J Seismol (2017) 21:1171–1184
DOI 10.1007/s10950-017-9658-0

ORIGINAL ARTICLE

Automatic detection of P- and S-wave arrival times: new strategies based on the modified fractal method and basic matching pursuit

Rodrigo Chi-Durán · Diana Comte · Marcos Díaz · Jorge F. Silva

Received: 7 February 2016 / Accepted: 15 March 2017 / Published online: 31 May 2017

Abstract In this work, new strategies for automatic identification of P- and S-wave arrival times from digital recorded local seismograms are proposed and analyzed. The database of arrival times previously identified by a human reader was compared with automatic identification techniques based on the Fourier transformation in reduced time (spectrograms), fractal analysis, and the basic matching pursuit algorithm. The first two techniques were used to identify the P-wave arrival times, while the third was used for the identification of the S-wave. For validation, the results were compared with the short-time average over long-time average (STA/LTA) of Rietbrock et al., Geophys Res Lett 39(8), (2012) for the database of aftershocks of the 2010 Maule $M_w = 8.8$ earthquake. The identifiers proposed in this work exhibit good results that outperform the STA/LTA identifier in many scenarios. The average difference from the reference picks (times obtained by the human reader) in P- and S-wave arrival times is ~ 1 s.

Keywords Autopicking · Fractal analysis · STA/LTA · Basic matching pursuit · Modified fractal method

R. Chi-Durán (✉) · M. Díaz · J. F. Silva
Departamento de Ingeniería Eléctrica and Advanced Mining Technology Center (AMTC), Facultad de Ciencias Físicas y Matemáticas, Universidad de Chile, Av. Tupper 2007, Santiago, Chile
e-mail: rchi@ing.uchile.cl; rodrigok.chi@gmail.com

D. Comte
Departamento de Geofísica and Advanced Mining Technology Center (AMTC), Facultad de Ciencias Físicas y Matemáticas, Universidad de Chile, Av. Blanco Encalada 2002, Santiago, Chile
e-mail: dcomte@dgf.uchile.cl

1 Introduction

The correct identification of the P- and S-wave arrival times is crucial for the joint determination of hypocenters and 3D body wave velocities (seismic tomography). Manual identification is a slow and expensive task; therefore, advances in the automatization of this process would allow for a more efficient procedure, without the lost of precision.

Since the late 1970s, automatization in the peak identification of seismic phases in time series associated to digital seismograms has been investigated. This procedure is usually known as "picking." Allen (1978) proposed the first automatic method for the identification of P-waves, which followed other advances in the 1980s, until Cichowicz (1993) who proposed the first method to detect the S-wave.

The automatic identification of phases continues to be an important challenge to the community due to the variability of the seismic signal. The morphology of signal can be depends on the distance of the source, where new seismic phases (P_n, P_g, or PmP, among others) appear when the duration of the event

increases. In this investigation, we will only be concerned with estimating P- and S-wave arrival times from local earthquakes.

Among the proposed solutions, the short-time average over long-time average (STA/LTA) algorithm created by Allen (1982) has been the most-used detection method, including some subsequent modifications made by Baer and Kradolfer (1987) or Earle and Shearer (1994). There are other STA/LTA derivations. For instance, there is the Jones and van der Baan (2015) hidden Markov models with STA/LTA to detect seismic in continuous data or Ross and Ben-Zion (2014a) who developed a pseudo-probability to determinate a seismic events and different phases using STA/LTA ratio (Ross and Ben-Zion 2014a). Besides the STA/LTA method, many other solutions have been implemented based on neural networks, Fourier analysis, wavelet analysis, statistical analysis, and fractal analysis (Dai and MacBeth 1995; Anant and Dowla 1997; Bai and Kennett 2000; Sabbione and Velis 2010). All of these methods provide different results depending on the nature and quality of the data collected: signal-to-noise ratio and models of the physical environment through which seismic waves travel (Ross and Ben-Zion 2014b), among others.

This work combines the fractal and spectrogram methods in order to estimate P-wave arrival times. On the other hand, we use basic matching pursuit algorithm to detect S-wave arrival times. Due to the nature of seismic signals, our approach exploits temporal changes of the P-wave to S-wave, in order to distinguish them from the noise. Both developed methods take advantage of this fact to identify the arrival of the P- and S-waves.

This work combines the fractal method and spectrogram method in order to estimate P-wave arrival times. For S-wave arrival time estimation, we proposed the use of the basic matching pursuit algorithm. For this reason only these methods are evaluated and compared to the performance of STA/LTA as standard method.

2 Mathematical background

2.1 The short-time Fourier transform

The short-time Fourier transform (STFT) consists of separating the Fourier transform in sections, applying it to a moving window $w(n, \tau)$, where τ is the window's position along the entire signal (Fig. 1).

The windows are usually overlapped in order to prevent discontinuities in the form of waves, for example, the Hamming window, which is given by the expression

$$w(n, \tau) = 0.54 - 0.4 \cos\left(\frac{2\pi(n-\tau)}{N_\omega - 1}\right). \tag{1}$$

For $0 \leq n - \tau \leq N_\omega - 1$ and a null sequence for any other value, where N_ω is the window's duration, the value N_ω is commonly chosen considering the trade-off between the width of the window and the resolution in frequency of the signal. The short-time Fourier transform (STFT) is defined by

$$X(\omega, \tau) = \sum_{n=-\infty}^{\infty} x(n, \tau) e^{j\omega n}, \tag{2}$$

where

$$x(n, \tau) = w(n, \tau) x(n), \tag{3}$$

which represents the window of the signal analyzed, centered at a time τ (Fig. 2). Note that $x(n, \tau)$ represents signal in time domain and $X(\omega, \tau)$ in the frequency domain for the same windows.

