Higher Order Statistics Based Pickers' Evaluation, Using Data from a Micrsoseismic Network

G-A. Tselentis [1], P. Paraskevopoulos [1], A. Lois [1] & N. Martakis[2]

[1](Seismological Laboratory, University of Patras)
[2] (LandTech Enterprises SA)

Introduction

The arrival time picking in passive seismic tomography is a tedious but important process and its accuracy is essential for the final result. Automatic picking algorithms' effectiveness can speed up this process but its application in a microseismic network, recording weak events with low signal to noise ratio, is an additional challenge.

The current study presents three methods, based on Higher Order Statistics (HOS), namely skewness, kurtosis and differential entropy (also known as negentropy), which were applied in automatic picking of P-wave arrival times for a number of selected events recorded in a microseismic network. Their performance is evaluated in comparison to the manual picks.

Theory

The first- and second–order statistics, [for example mean, variance, autocorrelation and power spectrum] are popular signal processing tools and are extensively used in describing linear and Gaussian processes. In practice, there are a lot of situations that the processes deviate for linearity and Gaussianity. Such processes can be studied using HOS. There are, in general, three reasons for using HOS in signal processing: 1) to extract information due to deviations from Gaussianity, 2) to recover the real phase character of the signals and 3) to detect and quantify nonlinearities in time series (Nikias et al. 1993).

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

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

The third- and fourth–order cumulant sequences of $\{X(k)\}$ are defined as:

$$C_3(m,n) = R_3(m,n)$$
$$C_4(m,n,l) = R_4(m,n,l) - 3(R_2(m))^2$$

and for the zero–lag case ($m = n = l = 0$) we obtain the skewness $sk(X) = C_3(0,0)$ and kurtosis $kur(X) = C_3(0,0,0)$ respectively. The estimators used are:

$$sk(X) = \frac{\sum_{i=1}^{N}\{(X(i) - \hat{m}_x)^3\}}{(N-1)\hat{\sigma}_x^{\,3}} \quad \text{and} \quad kur(X) = \frac{\sum_{i=1}^{N}\{(X(i) - \hat{m}_x)^4\}}{(N-1)\hat{\sigma}_x^{\,4}},$$

where $\hat{m}_x$ and $\hat{\sigma}_x$ are the estimates of mean and standard deviation of $\{X(k)\}$, respectively.

In our experiment we estimate the P-arrival time using these HOS parameters and, additionally, an estimation of the negentropy defined as a function of skewness and kurtosis (Jones et al. 1987):

$$J(X) \approx \frac{1}{24} sk^2(X) + \frac{1}{48} kur^2(X)$$

According to the implemented algorithm (Saragiotis et al. 2002) a moving window “slides” on the recorded signal, estimating skewness, kurtosis and negentropy. Since skewness provides a measure of symmetry of the distribution, and kurtosis a measure of heaviness of the tails, we take advantage of the fact that outliers, such as seismic events, have high values and appear in the tails of the distribution. Hence as these tails become heavier, skewness and kurtosis obtain high values due to the high degree of asymmetry of distribution and, therefore, present maxima in the neighborhood of the P-arrival. In order to avoid large delays on the estimation of P onset time, we evaluate the maximum slope and not the maximum values of the three HOS parameters’ curves. This is due to the fact that the maximum value of these parameters is reached only when a sufficient fraction of the time window contains the seismic signal, which is beyond the P-arrival.

Methodology of evaluation

A characteristic of the events recorded in a microseismic network for passive seismic tomography is that their epicentres are inside or close to the network and they mostly have low magnitudes. In order to evaluate the performance of the above three HOS based methods, 15 seismic events were selected. These were recorded by a micorseismic network in Delvina (SW Albania) using LandTech’s LT-S100 3-component velocity sensors with a sampling rate of 100 s/s. These events have magnitudes ranging from 1.3 to 2.4 M_D so their energy is relatively low. Their depths, also, range from 2.5 up to 11 Km. All records, having a P-wave arrival picked by an expert analyst, were utilised (353 arrivals). For each of these records the Signal to Noise ratio was calculated using a 3 seconds window before the first arrival as indicative for the noise, and a 3 seconds window for the signal.

