such a signal selection improves the *P* arrival detection (avoiding false arrivals) and reduces computational time. These scales correspond approximately to the frequency band between 1.6 and 12.5 Hz, for a sampling frequency of 100 Hz. The selection of these scales can be adjusted depending on the expected frequencies of the first *P* arrival (i.e. it may vary for local, regional and teleseismic events) and the sampling frequency.

2) An M-sample sliding window is applied to the selected signals $\lambda_{j,i}^{sta}(n)$ with an overlap of M-1 samples. In other words, the window is moved one sample at a time. For every window, the kurtosis, $K[\lambda_{j,i}^{sta}(n)]$ is calculated according to (4). The estimated value of the kurtosis is assigned to the end point of each window (sample M).

3) The rate of change of kurtosis is calculated through differentiation, obtaining the slope at every point, $K'[\lambda_{j,i}^{sta}(n)] \equiv \Delta K[\lambda_{j,i}^{sta}(n)]/\Delta n$. In real seismograms, the seismic signal and the noise may not follow a Gaussian; hence, a high kurtosis value may not be associated with the first *P* arrival. For this reason, the rate of change of the kurtosis becomes more suitable than the kurtosis itself for determining the different seismic arrivals, particularly the first arrival. We demonstrate this case in Fig. 2b, where the maximum kurtosis is at sample 7424 (scale 4), which does not correspond exactly to the beginning of the first *P* arrival.

4) The maximum slope is found for all $K'[\lambda_{j,i}^{sta}(n)]$ series obtained in the previous step, yielding the maximum value, $K'[\lambda_{j\_\max,i}^{sta}(n\_\max)]|_{\max}$, the position of the maximum, *n_max*, and the corresponding stationary wavelet scale, *j_max*. In Fig. 2c, the maximum slope of the kurtosis is located at sample 7421 (scale 4) which is closer to the analyst pick (at sample 7414) than that obtained using the maximum kurtosis alone.

5) In general, the sample *n_max* corresponds to the *P* onset of the seismogram, but not always. Sometimes, the maximum slope of the kurtosis is associated with secondary arrivals. That is why the procedure searches in all the selected scales for a sample previous to *n_max* with amplitude higher or equal to $Thr \cdot K'[\lambda_{j\_\max,i}^{sta}(n\_\max)]|_{\max}$, with $0<Thr<1$. If samples with these values exist, the one corresponding to the lower sample is chosen as the first *P* arrival. If not, the sample corresponding to $K'[\lambda_{j\_\max,i}^{sta}(n\_\max)]|_{\max}$ is chosen. Thus, the *P* onset is characterized by a new set of parameters, $K'[\lambda_{j\_thr,i}^{sta}(n\_thr)]|_{thr}$, *n_thr*, *j_thr*, rather than those of the

maximum value. Results obtained by testing different thresholds will be shown in section 4.3. In most cases, the sample obtained after the threshold check is equal, or very close, to *n_max*, In Fig.2c, the maximum slope of the kurtosis is at sample 7421 (scale 4), but when a threshold of 70% is applied, the picking is moved to the sample 7410 (scale 5), which is even closer to the analysts manual pick (at sample 7414).

4. Results and discussion

4.1. Dataset

Our method has been evaluated using three different datasets.

First, we applied the procedure to synthetic signals. The generation of the pure synthetic signal follows a simple scheme that includes body and Rayleigh waves. Sources were modeled following the Brune source function model [36, 37] and were located at a depth and epicentral distance of 7km and 13km respectively. The simulation involved a crustal model used for locating earthquakes in the Alicante area that consists of three homogeneous layers [38] .The synthetic seismogram was then added to real seismic noise recorded at stations of the seismic network of Alicante province, southeast Spain, and convolved with the instrument response for stations of the network. Real noise has been scaled by different factors in order that the ratio between the root mean square (rms) of the whole synthetic seismogram and that of noise recordings assumed fixed a priori values (14, 6 and 0dB). Fig. 3 shows an example of some of these synthetic signals. The sampling frequency was f_s = 100 Hz and the P onset of the signal is at sample 312.

The second dataset consists of 32 seismograms of local events recorded by the

Alicante seismic network (southeast Spain). In this case, data were collected at a sampling rate of 100 Hz.

The last dataset consists of seismograms of a local earthquake that occurred on 01:01:2004 at 21:10:6.3 UTC (M_L=2), and a regional event on 23:02:2004 at 17:31:20.0 UTC (M_L=5.9), recorded by an array of stations installed in the urban area of Bonn, Germany.

Our analysis was performed on a personal computer (Intel Pentium Centrino, 1.56Ghz, 256MB RAM), using Matlab 7.0 (Mathworks, Inc.).

4.2. Evaluation on synthetic seismograms

Our method has been evaluated using 12 synthetic signals contaminated with real noise of different characteristics and amplitudes. In all these signals, the *P* onset was fixed to be at sample 312. We show in table I, the results obtained without (column 4) and with (column 5) applying a threshold (step 5, section 3.2); those estimated through the MFV / MF-I criteria (column 3) are also shown, with the corresponding confidence factor indicated in brackets. If this is smaller than 0.6, then the MF-I criterion is applied [22] and the interval within which the *P* onset (with higher confidence factor) should lie is estimated and shown. For the application of this method, we have employed time-window lengths of T = 2, 3,…, 8s ($M=T\cdot f_s$) and probabilities of q = 0.70, 0.71,…,0.99, as suggested in [22]. The threshold was fixed for all our analyses to 70% of the maximum slope of the kurtosis. After testing with real seismograms (see section 4.3 for more details about the test), we show that this threshold provides the minimum overall mean bias, respecting to the results obtained with other threshold

values or even no-threshold. Our results have been obtained using a time-window length of 250 samples (2.50s), although other sizes were also tested. This period of time is enough to assure that more than one period of the dominant frequency in the signal will be contained in each window and therefore, the kurtosis values will be more reliable.

Results indicate that our method provides a nearly exact estimate of *P* arrival (at sample 312 or 313) for 8 out of 12 cases. Indeed, the results show the larger inaccuracy for only one case, with a deviation of 3 samples. For signal 6 (Fig. 3c), with very low SNR and a spurious low frequency oscillation of approximately 2Hz, we obtain a difference of 2 samples, which improves considerately the result obtained by the MFV / MF-I criteria for this event. For cases 10, 11 and 12 (contaminated by real noise recorded at one of the stations), results for all methods show that the accuracy of the picking algorithm is not only dependent upon the SNR, but also on the time-frequency features of the seismic noise. When some signals have a relatively high SNR, our method provides a *P* arrival with bias of 2 and 3 samples due to the time-frequency dependent characteristics of the seismic noise, a consequence of using real seismic noise instead of Gaussian or synthetic noise.

For the synthetic signals and the parameters used (time-window length of 250 samples and threshold of 70%) the results obtained with our method do not vary whether the threshold criterion is adopted or not. This is likely because the threshold criterion is used to improve the estimate accuracy when the first *P* arrival is very weak with respect to secondary waves arrivals (see Fig. 2), which is not the case for the synthetic seismograms at hand.