We define the spectrum $S(\omega, \tau)$ of the signal $X(\omega, \tau)$ as

$$S(\omega, \tau) = |X(\omega, \tau)|^2, \tag{4}$$

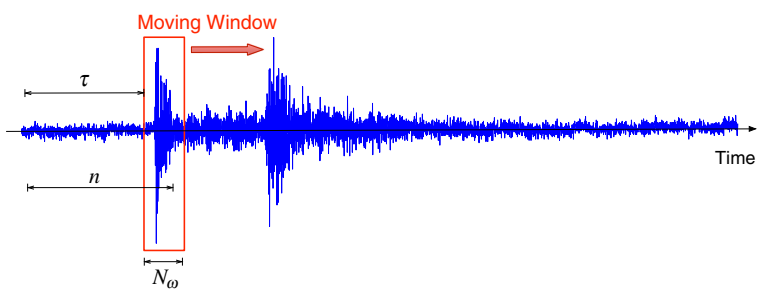

Fig. 1 Graphic explanation of the parameters τ, N_ω, and n

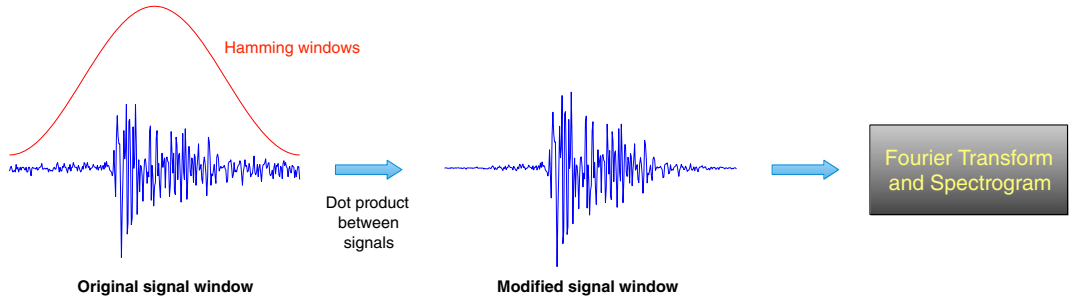

Fig. 2 Graphic explanation on the use of the Hamming window in the determination of the spectrogram

which can be interpreted as a two-dimensional energy density, relative to the one-dimensional spectral information received by the Fourier transform at different temporary locations.

2.2 Basic matching pursuit

Basic matching pursuit (BMP) was introduced by Mallat and Zhang (1993) and has been used for several applications with signal or image processing (Durka and Blinowska 1995; Phillips 1998; Wang 2006). The algorithm is based on the following: we considerer l observations $\{y_1, y_2, \ldots, y_l\}$ and a target function $f \in \mathcal{H}$ be evaluated at the points $x_1, x_2, \ldots, x_l$, where $\mathcal{H}$ is a Hilbert space. On the other hand, a finite dictionary $\mathcal{D} = \{d_1, d_2, \ldots, d_M\}$ of M functions belongs to $\mathcal{H}$. The algorithm looks for a sparse approximation of the function f of the form

$$f \approx f_N = \sum_{n=1}^{N} a_n g_n(t), \tag{5}$$

where N is the quantity of basic functions required to approximate the function f, $g_i \in \mathcal{D}$ to a basic function and $\alpha_i \in \mathbb{R}^N$ corresponds to the coefficients of the approximation. f_N represents an approximation of f using exactly N different basic functions taken from the dictionary $\mathcal{D}$.

The algorithm's objective is to find the combination of basic functions $\{g_1, g_2, \ldots, g_N\} \subseteq \mathcal{D}$ and its respective coefficients $\{\alpha_1, \alpha_2, \ldots, \alpha_N\} \subseteq \mathbb{R}^N$ solution of the following minimization problem.

$$\text{mín} \sum_{i=1}^{l} (y_i - f_N(x_i))^2 \tag{6}$$

In general, it is not possible to assume specific properties of the dictionary; thus, finding N optimal functions $\{g_1, \ldots, g_N\}$ implies an exhausting search for all the combinations of M basic functions. The procedure previously described is inefficient; therefore, BMP is a constructive solution to face this problem. The detailed description of the algorithm can be found in Vincent and Bengio (2002).

The dictionary can be built with different families of wavelet functions. An atom is a wavelet function and a dictionary is created by temporary shifts of a particular wavelet type. The number of translations will depend on the length of the signal. Thus, the dictionary could be significantly great if many kinds of family functions are used and the signal is long.

2.3 Fractal dimension

A fractal is a type of geometry that exhibits repetitive patterns at any scale. These models have been used in ecology (Sugihara and May 1990), economics (Peters 1994), and geophysics (Turcotte 1989), among other applications.

Mathematically, it is possible to define the fractal dimension as a metric set for which the Hausdroff-Besicovitch dimension exceeds strictly the topological dimension (Mandelbrot 1983). The fractal dimension generalizes the topological dimension from a typical set which is always a natural number 1, 2, or 3 to a nonnegative real number.

The fractal curves can be classified as autosimilar or self-affine (Turcotte 1997). A formal definition of an self-similar fractal in a two-dimensional space xy is a function such that $f(rx, ry)$ is statistically similar to $f(x, y)$, where r is a scale factor. On the other hand, a self-affine fractal is a function such that $f(rx, r^{H_a} y)$ is statistically similar to $f(x, y)$, where H_a is known

as the measurement of Hausdorff (Sabbione and Velis 2010). The definition of the fractal dimension for a self-affine fractal is given by

$$H_a = 2 - D. \tag{7}$$

For a time series $s(t)$, an equivalent definition for a self-affine fractal is as follows:

$$V(h) \sim h^{2H_a}, \tag{8}$$

where $V(h)$ is defined as

$$V(h) = \frac{1}{n_f - h} \sum_{j=t-n_f+h+1}^{t} (s(j) - s(j-h))^2, \tag{9}$$

where n_f is the size of time window where the fractal dimension is calculated. Various methods exist to determine the the fractal dimension, depending on whether the fractal is self-affine or self-similar (Klinkenberg 1994). The method of the variogram (Korvin 1992) is a efficient and robust way of determining the fractal dimension for self-affine curves. The method is based on the combination of the Eqs. 7 and 8 which delivers the following relationship on the

$$V(h) \sim h^{4-2D}. \tag{10}$$

Therefore, by applying logarithm function in both sides of the Eq. 10, it is possible to obtain the value of D.

$$\log V(h) = 4 - 2D \Longrightarrow D = 2 - \frac{\log V(h)}{2} \tag{11}$$

2.4 Short-time average over long-time average

Allen (1982) proposed the STA/LTA algorithm, and it is a commonly used identifier in comparison to new automatic identifiers. This algorithm utilizes the STA/LTA index, which identifies an increase in the increment of the STA window, with respect to the predecessor LTA window determined in a seismic trace. Mathematically, in a time series $s(t)$, we can define

$$s^{\mathrm{STA}}(t) = \frac{1}{T^{\mathrm{STA}}} \sum_{j=t-T^{\mathrm{STA}}}^{t} s(j)^2, \tag{12}$$

where T^{STA} is the size of short window. Similarly, we can define

$$s^{\mathrm{LTA}}(t) = \frac{1}{T^{\mathrm{LTA}}} \sum_{j=t-T^{\mathrm{LTA}}}^{t} s(j)^2, \tag{13}$$

where T^{LTA} is the size of long window. The STA/LTA ratio can be expressed by

$$S(t) = \frac{s^{\mathrm{STA}}(t)}{s^{\mathrm{LTA}}(t)}. \tag{14}$$

This method requires a previous calibration of its parameters which includes (1) the size of the STA and LTA windows and (2) the threshold value that the ratio must pass in order to be detected (see Fig. 3). The STA/LTA program, which was used to test the database, was previously programmed and documented by Nippress et al. (2010). The program

