GEOPHYSICS, VOL. 75, NO. 4 (JULY-AUGUST 2010); P. V67–V76, 8 FIGS., 3 TABLES.
10.1190/1.3463703

Automatic first-breaks picking: New strategies and algorithms

Juan I. Sabbione[1] and Danilo Velis[1]

ABSTRACT

We have developed three methods for the automatic picking of first breaks that can be used for marine, dynamite, or vibroseis shot records: a modified Coppens's method, an entropy-based method, and a variogram fractal-dimension method. The techniques are based on the fact that the transition between noise and noise plus signal can be automatically identified by detecting rapid changes in a certain attribute (energy ratio, entropy, or fractal dimension), which we calculate within moving windows along the seismic trace. The application of appropriate edge-preserving smoothing operators to enhance these transitions allowed us to develop an automated strategy that can be used to easily signal the precise location of the first-arrival onset. Furthermore, we propose a mispick-correcting technique to exploit the benefits of the data present in the entire shot record, which allows us to adjust the trace-by-trace picks and to discard picks associated with bad or dead traces. As a result, the consistency of the first-break picks is significantly improved. The methods are robust under noisy conditions, computationally efficient, and easy to apply. Results using dynamite and vibroseis field data show that accurate and consistent picks can be obtained in an automated manner even under the presence of correlated noise, bad traces, pulse changes, and indistinct first breaks.

INTRODUCTION

In seismic exploration, first-break picking is the task of determining, given a set of seismic traces, the onsets of the first signal arrivals as accurately as possible. In general, these arrivals are associated with the energy of refracted waves at the base of the weathering layer or to the direct wave that travels directly from the source to the receiver.

The accurate determination of the first arrivals onset (first-break times) is needed for calculating the static corrections, a fundamental stage of seismic data processing. Clearly, the effectiveness of reflection- and refraction-based methods of static corrections depends on the picking-process reliability (Yilmaz, 2001, p. 374). Also, applications such as near-surface tomographic static corrections (tomostatics) require a precise and rapid automated detection of the signal first arrivals.

Generally, first-break quality is related to the near-surface structure, source type, and signal-to-noise ratio (S/N) conditions. As a consequence, the automated picking of first breaks can be a very difficult task if data are acquired in complex near-surface scenarios or if the S/N is low. Moreover, if the source wavelet is zero-phase as when vibroseis sources are used, the sweep correlation often produces side-lobes that arrive before the first break, thus making the picking process even more difficult.

Traditionally, the determination of the signal advent was carried out by a visual inspection of the amplitudes and waveform changes ("manual" picking). Apart from being very time consuming, this strategy can lead to biased and inconsistent picks because it relies on the picking-operator subjectivity. With the development of modern computers, the use of dedicated software to carry out the picking interactively facilitated the process but in general the whole procedure is still very time consuming and subjective. Thus, the implementation of automated computer-based algorithms is very important for the rapid picking of first breaks in a consistent and objective way. Nowadays, the most common strategy is to use an automatic/semi-automatic picker as a first step and to correct the results interactively by visual inspection later. Often, this process needs to be repeated several times in certain difficult areas. As a result, when the data volume is large and the data quality is poor, the picking procedure can take up to 20–30% of the total processing time.

During the last few decades, numerous techniques have been developed for determining first breaks automatically or semi-automatically. First attempts were based on the crosscorrelation of adjacent traces to find the delay time between first breaks (Peraldi and Clement, 1972). However, these techniques tend to fail when the pulse shape changes from trace to trace and when bad or dead traces appear. Hatherly (1982) proposes some cumbersome statistical tests that would signal the advent of the first breaks. Gelchinsky and Sh-

Manuscript received by the Editor 8 July 2009; revised manuscript received 19 November 2009; published online 18 August 2010.
[1]Universidad Nacional de La Plata, La Plata, Facultad de Ciencias Astronómicas y Geofísicas, Argentina, and CONICET, Argentina. E-mail: jsabbione@fcaglp.unlp.edu.ar; velis@fcaglp.unlp.edu.ar.

tivelman (1983) present a technique based on a combination of the correlation properties of the signal and a statistical criterion. Other approaches, still included in various commercial software packages, have focused on the detection of a sudden increase in the signal energy (Coppens, 1985). As stated by Coppens (1985), this method is very robust when S/N is sufficiently high. As a first approximation to the location of the first arrivals, Spagnolini (1991) bases his adaptive picking method on the detection of abrupt changes in the energy. For convenience, the mentioned techniques will be referred to as conventional methods. Nonconventional methods include relatively new algorithms such as those based on neural networks. These approaches proved to be very useful for first-break detection (Murat and Rudman, 1992) but as it is well known an adequate training of the network often requires a considerable amount of operator time (and expertise). Without good network training, the results can require several time-consuming "manual" adjustments, especially when data quality is low. Other nonconventional methods involve the use of higher-order statistics (Yung and Ikelle, 1997; with limitations similar to that of the crosscorrelation-based approaches), fractal-dimension analysis (Boschetti et al., 1996; Jiao and Moon, 2000), and wavelet transform (Tibuleac et al., 2003). Unfortunately, when background noise is high and data quality is poor, these techniques tend to fail and the problem of obtaining consistent and reliable picks in an automated manner remains unresolved.

In this context, we propose some useful improvements for the Coppens and the fractal-dimension-based methods that help to significantly enhance their capability for detecting first arrivals automatically. We also present a new algorithm based on the entropy of curves. The entropy of curves has already been proposed for time- and spatial-series segmentation (Denis and Crémoux, 2002). Here, the trace entropy is used as a new seismic attribute that quantifies the statistical properties of the signal (variability and correlation structure).

These three algorithms rely on three seismic "attributes" (energy ratio, fractal dimension, and entropy), which we calculate within moving windows in a trace-by-trace process. The basic assumptions are that the selected attribute behaves differently whether data in the moving window contains noise only or noise plus signal and that the transition between these two regions is abrupt. Thus, the accuracy of the picking method is clearly related to the accuracy with which the attribute change can be determined. In this sense, we propose the use of an edge-preserving smoothing (EPS) filter (Luo et al., 2002) to significantly enhance this transition. As it will be shown, EPS becomes an essential step that significantly improves the results.