We tested the stability of our results to explore the effect of different time-

windows lengths (of 125, 70 and 40 samples). In general, for SNR = 14dB, the results appear to be independent of the window length; however, for SNR = 0dB, results obtained using window lengths of 40 or 70 samples show greater errors in the first *P*-wave arrival estimates. As the window length decreases, the value for the maximum slope of kurtosis also decreases, making it more difficult to separate this maximum from other peaks produced by noise or other spurious signals.

The MFV / MF-I criteria provide relatively good results, although the mean deviation with respect to the exact *P* onset is higher than that from our method (4%, excluding cases 6 and 9). The MFV criterion estimates the exact *P* onset, at sample 312, with a confidence factor of 0.47 in only one case. Moreover, in cases 5, 6 (Fig. 3c) and 9, this algorithm performed poorly, estimating the pick within an interval of several hundred samples.

4.3. Evaluation using real seismograms

We also applied our technique to real seismograms, and compared the results to those obtained by the MFV / MF-I criteria and the *P* onset estimated by an expert analyst of the Alicante Network (tables II and III). The parameters used for our method and for the MFV / MF-I criteria are the same as those used for the synthetic seismograms.

Taking as a reference the analyst picks, our method provides excellent results, with a mean bias of 8 samples, and a standard deviation of 14 samples for data listed in Table II, and a mean bias of 6 samples, and a standard deviation of 5 samples for data listed in Table III. The major bias in Table II is produced in case 26 (Fig. 4), which is a

very noisy seismogram with low frequency oscillations of high amplitude and where the frequency of the *P* arrival is mainly around 17Hz. This also suggests that the scale 2 of the SWT, which corresponds approximately to a frequency band between 12.5 and 25Hz, must be considered in the subsequent analysis. Increasing the number of scales in the analysis results in a selected *P* onset at sample 2968, 3 samples after the analyst's pick. With this modification to case 26, the mean bias and the standard deviation for the events of Table II are 6 and 5 samples, respectively.

It is important to note that our method provides estimates in good agreement with the analyst ones for all cases we have analyzed, including seismograms with very low SNR, low frequency oscillations and spurious signals before the first *P*-wave arrival. In contrast, the MFV / MF-I criteria is either not always able to correctly estimate the *P* onset, or able to locate the pick within a sufficient degree of confidence in an interval smaller than 100 samples. The large uncertainties in picks obtained by the MFV / MF-I criteria (cases 4, 7, 13, 19, 23, 27 and 28 of Table II and cases 8 and 12 of Table III) made it impossible for 9 out of 51 cases in Tables II and III to compare the results with those obtained by our method. For the remaining cases, adopting the MFV / MF-I criteria, the *P* onset was estimated, with respect to the analyst results, with a mean bias of 15 samples, a standard deviation of 19 samples for data from Table II, and a mean bias of 14 samples, a standard deviation of 17 samples for data from Table III.

Fig. 5 shows graphically the advantages of working in the time-frequency instead of only the time domain. For example, case 27, Table II, is analyzed both in the time domain (Fig.5b-c) and following the steps we proposed in the time-frequency domain (Fig. 5d-e). If the kurtosis and the slope are studied only in the time domain (i.e.

using a time-window length of 250 samples), the *P* onset estimation will be located at sample 3344 (the maximum slope of the kurtosis, which occurs before the maximum of the kurtosis, at sample 3382). This is very different from the analyst estimate (sample 2646). In contrast, the analysis in the time-frequency domain provides an estimate closer to that of the analyst. In fact, analyzing the kurtosis and the corresponding slope in scales 3, 4 and 5 of SWT, the P onset estimate is at sample 2651 (maximum slope of the kurtosis, at scale 3). This clearly shows the importance of the seismic noise spectral shape in affecting the picking results.

For the seismograms of Table II, our algorithm has also been evaluated for time-window lengths of 500, 125, 70 and 40 samples. 40- and 70-sample windows are ineffective even when the analyzed signals have a high SNR. The 125-sample window works fairly well only for seismograms with high SNR. Only 250- and 500-sample windows lead to correct estimates of first *P*-arrival in all cases. Since the calculation time depends directly on the window length, it is very important to select the minimum number of samples that guarantee satisfactory results under the worst noise conditions- As the window length decreases, the maximum slope of the kurtosis also decreases and the selected threshold could be raised.

We have also tested the effect of the threshold in the mean bias of our results. Obviously, if the threshold is equal to 1, we are choosing directly the value obtained for the maximum slope of the kurtosis. In the other way, if the threshold is too small, other peaks produced by noise or other spurious signals could be selected as the pick. In Table IV, we show the mean bias obtained for the 50 seismograms of Tables II and III when different thresholds are used. We indicate also the number of cases where the

difference between our estimates and the analysts pick is higher than 50 samples. Finally, last row compares these results with those obtained by the MFV/MF-I criteria. We can see that our method yields better estimates than those obtained using the MFV/MF-I criteria even when the threshold is not used (i.e. threshold fixed to 1). From this test, we can also observe that a threshold of 0.7 reduces considerably the mean bias to only 7 samples, that is much lower than the value of 33 samples obtained when no-threshold is used or than the 142 samples obtained by the MFV/MF-I criteria. Therefore, we can conclude that a threshold of 0.7 is well-adequate to estimate properly the first *P* arrival in our automatic *P*-wave picking scheme.

Another additional comparison can be carried out considering the computation time required by the different methods. For the seismograms in Table II, our method accomplishes the *P* onset estimation in an average of about 12 seconds, whereas use of the MFV / MF-I criteria, requires more than 8 minutes. This is because the algorithm based on the MFV / MF-I criteria repeats the *P* estimation process for different time-window lengths and probabilities to also provide a confidence interval.

5. Conclusions

In this paper, a new method for automatic *P* onset picking for very low SNR signals has been presented, based on the SWT and the kurtosis. The proposed method estimates properly the first *P* arrival for all analyzed seismograms, even in cases with very low SNR, low frequency oscillations or previous artifacts.

Our algorithm has been tested on synthetic signals contaminated with real noise

of different time-frequency characteristics and amplitudes, providing good results for all the analyzed signals, independent of the kind of noise. Moreover, results have been compared with those obtained by an algorithm based on the application of maximum kurtosis and k-statistics criteria in the time domain [22]. The comparison highlights how our method is less sensitive to both the SNR and the kind of noise than the time-domain methods.

Our scheme has been applied to real seismograms, providing for all of them good *P* onset estimations, even in cases for which other time-based methods could be not precisely estimate the onset or did so with a large uncertainty.

Our method is relatively fast computationally in comparison to the other algorithm we tested and to the time taken for the P onset determination by an expert analyst. This makes it particularly attractive for early warning and rapid response systems as well as for the fast and automated picking within large data sets collected by temporary field campaigns.

Finally, in the case of seismograms with good SNR, our method may support an expert analyst in identifying different arrivals for regional events (*Pn*, etc) (i.e. Fig. 6), since it also provides other peaks of amplitude lower than the associated to the selected pick. The potential of the method to this regard will be subject of future studies.