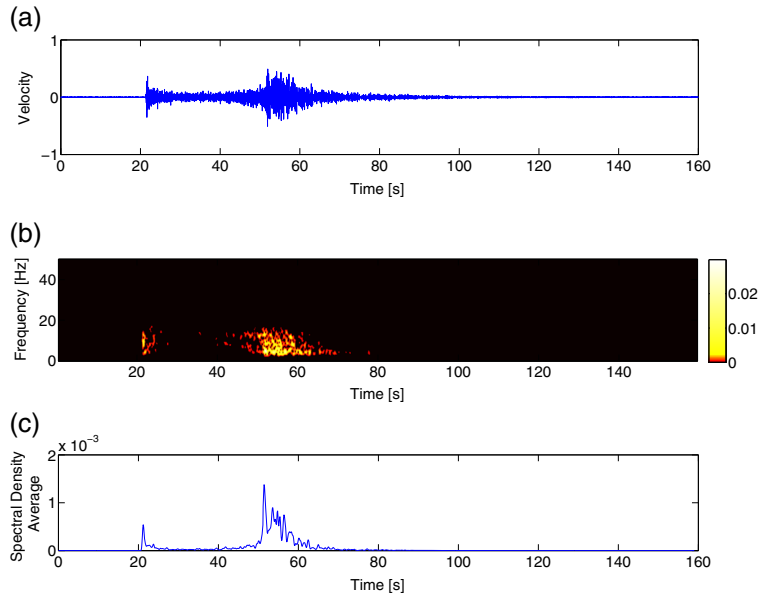

Fig. 3 **a** Seismogram. **b** Seismogram spectrogram, where the increase in energy identifies the P- and S-wave arrival times. **c** Average spectral density obtained from the spectrogram, where a clear increase can be associated with the arrival time of the P-wave

was also used to identify the hypocenters of the aftershocks after the Maule earthquake in 2010 and was developed by Rietbrock et al. (2012).

3 Proposed identifiers

In this section, we describe the identifiers proposed to determine the arrival time of the P- and S- waves on seismograms.

3.1 Spectrogram method

The spectrogram allows graphically displaying more energetic zones of the signal in some frequency bands, which is compatible with the characteristics of seismic signals. Joswig (1990) appreciated that seismic signals have similar and repetitive spectral characteristics. For this reason, we proposed a spectrogram method (SM) for detecting the arrival time of the P-waves.

In Fig. 3, the spectrogram and the average spectral density of a seismic signal are computed. As explained in Section 3.2, the spectrogram was obtained with a window of 1.2 s and an overlap of the Hamming windows of 90%. The increase in energy can be seen at the arrival of the P- and S-waves (see Fig. 3b). Finally, the spectral density average (SDA) was obtained by averaging out the values of the spectrogram in discrete frequencies for each time of the seismogram (see Fig. 3c). The threshold is set to standard deviation of SDA trace, σ_{SDA} (Fig. 3c). Then, the arrival time of P-wave is determined when the value of the spectral density average SDA at some time overcomes the threshold σ_{SDA}.

3.2 Modified fractal method

Korvin (1992) applied fractal dimensions to time series via the variogram method (Section 2.3). Figure 4 shows an example of use of this method. In this figure, the global minimum of fractal dimension derivative (Fig. 4c) coincides with P-wave arrival. However, there are cases, as shown in Fig. 5, where the global minimum of fractal dimension derivative does not match with P-wave arrival. One way to solve this issue is by using spectrogram method (Section 3.1) instead of FM. Nevertheless, the spectrogram method by itself has poor precision on identifying the arrival time. Therefore, we propose to combine both methods with the objective to improve the estimation of the P-wave arrival time.

The proposed modified fractal method (MFM) works as follows:

1. Carry out the spectrogram method with a relatively large time window (for example, 3 s). This allows a rough identification of the P-wave arrival

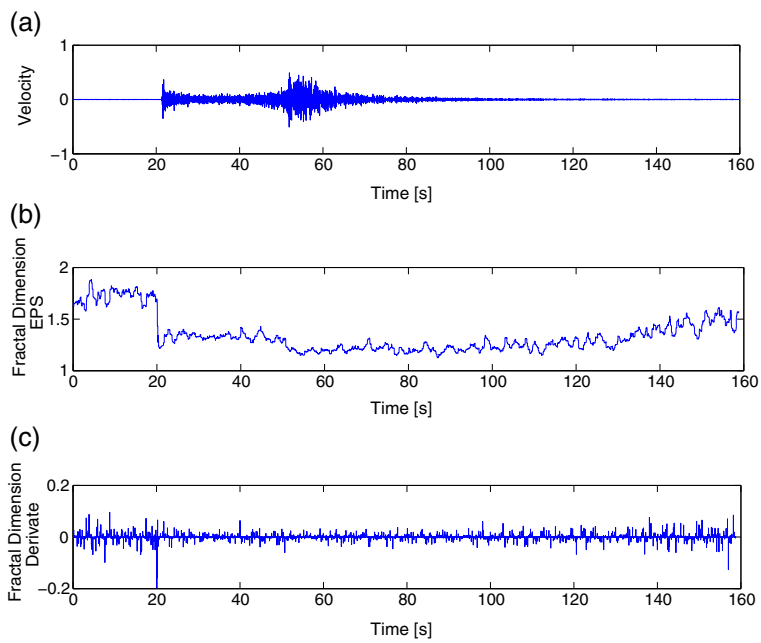

Fig. 4 **a** Seismogram. **b** Seismogram fractal dimension. **c** Discrete fractal dimension derivative (subtraction of consecutive values of seismogram fractal dimension, **b**), where it is possible to verify a minimum that coincides with the arrival time of the P-wave

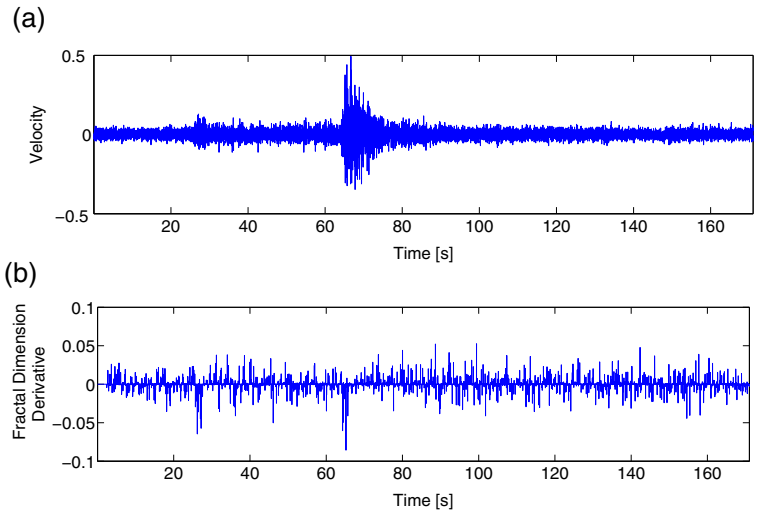

Fig. 5 **a** A noisy seismogram. **b** Derivative of the fractal dimension where the onset of arrival time of P-wave is not identifiable via MF because the global minimum in this case coincides with the arrival of the S-wave

time. In this point, the interval where the P-wave arrives can be identified.