Rather than limiting the picking to a trace-by-trace procedure, the proposed algorithms are devised to exploit the benefits of the whole shot gather in which first breaks, as it is well known, are approximately aligned along straight lines. So we develop a five-step mispick-correcting procedure to restrict the search of the first breaks in the proximity of these lines, which are built by least-squares using the trace-by-trace picks. As a result, the consistency and spatial coherence of the picks are significantly improved. On the other hand, mispicks associated with bad or dead traces and those associated with traces that present indistinct first breaks are corrected or discarded automatically using simple statistical criteria.

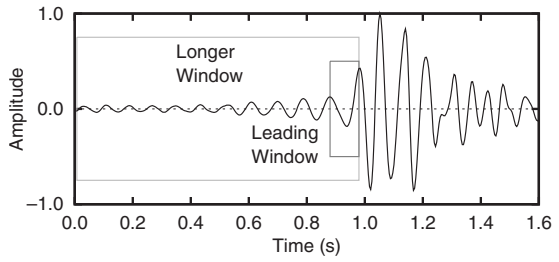

Figure 1. Windows used in the calculation of the energy-ratio attribute.

First, we illustrate the three proposed algorithms using a single trace. Then we test the complete strategies (including the mispick-correction procedure) under different S/N and source-type conditions using field-shot records. For these tests, we selected various dynamite and vibroseis noisy data sets. The results show that the proposed algorithms are in general very robust and lead to consistent and accurate picks even when using vibroseis data, in which the onsets are not so distinct. Likewise, they are very efficient in terms of computational effort and simple to use.

METHODS

In the proposed methodology, the picking procedure has two stages. The first stage is a trace-by-trace process aimed at identifying abrupt changes in a certain trace attribute (energy ratio, entropy, or fractal dimension), which we calculate as a function of time. To enhance the attribute changes, the selected attribute is filtered using EPS. Finally, we assign the first-break onset to the time when the derivative of the filtered attribute is at a maximum (or minimum in the case of the fractal-dimension method). In this section, we describe, illustrate, and compare the three proposed algorithms using the same single field-data trace. The selected trace is trace number 14 from the vibroseis-source shot-gather number 4 in Yilmaz (2001, p. 71), which was acquired with a sampling interval of 4 ms. For convenience, seismic amplitudes are normalized to $(-1,1)$ before processing.

The second stage involves the use of the complete set of picks of the first stage obtained for every trace in the shot gather. This procedure consists of simple statistical criteria that help improve the picking spatial coherency and automatically detect bad or dead traces. The objective of this process is to account for the expected approximate alignment of the picks along the refraction model straight lines. Mispicks can be adjusted or eliminated at this stage.

Modified Coppens's method (MCM)

The basic idea of this strategy is very simple and follows Coppens's method (Coppens, 1985). The aim of the method is to distinguish the signal from the incoherent or coherent background noise in terms of their energy difference. For this purpose, we calculate the energy of the seismic trace $s(t)$ within two nested windows

$$E_1(t) = \sum_{i=t-n_\ell+1}^{t} s_i^2, \tag{1}$$

$$E_2(t) = \sum_{i=1}^{t} s_i^2, \tag{2}$$

where n_ℓ is the length of the first (leading) window, a parameter that is fixed and selected a priori. Note that on the contrary the length of the second (longer) window increases with time (see Figure 1). Then

we calculate the energy ratio

$$ER(t) = E_1(t)/(E_2(t) + \beta), \tag{3}$$

where β is a stabilization constant that helps reduce the rapid fluctuations of $ER(t)$ that might lead to a false first-break pick, especially when the background noise is very low. This attribute is assigned to the last sample of the windows. The nested windows approach guarantees that $ER(t) < 1$, which is very convenient to control the magnitude of this attribute.

The energy-ratio attribute (or the entropy and fractal-dimension attributes described in next sections) exhibits a transition between noise and noise plus signal. Even so, it is a difficult task to determine when the seismic signal arrives because of the attribute variability and because often the transition is not sufficiently abrupt. To mitigate this problem, we propose to filter $ER(t)$ by applying an EPS operator (Luo et al., 2002). Edge-preserving smoothing can be understood as a simple modification of the running-average smoothing method and was devised to reduce noise while preserving the most noticeable changes of the data (EPS is described in Appendix A). Finally, we assign the first-break onset to the sample in which the derivative of the filtered attribute is largest. This strategy allows us to minimize the delay usually found between the attribute maximum and the actual first-break onset, improving the accuracy of the results. We will refer to this method as the "modified Coppens's method" (MCM).

From a practical point of view, the MCM requires the user to set three parameters: the leading window length n_ℓ, the EPS operator length n_e, and the stabilization constant β. Because the leading window is supposed to capture the first-arrival energy, we set its length equal to one period of the first-arrival waveform. In this work, the period is easily determined by measuring the time-distance between two wave crests or troughs right after the first-break onset. On the other hand, we found good results using EPS lengths between one and two signal periods so we fix n_e to one and a half periods in all cases. The selection of β provides a means to control the attribute sensitivity to energy changes avoiding false picks. Based on some experimental tuning, we found that the selection of β is not critical to the method and we decided to fix it at 0.2 (recall that the amplitudes are previously normalized to $(-1,1)$).

Despite being old, Coppens's method (CM) is very robust when the S/N is high. Coppens's method and its variations are usually included in various processing software. If the first break is impulsive (such as in dynamite or marine data), the performance of the CM is very good. However, if the first arrival is not very abrupt (such as in vibroseis data), usually $ER(t)$ attains its maximum at a later time, limiting the capabilities of the CM to detect the exact location of the first break. To illustrate the MCM and to compare it with the CM, we tested both methods using the selected trace shown in Figure 2. By visual inspection, first arrivals have a period of approximately 80 ms (20 samples) so we set $n_\ell = 20$ and $n_e = 30$. Note how the first-break determination is improved by the use of EPS. If the CM had been used without the proposed modifications, the first break would have been picked approximately 50 ms later.

