



**HAL**  
open science

## Automatic picking of P and S phases using a neural tree

Stefania Gentili, Alberto Michelini

► **To cite this version:**

Stefania Gentili, Alberto Michelini. Automatic picking of P and S phases using a neural tree. *Journal of Seismology*, 2006, 10, pp.39-63. inria-00367276

**HAL Id: inria-00367276**

**<https://inria.hal.science/inria-00367276>**

Submitted on 10 Mar 2009

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

The original publication is available at <http://www.springerlink.com/>

[www.springerlink.com/content/04128460kj6g5718/?p=ab78aa19086d40498605e5bd2bb7d850&pi=1](http://www.springerlink.com/content/04128460kj6g5718/?p=ab78aa19086d40498605e5bd2bb7d850&pi=1)

## Automatic picking of P and S phases using a neural tree<sup>1</sup>

*S. Gentili<sup>(1)</sup> and A. Michelini,<sup>(2)§</sup>*

(1) Istituto Nazionale di Oceanografia e di Geofisica Sperimentale - OGS, Dipartimento Centro Ricerche Sismologiche (CRS), Via Treviso 55 - 33100 Udine, Italy

(2) Istituto Nazionale di Geofisica e Vulcanologia, Centro Nazionale Terremoti, Via di Vigna Murata, 605, 00143 Roma

### Abstract.

The large amount of digital data recorded by permanent and temporary seismic networks makes automatic analysis of seismograms and automatic wave onset time picking schemes of great importance for timely and accurate event locations. We propose a fast and efficient P- and S-wave onset time, automatic detection method based on neural networks. The neural networks adopted here are particular neural trees, called IUANT2, characterized by a high generalization capability. Comparison between neural network automatic onset picking and standard, manual methods, shows that the technique presented here is generally robust and that it is capable to correctly identify phase-types while providing estimates of their accuracies. In addition, the automatic post processing method applied here can remove the ambiguity deriving from the incorrect association of events occurring closely in time. We have tested the methodology against standard STA/LTA phase picks and found that this neural approach performs better especially for low signal-to-noise ratios. We adopt the *recall*, *precision* and *accuracy* estimators to appraise objectively the results and compare them with those obtained with other methodologies.

---

<sup>1</sup> This paper has not been submitted elsewhere in identical or similar form, nor will it be during the first three months after its submission to Journal of Seismology.

<sup>§</sup> Previously at OGS-CRS, Trieste, Italy.

Tests of the proposed method are presented for 342 earthquakes recorded by 23 different stations (about 5000 traces). Our results show that the distribution of the differences between manual and automatic picking has a standard deviation of 0.064 s and 0.11 s for the P and the S waves, respectively. Our results show also that the number of false alarms deriving from incorrect detection is small and, thus, that the method is inherently robust.

## 1. Introduction

In recent years, a large effort has been put into finding algorithms able to pick the earthquake P- and S-wave onsets automatically. Many different approaches to the problem have been proposed. These include, methods based on the comparison between the short- (STA) and the long-term average (LTA) of the signal (Allen, 1978; McEvilly and Majer, 1982; Earle and Shearer, 1994), autoregressive methods (Leonard and Kennett, 1999; Leonard, 2000), algorithms that analyze wave-polarization (Cichowicz, 1993; Jurkevics, 1988), algorithms that adopt the wavelet transform (Anant and Dowla, 1997; Zhang *et al.*, 2003), and algorithms that use neural networks (Mousset *et al.*, 1996; Dai and MacBeth, 1995a; Dai and MacBeth, 1997; Zhao and Takano, 1999; Enescu, 1996; Dai and C. MacBeth, 1995b; Wang and Teng, 1995; Wang and Teng, 1997).

Neural networks have the advantage that they adopt data-driven learning schemes to find the solution to the problem. Therefore, it is not necessary to have a deep knowledge of the statistical distribution of the features to obtain a solution even if these features extracted are redundant or noisy. In addition, neural networks permit increased generalization and flexibility.

There are essentially two possible approaches when using neural networks for P- and S-wave phase onset picking: (1) direct signal classification, and (2) signal classification from previously extracted features.

The two approaches present advantages and disadvantages. The first has the advantage that it supplies to the network all the available information, but, owing to the large amount of information, the learning examples are mapped in the feature space very sparsely and, thus, the network can

converge to some erroneous solution (i.e., this is similar to an over-parameterized model space in geophysical inversion). In contrast, feature extraction can be very advantageous if the relevant features are selected, because the learning examples can be mapped more densely in the feature space. However, if a relevant part of the information is lost, this approach is inappropriate (i.e., like an under-parameterized model space in geophysical inversion).

The neural networks adopted in the literature for P- and S-wave phase picking are generally multi-layer, feed-forward neural networks, i.e., networks composed by several layers of neurons in which the signal passes from an input layer to some hidden layer of neurons, reaching, at the end, the output layer.

Dai and MacBeth (1995a and 1997), Zhao and K. Takano (1999) and Enescu (1996) all present algorithms based on the direct signal classification made by multi-layer, feed-forward neural networks. Some of these methods present a large number of false alarms, i.e., noise or parts of the earthquake far from the picking point confused as true arrival onset times. In order to eliminate some of these false alarms, Dai and MacBeth (1995b) present a hybrid method, in which the network described in their previous paper (1995a) is used to find the “candidate time” for the start of S or P phases, while a second network, trained by a feature, is used to classify these candidates in the three classes “noise”, “P pick” and “S pick”. Wang and Teng (1995) develop two neural network based picking methods for detecting P arrivals. These are based entirely on features and not on the seismogram signal. The same authors (Wang and Teng, 1996) present S-wave phase onset picker based on features.

In this paper, we present a new feature-based neural approach that uses an innovative model of neural network, called IUANT2 (Gentili, 2003b for details see Appendix A). The advantage of these neural structures, with respect to the multilayer neural networks used generally in literature, is that the structure of the network is not chosen empirically by the operator, but it is adapted automatically to the problem addressed. In this particular application, we demonstrate (see Appendix A and section 5 for details) that a good solution to the classification problem for both P

and S picking is simply a one-network neural tree. IUANT2 uses as neural networks the perceptrons, that are composed by two layers of neurons, one corresponding to the input patterns and the other corresponding to the output classes “phase pick” and “not phase pick”. The fact that, for the particular choice of features adopted and the problem addressed, a simple perceptron solves the classification problem, is by itself an important result, because it allows to have a higher generalization capability and a faster response with respect to a multilayer neural network (see appendix A for details).

The algorithm proposed here can be applied both to data acquired by fixed and temporary seismic networks, and to single-station data. We apply the technique to data acquired by a local, temporary seismic network. The overall performance, in terms of accuracy of the onset picking, is determined through comparison between automatically and the manually obtained P- and S-wave onset readings and we show the performance of the IUANT2 neural network as the S/N ratio of the test set is increased.

Comparison between different automatic picking systems proposed in the literature is made difficult because authors use different descriptions of the performance of the algorithm they present. For this reason, in this paper we have tried to define more systematically and rigorously some evaluation parameters. In particular, we defined the accuracy in the phase onset picking in terms of standard deviation of the distribution of the differences between the automatic and manual picks. In addition, we introduce the concept of “Recall” and “Precision”, two parameters that give a measure of the percentage of onset time picks detected by the system and of the percentage of false alarms, respectively. The P picking is also compared to the performance of the well known standard STA/LTA approach.

## **2. Method**

### **2.1 Preprocessing**

For data recorded by a local earthquakes network, the three component seismograms, after mean removal, are filtered between 2 and 50 Hz and 2 and 8 Hz for P- and S-wave onset detection,

respectively. This approach (e.g., Wang and Teng, 1997) allows to filter out the noisy signal at less significant frequencies. The three component (3C) seismograms are then transformed into two components: vertical (V) and horizontal (H) one. The vertical remains unchanged, whereas the horizontal consists of the module of the two horizontal components i.e.:

$$h_i = \sqrt{n_i^2 + e_i^2}$$

where  $h_i$  is the value of  $i^{th}$  sample of the new horizontal component,  $n_i$  and  $e_i$  are the values of the  $i^{th}$  sample of the north and the east components, respectively.

## 2.2 Extractor Modules

The system proposed here consists of two main modules: the *P-pick Extractor Module* and the *S-pick Extractor Module*.

For each earthquake, the *P-pick Extractor Module* extracts the P-wave phase onset time, together with an estimate of its accuracy. In addition, it supplies information on the time window where to search for the S-wave onset. If no P-wave onset time is found, the algorithm stops and the S picking is not carried out altogether. In this case, it is made the assumption that no earthquake occurred within the selected time window.

The *S-pick Extractor Module* searches the S-wave phase onset time and, if an onset time pick is found, this is supplied to the system together with an estimate of its accuracy. Again, if no onset time pick is found, the algorithm stops. Thus, the P and S picker used here require only single station data. If more than one station data are available, we have found it convenient to adopt some post-processing to “clean” the automatic onset picking (see section 4), and improve the overall performances of the system.

### 2.2.1 P-pick Extractor Module

P-picking is performed on single, 3C waveform data. We consider four features extracted from either H or V traces. They are variance (*Var*), absolute value of skewness (*Skew*), kurtosis (*Kurt*) and a combination of skewness and kurtosis and of their derivatives respect to time we call *IntegrInf*

(*Integrated-Information* – “*Integ*” hereafter). They are calculated at each sample using a 2.048s long sliding window and they are defined as follows:

$$\begin{aligned}
 Var_i &= \sigma_i^2 = \frac{1}{N} \sum_{j=-\frac{N}{2}}^{\frac{N}{2}} (x_{i+j} - \bar{x}_i)^2 \\
 Skew_i &= \left| \frac{1}{N+1} \sum_{j=-\frac{N}{2}}^{\frac{N}{2}} \left( \frac{x_{i+j} - \bar{x}_i}{\sigma_i} \right)^3 \right| \\
 Kurt_i &= \frac{1}{N+1} \sum_{j=-\frac{N}{2}}^{\frac{N}{2}} \left( \frac{x_{i+j} - \bar{x}_i}{\sigma_i} \right)^4 - 3 \\
 Integ_i &= Skew_i \cdot Kurt_i \left| \frac{d(Skew_i)}{dt} \cdot \frac{d(Kurt_i)}{dt} \right|
 \end{aligned} \tag{1}$$

where  $N+1$ ,  $\bar{x}_i$  and  $\sigma_i$  are respectively the number of points, the mean and the standard deviation of the signal in the chosen time window. The number of points of the sliding window depends on the sampling rate. The features  $Var_i$  is sensible to the increasing amplitude of the signal, while the other three features reflect, to different degree, the asymmetry of the signal distribution around the P and S onset time pick (Press *et al.* 1992). All the features are normalized between 0 and 1 before their use. In the following we set the letters H and V to indicate whether they are calculated on the horizontal or the vertical component, respectively (e.g.,  $VarH$ ). In Fig. 1 it is possible to see the vertical component of a large signal-to-noise (S/N) ratio seismogram and the corresponding first three features of equation (1). We see that the peak of the absolute value of the skewness is nearly aligned to the peak of the kurtosis. The time shift of about 1s, between the peaks of these features and the P wave onset time, corresponds to about one-half of the length of the sliding window.

The previous features are used i.) to find  $M$  temporal intervals spanning in time over one earthquake, and ii.) to provide a first order estimate of the P-phase onset time. In order to do this, the intervals  $t_{Ij}$  and  $t_{Fj}$   $j=1\dots M$  where  $VarV$  is above a given threshold are selected, and, inside these intervals, the time  $t_{Mj}$  that corresponds to the first maximum of  $VarH$  above a given threshold is searched (see Fig. 2). This allows to obtain the windows of interest which extremes are  $t_{Ij}$  and  $t_{Mj}$ .

Since  $t_{Mj}$  is the time in which  $VarH$  reaches the maximum, this time is delayed with respect to the S-phase onset time.

Therefore, we can say that the correct onset time pick of the P and S phases of the  $j^{th}$  candidate earthquake belongs to the interval  $[t_{Ij}-\Delta t+\delta t, t_{Mj}+\delta t]$ , where  $\delta t$  is the time shift due to the length of the sliding window (after several tests we found  $\delta t = 0.83$  s) and  $\Delta t$  is a shift we added (we adopted a time shift equal to the dimension of the sliding window - 2.048s), in order to overcome problems related to the inaccurate threshold method that could, in principle, detect the increase of the amplitude too late. In addition,  $t_{Ij}+\delta t$  can be considered as a rough estimate of the P-wave onset time pick for the successive analysis. In Fig 5a it is possible to see the previous features in the window of interest for a well picked earthquake. The *Integ* feature is calculated after the normalization in the window of interest of *Skew* and *Kurt*. It is possible to see that the peak of this feature is sharper than those of *Skew* and *Kurt*.

For each candidate earthquake, we consider *SkewV*, *KurtV* and *IntegV* as well as *VarH* and *VarV*. The first three features are useful to find the correct onset time pick whereas the variance is more useful to filter out noise or to avoid confusion between P and S phases since the variance increases generally rather slowly as the P-waves are recorded whereas it is generally sharper and larger for S arrival phases. A vector composed by 21 samples of the five features is extracted for every point of the selected time window  $[t_{Ij}-\Delta t, t_{Mj}]$  and considered for P-phase classification using the IUANT2 neural tree.

The output of the neural network, and the rough evaluation of the time pick ( $t_{Ij}+\delta t$ ) are used to infer the determined onset time pick (see section 3) and its accuracy. An onset time pick is considered incorrect if it differs from  $t_{Mj}$  by less than 0.4s (i.e., P pick very close to S maximum). In this case, the corresponding candidate interval is classified as noise and neglected for further analysis.

### **2.2.2 S-pick Extractor Module**

For S-picking we consider the horizontal components of the seismogram. From these signals, *VarH*, *SkewH*, *KurtH* and *IntegH* are extracted, together with two additional features we name *Varrot* and

*FeatBG*. As for the P-waves, 21 samples of these 6 features are extracted from every point of the selected time window and used for S phase classification, using the IUANT2 neural tree. When the S phase starts, *SkewH* and *KurtH* present a sharp peak, which corresponds to the onset pick time.