2. In order to refine the P-wave arrival time, run the fractal method for the whole signal.
3. Take the P-wave arrival time as the local minimum of the fractal dimension derivative (the time series obtained in point 2) within the time interval obtained in point 1.

Additionally, we apply an edge-preserving smoothing (EPS) filter to the fractal dimension series (AlBinHassan et al. 2006). The 1D EPS filter is useful for smoothing out the time series signal; in this case, it helps to detect the drastic change in time of the fractal dimension.

An example of an inaccurate identification of the P-wave resulting from the use of the global minimum (fractal method) can be observed on Figs. 5 and 6. In contrast, the MFM solves this problem by searching for the interval with the highest probability of arrival of the P-wave.

3.3 Basic matching pursuit

This scheme was explained in Section 2.2 and an application to seismogram is shown in Fig. 7.

From this example, it is possible to see that the basic matching pursuit (BMP) method completely eliminates the appearance of the P-wave within the

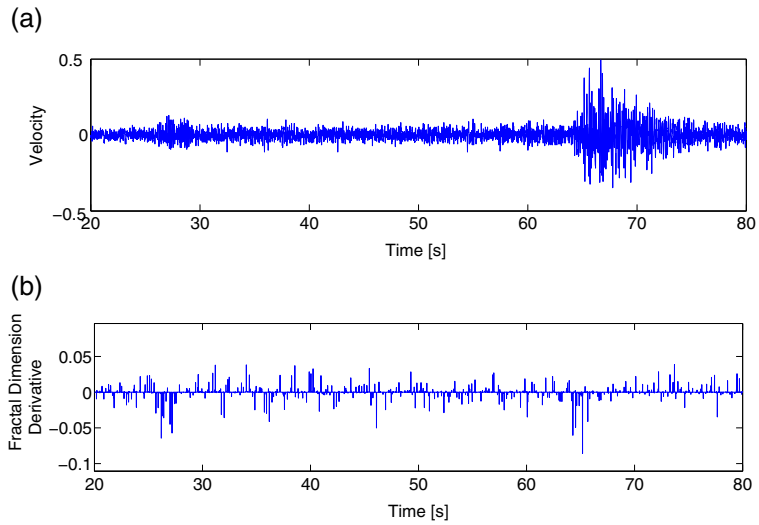

Fig. 6 Zoom in Fig. 5 between 20 and 80 s

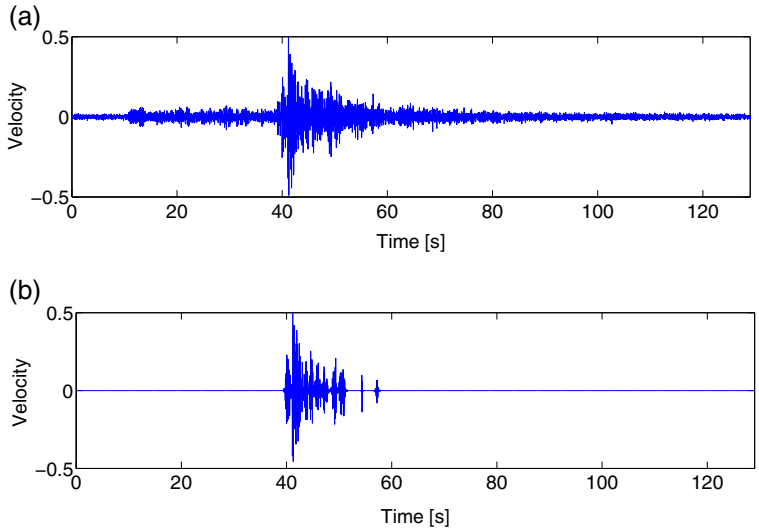

Fig. 7 **a** Complete seismogram. **b** Seismogram after the application of the BMP algorithm, where the complete disappearance of the P-wave can be observed

seismic trace. This is a consequence of the ability of the method to approximate the best shape of the most energetic parts of the seismogram using a finite number of wavelet functions, which generally coincides with the arrival of the S-wave. The time arrival is chosen selecting the first non-null value in filter signal (Fig. 7).

Two examples of a less desirable decomposition, which are recurrent in the database, are shown in Fig. 8. In the first case (Fig. 8a, b), the P-wave is more energetic than the S-wave, and consequently, the BMP algorithm decomposes some segments of the P-wave. This situation could depend on many factors like the paths of the body waves controlled by the material of the medium, the local anomalies associated with the specific seismic station, and the depth and the focal mechanism associated to the seismic

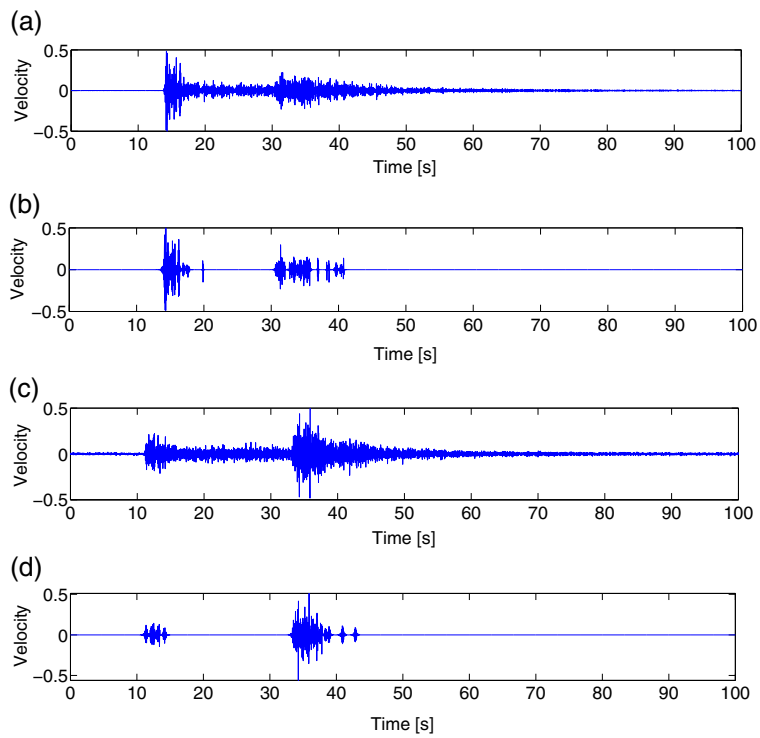

Fig. 8 **a–c** Complete seismogram. **b–d** Seismogram after the application of the BMP algorithm, where the partial disappearance of the P-wave can be observed