Acknowledgments

We are very thankful to P. Jáuregui for his help in analyzing the seismograms. K. Fleming kindly improved our English. The local seismograms recorded in Alicante

were provided by Local Seismic Network of the University of Alicante (supported by Diputación de Alicante). The Geophysical Instrument Pool Potsdam (GFZ) provided instruments for the experiment in Bonn. Finally, we would also like to thank the associate editor and the anonymous reviewers for their comments, which helped us to clarify and improve this article.

References

[1] R. V. Allen, "Automatic earthquake recognition and timing from signal traces," *Bull. Seismol. Soc. Amer.*, vol. 68, pp. 1521–1532, Oct. 1978.

[2] B. O. Ruud and E. S. Husebye, "A new three-component detector and automatic single-station bulletin production," *Bull. Seismol. Soc. Amer.*, vol. 82, pp. 221-237, Feb. 1992.

[3] P. S. Earle and P. M. Shearer, "Characterization of global seismograms using an automatic-picking algorithm," *Bull. Seismol. Soc. Amer.*, vol. 84, pp. 366-376, Apr. 1994.

[4] J. Saari, "Automated phase picker and source location algorithm for local distances using a single three-component seismic station," *Tectonophysics*, vol. 189, pp. 307-315, 1991.

[5] M. Berand and U. Kradolfer, "An automatic phase picker for local and teleseismic events," *Bull. Seismol. Soc. Amer.*, vol. 77, pp. 1437-1445, Aug. 1997.

[6] A. Jurkevic, "Polarization analysis of three component array data," *Bull. Seismol. Soc. Amer.*, vol. 78, pp. 1725-1743, Oct. 1988

[7] N. Magotra, N. Ahmed, and E. Chael, "Single-station seismic event detection and location," *IEEE Trans. Geosci. Remote Sensing*, vol. 27, pp. 15-23, Jan. 1989.

[8] N. Magotra, N. Ahmed, and E. Chael, "Seismic event detection and source location using single-station (three-component) data," Bull. Seismol. Soc. Amer., vol. 77, pp. 958-971, June 1987.

[9] S. Anderson and A. Nehorai, "Analysis of a polarized seismic wave model," *IEEE Trans. Signal Processing*, vol. 44, pp. 379-386, Feb. 1996.

[10] J. E. Vidale, "Complex polarization analysis of particle motion," *Bull. Seismol. Soc. Amer.*, vol. 76, pp. 1393-1405, 1986.

[11] Y. Zao and K. Takano, "An artificial neural network approach for broadband seismic phase picking," *Bull. Seismol. Soc. Amer.*, vol. 77, 89, pp. 670-680, June 1999.

[12] J. Wang and T. Teng, "Artificial neural network-based seismic detector," *Bull. Seismol. Soc. Amer.*, vol. 85, pp. 308-319, Feb. 1995.

[13] R. G. Robert, A. Christoffersson, and F. Cassidy, "Real-time event detection, phase identification and source location estimation using single station three-component seismic data," *Geophys. J.*, vol. 97, pp. 471-480, 1989.

[14] A. Christoffersson, E. S. Husebye, and S. F. Ingate, "Wavelet decomposition using ML-probabilities in modeling single-site 3-component records," *Geophys. J.*, vol. 93, pp. 197-213, 1988.

[15] C. –K. Chu and J. Mendel, "First break refraction event picking using fuzzy logic systems," *IEEE Trans. Fuzzy Syst.*, vol. 2, pp. 255-266, Nov. 1994.

[16] N. Maeda, "A method for reading and checking phase times in autoprocessing system of seismic wave data," *Jishin*, vol. 38, pp. 365-379, 1985.

[17] R. Sleeman, and T. van Eck, "Robust automatic P-phase picking: an on-line implementation in the analysis of broadband seismogram recordings," *Phys. Earth Planet, Interiors*, vol. 113, pp. 265-275, 1999.

[18] M. Leonard, and B. L. N. Kennet (1999), "Multi-component autoregressive techniques for the analysis of seismograms," *Phys. Earth Planet, Interiors*, vol. 113, pp. 247-264, 1999.

[19] M. Leonard, "Comparison of manual and automatic onset time picking," *Bull. Seismol. Soc. Amer.*, vol. 90, pp. 1384-1390, 2000.

[20] S. K. Yung and L. T. Ikelle, “An example of seismic time picking by 3rd order bicoherence,” *Geophysics*, vol. 62, no. 6, pp. 1947-1952, 1997.

[21] C. D. Saragiotis, L. J. Hadjileontiadis, and S.M. Panas, “PAI-S/K: a robust automatic seismic P phase arrival identification scheme,” *IEEE Trans. Geosci. Remote Sensing*, vol. 40, pp. 1395-1404, June 2002.

[22] C. D. Saragiotis, L. J. Hadjileontiadis, I. T. Rekanos, and S.M. Panas, “Automatic P phase picking using maximum kurtosis and *k*-statistics criteria,” *IEEE Geosci. Remote Sensing Letters*, vol. 1, no. 3, pp. 147-151, July 2004.

[23] C. D. Saragiotis, L. J. Hadjileontiadis, S. M. Panas, “A higher-order statistics-based phase identification of three-component seismograms in a redundant wavelet transform domain,” Proc. IEEE Workshop High-Order Statistic, pp. 396-399, June 14-16 1999.

[24] F. Poletto, "A blind interpretation of drill-bit signals," *Geophysics*, vol. 65, no. 3, pp. 970-978, May-June 2000].

[25] K. S. Anant and F. U. Dowla, “Wavelet transform methods for phase identification in three-component seismograms,” *Bull. Seismol. Soc. Amer.*, vol. 87, pp. 1598-1612, Dec. 1997.

[26] F. Botella, J. Rosa-Herranz, J. J. Giner, S. Molina, and J.J. Galiana-Merino, “A real-time earthquake detector with prefiltering by wavelets,” *Computers & Geosciences*, vol. 29, pp. 911-919, 2003.

[27] H. Zhang, C. Thurber, and C. Rowe, “Automatic P-wave arrival detection and picking with multiscale wavelet analysis for single-component recordings,” *Bull. Seismol. Soc. Amer.*, vol. 93, no. 5, pp. 1904-1912, Oct. 2003.

[28] G. P. Nason, and B. W. Silverman, “The stationary wavelet transform and some statistical applications,” *Lecture Notes in Statistics*, vol. 103, pp. 281-299, 1995.

[29] S. Solbϕ, and T. Eltoft, "A stationary wavelet-domain Wiener filter for correlated speckle," *IEEE Trans. Geosci. Remote Sensing*, vol. 46, no. 4, pp. 1219-1230, April 2008.

[30] G. B. Giannakis and M. K. Tsatsanis, "Time-domain tests for gaussianity and time-reversibility," *IEEE Trans. Signal Processing*, vol. 42, pp. 3460-3472, Dec. 1994.

[31] P. Steeghs, *Local Power Spectra and Seismic Interpretation*, Ph.D. Thesis, Delft University of Technology, The Netherlands, 1997.

[32] I. Daubechies, "Ten lectures on wavelets," in *CBMS-NSF Reg. Conf. Series Appl. Math.* Philadelphia, PA: SIAM, 1992, vol. 61.