The feature called “*Varrot*” (i.e., variance under rotation) is obtained by evaluating the variance of one of the horizontal components under rotation around the vertical axis. This rotation is applied at angular increments of 10° to encompass a total angle of 170° .

In other words, if  $p_{i\theta}$  is the projection for a single sample ( $p_{i\theta} = n_i \cos(\theta) + e_i \sin(\theta)$ ), the *Varrot* feature is defined as:

$$Varrot_i = \frac{1}{18 \cdot N} \sum_{\theta=0,10}^{170} \sum_{k=i-N/2}^{i+N/2} (p_{k\theta} - \bar{p}_i)^2 \quad (2)$$

where

$$\bar{p}_i = \frac{1}{18 \cdot (N + 1)} \sum_{\theta=0,10}^{170} \sum_{k=i-N/2}^{i+N/2} p_{k\theta}$$

is the mean of the value of  $p$  on a window  $N+1$  samples long and with 18 different rotations of 10 degrees.

This feature exploits the fact that S-waves have nearly horizontal polarization and that search for the S wave onset time can be performed by focusing on the time where the horizontal motion becomes dominant.

*FeatBG* is derived by one feature parameter developed by Bragato and Govoni (2000) for the Friuli automatic earthquake alert system. This feature follows from S waves having longer periods and larger amplitude than P-waves. Therefore, the absolute value of the integral of the horizontal traces, between one zero crossing and the next one, is larger for S- than for P-waves. We adopt the following formulation:

$$FeatBG_{i-sh} = \sum_{k=z_j}^{z_{j+1}-1} |h_k| \quad \forall i \in [z_j, z_{j+1}) \quad (3)$$

where  $z_j$  are the zero-crossing points. Since the picking on all the other features is shifted by  $\delta t$ , we insert in the definition of *FeatBG* a shift  $Sh$ , where  $Sh$  is the number of samples in the seismograms corresponding to the time interval  $\delta t$ . See Fig.5a for an example of the new features.

Since *Varrot* is generally quite accurate for S-wave phase onset reading, a rough evaluation of the time of the S pick is done by simply choosing a threshold on the value of *Varrot*. In order to do this, the algorithm starts from larger times and proceeds backward until the value of *Varrot* is less than a given threshold (we choose 0.08). If the interval is wrongly selected, it can occur that the extractor output is equal to  $t_{ij}$  or  $t_{Mj}$ . In these cases, it is assumed that the window is incorrect and the algorithm stops. Otherwise, *Varrot* and *FeatBG* are considered, together with the other four features, for S phase classification using the neural networks. The usefulness of the *SkewH*, *KurtH* and *IntegH* features in detecting the S onset time picks, even if they are calculated on the horizontal component of the signal, is generally decreased by the larger signal of the P-wave arrival. Thus, to improve the relevance of these features, the time window in which they are evaluated is shortened by removing the P-wave arrival (i.e., the time window starts 0.4 s after the estimated P pick).

The rough picks obtained for S waves are compared to the output of the neural network in order to determine the S final pick and its accuracy (see section 3).

### **3. Neural tree**

#### **3.1 Training**

Neural networks training, has been done by vectors (patterns) composed by 21 samples of vertical Var V, Skew V, Kurt V, Integ V and Var H for the *P Phases Neural Classifier* and by 21 samples of Var H, Skew H and Kurt H, *Varrot* and *FeatBG* for the *S Phases Neural Classifier*. The training has been done by supplying to the network, designated for the picking of the P phases, 21 pattern corresponding sliding windows centered on the P onset time pick and 60 pattern corresponding to sliding windows centered elsewhere. The network for S phases picking was trained by 44 pattern corresponding to sliding windows centered on the S onset time pick and 56 patterns corresponding to sliding windows centered elsewhere.

In both cases, the neural tree obtained were simple perceptrons, without children nodes. This result is interesting, because it demonstrates that the problem is linearly separable in the feature space and, therefore, it can be solved by a simple and fast neural network with a high generalization capability. Using a multilayer neural network in this case would be slower (a larger number of operations for every classification) and the performances could be worse due to the excessive number of neurons involved and the consequent loss of generality. Further details can be found in Appendix A.

### 3.2 Classification

Proper interpretation of the output of the neural network is critical in order to use correctly the neural classifier and for evaluating its performances. In many pattern recognition applications, the output of a neural tree, or of the feed forward neural network, is the object class, and not the evaluation of the quality of the classification. Usually, the neuron which has the higher output for the network (or for the last network, if a neural tree is adopted), called the “winning neuron”, is found, and the output class is the class to which this neuron corresponds.

However, in a continuous signal interpretation like that of this application, a continuous response of the output can be useful, to better define the phase onset time pick, among the different candidate pick times. This problem has been solved in different ways in the literature – e.g., by having one only output neuron and taking its output as the probability to have the phase pick (Wang and Teng, 1995; Wang and Teng, 1997), or by constructing the network so that every neuron corresponds to a different sample (Enescu, 1996).

A frequently used approach is to construct a two-classes neural network. The two neurons output a number close to unity, when the pattern is recognized in the corresponding class, and close to 0 when the pattern is not recognized. A unique response from two output neurons can be achieved by determining

$$f(o_1, o_2) = \frac{1}{2} [(o_1)^2 + (1 - o_2)^2] \quad (4)$$

where  $o_1$  is the response of the neuron corresponding to the class “phase detection” and  $o_2$  is the response of the other neuron (Dai and MacBeth, 1995a; Dai and MacBeth, 1995b; Dai and MacBeth, 1997; Zhao and Takano, 1999).

In this paper, we use an approach more general than equation (4). The equation, in fact, has several limitations: 1) It can be applied only if the output layer of the neurons is defined, which, in principle, is not always true in a neural tree (see Appendix A for details), although it is defined in this application case. 2) It can be applied only if there are only two output classes. 3) It treats in the same way both the well defined class “phase detection”, and the other class “not phase detection” that can correspond to noise, parts of the earthquake far from the onset time pick, other phases pick-times, or unrecognized signal. For this reason, a new equation has been proposed (5), derived from the analysis done in (Gentili, 2003a; Nebrensky *et al.*, 2002) about the “unknown class” (i.e. the undefined one).

In Fig 3 the steps of the classification are summarized. From the positions  $[i-10, i-9, \dots, i+10]$  of the sliding window, 21 samples for each of the  $n$  features of the classification are extracted, obtaining a  $21 \times n$  long vector that is the input of the neural network (one input neuron for each element of the vector). Since in this case we demonstrated that the neural networks are simple perceptrons (see Appendix A), the output of each network is the two activation values  $AV_{i1}$  and  $AV_{i2}$  of the two output neurons.

Let  $M_i$  be the maximum of the activation values of the output neurons and let  $m_i$  be the second maximum (in this simple two-classes problem it is simply the other activation value). A weighting parameter  $w_i$  is defined as:

$$w_i = \frac{M_i^2 + (M_i - m_i)^2}{2}$$

Let  $c_i$  be the class corresponding to the neuron presenting the maximum activation value;  $c_i$  is 1 if the network classification is “phase detection” and 0 otherwise. The output of the networks of the whole seismogram is reconstructed in the following way:

$$out_i = \begin{cases} w_i & \text{if } w_i > Th \text{ and } c_i = 1 \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

where  $Th$  is a threshold chosen in order to filter out uncertain classifications.

In this way, investigating 21 positions of the sliding window around the  $i^{th}$  position, the system is able to supply the output for the  $i^{th}$  sample of the seismogram.

By considering all the points of the seismogram inside the window of interest, it is possible to reconstruct two output signals  $out1$  and  $out2$  obtained respectively by the P and S neural classifiers.

In order to find the P and S onset time picks from these two signals, the following approach has been applied:

1. The neural network onset pick time is taken as that corresponding to the maximum of  $out1$  in the first interval where  $out1 > 0$  (for P pick – in order to avoid confusion with S picks) or as the time corresponding to the maximum of  $out2$  (S pick).
2. If the time in 1.) is equal, within a statistically evaluated error, to that of the rough pick, this is selected and its accuracy is considered best (i.e., it becomes “score 1 pick”).
3. If the time in 1.) is not equal, then another maximum of the output signals, still satisfying the previous condition, is searched and this result is output with score 1.
4. If no maximum satisfies the condition, the neural network classification is neglected, and the rough evaluation pick time is output and the labeled accuracy is lower (score 0).

This approach has been applied, since it has been shown statistically on a large earthquake dataset that the neural network pick-time is generally more accurate than the rough picking method. For example, on the dataset we have analyzed (see section 5), the standard deviation of the distribution of the time difference between manual and NN picks is 0.06 s for P waves and 0.11 s for S waves. The corresponding standard deviation using the rough methods is 0.15 s and 0.13 s, respectively. However, we have found a few cases where, because of very large errors due to noise, the rough picking method performs somewhat better.

#### 4. Post-processing and noise cleaning

The last module of the system can be applied only when the earthquakes have been recorded by more than one station. It has the advantage that erroneous picks can be easily identified and removed. In fact, some erroneous pick times are likely to remain after the processing because of incorrect event association.

The analysis described below is based on the determination of the differences between P and S arrival phases at different stations and it can be considered derived from the Wadati diagrams (Lay and Wallace, 1995). This analysis allows for the evaluation of the accuracy of the manual picks and it permits identification of the wrong ones.

P and S onset time picks at two different stations ( $i$  and  $j$ ) are compared two by two for the same earthquake by determining the differences  $\Delta P_{ij} = P_i - P_j$  and  $\Delta S_{ij} = S_i - S_j$ . It is easy to show that when the medium is homogenous and the velocities  $v_p$  and  $v_s$  of the P and S waves are constant, for all the possible combinations of  $i$  and  $j$  ( $i \neq j$  and  $P_i \neq P_j$ ), we obtain  $\Delta S_{ij} / \Delta P_{ij} = v_p / v_s$ .

This implies that by plotting the values of  $\Delta S$  vs.  $\Delta P$  on a diagram, for all the possible pair of stations, all the points should lie on a straight line through the origin with angular coefficient (slope)  $v_p / v_s$ .

The value of the angular coefficient does not change by plotting together points obtained separately by different earthquakes on the same set of stations.

In the Earth, seismic velocities are not constant, the raypath of the P and S waves can be different and there is an error in picking the onset time picks due, for example, to the finite sampling rate or, more often, to noisy traces. In general, however, the points on the  $\Delta S$  vs  $\Delta P$  diagrams lie on a narrow band around the diagonal straight line around the mean  $v_p / v_s$ .

In our implementation the order of the stations compared is chosen so that the values of  $\Delta P$  are always-positive.

It is interesting, from our point of view, that, when some pick-times are incorrect, the diagram presents some anomalies. From the analysis of these anomalies, it is possible the identification of the wrong pick-times. The data-points corresponding to incompatible onset readings are the following:

1. The points corresponding to values of  $\Delta P$  and  $\Delta S$  too far from the origin. Their time shift is too large compared to the inter-station distance. In the case analyzed in this paper, the largest distance among stations is less than 90 km. If we assume a minimum velocity for the P waves of about 5 km/s and for S waves of 3 km/s, we inferred that  $\Delta P > 20$  s or  $\Delta S > 30$  s must be attributed to event-phase misidentification.
2. The points with large negative values of  $\Delta S$ . Since, by construction,  $\Delta P > 0$ , also  $\Delta S$  should be greater than zero. In this paper, in order to make only a first order cleaning and to avoid to eliminate stations characterized by small  $\Delta P$  and only noisy picking, we chose to eliminate those points with  $\Delta S < -2$ s.
3. The points too far from the straight, diagonal line  $\Delta S = (v_p/v_s)\Delta P$  are not compatible with true physical values of  $v_p/v_s$ . After elimination of the outliers of point 1 and 2, it is possible to fit linearly the points of the diagram and eliminate the points, which lie too far from the fitted straight line. In our application, the removal of the outliers has been applied twice - the first to all the pick-times found featuring large delays (2s off the straight line), the second only to the more accurate picks (1s out of the straight line).

Once the incorrect points have been detected, the algorithm steps through to analyze which picks have caused them, since every point is the result of the P and S picks of two different stations. A pair of P and S picks of a station is considered wrong if it causes at least  $N/2$  wrong points in the  $\Delta P/\Delta S$  diagram (where  $N$  is the number of stations recording the same earthquake). If a pair is found wrong, the S pick is always rejected, while the P pick is rejected only if its score is 0, or if it corresponds to a point in the  $\Delta P/\Delta S$  diagram with values too large of  $\Delta P$ .

## 5. Application

The set of seismograms adopted here consists of 5068 time windows containing one or more earthquakes detected by a 23 stations network installed during the 1997 Umbria-Marche earthquake sequence. The total number of earthquakes analyzed is 342. The windows are obtained using an event association algorithm that scans the continuously acquired time series and identifies an earthquake (and the associated time window) only in presence of a simultaneous amplitude grow over a given threshold on at least 5 stations (Govoni *et al.*, 1999). This method is fast but inaccurate, and more than one earthquake can be selected within a single window. In these cases, the first event in the selected window is that manually picked (when it is not too small in amplitude). In Fig. 4, the map of the stations of the test database and of the epicenters obtained by manual pick is presented. The magnitude of the events ranges from 1 to 4, the epicenters are located in a 50x10 km area, oriented NW-SE and the depth ranges from 1.5 to 10 km. Due to the range of distance, the P and S phases picked are Pg and Sg phases.

Some examples of the performance of the method on earthquakes featuring diverse waveforms and large amounts of noise are presented in Fig 5a and 5b. In these cases, the final pick of P and S phases correspond to the neural networks picking. The figure shows the three components of the seismogram together with the features adopted in the window of interest. Fig 5c and 5d present examples in which, even if the neural classifier fails in picking P (Fig 5c) or the S onset phases (Fig 5d), the pick is nevertheless accurate, due to the overall robustness of the rough pick. This is an example of the robustness of the method, which is able to detect correctly the onset time pick even when neural classification fails. In order to show the limitations of the method and therefore where it can be improved, Fig. 5e shows one of the cases in which the system fails completely, due to the large amount of noise and to having two earthquakes close together in time.