Entropy method (EM)

We propose a new picking method called the entropy method (EM), which is based on the entropy of curves, a concept used by Denis and Crémoux (2002) in the context of the segmentation of time or spatial series. In their work, the entropy of a curve is viewed as a measure of the variability and correlation structure of a time series. This permits detecting changes in the statistical properties of the signal and thus dividing it into local stationary segments. In this work, we apply the same concepts to the picking of first arrivals because a rapid change in the statistical properties of the seismic trace is expected when the first break arrives.

Denis and Crémoux (2002) compute the entropy of a time series as a function of time (or space) by means of

$$H(t) = \log(L(t)/t), \tag{4}$$

where $L(t)$ is the "length" of the time series and is approximated by the sum of absolute values of first differences.

In computing entropy as an attribute of the seismic trace $s(t)$, we estimate $H(t)$ within a moving window of fixed length n_h and assign its value to the last point of the window. Therefore,

$$H(t) = \log\left(\frac{1}{n_h}\sum_{i=t-n_h+1}^{t-1}|s_{i+1} - s_i|\right). \tag{5}$$

In general, the entropy varies significantly when the moving window encompasses noise only or signal plus noise and thus the advent of a first break can be identified by detecting rapid changes in this attribute.

It is worth mentioning that if the moving window is too short the variability of the entropy would be very large. On the other hand, if it is too long, time resolution would be diminished. Besides, to completely capture the statistical properties of the first arrival it seems reasonable to select n_h as a multiple of the first-arrival period. In all the tested seismic traces we found good results setting n_h equal to twice the main period of the signal.

Figure 3 illustrates the behavior of the EM when applied to the same trace used in the MCM. Because the first-arrival period is ap-

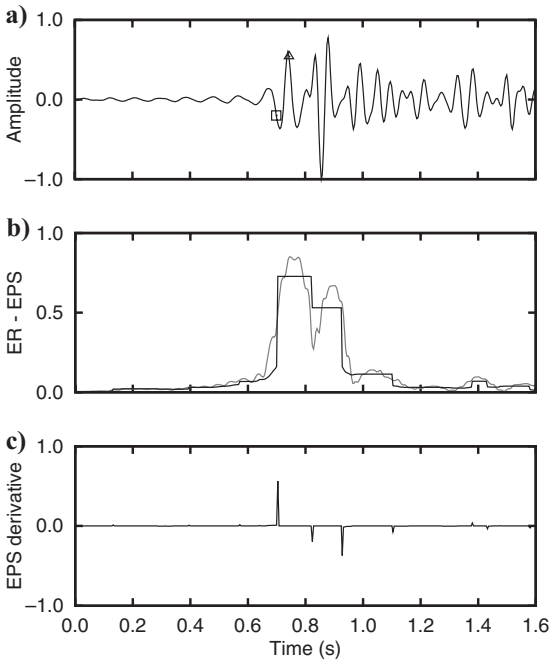

Figure 2. (a) Seismic trace and first break picked by the CM (triangle) and the MCM (square); (b) raw $ER(t)$ attribute (dashed) and filtered attribute (solid); (c) derivative of the filtered attribute. Its maximum signals the first-arrival onset.

proximately 20 samples, we set $n_h = 40$. We can observe that the entropy attribute increases significantly at approximately 0.7 s. Again, this change is then enhanced after the edge-preserving filtering (Figure 3b). The EPS operator length was fixed to one and a half periods (i.e., 30 samples), as in the MCM case. The maximum of the derivative of the filtered attribute clearly signals the onset of the first arrival, which turns out to be very similar to the pick obtained using the MCM (see Figure 2).

Fractal-dimension method (FDM)

A fractal is by definition a set in a metric space for which the Hausdorff-Besicovitch dimension strictly exceeds the topological dimension (Mandelbrot, 1983). The Hausdorff-Besicovitch dimension generalizes the topological notion of the set dimension (a natural number) to nonnegative real values. In the case of a curve on the plane, whereas topological curves are one dimensional, a fractal curve has a fractal dimension D that is in the range $1 \le D \le 2$. It can be said that the fractal dimension quantifies the degree of complexity of a fractal curve. The theoretical basis of the fractal theory can be found in Mandelbrot (1983), Feder (1988), and Peitgen et al. (1992). The use of fractals in geophysics is fully described in Turcotte (1997) and Korvin (1992).

Fractal curves can be classified as self-similar or self-affine. According to Turcotte (1997), a formal definition of a self-similar fractal in a 2D xy-space is that $f(rx,ry)$ is statistically similar to $f(x,y)$, where r is a scaling factor. On the other hand, a formal definition of a self-affine fractal is that $f(rx,r^{H_a}y)$ is statistically similar to $f(x,y)$, where H_a is known as the Hausdorff measure. The definition of the fractal dimension of a self-affine fractal is given by

$$H_a = 2 - D. \tag{6}$$

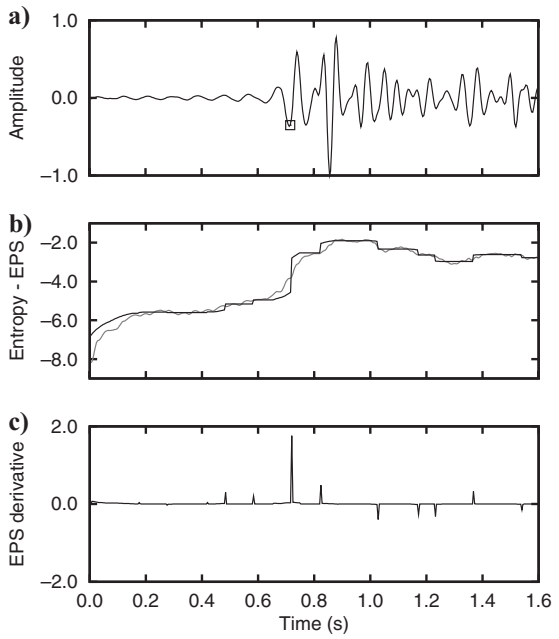

Figure 3. (a) Seismic trace and first break picked by the EM (square); (b) raw entropy attribute (dashed) and filtered attribute (solid); (c) filtered attribute derivative. Its maximum signals the first-arrival onset.