[33] M. V. Wickerhauser, *Adapted Wavelet Analysis from Theory to Software*, A. K. Peters, Ltd., Wellesley, 1994.

[34] S. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, pp. 674-693, July. 1989.

[35] P. J. Bickel and K. A. Doksum, *Mathematical Statistics*. San Francisco, CA: Holden-Day, 1977.

[36] J. Brune, "Tectonic stress and the spectra of seismic shear waves from earthquakes." J. Geo. Res., vol. 75-26, pp. 4997-5009, 1970.

[37] J. Brune, "Correction to Tectonic stress and the spectra of seismic shear waves from earthquakes," J. Geo. Res., vol. 76-20, pp. 5002, 1971.

[38] P. Jáuregui, *Estudio de la operatividad de redes sísmicas locales aplicado a la optimización de los recursos de la Red Sísmica Local de la Universidad de Alicante*, Ph.D. Thesis, University of Alicante Spain, 1997 (in Spanish).

Biographies

Juan J. Galiana-Merino (M'06) was born in Alicante, Spain in 1972. He received a Diploma in electrical engineering in 1995 and a Diploma in physics in 1996, both from the University of Valencia, Valencia, Spain, and a Ph.D. in computer engineering from the University of Alicante, Alicante, Spain, in 2001.
He has been professor at the University of Alicante since 2000. He was also a professor at the University Miguel Hernandez, Spain, in 1999. Since 1998, he has been member of the Local Seismic Network of the University of Alicante. His research interests are in digital signal processing, wavelets and time-frequency analysis for geophysical applications, and seismic signal processing.
Dr. Galiana-Merino is a member of the American Geophysical Union. He received a PhD grant by the Local Government of Alicante.

Julio Luis Rosa Herranz was born in Segovia, Spain in 1956. He received a Diploma in physics in 1979 from the University of Valencia, Valencia, Spain, and a Ph.D. in computer engineering from the University of Alicante, Alicante, Spain, in 1997.
He is a Professor at the University of Alicante since 1993. He previously worked for Nixdorf Computer S.A y Siemens-Nixdorf S.I companies as a data transmission specialist. Since 1998, he has been a member of the Local Seismic Network of the University of Alicante. His research interests are data acquisition and data transfer for geophysical applications and seismic signal processing.

Stefano Parolai was born in Genova, Italy in 1968. He received the Diploma in Geological Sciences in 1993, and a Ph.D. in Geophysics in 1997, both from the University of Genova, Genova, Italy.
He has been a senior scientist at the GeoForschungsZentrum Potsdam since 2004. He was also a researcher at the Osservatorio Geofisico di Macerata, Italy, in 1999. His main research interests are in engineering seismology, and time-frequency analysis for geophysical applications.

TABLE I
P ARRIVAL ESTIMATES FOR THE SYNTHETICS SEISMOGRAMS (ANALYST PICK EQUALS TO 312)

Case	Station Id. / SNR (dB)	MFV / MF-I Criteria	*j_max* / *n_max*	*j_thr* / *n_thr*
1	crev / 14	314 (.89)	3 / 312	3 / 312
2	crev / 6	317 (.60) / [317-318]	3 / 313	3 / 313
3	crev / 0	318 (.61)	3 / 313	3 / 313
4	maig / 14	315 (.84)	3 / 312	3 / 312
5	maig / 6	317 (.60) / [225-678]	3 / 313	3 / 313
6	maig / 0	754 (.18) / [551-1100]	3 / 314	3 / 314
7	puja / 14	315 (.78)	3 / 312	3 / 312
8	puja / 6	317 (1.0)	3 / 313	3 / 313
9	puja / 0	754 (.36) / [317-755]	3 / 313	3 / 313
10	univ / 14	312 (.47) / [312-317]	3 / 314	3 / 314
11	univ / 6	316 (.55) / [316-317]	3 / 314	3 / 314
12	univ / 0	319 (1.0)	3 / 315	3 / 315

TABLE II
P ARRIVAL ESTIMATES FOR THE LOCAL SEISMOGRAMS (ALICANTE PROVINCE, SPAIN)

Case	Event , Station Id.	Analyst	MFV / MF-I Criteria	*j_max* / *n_max*	*j_thr* / *n_thr*	Analyst comments
1	02/09/22 , maig	2355	2392 (.99)	3 / 2363	4 /2350	
2	02/09/22 , crev	2297	2213 (.94)	3 / 2295	3 / 2295	low frequency oscillations
3	02/09/22 , puja	2339	2378 (.92)	4 / 2338	4 / 2337	low frequency oscillations, weak P arrival
4	02/09/22 , univ	2717	4051 (.38) [4047-4094]	4 / 2725	4 / 2724	low frequency oscillations
5	02/09/22 , unip	2626	2586 (.95)	3 / 2650	3 / 2624	Noisy signal, not clear P arrival
6	02/12/24 , maig	2434	2424 (.40) [2424-2438]	3 / 2446	3 / 2446	weak P arrival
7	02/12/24 , crev	2500	3553 (.74)	3 / 2502	3 / 2502	low frequency oscillations, weak P arrival
8	02/12/24 , puja	2639	2622 (.53) [2622-2623]	4 / 2656	4 / 2656	low frequency oscillations
9	02/12/24 , univ	2161	2170 (.50) [2170-2171]	3 / 2164	4 / 2159	low frequency oscillations of high amplitude
10	02/12/24 , unip	2064	2075 (.28) [1987-2083]	3 / 2074	4 / 2062	very noisy signal
11	02/12/26 , crev	2218	2224 (.78)	3 / 2254	3 / 2219	
12	02/12/26 , puja	2640	2633 (.75)	3 / 2651	3 / 2651	
13	02/12/26 , univ	3120	4182 (.14) [3118-4912]	4 / 3117	4 / 3116	low frequency oscillations of high amplitude
14	02/12/26 , unip	3547	3581 (.31) [3578-3581]	5 / 3532	5 / 3531	very noisy signal
15	03/02/07 , maig	2763	2761 (.61)	3 / 2773	3 / 2773	
16	03/02/07 , crev	2244	2246 (.51) [2235-2246]	3 / 2241	3 / 2241	
17	03/02/07 , puja	2873	2878 (.92)	4 / 2866	4 / 2866	
18	03/02/07 , univ	2994	3002 (.54) [3001-3004]	3 / 3016	3 / 3015	
19	03/02/07 , unip	2807	2807 (.47) [2807-3016]	4 / 2811	4 / 2800	noisy signal
20	03/02/16 , maig	2853	2859 (.55) [2858-2859]	4 / 2854	4 / 2854	
21	03/02/16 , crev	2335	2337 (.51) [2335-2337]	3 / 2331	4 / 2324	
22	03/02/16 , puja	2977	2982 (.56) [2982-2983]	3 / 2978	3 / 2978	low frequency oscillations
23	03/02/16 , univ	3097	2839 (.08) [1735-2839]	3 / 3105	3 / 3105	low frequency oscillations of high amplitude
24	03/02/27 , maig	2303	2304 (.83)	3 / 2312	3 / 2299	
25	03/02/27 , crev	2420	2428 (.79)	3 / 2423	3 / 2423	low frequency oscillations
26	03/02/27 , puja	2965	2999 (.65)	3 / 3044	3 / 3044	very noisy signal, low frequency oscillations
27	03/02/27 , univ	2646	3342 (.38) [2651-5205]	3 / 2651	3 / 2650	low frequency oscillations of high amplitude
28	03/02/27 , unip	2568	3105 (.71)	4 / 3079	3 / 2566	very noisy signal, poor quality
29	03/08/26 , crev	2392	2396 (.35) [2395-2396]	5 / 2390	5 / 2389	
30	03/08/26 , puja	2273	2266 (.99)	3 / 2272	3 / 2272	
31	03/08/26 , univ	3239	3242 (.29) [3241-3244]	3 / 3248	4 / 3243	low frequency oscillations
32	03/08/26 , unip	3011	3014 (.43) [3013-3016]	3 / 3020	3 / 3019	very noisy signal