## 5.1 Overall system performances

In order to compare automatic and manual picking procedures, we have determined the time-difference  $\Delta P$  between the two types of time picks and then fitted a gaussian function to the distribution. The value of the standard deviation  $\sigma_t$  of the obtained function provides a measure of the accuracy of the method. In particular, we define the accuracy in the following way:

$$Accuracy = \frac{1}{\sigma_t} \quad (6)$$

Note that because the manual pick has also an error (i.e., only a rough estimate on its quality is available), the value of  $\sigma_t$  supplied before is the result of the inaccuracy in picking of the system combined with that of the manual pick. In other words, our estimates are relative to a manual pick which we assume to be correct though not necessarily in all instances.

In order to measure the capability of the system to pick a large percentage of the waves and of rejecting false alarms, it is necessary to define an automatic criteria to understand which, between the picks not coincident with the manual one, are definitely wrong (i.e., outliers) and which are just inaccurate. To do that, we have adopted the Chauvenet data rejection method (Taylor, 1982). This method assumes that the available sample contains a set of data normally distributed data plus a few outliers. For a normal distribution  $N(\mu, \sigma)$ , the probability to observe a value that differs from the mean  $\mu$  for more than a quantity  $\delta$  is:

$$Prob(|X - \mu| > \delta) = 2 \cdot [1 - F(\mu + \delta)] \quad (7)$$

where  $F$  is the cumulative density function (cdf) of the normal distribution. Given a set of  $M$  observations, the expected number of data that differ from  $\mu$  for more than  $\delta$  is

$$K = M \cdot Prob(|X - \mu| > \delta) \quad (8)$$

The data rejection method fixes  $\delta$  so that  $K=1/2$ , and recognizes as the outliers the data that differs from  $\mu$  for more than  $\delta$ . For our data set, which contains about 5000 onset time picks, we obtain

$\delta=4\sigma$ . The picks that lie outside the 4 sigma from the mean are labeled as “false”. The other picks are considered correct, even if they may be inaccurate, and labeled as “true”.

To measure the ability of the system to detect the onset of seismic waves and of rejecting false alarms, two features are evaluated - *precision* and *recall* (Aviles-Cruz *et al*, 1995). They are defined as:

$$Precision = \frac{t}{t + f} \quad (9)$$

$$Recall = \frac{t}{T} \quad (10)$$

T is the total number of manual picks, f is the number of false automatic picks and t is the number of the true/correct automatic picks.

The ideal picker has both precision and recall equal to 1. A not very sensitive picker, that does not detect a large part of the events, but has few false alarms, is characterized by a high precision and a small recall. Conversely, a sensitive picker that detects all phases pick times, but it is also sensitive to noise, will result into high recall values but low precision.

In Table 1, the values of  $f$ ,  $t$ ,  $\sigma_t$ , precision and recall are presented for all P and S picks, and for those selected automatically by the system as the more accurate (here and after score 1 P and S). It is possible to see that the quality of score 1 P picks is larger than that of all the P picks (i.e.,  $\sigma_t$  and the recall are equal within the error, but the precision is higher).

In order to make a comparison between our method and a well-known and simple, standard methodology which does not avail of neural networks, we have determined the P picks also using the STA/LTA method. In particular, among different implementation of the STA/LTA, we have chosen, because of its high accuracy, that proposed by Earle and Shearer (1994). The method calculates the STA/LTA of an envelope function that depends on the seismogram and on its Hilbert transform and smoothes the resulting STA/LTA by convoluting it with an Hamming function before detecting the onset time pick. The performances of this method are presented for comparison in the

fourth column of Table 1. It is possible to see that all the P picks obtained by the method proposed in this paper present larger values of recall and precision and smaller values of  $\sigma_t$  than those obtained by the STA/LTA method

In Table 1, the performances of the method for S waves are also presented. It is possible to see that  $\sigma_t$  and precision of all picks and score 1 picks are equal within the error, whereas the recall is smaller for score 1 S, since they are a sub-set of all the S picks. This owes to the fact that the quality of the picks of the neural network and of the rough method are comparable. In fact, the contribution of *Varrot* is very relevant for S-phase picking and, in addition, the neural classification of the S phase is less accurate than the neural classification of the P phase because it operates in the coda of the P waves. For this reason, we assigned the same score (i.e., 1) to all the S picks of the system. When comparing the performances on P and S waves, it emerges that the precision of S-waves picking is equal, within the error, to that of score 1 P picks, whereas the recall is larger for S picks. This is an effect of the smaller standard deviation of P distributions and therefore of the higher accuracy requirements for P picks.

## **5.2 Performances as a function of S/N**

In order to evaluate the performances of the method as function of the S/N ratio, it is necessary to define clearly what is meant for S/N ratio. The definition we have found in the literature differs depending on the authors. For example, Wang and Teng (1997) use the STA/LTA ratio, Enescu (1996) supplies a logarithmic definition, Dai and MacBeth (1995a) define the signal-to-noise ratio as the ratio of the maximum value of the modulus of the signal before and after the onset time. Dai and MacBeth (1997) define the *mean signal-to-noise ratio* as the ratio between the mean absolute amplitude after and before the onset.

Since the mean signal-to-noise ratio is a measure less sensitive to small noisy spikes, and, therefore, it is more reliable than the simple ratio among values while being dependent linearly on the signal,

we define the S/N ratio as the ratio between the mean absolute values of the amplitude taken after and before the onset time. The mean is calculated on a window 1 s long.

In order to outline how different definition of S/N do affect the estimates of the performance of the methods, we plot in Fig 6 the S/N as defined by Dai and MacBeth (1995a) versus the S/N defined in this paper. To this purpose, we have used a dataset of more than 2000 earthquakes randomly chosen from our dataset. If the two estimates had been similar, the points on the graph would have lied on the diagonal. In contrast, we note a large spread in the values that is to be attributed to the sensitivity to noise of the Dai and MacBeth estimate. In addition, after best fitting the points to a straight line, it is possible to see that the Dai and Mac Beth method supplies S/N values which are generally smaller than those of our method. This will result in fictitious better performances if the accuracy is estimated as a function of the S/N.

In Fig. 7, the performances of the system are plotted as function of the S/N ratio for all the P picks and the score 1 picks.

We see that, when the S/N ratio is smaller than 2, the results of the system are unreliable (i.e., up to a few seconds of error in pick time evaluation). For  $S/N > 8$  the error reaches values in the range [0.06, 0.14] s and [0.06, 0.12] s for all P and for score 1 P picks, respectively. These values are comparable with manual picks. For  $S/N < 8$ , score 1 P perform better, because a large part of the inaccurate picks correspond to rough picks, and therefore have score 0. However, some problems still remain, due to the possible confusion between different earthquakes within the same window. In analyzing these data it is important to remark, however, that the smaller the S/N is, the larger is the error also in manual pick, and, therefore, the results we show are affected by both automatic and manual error. We should note that the method that we adopted here does not filter out the outliers thus affecting inevitably its statistics.

In Fig. 8, a detailed analysis is performed for  $S/N < 15$ . In practice, we have removed the outliers by considering as incorrect picks those with errors larger (in absolute value) than 1 s. The performance is compared to that of the STA/LTA method. For this comparison and in order to remove the effects

of inaccurate manual picks, we have selected 25 traces having high signal to noise ratio ( $> 50$ ) for which we expect manual picks to be “exact”, added real noise multiplied by a constant to simulate lower signal-to-noise-ratio signals and, finally, performed the analysis using both our method and the STA/LTA approach. In figure 8, it results that, at  $S/N=15$  the performances of the two methods are comparable, but, at lower values, the error of the STA/LTA picks increases dramatically. In contrast, the performance of the neural approach proposed here stays stable or increases slightly. In section 5.3, we present the approach followed to eliminate automatically the incorrect picks.

In Fig 9, the value of  $\Delta S$  for different values of  $S/N$  is presented. The  $S/N$  of S waves is generally smaller, due to the presence of P coda. In this case, the results for  $S/N < 2$  are poor, while for  $S/N > 4$  the error is in the range 0.1- 0.15s. The performance of the automatic picker for the S-waves on the overall dataset is worse than that of the Ps, because the  $S/N$  for the S waves is generally smaller. Also in this case, there are not relevant differences among score 0 and score 1 S picks, confirming the idea that setting all the S picks to the same score is correct.

### **5.3 Data Post-processing**

The technique described above for the post-processing has been applied to the entire data set. The following constraints have been set in the application of the post processing.

1.  $\Delta P < 20s$  and  $-2s < \Delta S < 30s$ .
2. The distance from the best-fit straight line  $< 2s$  for all the picks.
3. The distance from the best-fit straight line  $< 1s$  for the score 1 picks.

In Fig.10a the initial  $\Delta P/\Delta S$  diagram is presented, while in Fig. 10b and 10c we present the diagram after the post-processing has been applied for all the P picks and the score 1 P picks, respectively. At the end of the post-processing, we have found that the algorithm rejected 282 P picks (102 score 1 P picks) and 621 S out of the 5059 and 4933 P and S picks detected originally, respectively.

For comparison, the same approach has been applied to the manual picks while assuming all P picks as score 1 P picks. The results after the post-processing are shown in Fig.10d.

The procedure eliminates only one P pick and 336 S picks of the 5068 manually picked. Comparing Fig. 10c with Fig.10d, it is possible to see that at the end of the post-processing phase, manual and automatic picks display similar features at least from the  $\Delta P/\Delta S$  diagram point of view.

In Table 2, the performances of the system after the post-processing are presented. Comparing this table with Table1, we see that while the accuracy is equal within the error limits, the precision of the picks has improved after the post-processing, whereas the recall has somewhat decreased because of the rejection.

In Fig.11, the mean error of the P-wave picking of the system after the post-processing is presented as a function of the S/N ratio. From the comparison of Fig. 7 and Fig. 11, it is possible to see that, after post processing, the performances of both all P picks and of score 1 P picks have improved at low S/N ratio, because a large part of wrong picks has been detected and removed. For larger values of the S/N ratio, the results are nearly unchanged. In addition and because of the elimination of erroneous picks due to noise, the performances of all picks and score 1 picks are now similar. In Fig. 12, the performances of the system after the post processing for S waves are presented. Comparing the results with those obtained before the post-processing (see Fig. 9) it is possible to see that, also in this case, the performances have slightly improved for  $2 < S/N < 8$  . However, due to the shorter, predefined window that is searched, the number of erroneous picks for S waves is smaller and, therefore, the post-processing does not affect significantly the picking performances of the S-waves.

#### **5.4 Discussion**

Summarizing the results published in the literature, while making a thorough comparison with those obtained here, is difficult. This follows mainly from the use of different data sets by the different authors. In addition, the use of different approaches for testing the algorithms does not easy a thorough comparison. In particular, we have found the following problems when comparing the results.

1. Some of the studies are more related to labeling rather than to picking. This because they supply as results only the recall, rather than the time shift between some “ideal” picking and the automatic one (Dai and MacBeth, 1995b; Wang and Teng, 1995).
2. Even if some authors outline that the S/N ratio affects the performance of the picker (e.g., Wang and Teng, 1995; Wang and Teng, 1997), they do not supply enough information on how the performance of their picking scheme varies while changing the S/N ratio. In addition, no information is supplied on the S/N ratio of the test set (only Wang and Teng (1997) supply some information, although in the form of STA/LTA ratio). This approach will generally overestimate the performance of the picker. In fact, it is generally found that even a not very sensitive picker can determine the onset phases from signals having large S/N.
3. The definition of S/N is not always provided, or it is different depending on the authors (see section 5.2).
4. In many studies, the recall is supplied, but not the precision (Dai and MacBeth, 1995a; Zhao and K. Takano, 1999; Wang and Teng, 1995), or it is given only the number of misidentifications with noise but not the misidentifications between P and S arrivals (Wang and Teng, 1997). This can overestimate the performance of the pickers in which every small variation of the signal (that can be also due to noise) is recognized as potential phase. In addition, proper identification of the phase type is essential to avoid confusion and to be of any use.
5. In some papers, the test set is too small and/or the seismograms are all obtained from one or two stations (Dai and MacBeth, 1995a; Dai and MacBeth, 1995b; Dai and MacBeth, 1997; Enescu, 1996; Wang and Teng, 1995; Wang and Teng, 1997). A small and homogeneous test set is easier to be classified and this, again, will lead to overestimation of the picker performance.

Regardless of all these limitations, in the following we will make an attempt to provide at least a qualitative comparison and our aim is of understanding the advantages and disadvantages of the presented technique.

#### **5.4.1 Accuracy estimate**

Estimation of the accuracy of the different methods existing in literature is generally performed by supplying the percentage of pickings which difference with the manual ones is below a given threshold. For local earthquakes, as those analyzed in this paper (i.e., epicentral distance <80 km, approximately), the threshold has been chosen to 0.1 s (Dai and Mac Beth, 1995a; Dai and Mac Beth, 1997; Wang and Teng, 1997).

The percentage of P-wave picks which difference with the manual picks is below 0.1 s amounts, for our method, to 63% for all P picks and the 69% for score 1 P picks. For S picks the same estimate amounts to 36%.

All this would suggest that our method is less accurate than the others proposed in the literature (see Table 3). We remark, however, that the evaluations proposed in the literature are performed on small data sets, or on “ad hoc” subsets of larger ones. In this case, there is a prior selection phase that is not accounted in the performance estimates. In addition, the proposed methodologies adopt seismograms from very few stations and we have not found thorough information on the S/N level. For these reasons and in order to make a fair comparison between our results and those in the literature, we found it necessary to undergo our analysis to some additional processing in order to make the results of the different methodologies inter-consistent.

For example, some papers neglect the outliers and focus their analysis only on the data where the pick is correct but with different level of accuracy. For our purposes and in order to neglect the outliers while following an objective statistical approach, we make the hypothesis that the data are distributed normally and we adopt only the information deriving from the standard deviation and the mean of the Gaussian distribution that best fits our data. When this approach to data analysis is followed, the percentage of picks differing from those obtained manually (< 0.1 s) increases to 87% and 63% for P- and S-wave picks, respectively.