For time series, an equivalent definition of a self-affine fractal is the requirement that the variogram $V(h)$ (or the semivariogram, which is half the variogram) scales so that (Turcotte, 1997)

$$V(h) \sim h^{2H_a}, \tag{7}$$

where $V(h)$ is defined as the expected value of the squared differences of the samples of the time series $s(t)$ that are separated by a lag distance h. In agreement with other authors (e.g., Tosi et al., 1999; Jiao and Moon, 2000), we found that seismic traces satisfy this requirement and can be classified as self-affine fractals.

There are different methods to extract the fractal dimension of a curve, depending on whether the fractal is self-similar or self-affine (Klinkenberg, 1994). Boschetti et al. (1996) develop a fractal-based analysis for detecting first arrivals of seismic traces using the "divider method," which is devised for self-similar curves. Conversely, we strongly recommend using methods developed for self-affine curves when processing seismic traces (Sabbione and Velis, 2008). Likewise, Jiao and Moon (2000) use the "variance method" to detect refraction signals and Tosi et al. (1999) use a variogram analysis to extract the fractal dimension of seismograms for the detection of seismological events. These two techniques are devised for self-affine fractals.

The "variogram method" (Korvin, 1992) is one of the most suitable and robust methods for calculating the fractal dimension of self-affine curves. The method is based on the power law between the variogram and the lag distance h, which — combining equation 6 with equation 7 — can be written as follows:

$$V(h) \sim h^{4-2D}. \tag{8}$$

In practice, $V(h)$ is calculated for different lag distances h and then plotted on a log-log plot (known as Mandelbrot-Richardson plot). Equation 8 states that the fractal dimension D is given by the slope b of the straight line defined in the Mandelbrot-Richardson plot

$$D = 2 - b/2. \tag{9}$$

We developed a fractal dimension method (FDM) for first-break picking based on the fact that random noise exhibits a higher fractal dimension than the signal. White noise fractal dimension is two whereas the fractal dimension of a correlated theoretic signal is one. Thus, the onset of a first arrival can be determined by detecting the fractal-dimension transition between noise and noise plus signal.

In the method, the fractal dimension is estimated within a sample-by-sample moving window of length n_f and its value is assigned to the last sample of the window. Experience shows that if n_f is too small the variogram (and the fractal dimension) cannot be estimated adequately. On the other hand, if n_f is too large rapid changes in the fractal dimension would not be detected properly. Jiao and Moon (2000) analyze different window lengths and recommend using a window of 48 samples. Considering this information and taking into account the signal period T as in the MCM and the EM, we fixed the window size to $n_f = kT$, where k is the lowest integer that yields $n_f \ge 48 + T/2$.

To complete the description of the FDM, let us analyze the variogram fractal-dimension behavior when low-energy random noise is added to the data. This point is shown in Figure 4, in which we used the same trace as in the MCM and the EM. Figure 4a shows the case without noise added whereas Figure 4b shows the case with low-energy white noise added. Clearly, the fractal dimension before the first arrival increased after adding the random noise, approaching the

white-noise theoretical value of two. On the other hand, the effect after the first arrival was much smaller. As a result, the transition between noise and noise plus signal became more apparent, thus making the first-break detection easier. This effect, which was also noted by Jiao and Moon (2000), can be explained by the fact that the added noise (with energy similar to that of the background noise) tends to destroy the correlation structure of the (correlated) noise present in the trace before the first arrival. Therefore, we systematically add low-amplitude white noise to the data when using the FDM. The amount of added noise must be manually tuned for the given data because it depends on the background noise of the seismic survey. This simple strategy significantly improves the capabilities of the FDM for picking first breaks particularly when processing vibroseis data, in which high-correlated noise is induced before the first arrival by the source.

The FDM can be summarized as follows: We consider a sample-by-sample moving window within the seismic trace $s(t)$ (with noise added) and estimate its variogram $V(h,t)$ using

$$V(h,t) = \frac{1}{n_f - h} \sum_{i=t-n_f+1}^{t-h} (s_{i+h} - s_i)^2, \qquad (10)$$

for four lag distances $h = 1, 2, 3$, and 4 samples, which are adequate to capture the roughness of a noisy seismic trace. This lag distance range is in agreement with Jiao and Moon (2000). After fitting a straight line to the log-log Mandelbrot-Richardson plot, we estimate the fractal dimension using equation 9 and assign its value to the last point of the window. We subsequently use EPS to enhance the fractal-dimension transition between noise and noise plus signal. As in the case of the MCM and the EM, we set the filter length equal to one and a half periods. Finally, because the fractal dimension is supposed to decrease when the signal arrives, we assign the first arrival onset to the sample in which the derivative of the filtered attribute is at a minimum. The FDM is illustrated in Figure 5 using the same trace used in the MCM and the EM. In this case, we added white noise so that $S/N = 20$, where S/N is calculated as the ratio between the trace energy and the noise energy. Because the period of the first-arrival is approximately 20 samples, we set $n_f = 60$ and $n_e = 30$.

At this point, we would like to stress the differences between the FDM proposed in this paper and other fractal-dimension approaches. Boschetti et al. (1996) use the divider method to estimate the fractal dimension, which yields smooth transitions between noise and noise plus signal. They use a complicated three-segment scheme to fit the fractal-dimension curve in the proximity of the first arrival (this region is selected manually). Then they place the first break a few samples before the intersection of two of these lines. On the contrary, the use of the variance FDM leads to sharp transitions (Jiao and Moon, 2000), which makes the picking of first arrivals easier. Unfortunately, the authors do not explain how they determine the onset of the first break. Presumably, it is placed at the point where the transition begins (though this point is not easily identified), as observed in Figure 12 of their article. In contrast, the use of the variogram FDM together with the edge-preserving filtering allows us to easily signal the onset of the first arrivals in an automated way.