TABLE III
P ARRIVAL ESTIMATES FOR THE LOCAL AND REGIONAL EARTHQUAKES (BONN, GERMANY)

Case	Analyst	MFV / MF-I Criteria	*j_max* / *n_max*	*j_thr* / *n_thr*	Analyst comments
1	1520	1529 (.96)	3 / 1525	3 / 1525	low frequency noise
2	1480	1483 (.24) [1464-1493]	3 / 1492	4 / 1479	noisy signal
3	1547	1553 (.89)	3 / 1553	3 / 1553	very noisy signal
4	1535	1537 (.51) [1537-1540]	3 / 1536	3 / 1536	noisy signal
5	1562	1564 (.68)	3 / 1572	3 / 1572	noisy signal
6	1483	1511 (47%) [1502-1512]	3 / 2155	3 / 1497	noisy signal, not clear P arrival
7	1839	1840 (.99)	3 / 1836	3 / 1836	very noisy signal
8	1540	485 (.55) [485-1551]	3 / 1547	3 / 1547	very noisy signal, poor quality
9	13583	13569 (.76)	4 / 13602	4 / 13601	Very noisy signal
10	13462	13512 (.23) [13511-21400]	4 / 13475	5 / 13450	very noisy signal, poor quality
11	13526	13529 (.41) [13512-13530]	3 / 13530	4 / 13526	
12	13431	12889 (.55) [11629-12889]	5 / 13427	5 / 13426	previous artifacts, not clear P arrival
13	7433	7423 (.38) [7422-7423]	5 / 7433	5 / 7432	
14	7430	7432 (.48) [7431-7432]	4 / 7433	4 / 7432	
15	7470	7478 (.33) [7471-7483]	3 / 7479	3 / 7478	noise signal,
16	7414	7421 (.45) [7421-7422]	4 / 7421	5 / 7410	previous artifact
17	7502	7473 (.34) [7471-7474]	3 / 7499	3 / 7499	noisy signal
18	7402	7423 (.31) [7422-7436]	4 / 7441	5 / 7409	very noisy signal, low frequency oscillations

TABLE IV
MEAN BIAS BETWEEN THE PICK ESTIMATED FOR DIFFERENT THRESHOLDS AND ALSO FOR THE MFV/MF-I CRITERIA, AND THE PICK ESTIMATED BY THE ANALYST

Threshold	Mean Bias	Number of events with a difference between the P estimated and the analyst pick higher than 50 samples
<=0,3	>200	>4
0,4	59	4
0,5	43	3
0,6	8	2
0,7	7	1
0,8	19	2
0,9	33	3
1	33	3
MFV/MF-I	142	10

Figure captions:

Fig. 1. (a) An example of a real, very noisy seismogram (case 8, Table III), with some previous transients, and (b) the stationary wavelet coefficients of the seismogram. 5 levels and a 'haar' mother wavelet have been used for the SWT analysis.

Fig. 2. (a) Zoomed version around the *P* arrival of a real seismogram (case 16, Table III), with the pick determined by the analyst (solid line) and the ones obtained following the maximum kurtosis (dashed line), the maximum slope of the kurtosis (dotted line) and the 70% of the maximum slope of the kurtosis (dot-dashed line) criteria. In this case, the pick estimated by the MFV/MF-I criteria coincides with the maximum slope of the kurtosis (dotted line) (b) Analysis of the kurtosis and (c) the slope of the kurtosis in the stationary wavelet domain.

Fig. 3. Examples of synthetic seismograms contaminated with real noise of different characteristics and SNR. (a) SNR=14dB, (b) SNR=6dB, and (c) SNR=0dB.

Fig. 4. (a) Real seismogram (case 26, Table II), and (b) the Fourier amplitude spectrum of the first *P* arrival.

Fig. 5. (a) Real seismogram (case 27, Table II), (b) kurtosis on the time domain, (c) Slope of the kurtosis in the time domain, (d) kurtosis on the time-frequency domain, scales 3, 4 and 5 of the SWT, and (e) slope of the kurtosis on the time-frequency domain, scales 3, 4 and 5 of the SWT. The picks determined by the analyst (solid line), the MFV/MF-I criteria (dotted line) and our method (dot-dashed line) are also compared.

Fig. 6. (a) Zoomed version around the *P* arrival of a real seismogram (case 1, Table II), with the *P*-wave (solid line) and *Pn* wave (dashed line) estimated by the analyst and the picks estimated by the MFV/MF-I criteria (dotted line) and the proposed method (dot-dashed line). (b) Slope of the kurtosis in the stationary wavelet domain.