Another important aspect when comparing the result is the S/N level of the test data set. Any evaluation of the accuracy will, in fact, be affected by the S/N of the test set – the lower is the S/N,

the lower is the accuracy of the method. In addition, we have found that, for large values of S/N (>50), also the standard deviation of the distribution of the difference between manual and automatic picks is affected by the S/N ratio.

In Fig.13a, we show (for both all the P picks and the score 1 P picks) the ratio between the number of picks with error smaller than 0.1 s and all the picks as the minimum S/N level is varied. If the minimum S/N ratio lies between 50 and 256, the percentage of picks within 0.1 s is above 80 and 85% for all and score 1 P picks, respectively. It is interesting to notice that, while the S/N ratio is relevant for the performances in accuracy of the P picks, it seems to be less relevant for the S, causing a not-monotone behavior (see Fig.13b). We think that this depends on the fact that the accuracy of the picking of S waves depends not only on its S/N ratio, but also on the choice of the window of interest determined by the *P-pick Extraction Module*. However, also in this case, when increasing the minimum S/N level to 19, we obtain 57% of the data within 0.1 s of the manual pick. Overall, our method is general, because it can be applied to any number of stations, while many other methods have been trained and tested only on very few stations. In particular, if we select the stations in which the performances with our methodology are best, the results improve quite significantly. For example, by selecting the station BETT, CASC, MVL and NRC for the P waves and the stations ARM1 and LAVE for S waves, the percentage of picks within 0.1 s reaches the 97% for the P and 86% for the S waves, respectively.

In Table 3, the performances of the methods existing in literature are summarized, when available, together with the information on the method adopted for the tests (“?” means that it has not been specified). The errors on the measures have been calculated in the hypothesis of Poissonian distribution (the error on a number of picks equal to  $N$  is  $\sqrt{N}$ ) and propagated. From the table, it is possible to see how our algorithm is comparable or better than those existing in literature from the accuracy point of view, if the same analysis method is applied.

#### **5.4.2 Precision Estimate**

The precision of our method is high, especially after the post processing procedure. In particular, before the post processing, the precision ranges between 0.87 and 0.96 depending on the wave picked and on the score of the pick. After the post processing, values as high as 0.90-0.97% are reached. The information on the precision of the methods proposed in literature is often not available, and, sometimes, only the number of noisy signals confused with P and S phases is presented. This, however, neglects possible misidentifications of P with S waves.

The only information available is presented in Table 4. Also in this case, the method proposed here performs either better or equally well (within the error bounds) to those presented in the literature. This result is valid regardless of the selected “best stations”, or of the minimum S/N level.

#### **5.4.3 Recall estimate**

In order to make correct comparisons with methods existing on literature, we compared the results of our method with those in which rejection of the outliers is not carried out.

In Table 5, we find that our results are generally comparable (within the error bounds) or better than those presented in the literature. The only method in this table, that results into better recall estimates when compared to the method proposed here is the AND-B neural network introduced by Wang and Teng (1995) for P pick time detection. However, it should be pointed out that the network of this study was tested only on one station. Out of all our stations, the recall estimates of at least 13 of them provided comparable results to the AND-B methodology. In Table 5 the performances of higher recall stations for P (APPE and CASC) for S (GUA) are presented.

#### **5.4.4 Processing times**

The analysis of our data set was performed on current personal computers. Although the code that we have developed is not yet fully optimized, the time necessary to analyze 3C data (in ASCII format) on a time window of about 6000 samples (i.e. about 48 s long or more, depending on the sampling rate) is about 0.6 s (0.024 s for each neural classification and the remaining time for the other modules of analysis and for checking the I/O file). In detail, the software has been written in

C++ and the processing was performed on a 2.40 GHz Pentium 4 with 1 GB of RAM. A basic draft Matlab interface has been developed for data management and graphics. The interface uses files as I/O with the C++ code, allowing to calculate and plot the results into a mean of 2.2 s for a 6000 samples time window. Since these performances will be improved in the final version of the code, they should be seen as an upper bound on the processing time.

## **6. Conclusions**

In this paper, we have presented a general and robust P- and S-wave onset times, automatic picking methodology that is based on neural networks. In particular, the method uses a neural tree, called IUANT2, capable to infer the best structure of the network during the training phase, while allowing for noise filtering and high generalization capability. Comparison between standard, manual onset time picks and the results of the neural network has shown that the latter generally rejects the errors deriving from record noise and it is capable to accurately determine the onset times and their accuracies. In addition, we have compared the technique to the results obtained using the standard STA/LTA onset picking methodology on the same data set and found that our neural approach is far more accurate for signal-to-noise ratio values less than 12.

We have tested the methodology to a data set consisting of 342 local earthquakes recorded by a 23 station networks during the Umbria-Marche sequence in 1997. The mean epicentral distance of the earthquakes was about 20 km. In order to appraise the results, we have adopted the precision, accuracy and recall estimators. These estimators provide quantitative measures of the goodness of the methodology..

We have found that the distribution of the difference between manual and automatic pick times has a standard deviation of 0.064 s for the P waves and 0.11 s for the S. The precision is 0.87 for all the P waves, 0.96 for the P-waves selected by the system as those most accurate (i.e., score 1 P), and 0.96 for the S waves. These values show well that the number of false alarm is small and that the method is robust.

Finally, a post-processing procedure has been added, which allows for the identification and rejection of erroneous picks. We find that after post-processing, the precision is improved. The recall after the post-processing is larger than 0.8 for all the types of waves independently on the accuracy estimated by the method. This confirms that the method is capable of detecting and identifying correctly a very high percentage of the waves.

The advantage of the proposed method lies in its generality, in the speed of the computation, in the high percentage of earthquake detected, in the low number of false alarms and in the accuracy in detecting the wave onset pick times.

The method is designed to be used for the automatic picking of large, off-line, data sets recorded, for example, during continuous data field campaigns of temporary networks. Similarly, the method could be also used for real-time data analysis of earthquake monitoring networks.

## 7. Acknowledgements

We would like to thank Carla Barnaba, Pier Luigi Bragato, Gianni Bressan, Aladino Govoni, Lara Lovisa, Enrico Priolo and, in general, all the CRS staff for their valuable discussions and suggestions. This research has been funded through the FIRB contract n. [RBAU0155F9](#).

## 8. References

- Allen, R. V. (1978). Automatic earthquake recognition and timing from single trace, *Bull. Seism. Soc. Am.* **68**, 1521-1532.
- Aviles-Cruz, C., A. Guèrin-Duguè, J.L. Voz, D. Van Cappel (1995). Deliverable R3-B1-P Task B1: Databases ELENA-NERVES-2 “Enhanced Learning for Evolutive Neural Architecture”, ESPRIT –Basic Research Project Number 6891.
- Anant, K. S. and F. U. Dowla (1997). Wavelet transform methods for phase identification in three-component seismograms, *Bull. Seism. Soc. Am.* **87**, 1598-1612.
- Bragato, P. L. and A. Govoni (2000). The Friuli automatic earthquake alert system *Bollettino di Geofisica Teorica ed Applicata* **41**, 1, 69-77.
- Cichowicz, A. (1993). An automatic S-phase picker, *Bull. Seism. Soc. Am.*, **83**, 1, 180-189.

- Dai, H. and C. MacBeth (1995a). Automatic picking of seismic arrivals in local earthquake data using an artificial neural network, *Geophys. J. Int.* **120**, 758-774.
- Dai, H. and C. MacBeth (1995b). Identifying P- and S-Waves Using Artificial Neural Network, 57<sup>th</sup> Conference and Technical Exhibition, Glasgow, Scotland.
- Dai, H. and C. MacBeth (1997). The application of back-propagation neural network to automatic picking seismic arrivals from single-component recordings, *J. Geophys. Res.*, **102**, B7, 105-115.
- Earle, P. and P. Shearer (1994). Characterization of global seismograms using an automatic picking algorithm, *Bull. Seism. Soc. Am.* **84**, 2, 366-376.
- Enescu, N. (1996). Seismic Data Processing Using Nonlinear Prediction and Neural Networks, IEEE NORSIG Symposium, Espoo, Finland.
- Gentili, S. (2001). Information Update On Neural Tree Networks, *2001 International Conference On Image Processing (ICIP 2001)* Thessaloniki, Greece, 505-508.
- Gentili, S. (2003a). A New Method For Information Update In Supervised Neural Structures, *Neurocomputing*, **51C**, 61-74.
- Gentili, S. (2003b). Retrieving Visual Concepts in Image Databases, Ph.D. Thesis Series, Computer Science 2003/3, Forum eds., Udine, Italy, ISBN 88-8420-152-7.
- Govoni A., Spallarossa D., Augliera P. and L. Trojani, (1999). The 1997 Umbria-Marche Earthquake Sequence: the combined data set of the GNDT/SSN temporary and the RESIL/RSM permanent seismic networks, CD-ROM.
- Haykin, S. (1994). *Neural networks a Comprehensive Foundation*, Macmillan College Publishing Company.
- Jurkevics, A. (1988). Polarization Analysis of Three-Component Array Data, *Bull. Seism. Soc. Am.*, **78**, 5, 1725-1743.
- Lay, T. and T. C. Wallace (1995). *Modern Global Seismology*, 218-221, Academic Press

- Leonard, M. and B. L. N. Kennett (1999). Multi component autoregressive techniques for the analysis of seismograms, *Phys. Earth Planet. Int.* **113**, 2, 247-264.
- Leonard, M. (2000). Comparison of Manual and Automatic Onset Time Picking, *Bull. Seismol. Soc. Am.* **90**, 6, 1384-1390.
- McEvelly, T.V. and E.L. Majer (1982). ASP: An automated seismic processor for microearthquake networks, *Bull. Seism. Soc. Am.*, **72**, 1, 303-325.
- Minsky, M. L. and S.A. Papert (1969). *Perceptrons*. Cambridge MA MIT Press.
- Mousset, E., Y. Cansi, R. Crusem and Y. Souchet (1996). A connectionist approach for automatic labeling of regional seismic phases using a single vertical component seismogram *Geophysical Research Letters* **23**, 6, 681-684.
- Nebrensky, J.J., G. Craig, G. L. Foresti, S. Gentili, P. R. Hobson, H. Nareid, G. G. Pieroni, J. Watson (2002). A Particle Imaging and Analysis System for Underwater Holograms, Chapter 8 in the book "Optical Methods and Data Processing in Heat and Fluid Flow" eds. C. Greated, J. Cosgrove and J.M. Buick, Professional Engineering Publications, 79-91.
- Press, W. H., S. A. Teukolsky, W. T. Vetterling, B. P. Flannery (1992). Numerical Recipes in C. The Art of Scientific Computing, Second Edition, 610-612.
- Roseblatt, F. (1958). The Perceptron: A probabilistic model for image storage and organization in the brain *Psychological review*, **65**, 386-408.
- Taylor, J. R. (1982). *An Introduction to Error Analysis, The Study of Uncertainties in Physical Measurements*, University Science Books.
- Utgoff, P. E. (1988). Perceptron tree: a case study in hybrid concept representation. *Proceeding of the VII National Conference on Artificial Intelligence*, 601-605.
- Utgoff, P. E., N. C. Berkman and J. A. Clouse (1997). Decision Tree Induction Based on Efficient Tree Restructuring, *Machine Learning Journal* **29**, 5-44.
- Wang, J. and T.L. Teng (1995). Artificial Neural Network-Based Seismic Detector, *Bull. Seism. Soc. Am.*, **85**, 1, 308-319.

- Wang, J. and T.L. Teng (1997). Identification and Picking of S Phase Using an Artificial Neural Network, *Bull. Seism. Soc. Am.* **87**, 5, 1140-1149.
- Zhang, H., C. Thurber and C. Rowe (2003). Automatic P-wave Arrival Detection and Picking with Multiscale Wavelet Analysis for Single-Component Recording **93**, 5, 1904-1912.
- Zhao, Y. and K. Takano (1999). An Artificial Neural Network Approach for Broadband Seismic Phase Picking, *Bull. Seismol. Soc. Am.* **89**, 3, 670-680.

## Figures Captions

Fig 1: Example of the feature calculation on a seismogram vertical component (V), variance, module of the skewness and kurtosis. Note that both skewness and the kurtosis present a sharp peak about 1 second before the onset time pick.

Fig 2: The vertical component of the seismogram (V) together with the two horizontal ones (N and E), the vertical variance (VarV) and the horizontal variance (VarH). It is possible to see the position of  $t_{11}$ ,  $t_{F1}$  and  $t_{M1}$ .

Fig 3: Classification scheme: from the sliding window in the positions  $[i-10, i+10]$ , 21 elements for each of the  $n$  features are extracted. This results in a  $21 \times n$  long vector that is sent to the  $21 \times n$  input neurons of the neural network. The signal is then processed by the network obtaining two activation values,  $AV_{i1}$  and  $AV_{i2}$ . From the two activation values, a new value  $out_i$  is obtained, that corresponds to the output of the classification procedure for the  $i^{th}$  sample of the signal.

Fig 4: Map of the epicenters obtained using the manual picks from the 1997 Umbria-Marche earthquake sequence dataset (Govoni *et al.*, 1999). The dimension of the circles scale with the magnitude of the earthquakes, whereas the darkness scales with depth. Triangles, squares and rhombs of different grey level correspond to the position of the stations belonging to the different research agency that participated to the acquisition experiment or of different instrumental characteristics.

Fig 5: Examples of the accuracy of the method for both P and S onset time picks and for different waveform types and noise levels. PR and PNN (and analogously SR and SNN) refer to rough and neural network onset time picks, respectively. The onset time picks discarded automatically are shown as vertical thin dotted lines whereas the final onset time picks are shown as solid lines (in Fig. 5d they coincide). The manual, onset time picks are shown as solid squares in the top traces (observed data). In the bottom traces (features), the final onset

time picks are shown as circles. (a) good performance of the procedure deriving from proper neural tree classification; (b) good performance on the P-wave induced by appropriate “rough” picking, which has been preferred to the neural pick because none of the peaks of the neural network output was compatible with the rough pick. The small error on S-wave time picks results from neural tree classification; (c) small error of the method in P wave picking because of proper neural network classification. In this case, the error on the S is small because the rough S pick has been preferred to the neural pick (i.e., none of the peaks of the neural network output was compatible with the rough pick); the incorrect neural S pick has been indicated by a dotted line. (d) incorrect P-wave onset time picking resulting from large noise and at least two earthquakes within the selected time window. However, the system picks accurately the S-wave of the second earthquake.