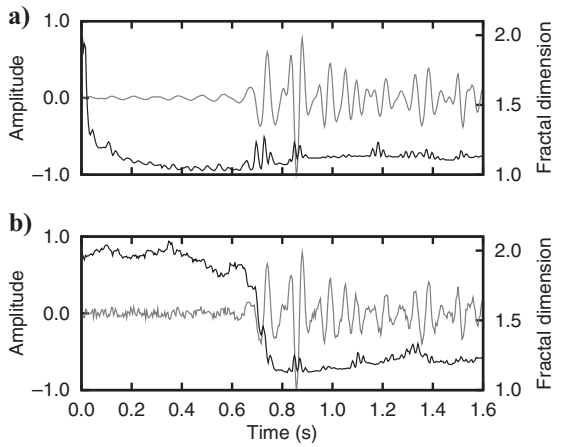

Figure 4. Seismic trace (dashed) and variogram fractal-dimension (solid). In (a) the seismic trace was not contaminated with noise whereas in (b) low-amplitude white noise was added ($S/N = 20$).

Parameter selection

In the previous section, we recommended a set of parameters for each picking method. Table 1 summarizes these quantities together with the actual values used for the automatic first-break picking illustrated in the next section. Note that most of the parameters are defined in terms of the approximate main period T of the first breaks, which is obtained directly from the data (a single approximate value of T is required for the whole shot). Thus, the selection of these parameters is straightforward. Note also that the EPS window length is the same for the three methods and the stabilization constant β used in the MCM is the same for all data sets (provided all traces are normalized as described before). In the particular case of the moving window length of the FDM, the integer k must be selected to guarantee that kT is at least $48 + T/2$ samples. In the end, the S/N is the only parameter that must be tuned for the FDM. In general, the smaller the

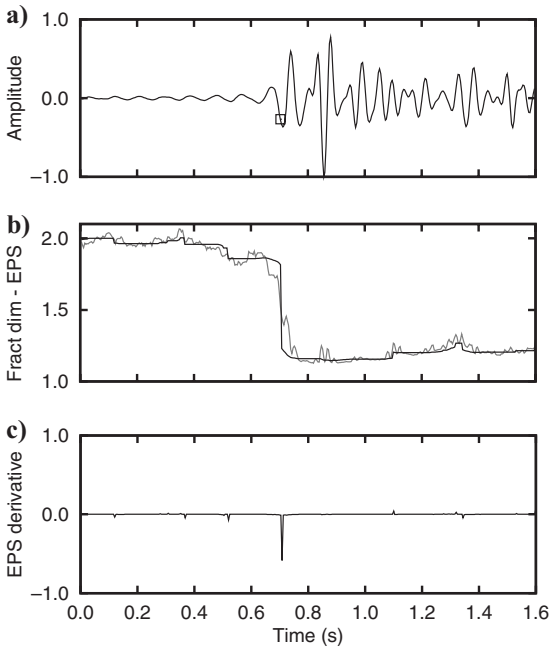

Figure 5. (a) Seismic trace and first break picked by the FDM (square); (b) raw fractal-dimension attribute (dashed) and filtered attribute (solid); (c) derivative of the filtered attribute. Its minimum signals the first-arrival onset.

S/N is, the more delayed the pick. This behavior facilitates the tuning process (usually carried out by visual inspection) required to select this parameter.

Correcting mispicks and final algorithm

The algorithms described in the previous sections search for the first arrival on a single trace but a human first-break picker would take into consideration the entire shot gather to select the picks. In doing so, he or she would certainly pay attention to common-sense factors such as constant time delays between adjacent traces, time increase with offset distance, first-break alignment along straight lines, etc. Any pick not meeting these criteria would need to be re-evaluated. These common-sense factors can be included into an automated picking algorithm by using constraints. For this purpose, we developed a mispick-correcting procedure to restrict the search of the first breaks in the proximity of these lines, which are built by least-squares using the trace-by-trace picks. As a result, mispicks can be corrected or discarded (e.g., picks associated with bad or dead traces) using simple criteria.

The mispick correction is implemented via a five-step process. The first step consists of a trace-by-trace picking using one of the three algorithms described in the previous sections. Depending on data quality, some of the automatic picks of this first step might be wrong (mispicks). For example, in a few traces the actual first break could exhibit the second or third largest value (or smallest in the FDM) of the filtered attribute derivative instead of the absolute maximum (or the absolute minimum in the FDM). Besides, there might be some bad traces whose associated picks are naturally wrong and must be rejected. To avoid these problems automatically, we fit all possible models of two straight lines per gather flank to the current picks via least-squares regression. This refracted model is devised to accommodate the expected direct and refracted arrivals. Then the χ^2 goodness-of-fit is evaluated for every model and the one with the lowest χ^2 is selected as the most probable. Next, if the a posteriori error of any point (pick) of the least-squares fit is greater than 3σ (where σ is the standard deviation of the fit) the pick is rejected and the straight lines are recalculated accordingly. This process is repeated until there are no points (picks) with errors greater than 3σ (in general, one to two iterations are enough and only a few step 1 picks are temporarily rejected). The final result of this second step is a (preliminary) refraction model consisting of two straight lines per gather flank. As an example, Figure 6a shows the results of the picking using the FDM (step 1) and the preliminary refraction model (step 2) for the field data shown in Figure 7b (next section). Note that in this particular case the picks associated with traces 4, 12, 13, 14, 27, and 37 were not considered in the preliminary refraction model because they did not pass the 3σ test.