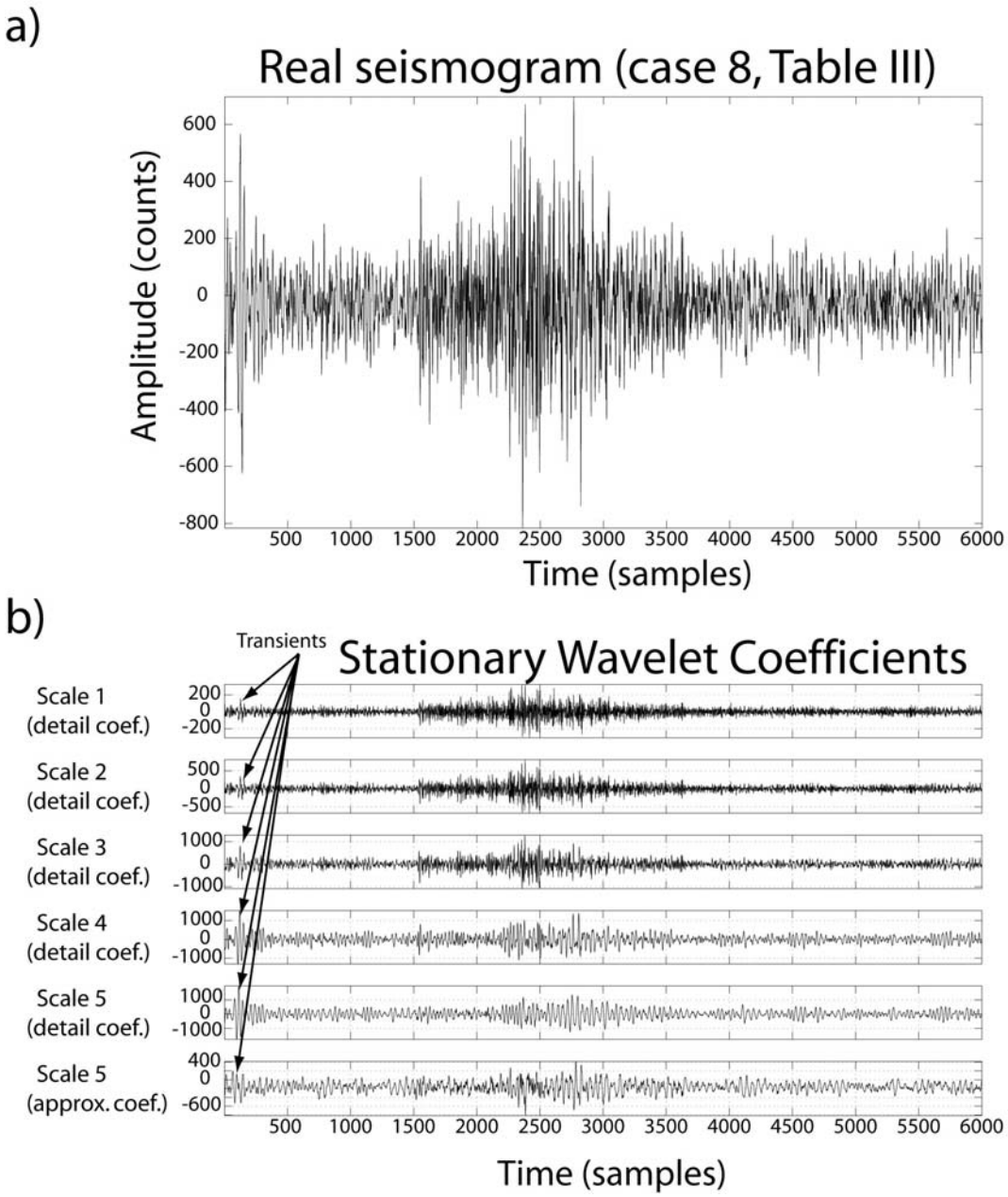

Fig. 1. (a) An example of a real, very noisy seismogram (case 8, Table III), with some previous transients, and (b) the stationary wavelet coefficients of the seismogram. 5 levels and a 'haar' mother wavelet have been used for the SWT analysis.

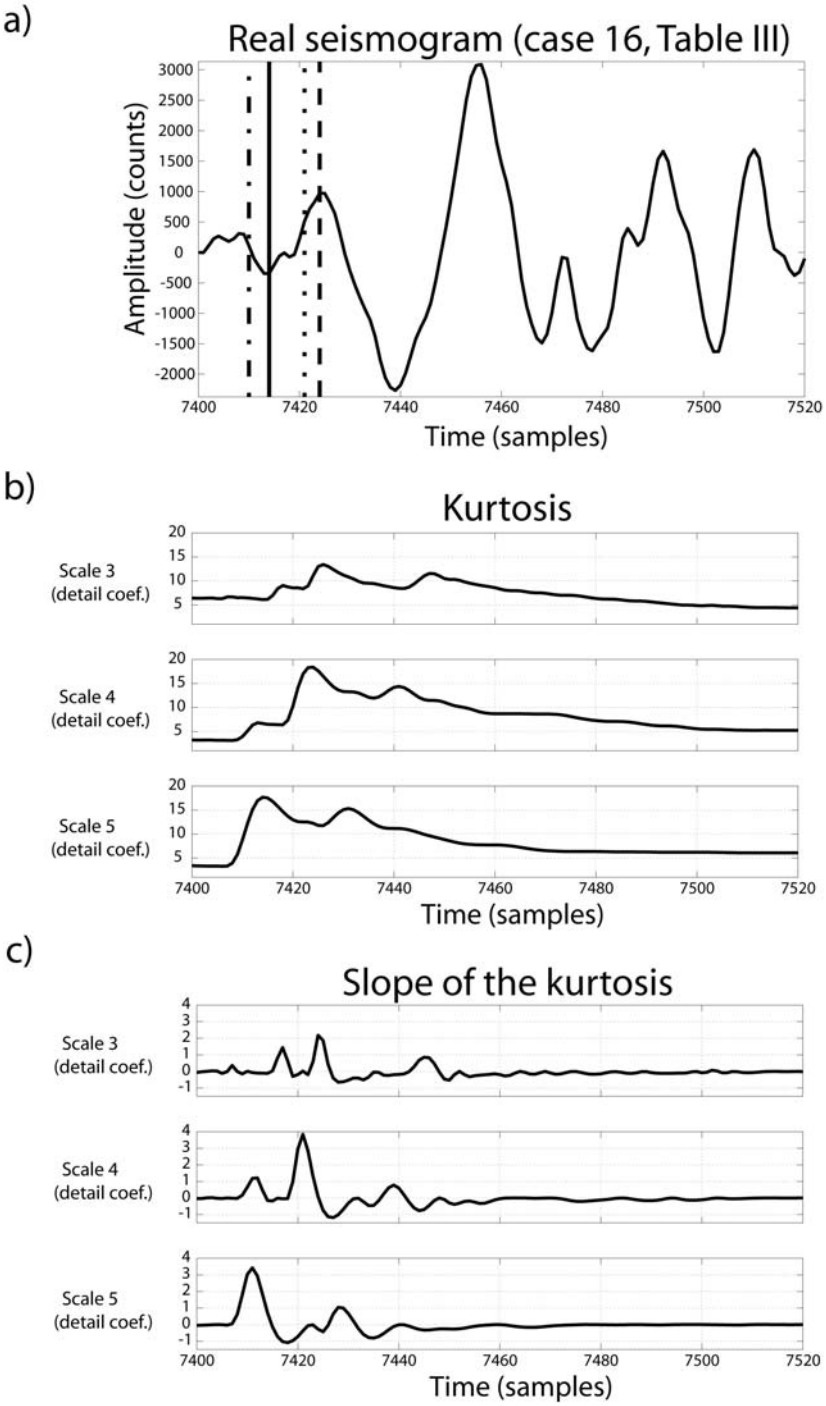

Fig. 2. (a) Zoomed version around the *P* arrival of a real seismogram (case 16, Table III), with the pick determined by the analyst (solid line) and the ones obtained following the maximum kurtosis (dashed line), the maximum slope of the kurtosis (dotted line) and the 70% of the maximum slope of the kurtosis (dot-dashed line) criteria. In this case, the pick estimated by the MFV/MF-I criteria coincides with the maximum slope of the kurtosis (dotted line) (b) Analysis of the kurtosis and (c) the slope of the kurtosis in the stationary wavelet domain.