Fig 6: Comparison between Dai and MacBeth (1995) S/N definition (vertical axis) and that adopted in this study. More than 2000 real seismograms have been used randomly chosen by our dataset. It is possible to see a large dispersion in the values due to the sensitivity to noise of Dai and Mac Beth definition. In addition, the fit (solid line) shows how Dai and MacBeth definition tends to supply lower values of S/N for the same data. The dashed line, for comparison, corresponds to equal values of the two S/N definitions.

Fig 7: Performances as a function of the S/N ratio for (solid circles) all the P picks and (open circles) score 1 P picks. The ordinate y of each circle refers to the median of the absolute values of  $\Delta P$  for the earthquakes whose S/N is in the interval  $[x-1, x+1]$ , where x is its abscissa.

Fig 8: Comparison between the the performances of our method and the STA/LTA methodology. Multiple amounts of recorded noise has been added to 25 earthquake seismograms featuring very high ( $> 50$ ) S/N ratio. This insures very high accuracy of the original picks performed on very low noise traces. The picks characterized by error larger than 1s are considered wrong picks and eliminated from the comparison.

Fig 9: Performances as a function of the S/N ratio for (solid circles) all the S picks and (open circles) score 1 picks. The ordinate y of each circle represents the median of the absolute values of  $\Delta S$  for the earthquakes whose S/N is in the interval  $[x-1, x+1]$ , where x is its abscissa. Fig 10:  $\Delta P/\Delta S$  diagram of (a) all the P and S picks of the system (b) all the picks after the post processing (c) score 1 P picks after the post processing (d) the manual data after the same processing. Note that nearly the same results are obtained in (c) and (d).

Fig 11: Performances after the post-processing as a function of the S/N ratio for (solid circles) all the P picks and (open circles) score 1 P picks. To see the improvement after post-processing compare this graph to Fig. 7.

Fig 12: Performances after the post-processing as a function of the S/N ratio for (solid circles) all the S picks and (open circles) score 1 S picks.

Fig 13: Ratio between the number of samples with error  $<0.1s$  and all the samples for S/N ratios larger than those listed on the abscissa. For (a) P picks (grey: all P; black: score 1 P) and (b) S picks.

## Tables

	All P	Score 1 P	STA/LTA	All S	Score 1 S
f	680	176	866	212	210
t	4379	4268	4226	4721	4121
$\sigma_t$	0.064±0.006s	0.064±0.005s	0.084±0.009	0.11±0.01	0.11±0.02
Precision	0.866±0.006	0.960±0.003	0.830±0.008	0.957±0.003	0.951±0.004
Recall	0.86 ±0.01	0.84±0.01	0.83±0.01	0.93±0.01	0.81±0.01

Table 1: Performance of the system (f = number of false alarms, t = number of correct picks,  $\sigma_t$  = standard deviation of the distribution of difference between manual and automatic picks, precision and recall) for all the P picks (All P), for the P picks labeled with score 1 (i.e., more accurate; Score 1 P); for all the S picks (All S) and for the score 1 S picks. The fourth column presents the performances of a STA/LTA method.

	All P	Score 1 P	All S	Score 1 S
f	449	119	87	94
t	4142	4038	4232	4225
$\sigma_t$	0.065±0.006 s	0.064± 0.004 s	0.11±0.01	0.11±0.02
Precision	0.902±0.003	0.971±0.003	0.980±0.002	0.978±0.002
Recall	0.81±0.01	0.80±0.01	0.84±0.01	0.83±0.01

Table 2: Performance of the system after the post-processing analysis (f = number of false alarms, t = number of correct picks,  $\sigma_t$  = standard deviation of the distribution of difference between manual and automatic picks, precision and recall). See Table 1 for detail.

Author	Outliers rejection	Number of stations	Minimum S/N P	Minimum S/N S	P (%) <0.1s	S (%) <0.1s
Dai & MacBeth 1995a	Y	2 (P & S)	?	?	94±4	90±4
Dai & MacBeth 1997	Y	2 (P & S)	?	?	90±4	83±4
Wang & Teng 1997	N	2 (S)	?	?	-	70±10
Gentili & Michelini (this paper)	N	23	0.5	0.7	63±1	36±1
	Y	23	DOOC	DOOC	87±2	63±1
	N	23	50 (1) - 256 (all)	19	86±2	60±20
	Y	4 (P) – 2(S)	DOOC	DOOC	97±4	86±4

Table 3: Comparison of the *accuracy* obtained using different methods and different data analysis. “Outliers rejection” means whether very erroneous picks have been removed before the analysis. The last two columns indicate the percentage of P and S picks within 0.1 residual from manual pick. “DOOC” stands for “Depending On Other Choices”, “?” means that the datum is not available, “-” means that the method has not applied to that particular picking (see text for detail).

Author	Outliers rejection	Number of stations	Minimum S/N P	Minimum S/N S	Precision P	Precision S
Dai & MacBeth 1995b	N	1 (P & S)	?	?	0.69±0.03	0.62±0.04
Wang & Teng 1997	N	2 (S)	?	?	-	0.98±0.03
Gentili & Michelini (this paper)	N	23	0.5	0.7	0.866±0.006 (all)	0.957±0.003 (all)
			0.5	0.7	0.960±0.003 (1)	0.951±0.004 (1)
			DOOC	DOOC	0.902±0.003(all-pp)	0.980±0.002 (all-pp)
			DOOC	DOOC	0.971±0.003 (1-pp)	0.978±0.002 (1-pp)

Table 4: Comparison of the *precision* obtained using different methods and data analysis. (see text and Table 3 for detail).

Author	Number of stations	Minimum S/N P	Minimum S/N S	Recall P	Recall S
Dai & MacBeth 1995b	1 (P & S)	?	?	0.82±0.05	0.54±0.04
Wang & Teng 1995	1 (P)	?	?	0.9±0.1 AND-A 1.0±0.1 AND-B	-
Wang & Teng 1997	2 (S)	?	?	-	0.9±0.1
Gentili & Michellini (this paper)	23	0.5	0.7	0.86 ±0.01 (all)	0.93±0.01 (all)
	23	0.5	0.7	0.84±0.01 (1)	0.81±0.01 (1)
	23	0.5	0.7	0.81±0.01 (all-pp)	0.84±0.01 (all-pp)
	23	0.5	0.7	0.80±0.01 (1-pp)	0.83±0.01 (1-pp)
	2 P & 1 S	DOOC	DOOC	0.97±0.05	0.96±0.09

Table 5: Comparison of the *recall* obtained using different methods and data analysis. In detail, “pp” refers to after the post processing, “1” means “score 1 pick”, “all” means “all the picks” (see text and Table 3 for further detail).

## Appendix A

The neural classifier adopted here is an innovative model of neural network, named IUANT2 (Gentili, 2003b), based on neural trees (Utgoff, 1988). A neural network is a massively parallel distributed processor storing experiential knowledge into inter-neuron connection strengths and making it available for use. The knowledge is acquired through the learning process (Haykin, 1994).

The processors, called “neurons” or “nodes”, work in parallel. The single neuron performs a very simple task, that is to supply a numerical answer, which is a function of the intensity of the input signal that reaches the neuron from the other neurons through the connections (or synapses).

The mathematical model is the following: let  $x_i$  be the output signal of the  $i^{th}$  neuron connected with the  $j^{th}$  one. The output signal  $y_j$  of the  $j^{th}$  neuron is:

$$y_j = \Phi \left( \sum_i w_{ij} x_i - \theta_j \right) \quad (11)$$

where  $w_{ij}$  is the strength of the connection (the weight) between the  $i^{th}$  and the  $j^{th}$  neuron,  $\theta_j$  is a threshold and  $\Phi$  is a function of the input signal called the *activation function* of the neuron. The most frequently used activation function is the *sigmoid function*, defined as  $\Phi(s) = (1 + e^{-ks})^{-1}$  where  $k$  is a constant; however, also other functions, like e.g. the gaussian one, are frequently used.

In the network, the neurons can be organized in a single layer, on more layers connected one to the other or in more complex structures. The simplest model of neural network, is the perceptron, developed by Roseblatt in 50’-60’ (Roseblatt, 1958). It is the less computationally complex model and so is faster than the others, but the applicability is limited; in fact it is possible to demonstrate (Minsky and Papert, 1969) that a network characterized by one single layer of neurons is not able to solve problems not linearly separable (i.e. a problem in which the classes can not be divided by an hiper-plane in the feature space).

The approaches to solve not linearly separable problems, maintaining the advantages of neural networks, i.e. the learning by examples and the generalization capability, are essentially two: (1)

developing networks with more than one layer of neurons, like e.g. multilayer feed-forward neural network (2) developing an adaptive networks whose structure depends on the problem addressed.

The multilayer feed forward neural networks are the networks most frequently used till now in earthquake phases picking algorithms; they have a structure composed by several layers of neurons.. The information moves from the input neurons (input layer) to the output ones (output layer). The layers between the input and output layer are called “hidden layers”. The main disadvantage of these networks is that the number of hidden neurons affects the performances of the network: if it is too small, the network can not solve some problems, while if it is too high, the network loses its generalization capability. Since the “ideal” number of hidden layers and the number of neurons for every hidden layer is unknown, and changes depending on the problem addressed, the approach used in literature is performing several attempts and choosing the networks whose performances are best (Haykin, 1994). Unfortunately this approach is empiric and slow and the ideal structure may not be found.

The neural trees, vice-versa, have a structure that is determined during the training and depends on the problem addressed. They are composed by a variable number of very simple neural networks and sometimes some decision nodes. In IUANT2 neural trees, perceptron network have been chosen, because for these networks the structure is determined simply by the length of the input patterns and by the number of output classes. Even if a single perceptron is not able to solve a problem if it is not linearly separable, it is possible to show that a set of perceptron in cascade with some decision nodes can do it. The perceptron is a supervised neural network. This means that a set of examples (training set) of the desired performances is defined for the training. The examples are vectors (patterns) corresponding to the features of what has to be classified, together with the information of the desired output of the network (target). The target is a vector with all the elements equal to 0 except the one corresponding to the output class, that is equal to 1.

The training of the perceptron is generally made (and this is the approach used by IUANT2 neural tree) recurrently by using a gradient descent optimization algorithm (perceptron learning rule or delta rule) defined as:

$$\Delta w_{ij} = \eta \sum_{\mu=1}^N (t_i^{\mu} - y_i^{\mu}) \dot{\Phi} \left( \sum_{k=0}^M w_{ik}^{\mu} x_k^{\mu} \right) x_j^{\mu} \quad (12)$$

where  $N$  is the number of training patterns,  $M$  is the number of elements of the patterns,  $t_i^{\mu}$  is the  $i^{\text{th}}$  component of the target output of the  $\mu^{\text{th}}$  input pattern (1 if the target class is the  $i^{\text{th}}$ , 0 otherwise),  $w_{ij}$  is the value of the weight between the  $i^{\text{th}}$  and the  $j^{\text{th}}$  neuron,  $\Delta w_{ij}$  is its change,  $x_j^{\mu}$  is the  $j^{\text{th}}$  component of the  $\mu^{\text{th}}$  input pattern,  $\dot{\Phi}$  is the derivative of the activation function (we choose a sigmoid) and  $\eta$  is the learning rate.

The training of the IUANT2 neural tree is based on the concept of Minimal Description Length (MDL) (Utgoff *et al.*, 1997), of Minimum Tree Description Length (MTDL) (Gentili, 2001; Gentili 2003a; Gentili 2003b) (see after for definition) and on the perceptron learning rule.

The training of the IUANT2 neural tree is the following:

1. The training set  $P$  is processed by the perceptron in the root node, which tries to subdivide it into groups corresponding to the  $J$  classes of the problem. The training of the perceptron continues until a stopping criterion is met. In this case, the stopping occurs when the error does not decrease by a minimum amount in a given number of cycles. If the problem is not linearly separable, some of the training patterns are classified in the wrong class.
2. A new level of  $J$  neural networks (the children nodes) is added to the tree. If one or more subsets are entirely assigned to a class  $\Omega_i$ , the corresponding children node becomes a leaf node labeled with the class  $\Omega_i$  and the training ends for that subset.
3. The children nodes to which no subset is assigned are ignored, while the leaf nodes whose corresponding subsets are composed of different classes, are marked as “wrong leaves”.

4. If the MDL of the wrong leaf is lesser or equal to its MTDL, the wrong leaf node becomes a leaf labeled with the class with higher cardinality in the leaf.
5. Otherwise, the corresponding subset is used to train another perceptron that tries to subdivide the subset into its  $J_1$  classes. The procedure continues in a recurrent way following the steps 1-5 for all children. If the training set cannot be divided in any group (the perceptron repeats the same classification of the parent node), a splitting rule (decision node) is applied, dividing the feature space by an hyper-plane perpendicular to the straight line passing from the baricenters of two classes or by a random hyper-plane if the baricenters are all coincident.
6. When all patterns reach a leaf, the algorithm ends.

The MDL of a wrong leaf is defined as

$$\text{MDL} = 1 + \log_2(c) + x[\log_2(n) + \log_2(c+1)] \quad (13)$$

where  $c$  is the number of classes observed at the leaf,  $n$  is the total number of patterns in the leaf and  $x$  is the number of patterns wrongly classified at the leaf.

The MTDL is defined as:

$$\text{MTDL} = 1 + \log_2(t) \quad (14)$$

where  $t = c \cdot v$ ,  $c$  is the number of classes of set of patterns reaching the wrong leaf and  $v$  is the number of elements of the pattern.

The comparison between the MDL and the MTDL gives a measure of the opportunity to substitute the wrong leaf with the smaller tree that can be constructed, i.e. a perceptron.