In the third step, the picking process is repeated but the analysis is now restricted to a tolerance window of size n_{tol} centered at the preliminary refraction-model straight lines obtained in step 2. We require that $n_{\text{tol}}/4$ be greater than the largest static delay that one would expect on the data, as explained below when describing step 5. The aim of this local repicking is to reevaluate those traces that might have been temporarily discarded during step 2 (because of the 3σ test) and whose actual first arrivals are not associated with the global maximum (or minimum for the FDM) of the derivative of the filtered attribute but to a local maximum (or minimum) within the window n_{tol}. Thus, some picks that might have been interpreted as mispicks in step 2 are re-evaluated and corrected during step 3 and a new set of picks is obtained.

The next step (step 4) consists of the same procedure devised for step 2 but now using the new set of picks derived after step 3. As a result, a final refraction model is obtained. Figure 6b shows the results after steps 3 and 4. The picks associated with traces 4, 14, and 37 were corrected and included on the final refraction model. The pick of trace 27 was not corrected but passed the 3σ test for this model.

Finally, in the fifth and last step we analyze the picks trace-by-trace and decide whether to readjust the pick accordingly to the final refraction model or to reject it. The rejected picks are associated with bad or dead traces or with traces for which the selected method (i.e., the MCM, the EM, or the FDM) was not able to detect the actual first arrival. In this sense, with the assumption that the first breaks will approximately follow the straight lines obtained on step 4, we define a new tolerance window of half the previous tolerance window-length centered at the final refraction-model straight lines. Then, the local maximum (or minimum for the FDM) within this narrower tolerance window is picked. Note that the picking is carried out even if the first arrival is not very clear, thus simulating the way a human picker would proceed by following the straight lines of a hypothetical refraction model. On the other hand, because the filtered attribute is constant for those traces in which the attribute exhibits no abrupt

Table 1. The three field data sets used to illustrate the picking process and the selected parameters for each method. The sampling interval Δt is given in milliseconds. The other quantities except β and S/N are given in samples.

Method	Parameter	Equal to	Dataset1 $\Delta t = 4$ $T = 13$	Dataset2 $\Delta t = 2$ $T = 32$	Dataset3 $\Delta t = 4$ $T = 20$
	n_l	T	13	32	20
MCM	n_e	$1.5T$	20	48	30
	β	0.2	0.2	0.2	0.2
EM	n_h	$2T$	26	64	40
	n_e	$1.5T$	20	48	30
	n_f	kT	65	64	60
FDM	n_e	$1.5T$	20	48	30
	S/N	(tune)	50	70	20

changes within the short tolerance window, there is no local maximum or minimum to select and thus those traces are rejected.

In other words, the final first breaks are picked within the interval

$$|t - t_r| < n_{\text{tol}}/4, \tag{11}$$

where t_r is the time associated with the straight lines of the final refraction model obtained in step 4. Thus, n_{tol} must be selected to guarantee that $n_{\text{tol}}/4$ is greater than the largest static correction expected on the data. Note that in the final results shown in Figure 6b, picks of traces 12 and 13 were associated with bad or dead traces and rejected.

The proposed final five-step picking algorithm can then be summarized as follows:

Step 1. For every trace on a shot gather, the corresponding attribute time series (energy ratio, entropy, or variogram fractal-dimension) is calculated. Then the selected attribute is filtered using EPS and the preliminary first-break picks are set at the maximum of the filtered attribute derivative (the minimum in the case of the FDM).

Step 2. All picks of the previous step are fitted to the best two straight lines per flank model. Then those picks with errors larger than 3σ are temporarily rejected and a new model with two straight lines per flank model is calculated. The process is repeated until there are no picks with errors larger than 3σ. As a result, the preliminary refraction model is obtained (see Figure 6a).

Step 3. First breaks are now repicked locally by searching the maximum (or minimum for the FDM) of the filtered attribute derivative within a window of size n_{tol} centered at the straight lines of the preliminary refraction model obtained in the previous step. This process usually leads to the correction of some picks.

Step 4. Step 2 is repeated using the updated set of picks of step 3. Some mispicks that had been corrected in step 3 (and temporarily discarded in step 2) are taken into account and the final refraction model is obtained (e.g., picks of traces 4, 14, and 37; Figure 6b).

Step 5. Finally, a narrower window of half the tolerance window length is fixed and centered on the final refraction model of step 4. The picks are adjusted within this new window or rejected if no local maximum (or minimum in the FDM) is found. Seismic traces with rejected picks are interpreted as bad or dead traces or as traces for which the selected method has failed (e.g., picks of traces 12 and 13, Figure 6b).

It is worth mentioning that this five-step process, which is fully automatic, is computationally efficient because the seismic attribute is calculated only once at step 1. The simplicity of this correction procedure provides great robustness to the final algorithm. The mispick-correction stage is shot consistent. Thus, problems related to inhomogeneities that can lead to different refraction models at different regions of the same survey are of no concern. Tests with several field-data records show that this strategy is very useful when data quality is poor or difficult to pick. However, if a large number of picks after step 1 are wrong (more than 35%, approximately), a reasonable refraction model might not be obtained and the five-step mispick-correction process described above might fail.

Finally, we would like to remark that a later process can be added to the proposed methods to finely adjust the picks and to follow some predefined criteria (e.g., correcting every pick to the closest inflection or zero crossing point or moving the pick to nearest trace maximum or minimum). This later adjustment of the picks is viewed as a second-order process that in any case represents a criterion that is as arbitrary as picking the maximum (or minimum) of the derivative of the filtered attribute.

FIELD-DATA TESTS

In general, marine field data exhibit clear and distinct first breaks and thus offer no difficulties for the proposed automated picking algorithms. For this reason we do not show examples using this type of data. Instead, we illustrate the behavior of the MCM, the EM, and the FDM together with the mispick-correction stage using dynamite and vibroseis field records. We have selected three shot gathers with different background noise levels that we believe are not easy to pick because of the presence of correlated noise, bad traces, pulse changes, and not-so-distinct first breaks. The field data were taken from Yilmaz (2001), available at http://www.cwp.mines.edu/data/oz.original/. Table 1 shows the sampling interval and the signal approximate periods for each data set with the parameters we used for the different picking methods. Regarding the mispicks correction stage, we set n_{tol} equal to four periods for the three shot gathers and for the three methods (i.e., 52 samples for data set 1, 128 samples for data set 2, and 80 samples for data set 3).