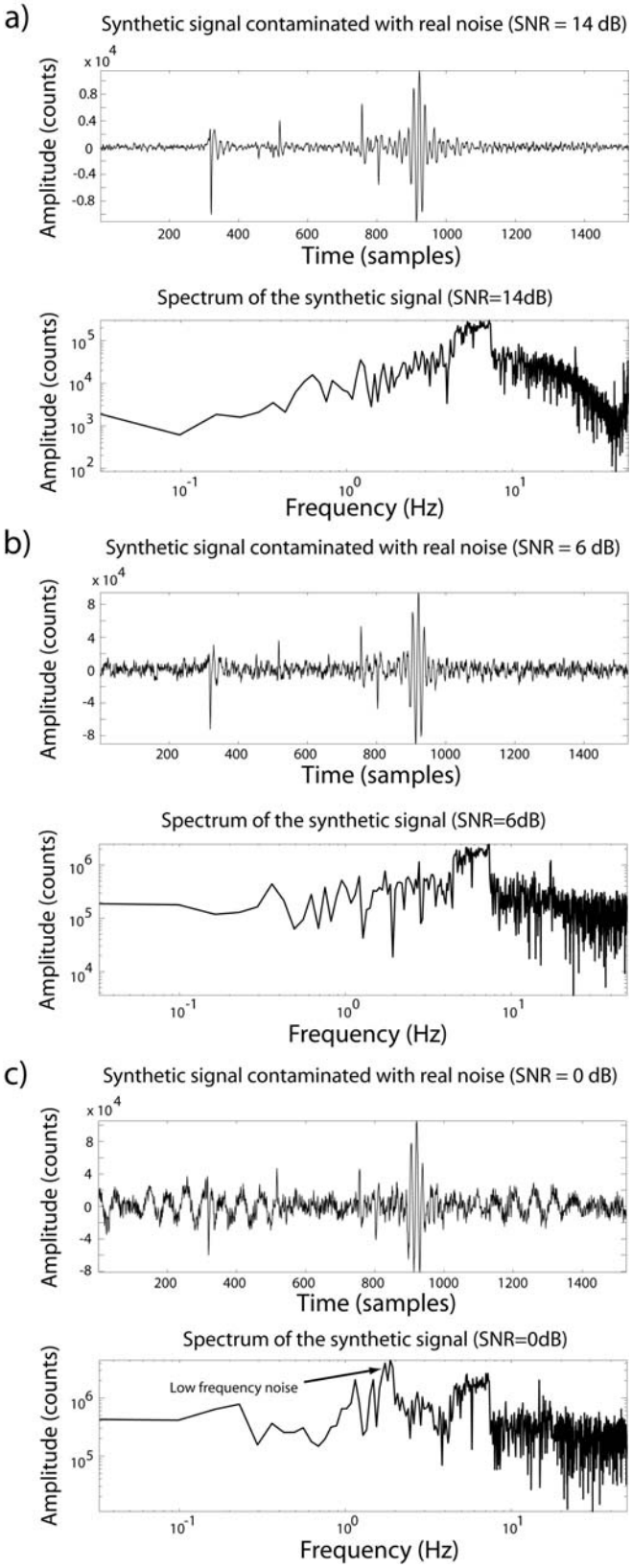

Fig. 3. Examples of synthetic seismograms contaminated with real noise of different characteristics and SNR. (a) SNR=14dB, (b) SNR=6dB, and (c) SNR=0dB.

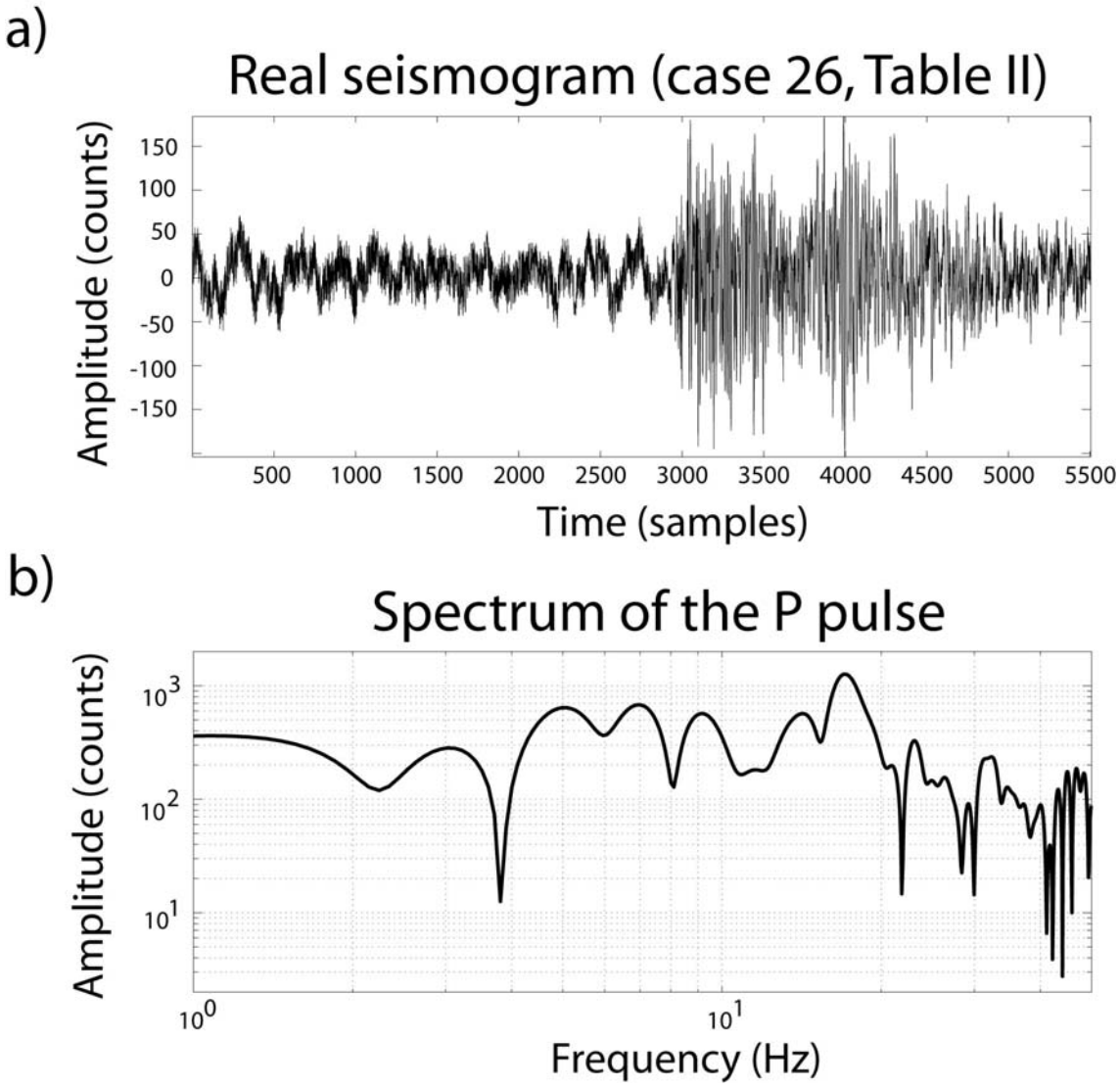

Fig. 4. (a) Real seismogram (case 26, Table II), and (b) the Fourier amplitude spectrum of the first *P* arrival.