After the training, the network can be used to classify new patterns.

The classification of a new pattern is made by moving toward the tree in a top-down way, following a path determined by the classification given by each considered node, till a leaf node is reached; the pattern is then labeled with the classification provided by the leaf node. The classification of each node corresponds to the neuron giving higher output value.

The advantages of IUANT2 approach with respect to other neural networks are the following:

1. The network structure is determined during the training and depends on the problem addressed;

2. A noise filtering, during the training, avoids a grow of the tree into wrong directions;
3. The IUANT2 neural network can be trained again without losing the existing stored information.

The updating procedure of the network (point 3) is done by the addition of new networks to the existing tree in places determined by a new training. However, since this feature of the network has not been used here, we refer for a detailed discussion to (Gentili, 2001a; Gentili 2003; Gentili, 2003b). This feature is useful if we already have a network supplying good performances and we want to update it for different typology of earthquakes, maintaining the same performances on the most of the data.

The capability of this network to adapt its structure to the problem (point 1), has been found of great advantage in this application. We have found that both the P and S pick problems are linearly separable, since they can be solved by only one perceptron. This allows having a fast classification system, with a high generalization capability, that a multilayer perceptron may not have, owing to the very large number of neurons. Also the noise filtering capability (point 2), obtained by the comparison among MTDL and MDL, has been found useful during this application, because it allowed to avoid a wrong branch, due to noise, during the S phases neural classifier training phase.

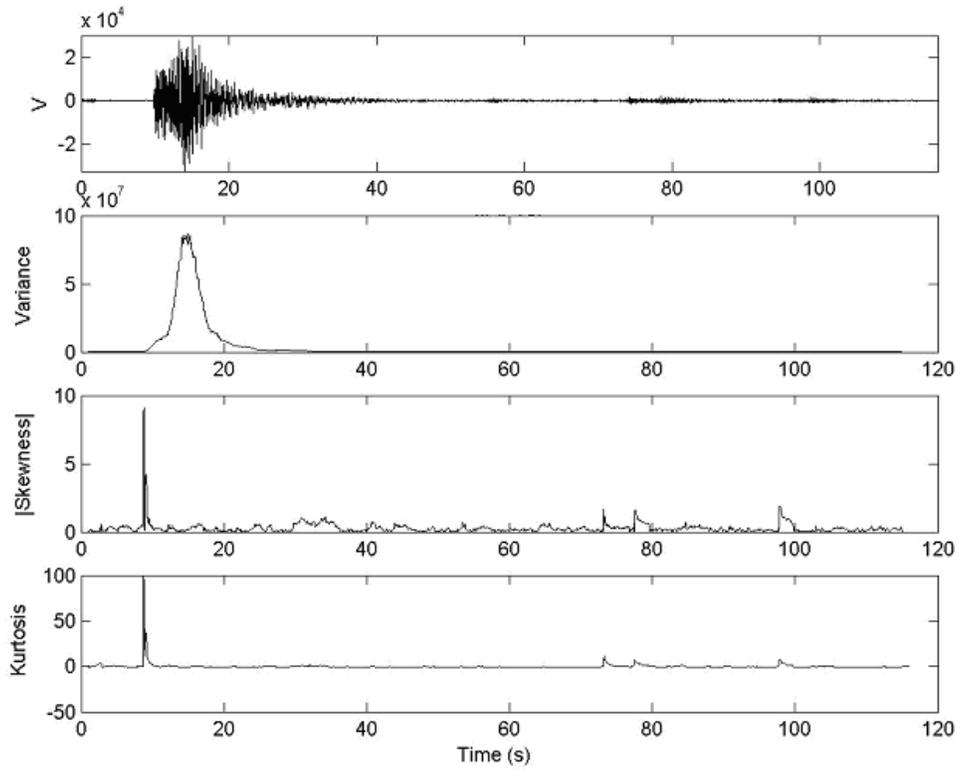


Fig. 1

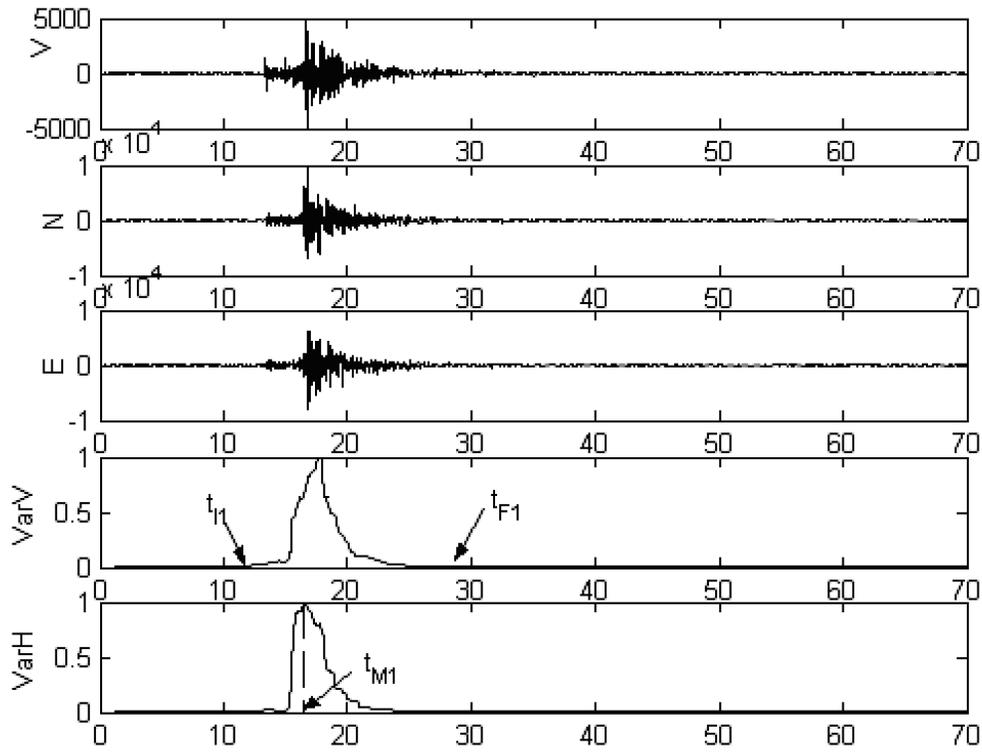
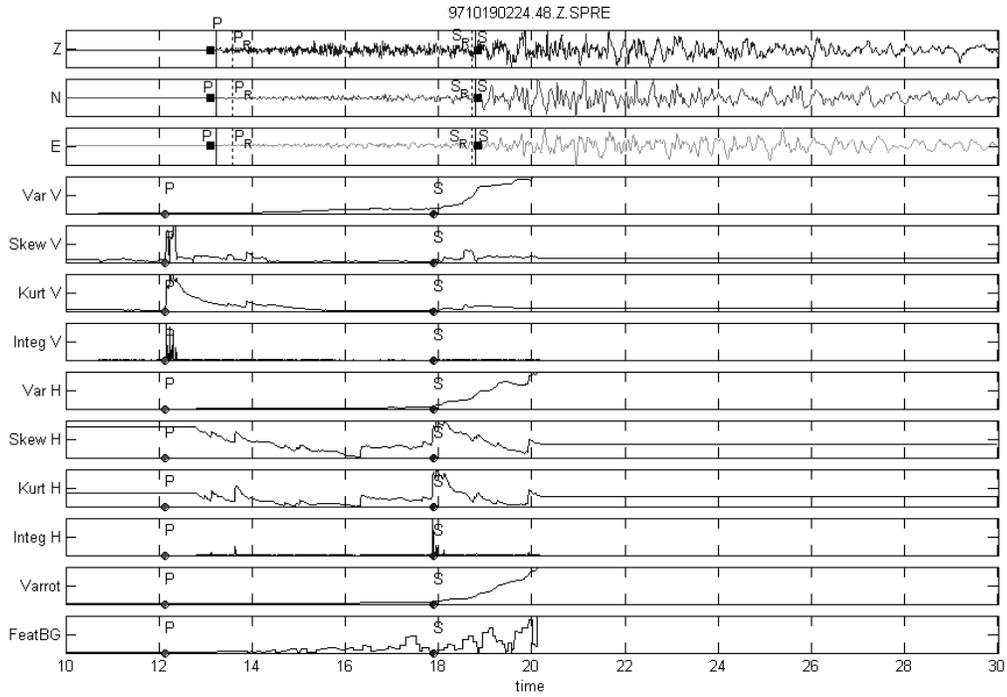
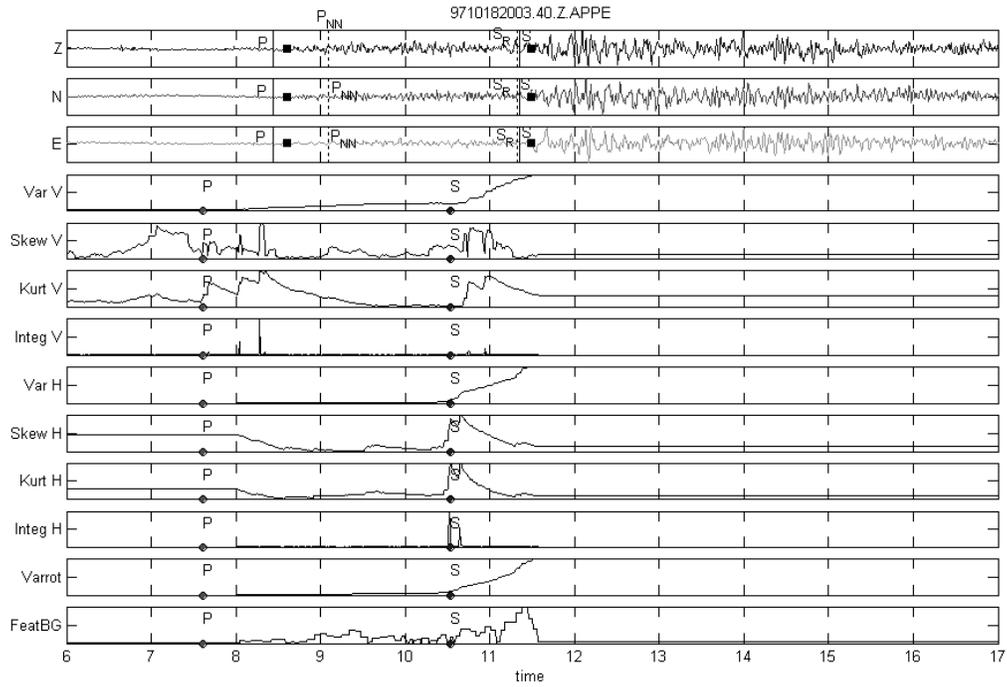


Fig. 2

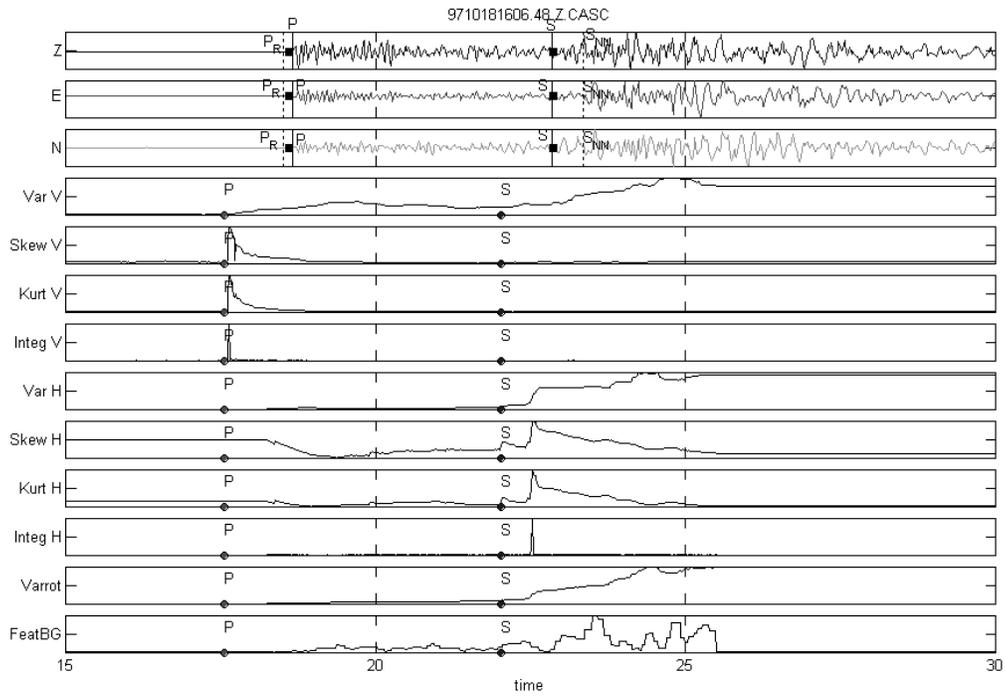




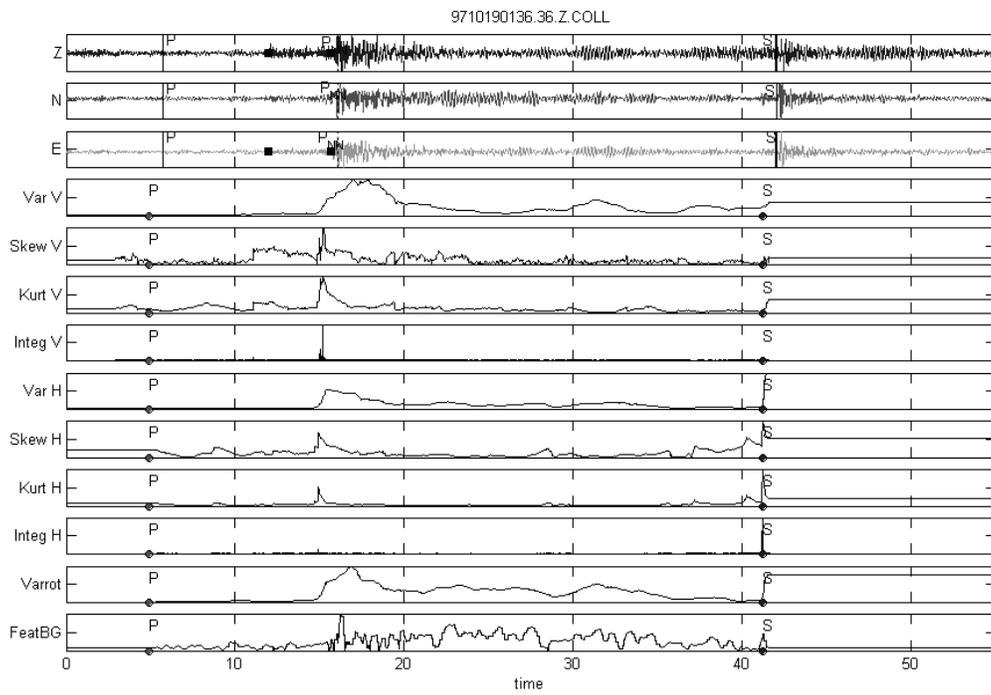
(a)