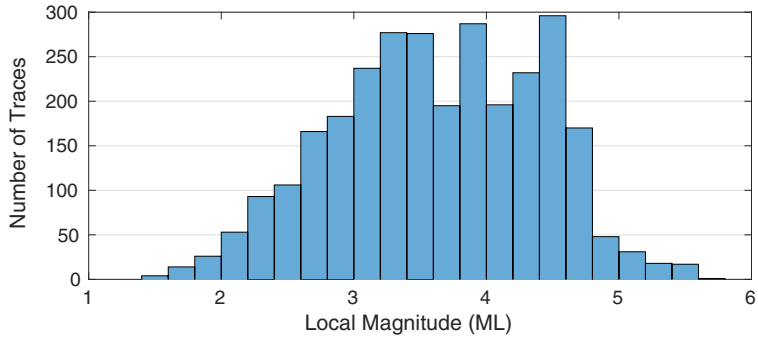

Fig. 9 Distribution of seismic magnitudes in the databases

event. Figure 8c, d shows another case, where P-wave is detected after BMP method. These situations create an ineffective selection of the S-wave, that produces miss-detection between the detection of the P- and S-wave (see Fig. 8). In order to take these limitations into account, we propose some modifications that consist of the following:

- running the filtered signal and identifying the "gaps," which are defined as time intervals where the filtered signal is completely null.
- ignoring the initial and final gaps, it appears that there is a sufficiently long gap (based on the threshold) which is considered to be a separation between the P- and S-wave arrival times. In the event of a separation, the first non-null value after the end of this gap is selected.
- if the gap is not long enough or there is no gap in the filtered signal, the first non-null value is selected as the arrival time of the S-wave.

Figure 8 shows that at 50 s, the signal is divided by the largest gap, which also separates the P- and S-wave arrival times. With the modification of the algorithm, the correct form of the arrival time of the S-wave would then be recognized.

4 Database and pre-processing

In this study, the database corresponds to seismic events recorded by a network of 26 seismological stations. The stations are short-period, three-component continuous recording seismological stations located around Pisagua, Northern Chile (FONDECYT 1130071). These 528 local seismic events occurring between June 1st and June 9th 2013 were chosen to serve as test bed for this work. The stations operated for 4 months (from June to October 2013), with a sampling rate of 100 Hz. Total measured traces for the stations correspond to 2926 analyzed signals, because there are stations that do not record all events in all components. The lengths of the traces range from 30 to 180 s.

The magnitude of the events ranging from 1.5 to 5.2 M_L (Fig. 9) and their depths are from 0.43 to 175.6 km (Fig. 10). All the traces utilized in this study have had the onset of P- and S-wave arrival times manually identified by an expert. The pre-processing of the data consists on subtracting the arithmetic mean of the signal to its standardization, leaving a range of velocity between ± 1, according to the following formula:

$$S_N(t) = \frac{s_0(t)}{\max\{s_0\} - \min\{s_0\}}, \tag{15}$$

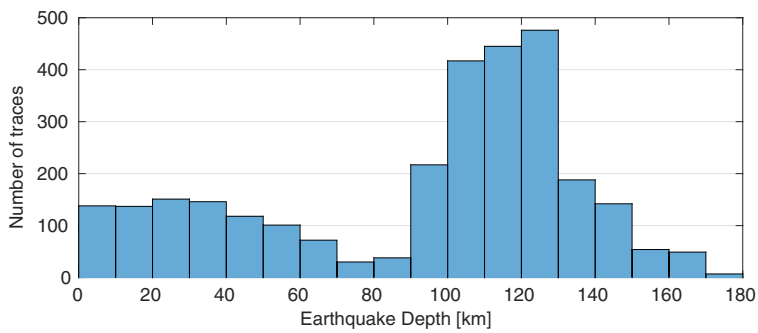

Fig. 10 Distribution of depth of the earthquakes

where

$$s_0(t) = s(t) - \bar{s} \tag{16}$$

and $\bar{s}$ is average of original signal $s(t)$, for any time t.

5 Methodology and parameter identification

For the identification of P-wave arrival times, vertical components were used. On the other hand, the horizontal components were used for identifying S-wave arrival times (Rawles and Thurber 2015).

The parameters used are shown in Table 1. For the MFM, the values recommended by Sabbione and Velis (2010) were used. On the other hand, the BMP parameters were determined by iterating through various combinations of them.

Additionally, Nippress et al. (2010) recommend a range of optimal values for the parameter of STA/LTA detector. They were found by trying different values within this range in a incremental manner from the lower limit to the upper limit by using equally spaced steps or increments (Nippress et al. 2010).

6 Results and analysis

In the following section, the results obtained from identifiers described above are shown. Given that the database had previously identified the P- and S-wave arrival times by a human expert, the residual time was defined by

$$t_{\mathbf{res}} = t_{\mathbf{auto}} - t_{\mathbf{manual}}, \tag{17}$$

where $t_{\mathbf{auto}}$ is the time obtained by automatic identifiers and $t_{\mathbf{manual}}$ is the time obtained by an expert.

Table 1 Used parameters of the different identifiers

Method	Parameters	Value/s
MFM	Size of spectrogram windows	5
	% of overlap between windows	20
	Filter length EPS (s)	0.15
	Fractal dimension window size (s)	12
BMP	Number of atoms	25
	Type of wavelet	Daubechies lv5
	Duration gap (s)	180
STA/LTA	Window size STA (s) Increments: 8	(1, 8)
	Window size LTA (s) Increments: 23	(8, 30)
	Threshold increments: 9	(4, 12)

Table 2 Parameter STA/LTA for the identification of the P-wave arrival time

Method	Parameters	Value
STA/LTA	Window size STA (s)	1
	Window size LTA (s)	10
	Threshold	4

6.1 P-wave detection

Regarding the detection of P-waves, various iterations were carried out to find the parameters that obtain the best detection performance for the STA/LTA method for further comparison with the MFM.