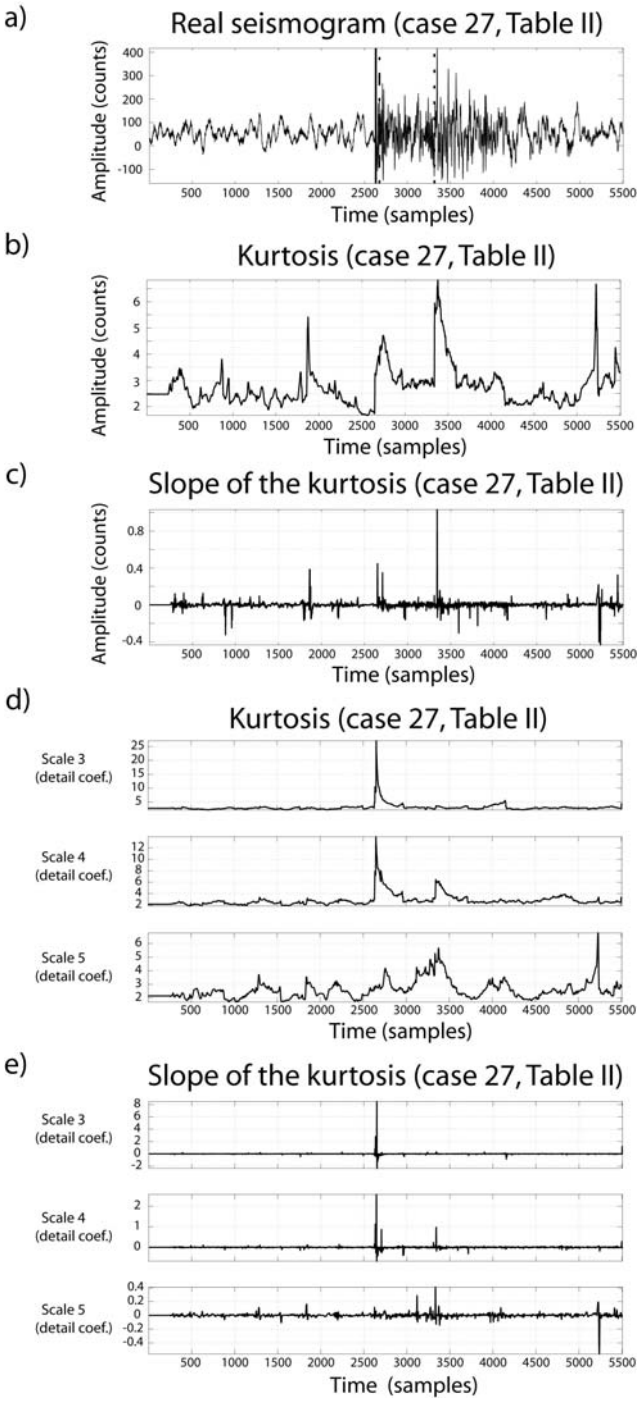

Fig. 5. (a) Real seismogram (case 27, Table II), (b) kurtosis on the time domain, (c) Slope of the kurtosis in the time domain, (d) kurtosis on the time-frequency domain, scales 3, 4 and 5 of the SWT, and (e) slope of the kurtosis on the time-frequency domain, scales 3, 4 and 5 of the SWT. The picks determined by the analyst (solid line), the MFV/MF-I criteria (dotted line) and our method (dot-dashed line) are also compared.

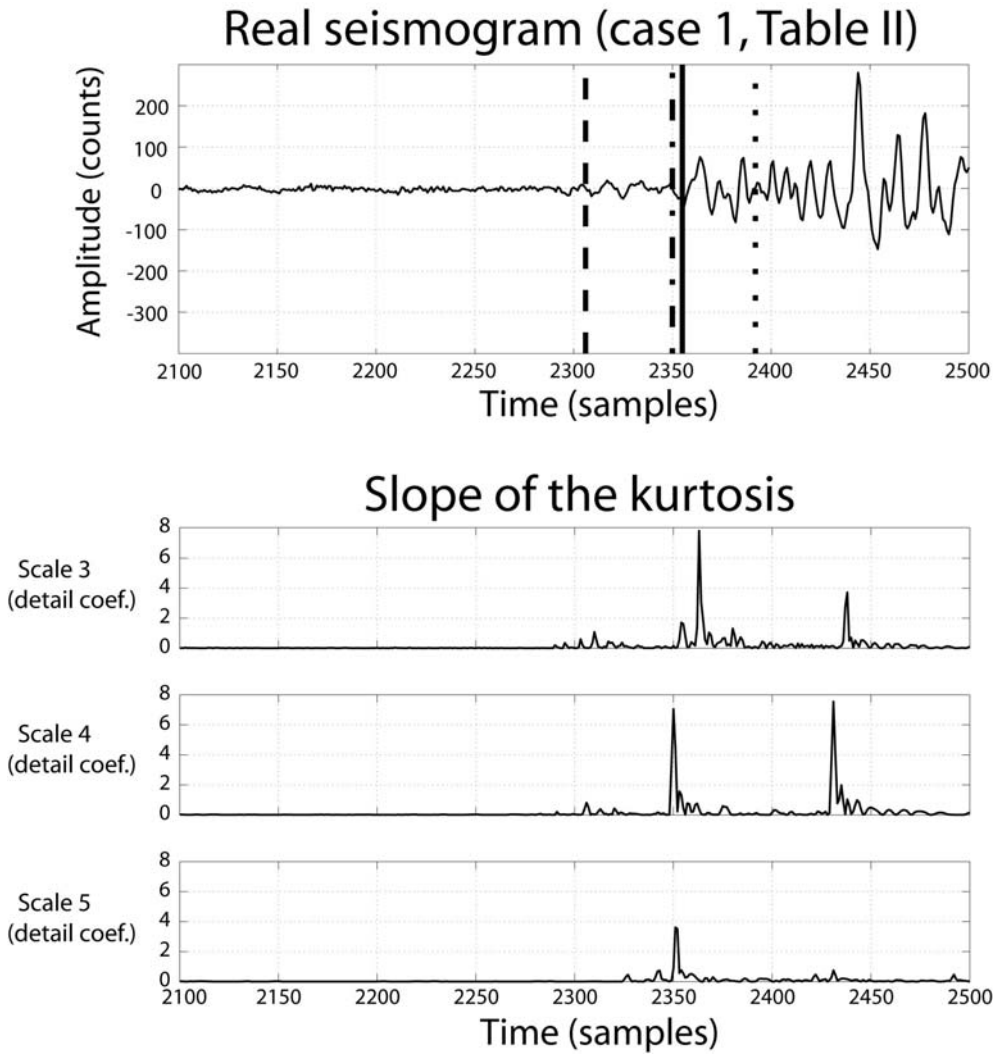

Fig. 6. (a) Zoomed version around the *P* arrival of a real seismogram (case 1, Table II), with the *P*-wave (solid line) and *Pn* wave (dashed line) estimated by the analyst and the picks estimated by the MFV/MF-I criteria (dotted line) and the proposed method (dot-dashed line). (b) Slope of the kurtosis in the stationary wavelet domain.