(b)



(c)



(d)  
Fig. 5

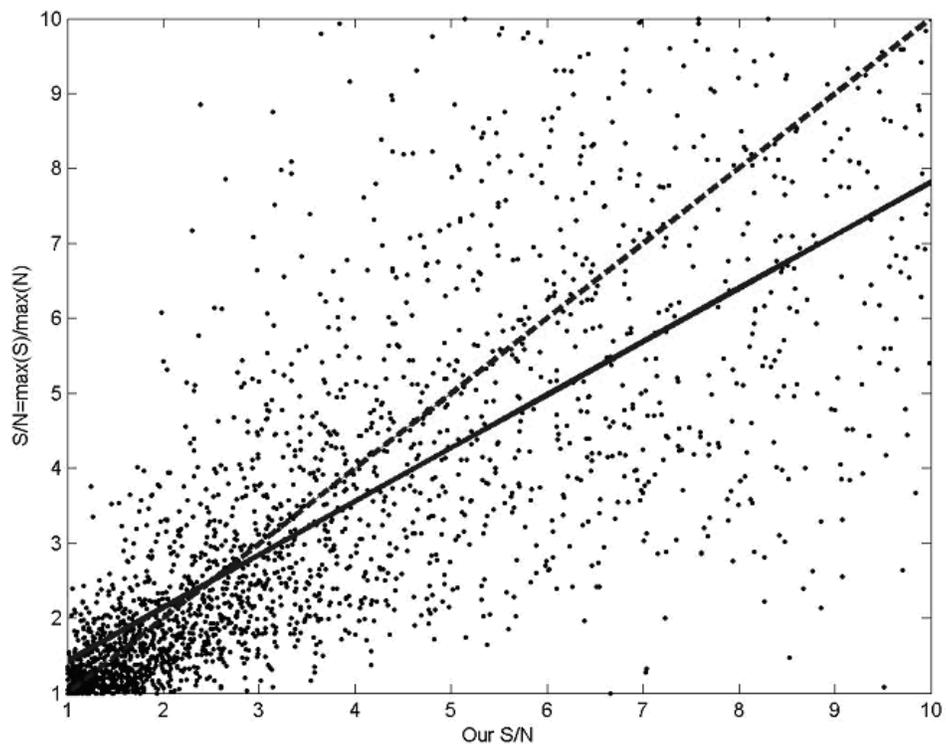


Fig. 6

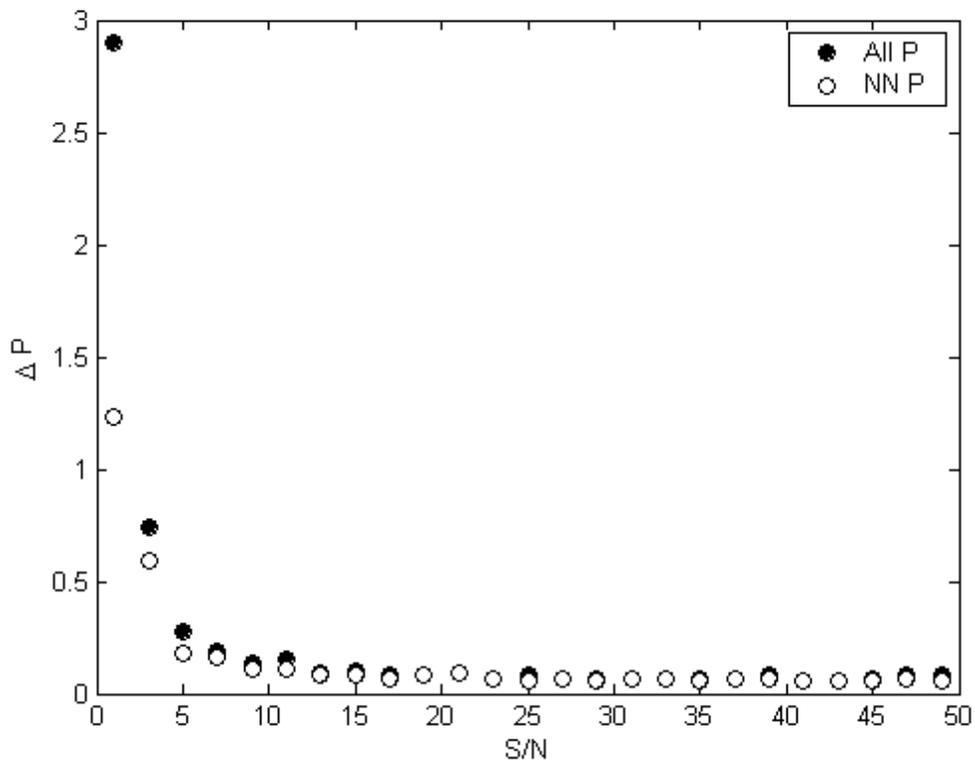


Fig. 7

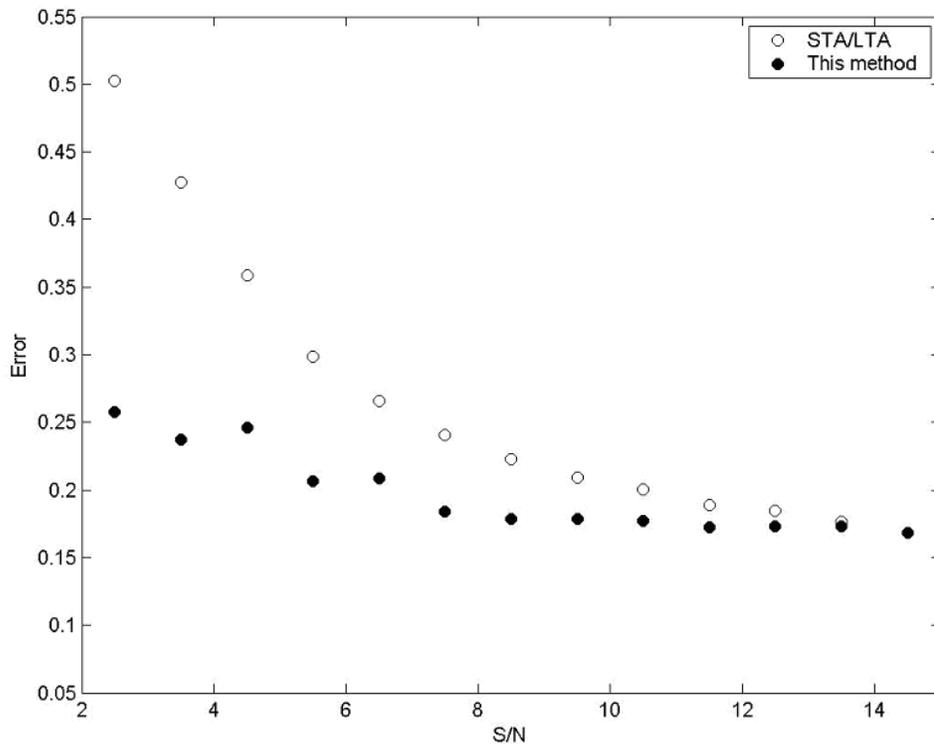


Fig. 8

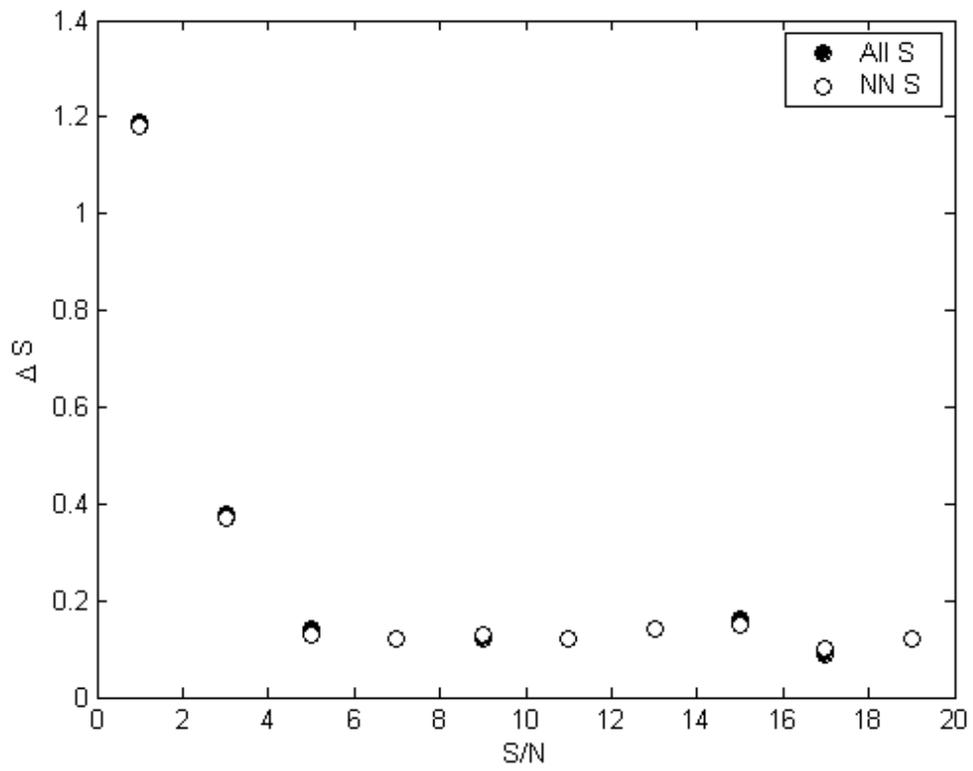
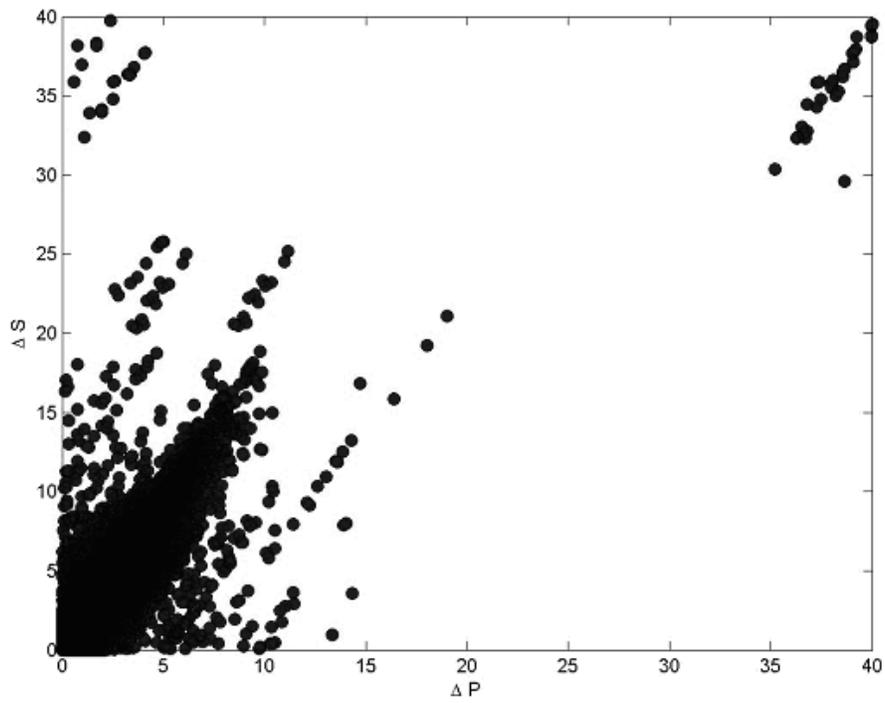
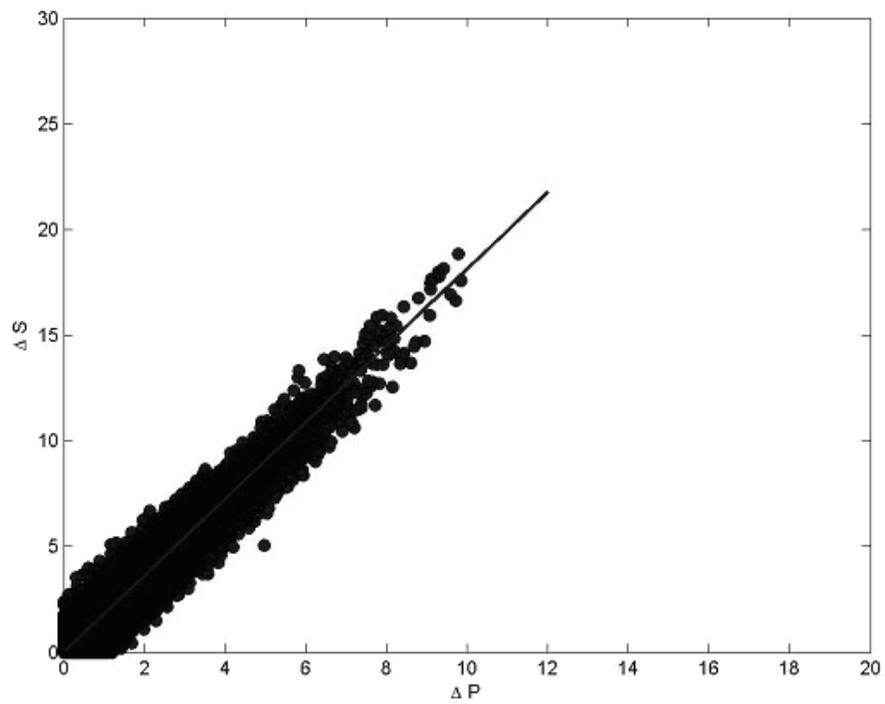