The first selected data set (Figure 7a) is shot number 6 in Yilmaz (2001, p. 72). The shot was acquired in the Far East with a dynamite source. It consists of 48 traces with 100-m trace spacing. Figure 7a shows the resulting first breaks picked by the MCM (red), the EM (green), and the FDM (blue). Some traces show no clear first break and in particular traces 2, 3, 4, 27, and 28 have some problems. In some of these traces, the methods could not find an acceptable pick after step 5 and thus the traces were automatically marked as bad traces. Despite these facts, all three methods picked the first arrivals correctly in most of the traces. The dynamite source provides impulsive first breaks and although some traces are rather noisy and some first breaks are not distinct they tend to pose no major difficulties for the proposed methods.

The next shot gather (Figure 7b) is a field record from the San Joaquin Basin (shot number 23 on Yilmaz [2001, p. 76]), which was

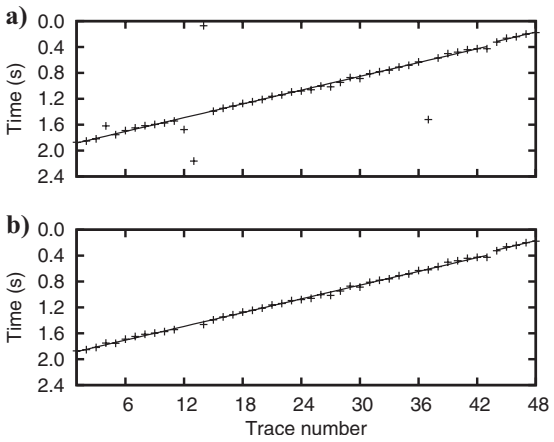

Figure 6. Mispick correction procedure. The results correspond to the FDM applied to the field data shown in Figure 7b. (a) Trace-by-trace picks (step 1) and fitted straight lines corresponding to the preliminary refraction model (step 2); (b) corrected picks (step 3) and fitted straight lines for the final refraction model (step 4). Picks of traces 12 and 13 were associated with bad or dead traces and rejected (step 5).

acquired with a vibroseis source. The 48 traces are separated by 67 m (220 ft). The shot gather and the corresponding first breaks detected by the MCM, the EM, and the FDM are plotted in Figure 7b. Some traces are problematic (namely traces 4, 12, 13, and 14). Here, some methods succeeded to pick the first breaks and some methods failed, as expected due to the presence of bad traces. Also, data exhibit lobes before the first-arrivals onset (because of the zero-phase wavelet of the source) and the first breaks are not very clear. This is a common issue with vibroseis records. The data complexity makes the methods miss the first break by approximately one period in a few traces (see traces 22, 24, 43, and 46) but the results are generally very good.

The last example and perhaps the most challenging data set shown in this paper corresponds to a vibroseis field record collected in Turkey (shot number 4 in Yilmaz [2001, p. 71]). It consists of 48 traces with a 100-m trace interval. The data and the resulting picks are shown in Figure 7c. Here, conventional automated first-break-picking algorithms are prone to fail. Recall that the S/N of the added noise for the FDM was approximately 20 to decorrelate the precursor energy of the vibroseis source. Because of the presence of quasi-monochromatic noise before the first breaks, first arrivals are very difficult to detect. Even a trained human picker would find it very difficult to decide where to locate the first breaks in some of the traces (see, for example, traces 29–33). Yilmaz (2001) suggests the presence of near-surface irregularities, which are more evident by inspecting the whole record at later times. Despite the data complexity, the results of the automated picking are generally very good. The EM tends to delay the onset of the first breaks.