The values of the parameters which give the best performance in STA/LTA are shown in Table 2. In Table 3, the effectiveness of the MFM detection versus the results from the best STA/LTA can be observed. It should be noted that MFM is more accurate than STA/LTA in all intervals. Nevertheless, both methods were correct in more than 85% with residual times within ±1-s interval.

The histogram of the distributions of time in the interval ±2 s can be observed in Fig. 11, where the distribution of the residual times in this interval is shown in detail. It is possible to verify that the MFM residual

Table 3 Proportion of P-wave arrival times detected within different residual time intervals using MFM and STA/LTA methods

Residual interval time (s)	MFM (%)	STA/LTA (%)
(−0.1, 0.1)	47.3	10.9
(−0.5, 0.5)	83.9	68.1
(−1, 1)	93.8	85.4
(−2, 2)	98.2	90.7
Phase not identified	0	0.06

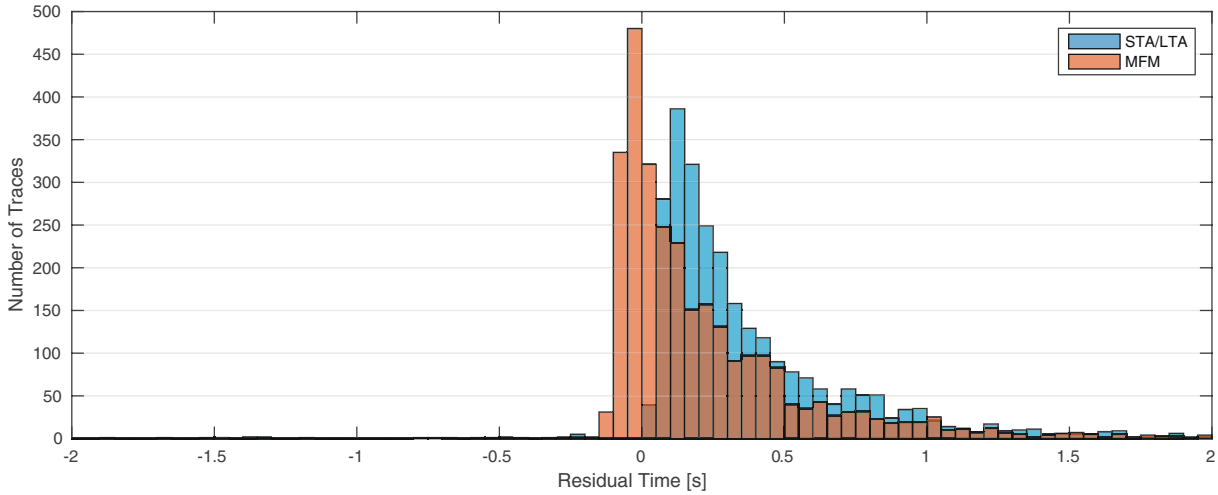

Fig. 11 Histogram of the distribution of residual times for the P-wave in the interval ±2 s (MFM and STA/LTA methods)

times are closest to zero, so that 47% of the traces have a residual time less than 0.1 s. In general for all the traces in the database, 80% of the residual errors obtained by MFM are smaller than those found by the STA/LTA method.

Another characteristic of both methods is that they tend to have a positive residual error, that is $t_{\mathbf{manual}} < t_{\mathbf{auto}}$. Consequently, both methods tend to estimate later the arrival of P-waves. For the MFM, a possible explanation for the overestimation might be related to the fact that fractal dimension needs a moving window of 12 s in order to be determined; therefore, this provokes a difference between the arrival of the wave and the time of detection.

Physically, the overestimation of the P-wave arrival times could propagate as an error in the determination of hypocenters and the structure of wave velocities. Miscalculating the P-wave arrival time by just 1 s could translate into approximately 7 km of difference in the hypocenter's actual location.

Comparing the MFM method to other methods, Küperkoch et al. (2010) implement a high-order statistics to estimate arrival time of the P-wave. On the other hand, Rawles and Thurber (2015) use the nearest neighbor algorithm and build a model based on non-parametric time-series classification. In this way, they estimate the parameters from a set of reference data that transforms the original signal and obtains the arrival time of the wave based on the analysis of the processed signal. The methods previously mentioned use very sophisticated statistical tools, whereas there are other methods like Rawles and Thurber (2015) who proposed a simple methodology to determine P- and S-waves.

Our average deviation from the reference picks is 0.18 ± 0.62 s. This result is more accurate in comparison to that of Küperkoch et al. (2010), because they have 0.26 ± 0.64 and 0.38 ± 0.75 s in their P-pickers. However, our identifier is less accurate than Rawles and Thurber (2015) method, which was 0.04 ± 0.51 and 0.1 ± 0.57 s. Although the ideal is that the methods can be compared with the same data, we can verify that our P-wave detector has acceptable performance with respect to other methods.

Figure 12 shows that the method is sensitive to SNR of the P-wave signal. It is possible to see the residual time tends to zero when the SNR increases. For example, for SNRs greater than 15 dB, most of the P-wave signals have a residual time less than 0.5 s. The result is coherent with the fact that high SNR implies a clearer signal, so identification is more accurate.

With respect to the computational cost of the process, currently, MFM requires much more time (factor of 4) than the STA/LTA method. However, the comparison is difficult since the languages used are different: C for STA/LTA and Matlab for MFM. Nevertheless, we anticipate that MFM requires more computational time since the determination of the fractal dimension is a sequential process of moving time windows. Additionally, the application of EPS filter is slow too; thus, this process could be parallelized.

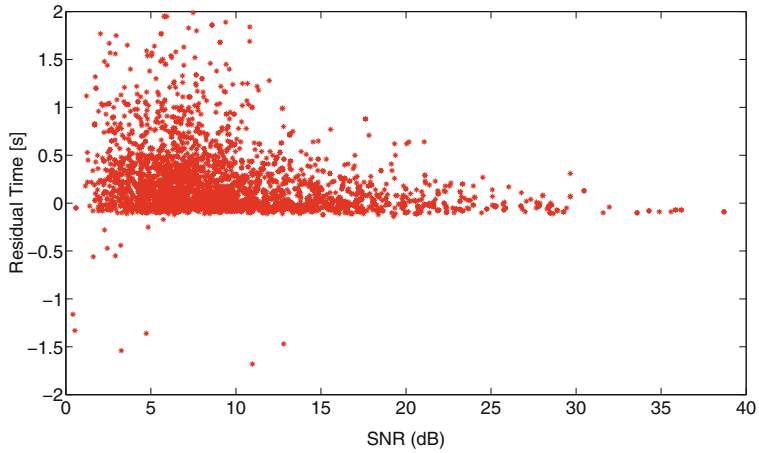

Fig. 12 Comparison between SNR and residual times for the MFM (P-wave detection)

6.2 S-wave detection

In the case of the S-wave arrival time detection, we iterated using the range of values that is specified in Table 4.