Moreover, from each station’s continuous record we properly selected a sufficiently large segment, part of which contained the seismic event. The vertical components of these records were bandpass filtered to remove the very low frequency content that was present in the record. Three sets of automatic picks where calculated by applying the three HOS based picking algorithms. Aiming in having a clear view of the algorithm’s performance, no artificial corrections were applied to the estimated P-onset times.

Finally, automatic picks are compared against the manual picks calculating the residual times as a measure of their performance.

Results

The application of the automatic HOS based pickers had low computational requirements for the time windows selected, and was not computationally intensive. Our Matlab implementation of the algorithms runs for several seconds only, for all stations recording an earthquake on a standard personal computer.

As it can be seen, by examining the Signal to Noise Ratio (SNR) versus the residual times (figure 1), the quality of the picks depends on the SNR of the record. In most cases, as the SNR increases the P arrival times become quite accurate and with low residual times compared to the manually picked arrivals. On the other hand, as the SNR becomes lower the accuracy decreases, as the auto pickers start missing the P- wave arrivals selecting either secondary arrivals or S- waves or noise bursts (eg anthropogenic noise, electronic noise) in the record.

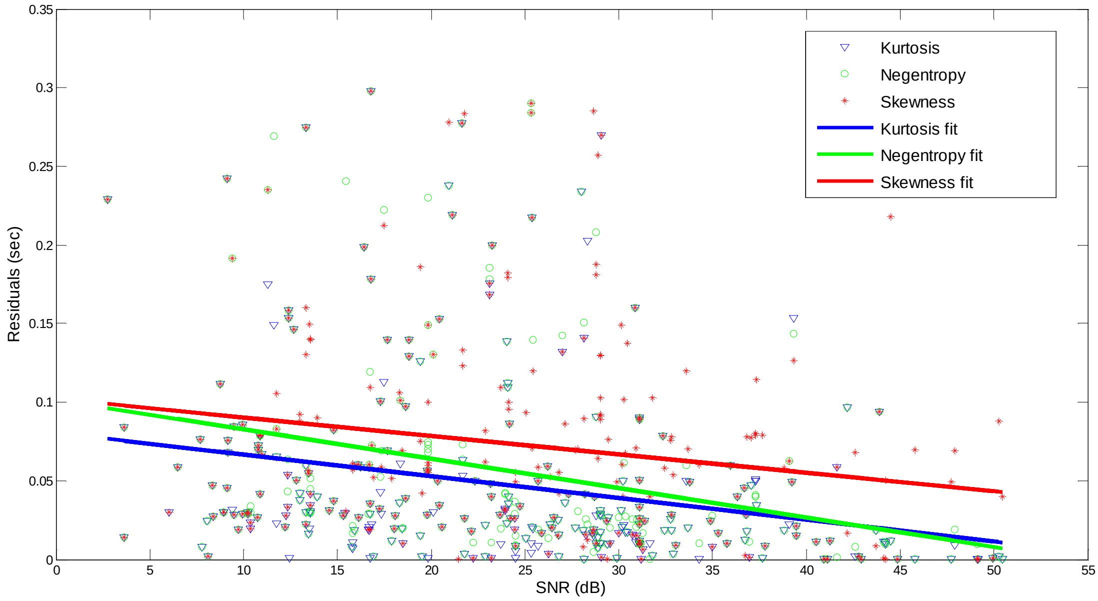

Figure 1 *Diagram of SNR versus residual times for the results of each algorithm. The results are fitted with straight lines in a least squares' sense.*

We consider that the data with residuals above 0.5 sec have missed the P wave arrival pulse and were ignored for the rest of the analysis. These criteria were fulfilled by about 85% of the picks (298 picks for skewness, 302 for kurtosis and 301 for negentropy) and were subsequently used. It should be noted that about 81% of the picks (depending on the method) had residual times below 0.3 sec. In order to obtain a better visualization of residual times, we construct the corresponding histograms (Figure 2).