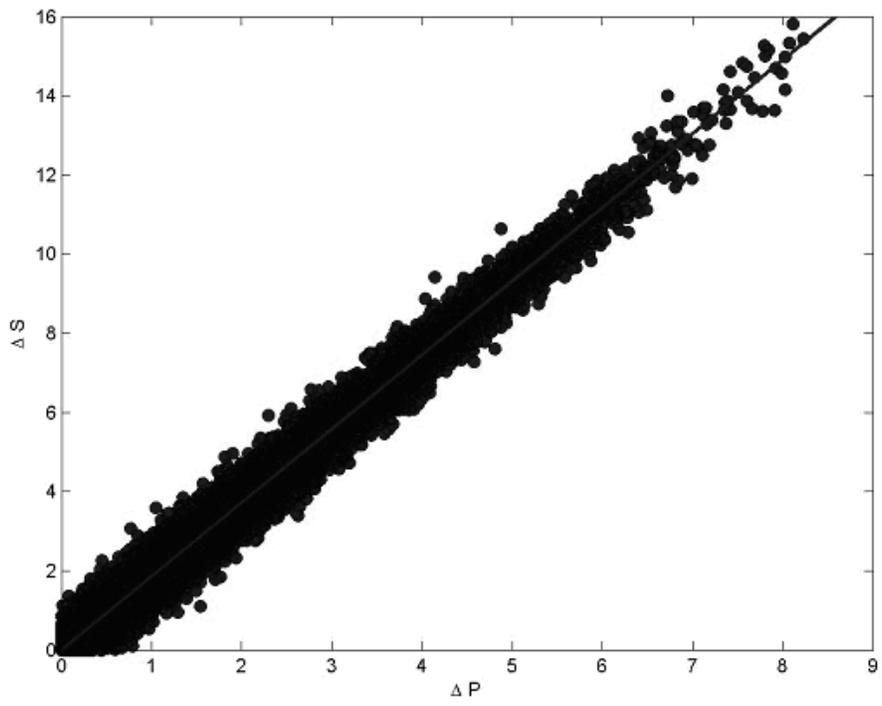
Fig. 9



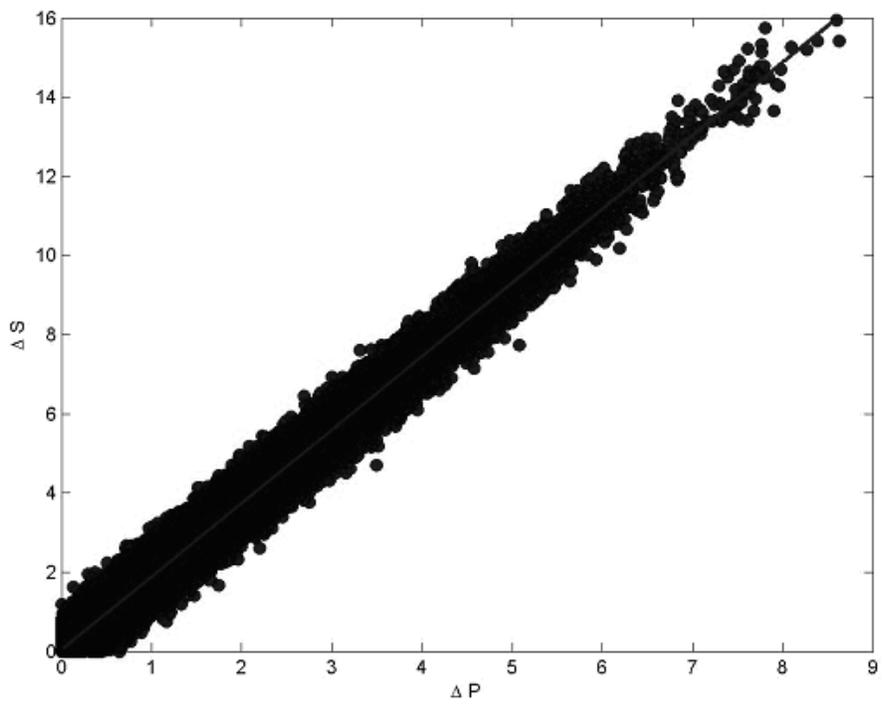
(a)



(b)



(c)



(d)  
Fig. 10

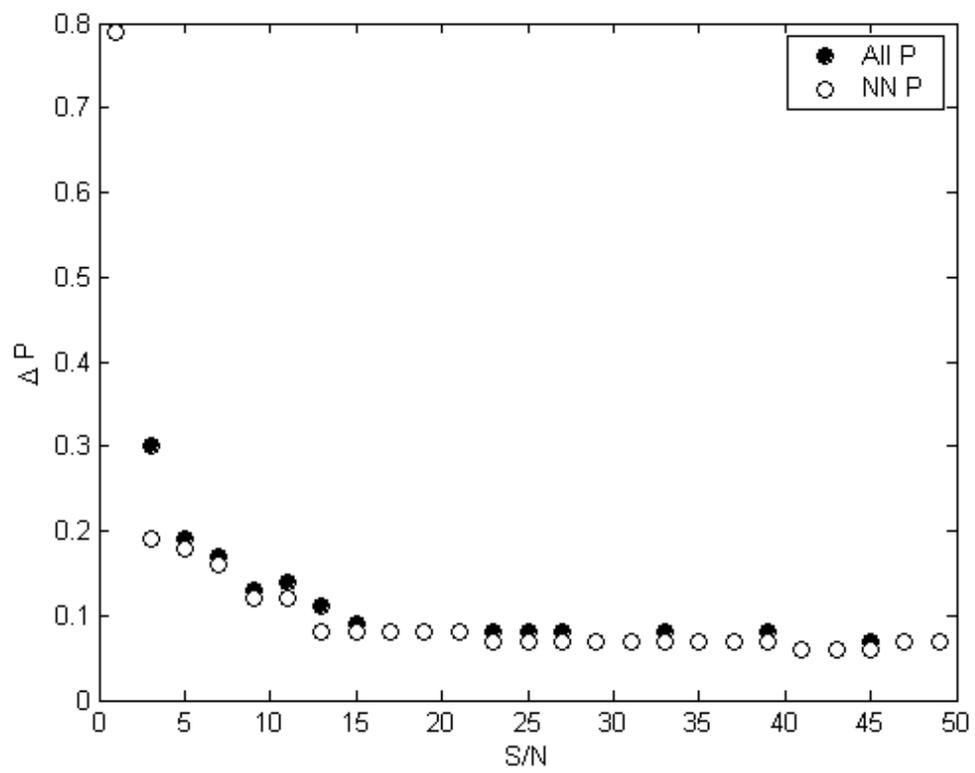


Fig. 11

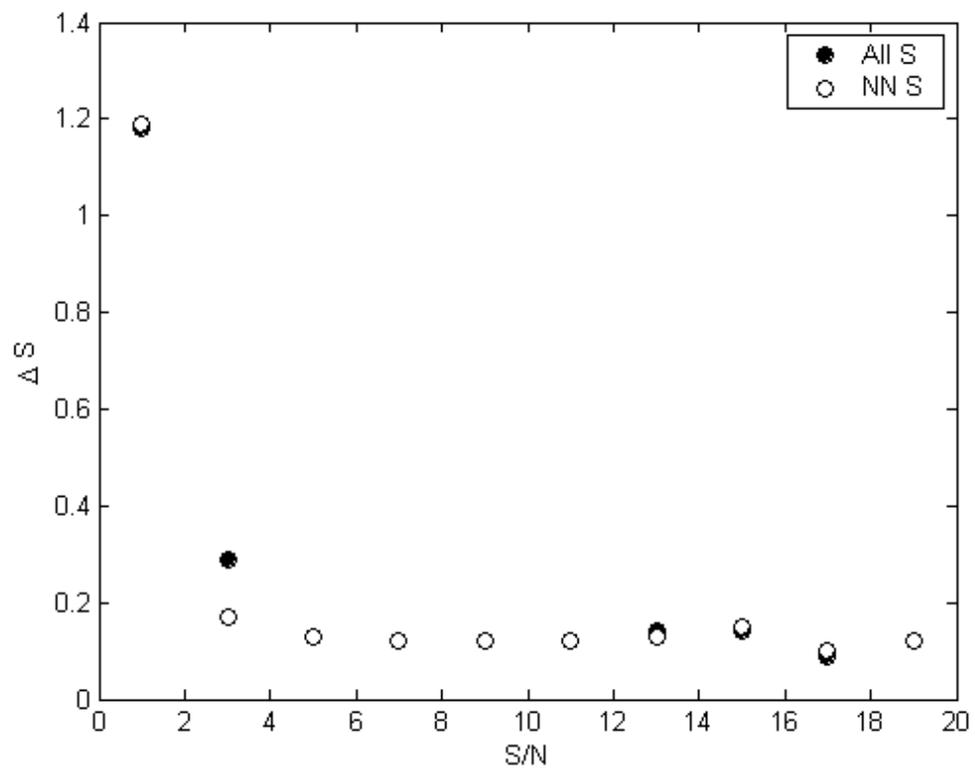
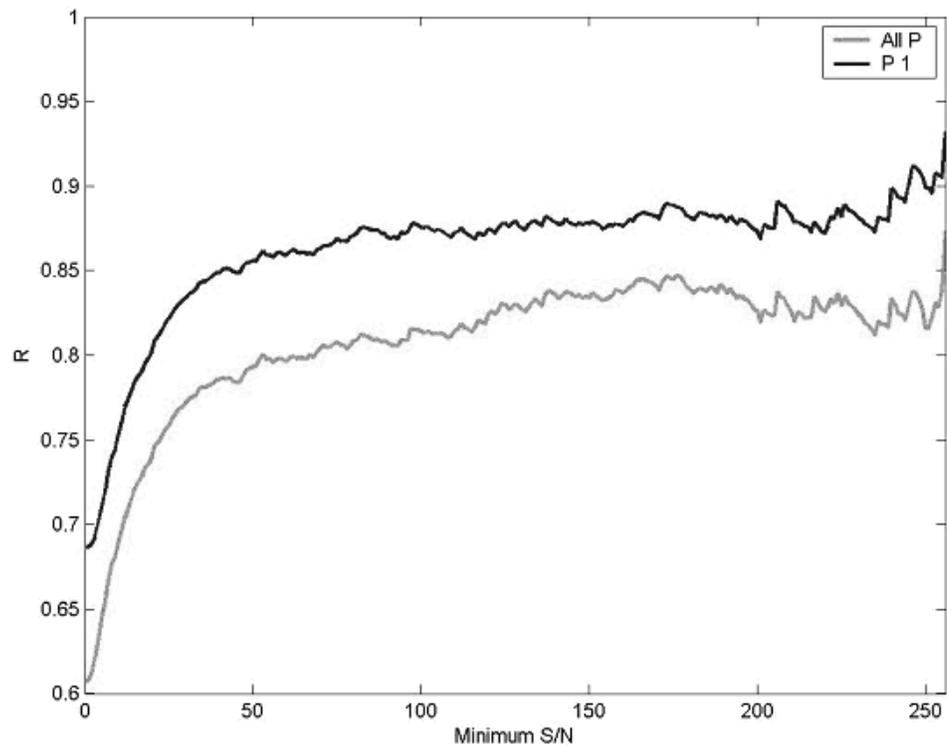
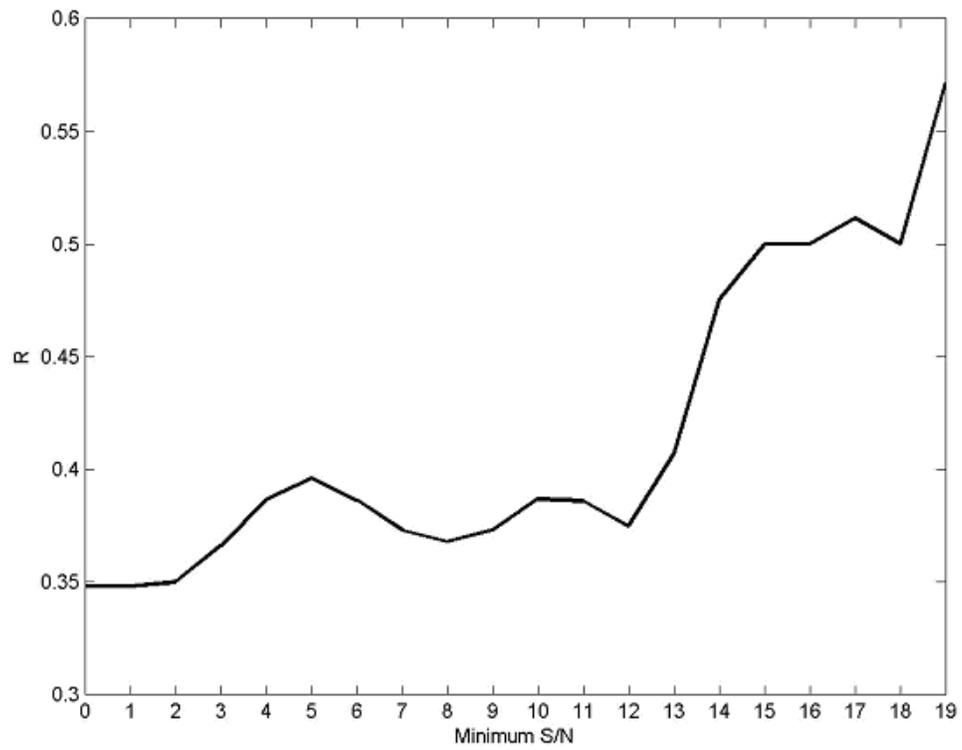


Fig. 12



(a)



(b)

Fig. 13