The maximum values found for the performance of the picker in the horizontal components are shown in Table 4 (i.e., those which have the largest quantity of identifications close to the residual null time). These values are similar to those found for the P-wave.

As shown on Table 5, the results of the STA/LTA picker are far below the results from the BMP. In general, the majority of traces observed are not capable of being identified by STA/LTA in the range of parameters recommended by Nippress et al. (2010).

In Figs. 13 and 14, it is possible to observe the distribution of residual times for S-waves using the North-South and East-West components, respectively. Those results are consistent with what is shown in Table 5, that the detection of the S-wave is similar in the STA/LTA indicator.

The STA/LTA method has miss-picking between 25 and 32% and it is more inaccurate than BMP in all intervals (Table 5). It is interesting that, for the same range of parameters, the detection of the P-wave is considerably more precise and effective than that of the S-wave.

Table 4 STA/LTA parameters for the identification of the S-wave arrival times

Method	Parameters	North-South value	East-West value
STA/LTA	Window size STA (s)	1	1
	Window size LTA (s)	10	10
	Threshold	4	4

On the other hand, the BMP detector shows a very good approximation of the arrival time of the S-wave, even though its displacement to the left -0.6 ± 0.97 s in the North-South component and -0.53 ± 0.92 s in the East-West component. Additionally, this method could continue improving the creation of dictionaries of functions specialized in seismic signals, and by this way, the decomposition would be much more exact in the arrival times of the P- and S-wave. In contrast to S-wave arrival identifiers, the BMP is less precise than other methods; for instance, Rawles and Thurber (2015) introduce a detector with -0.13 ± 1.58 and 0.06 ± 1 s or Küperkoch et al. (2012) who have reported detection results of 0.5 ± 0.8 s.

Furthermore, BMP gives a good result indentifying the arrival time of the S-wave despite we did not apply any kind of sophisticated preprocessing to the original data

Furthermore, Fig. 15 shows that it is possible to visualize the behavior of the S-wave's residual time against the SNR. In this figure, it is possible to notice that the identification of the S-wave is not clear in a noisy environment.

At this point, it is important to highlight the fact that the evaluation of BMP has been performed only with this data base, which implies seismic signals with high SNR, near to the source (local events) and with measurements in three components. Therefore, generalization of the performance of the method requires

Table 5 Proportion of S-wave arrival times detected within different residual time intervals using BMP and STA/LTA methods

	BMP (%)		STA/LTA (%)	
Residual interval time (s)	North-South	East-West	North-South	East-West
(−0.1, 0.1)	6.29	6.56	1.57	1.57
(−0.5, 0.5)	35.17	36.43	34.93	31.95
(−1, 1)	79.87	82.54	59.71	54.20
(−2, 2)	94.74	95.42	70.43	63.74
Phase not identified	0	0	25.46	32.5

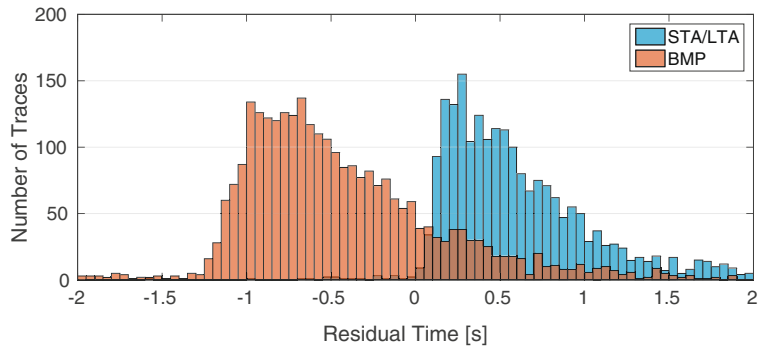

Fig. 13 Histogram of the distribution of residual times for the S-wave in the interval [−2, 2] seconds. North-South component

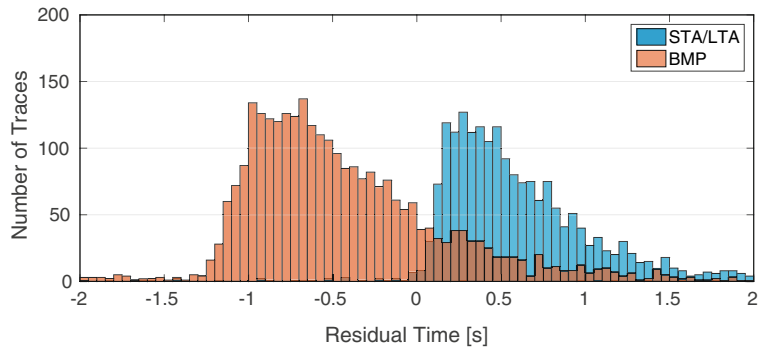

Fig. 14 Histogram of the distribution of residual times for the S-wave in the interval [−2, 2] seconds. East - West Component

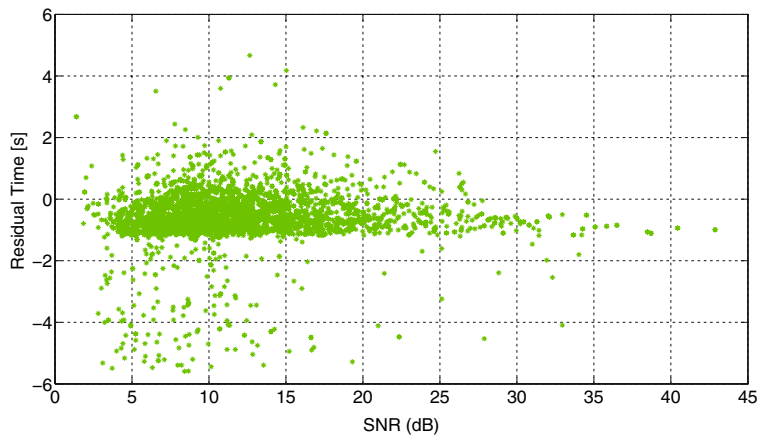

Fig. 15 Comparison between SNR and residual times for the BMP in North component (S-wave detection)

further study. The used signal duration goes from 30 to 180 s, and we used two kinds of wavelets to decompose the signal:

- Daubechies wavelet with four vanishing moments at level five.
- Wavelet packet Symlet with four vanishing moments at level five (modified version of Daubechies wavelets).

These kinds of waveles were chosen because the morphology of these functions is similar to that of seismic traces (Rinehart et al. 2016).