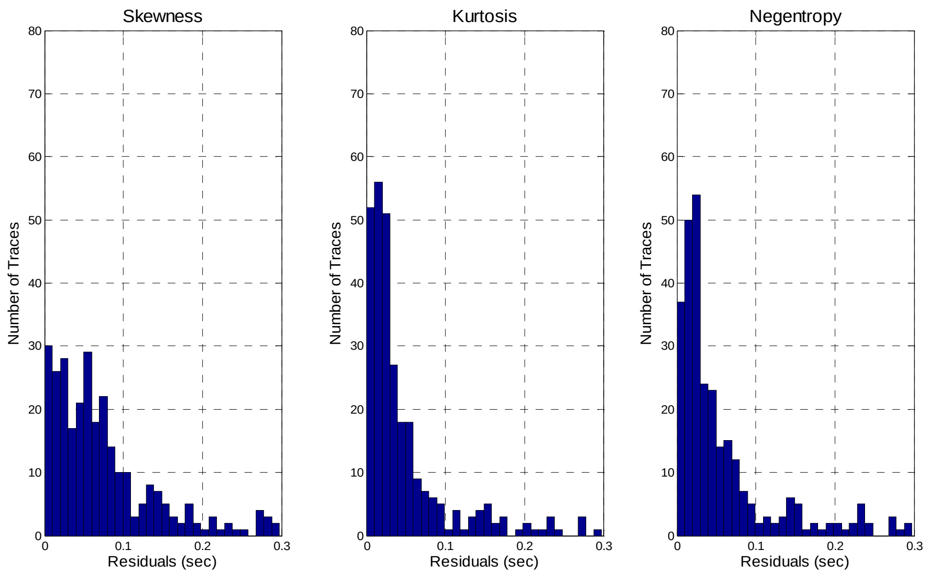

Figure 2 *Histograms of the residuals for each of the HOS based criteria.*

The mean values of the residuals for skewness, kurtosis and negentropy are 0.0733, 0.0469 and 0.0559 seconds respectively, with standard deviations at 0.0658, 0.0571 and 0.0638 seconds.

Comparing the three sets of automatic picks, the kurtosis criterion provided marginally better results than the negentropy criterion, and the skewness criterion had the least accurate results. The performance of these three algorithms is in accordance to Lois et al., (2010).

Conclusions

The performance of HOS based pickers on data from a microseismic network for passive seismic tomography was studied. Comparing the automatic picks to the manual ones made by an analyst, the dependence of the pickers' performance on the signal quality has been shown. For the low magnitude events examined and for offsets, in the order of several tens of kilometres, the kurtosis criterion provides the best results.

The automatic pickers used can be very useful in speeding up the process of P-wave arrivals by quickly estimating the P onset time. Once the segment of the record containing the earthquake has been identified (e.g using the STA/ LTA algorithm), they can be very fast as they have uncomplicated structure and their estimation requires very low computational resources. They are useful but there must be personal inspection and control of the results for improving the accuracy or picking the more noisy records.

Acknowledgments

The authors want to thank Stream Oil &Gas Ltd and LandTech Enterprises SA for providing the seismic data.

References

Jones, M. C., and Sibson, R. [1987] What is projection pursuit. Journal of the Royal Statistical Society: Series A, **150**(2), 1–36.

Lois, A., Psarakis, E.Z., Pikoulis, V., Sokos, E., and Tselentis, G.-A. [2010] A New Chi-Squared based Test Statistic for the Detection of Seismic Events and HOS based Pickers' Evaluation, 32nd ESC Assembly, Montpellier.

Nikias, Ch. L. and. Petropulu, A. M. [1993] Higher-Order Spectra Analysis: A nonlinear signal processing framework. PTR Prentice-Hall, New Jersey.

Saragiotis, C. D., Hadjileontiadis, L. J. and Panas, S. M. [2002] PAI-S/K: a robust automatic seismic P phase arrival identification scheme. IEEE Transactions on Geoscience and Remote Sensing, **40**, 1395–1404.