A major drawback of BMP is its computational cost during the creation of the dictionary. MATLAB was presented problems for the creation of dictionaries for signals that have more 18000 samples (180 s). Although the BMP algorithm does not use brute force, it is slower if you use larger dictionaries with different kinds of wavelets. We recommend using a language like C or Fortran if large databases are analyzed.

7 Conclusions

The methods presented in this study showed an improved performance when it is compared with the STA/LTA identifier. In this work, the estimated arrival time obtained with all methods were compared with times obtained by a human expert.

For the identification of the P-wave, the modified fractal method (MFM) proves to be quite precise in predicting the first arrivals of seismic waves compared to STA/LTA; 78.5% of the traces got a more precise time arrival with MFM than with STA/LTA.

For the identification of the S-wave, the basic machine pursuit (BMP) algorithm method is an accurate approximation in an interval close to the value detected by an expert of the S-wave arrival times. The BMP method outperform the STA/LTA in all time residual intervals of comparison; additionally, the STA/LTA method does not able to detect S-wave in 25 to 32% of cases.

The data presented in this work have been evaluated with a data base where seismic signals are similar in magnitude, depth, and hypocentral distance, and are for local events. Additionally, the measurements have three components; thus, in the vertical component, it is possible to see more clearly arrival of the P-wave and the S-wave in the horizontal components. The proper visualization of the arrival of the P-wave allows to see clearly the precipitous change of the derivative of fractal dimension. On the other hand, an energetic S-wave allows a better wavelet decomposition in the seismic signal. Our data also had good SNR. In general, it was shown that the results from MFM improve if the SNR of the signal is high, for which an abrupt change in the fractal dimension is clearer and the identification becomes more precise. On the other hand, MBP is presented as a method capable of identifying the S-wave as long as the earthquakes are temporally bounded. Therefore, in order to verify the validity of the results obtained in this work to other cases (generalizing the results), further studies need to be performed with different type of data.

Acknowledgements We thank Andreas Rietbrock for the use of his STA/LTA software in this study, and also the anonymous reviewers who contributed to the improvement of our work. This research was supported by FONDECYT, Project 1130071. J. F. Silva acknowledges support from the Advanced Center for Electrical and Electronic Engineering, Basal Project FB0008.

References

AlBinHassan NM, Luo Y, Al-Faraj MN (2006) 3D edge-preserving smoothing and applications. Geophysics 71(4):P5–P11

Allen R (1978) Automatic earthquake recognition and timing from single traces. Bull Seismol Soc Am 68(5):1521–1532

Allen R (1982) Automatic phase pickers: their present use and future prospects. Bull Seismol Soc Am 72(6B):S225–S242

Anant KS, Dowla FU (1997) Wavelet transform methods for phase identification in three-component seismograms. Bull Seismol Soc Am 87(6):1598–1612

Baer M, Kradolfer U (1987) An automatic phase picker for local and teleseismic events. Bull Seismol Soc Am 77(4):1437–1445

Bai CY, Kennett B (2000) Automatic phase-detection and identification by full use of a single three-component broadband seismogram. Bull Seismol Soc Am 90(1):187–198

Cichowicz A (1993) An automatic S-phase picker. Bull Seismol Soc Am 83(1):180–189

Dai H, MacBeth C (1995) Automatic picking of seismic arrivals in local earthquake data using an artificial neural network. Geophys J Int 120(3):758–774

Durka P, Blinowska K (1995) Analysis of EEG transients by means of matching pursuit. Ann Biomed Eng 23(5):608–611

Earle PS, Shearer PM (1994) Characterization of global seismograms using an automatic-picking algorithm. Bull Seismol Soc Am 84(2):366–376

Jones JP, van der Baan M (2015) Adaptive STA–LTA with outlier statistics. Bull Seismol Soc Am 105(3):1606–1618

Joswig M (1990) Pattern recognition for earthquake detection. Bull Seismol Soc Am 80(1):170–186

Klinkenberg B (1994) A review of methods used to determine the fractal dimension of linear features. Math Geol 26(1):23–46

Korvin G (1992) Fractal models in the earth sciences. Elsevier Science Ltd

Küperkoch L, Meier T, Brüstle A, Lee J, Friederich W (2012) Automated determination of S-phase arrival times using autoregressive prediction: application to local and regional distances. Geophys J Int 188(2):687–702

Küperkoch L, Meier T, Lee J, Friederich W, Group EW et al (2010) Automated determination of P-phase arrival times at regional and local distances using higher order statistics. Geophys J Int 181(2):1159–1170

Mallat SG, Zhang Z (1993) Matching pursuits with time-frequency dictionaries. IEEE Trans Signal Process 41(12):3397–3415

Mandelbrot BB (1983) The fractal geometry of nature: W. H. Freeman and Company

Nippress S, Rietbrock A, Heath A (2010) Optimized automatic pickers: application to the ANCORP data set. Geophys J Int 181(2):911–925

Peters EE (1994) Fractal market analysis: applying chaos theory to investment and economics, vol 24. Wiley

Phillips PJ (1998) Matching pursuit filters applied to face identification. IEEE Trans Image Process 7(8):1150–1164

Rawles C, Thurber C (2015) A non-parametric method for automatic determination of P-wave and S-wave arrival times: application to local micro earthquakes. Geophys J Int 202(2):1164–1179

Rietbrock A, Ryder I, Hayes G, Haberland C, Comte D, Roecker S, Lyon-Caen H (2012) Aftershock seismicity of the 2010 Maule mw = 8.8, Chile, earthquake: Correlation between co-seismic slip models and aftershock distribution Geophys Res Lett 39(8)

Rinehart AJ, McKenna SA, Dewers TA (2016) Using wavelet covariance models for simultaneous picking of overlapping P- and S-wave arrival times in noisy single-component data. Seismological Research Letters

Ross ZE, Ben-Zion Y (2014a) An earthquake detection algorithm with pseudo-probabilities of multiple indicators. Geophys J Int 197(1):458–463

Ross ZE, Ben-Zion Y (2014b) Automatic picking of direct P, S seismic phases and fault zone head waves. Geophys J Int 199(1):368–381

Sabbione JI, Velis D (2010) Automatic first-breaks picking: new strategies and algorithms. Geophysics 75(4):V67–V76

Sugihara G, May RM (1990) Applications of fractals in ecology. Trends Ecol Evol 5(3):79–86

Turcotte DL (1989) Fractals in geology and geophysics Fractals in geophysics. Springer, pp 171–196

Turcotte DL (1997) Fractals and chaos in geology and geophysics

Vincent P, Bengio Y (2002) Kernel matching pursuit. Mach Learn 48(1-3):165–187

Wang Y (2006) Seismic time-frequency spectral decomposition by matching pursuit. Geophysics 72(1):V13–V20