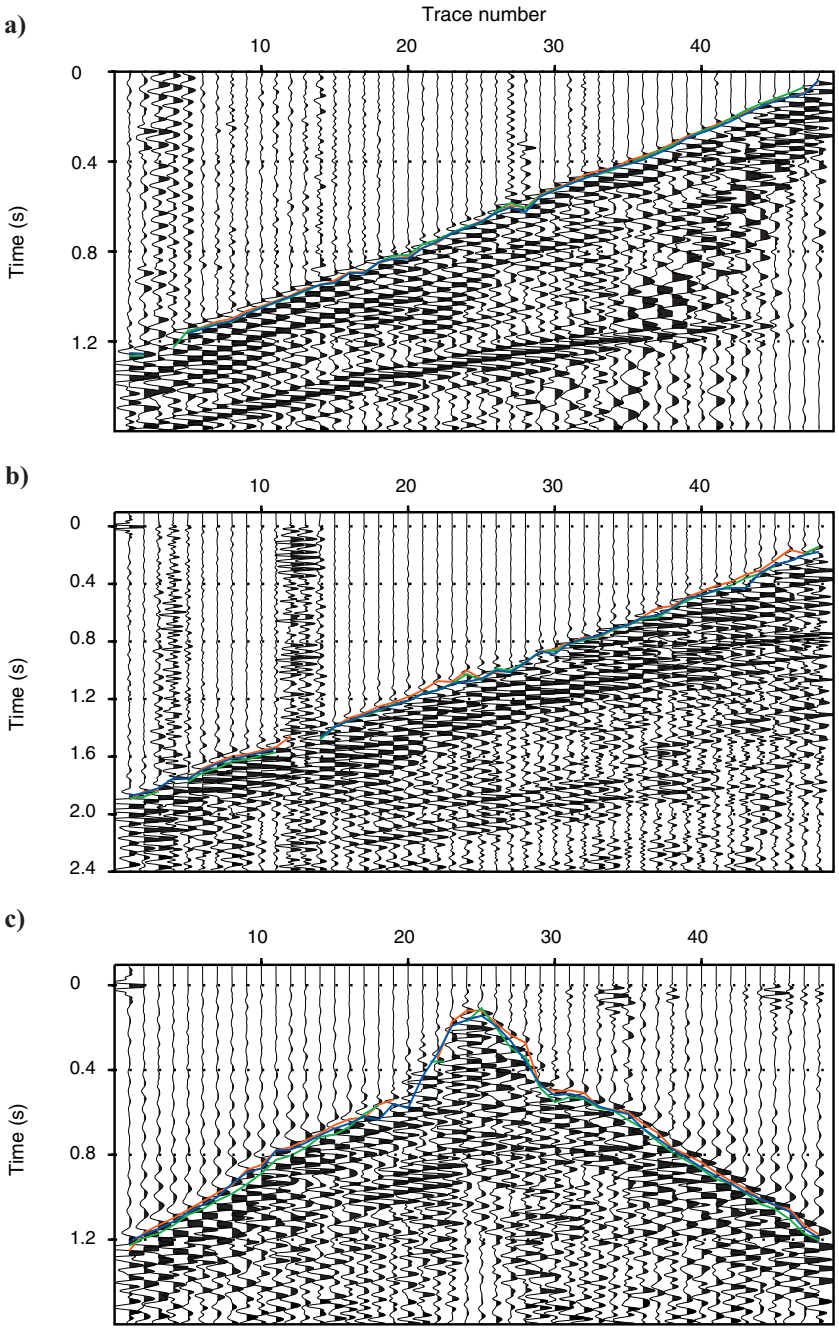

Figure 7. Field-data tests: (a) dynamite shot record number 6; (b) vibroseis shot record number 23; and (c) vibroseis shot record number 4. First breaks picked by the MCM (red), the EM (green), and the FDM (blue).

Table 2. Variation ranges for the different methods parameters for the sensitivity analysis.

Method	Parameter	Ranges
MCM	n_ℓ	$0.5T$–$1.5T$
	n_e	T–$2T$
EM	n_h	$1.5T$–$2.5T$
	n_e	T–$2T$
FDM	n_f	$(k-0.5)T$–$(k+0.5)T$
	n_e	T–$2T$

Table 3. Mean standard deviation of the picks. Values are given in milliseconds.

Method	Data set 1	Data set 2	Data set 3
MCM	3.7	7.9	8.9
EM	4.8	7.0	12.2
FDM	4.1	7.6	12.8

Sensitivity analysis

In this section, we carry out a sensitivity analysis by repeating the picking of each shot gather using a large number of different parameter sets. In this sense, for every trace in a given shot we calculated the standard deviation of the picks, in which each pick was obtained using a different set of parameters. The leading window length (for the MCM) or the moving windows length (for the EM and the FDM) and the EPS operator length were varied in a one-period range centered at the values shown in Table 1. Both β for the MCM and S/N for the FDM were fixed to the values indicated in the same table. The ranges used for the sensitivity analysis are shown in Table 2.

Finally, we calculated a mean standard deviation by averaging all the standard deviations of the picks associated with each individual trace. This quantity is viewed as an indicator of the variability of the picks for the whole shot gather. Table 3 shows the mean standard deviation for each case. Note that the variability of the picks is very small, thus validating the parameter-selection criteria. Furthermore, the parameter selection is not critical within the variation ranges analyzed in this section.

CONCLUSION

We present three new methods to automatically determine the onset of the first arrivals in either marine, dynamite, or vibroseis shot records. The methods, which are the modified Coppens's method (MCM), the entropy method (EM), and the fractal dimension method (FDM), are based on the analysis of certain trace attributes that are especially sensitive to the advent of a signal within background noise. Attributes include the energy ratio, entropy, and the fractal dimension, which are calculated along the seismic trace within moving windows and analyzed to detect abrupt changes when the signal arrives. The transition between noise and noise plus signal is significantly enhanced using an EPS filter, leading to an automatic strategy used to easily signal the exact location of the onset (maximum or minimum of the filtered-attribute derivative). Edge-preserving smoothing represents a very useful tool for tackling the first-break–picking problem. Furthermore, we propose a mispick-correcting procedure that allows us to exploit the benefits of the data present in the entire shot record, to adjust the trace-by-trace picks, and to discard picks associated with bad or dead traces. As a result, the accuracy and consistency of the first-break picks are significantly improved.

The proposed methods are robust for noisy data and provide accurate and consistent picks even under the presence of correlated noise, bad or dead traces, pulse changes, and indistinct first breaks. Besides, the methods are computationally efficient and easy to apply because the user needs to select only two or three parameters, depending on the selected attribute. Most of the parameters are set based on the period of the first-arrival waveforms, which is easily determined by visual inspection, thus their selection is straightforward. Moreover, a sensitivity analysis shows that the variability of the picks is very small when different parameter sets are used.

Results show that the performance of the MCM considerably exceeds the performance of the traditional CM, mainly due to the use of the EPS. The EM, on the other hand, is a new approach for picking first breaks that succeeds on most of the tests but tends to detect the first breaks a few samples before or after the actual arrival. The FDM yields very consistent results and overcomes the weaknesses of other fractal-based methods published in the literature.

In general, we have observed that the MCM is very effective for picking first breaks that are characterized by a strong energy arrival. On the other hand, the FDM provides excellent results when dealing with vibroseis data when other methods tend to be less effective. However, because the three methods perform very well and are computationally efficient, we recommend running the three methods and selecting the one that appears to yield the best results.

ACKNOWLEDGMENTS

We are grateful to Daniel Lorenzo (Repsol-YPF, Argentina) for sharing his data-processing experience. We deeply appreciate his advice and suggestions. In addition, we thank Ricardo Rebollo for helpful discussions on picking first breaks. This work was partially financed by Agencia Nacional de Promoción Científica y Tecnológica, Argentina (PICT 03-13376) and CONICET (PIP 04-5126).

APPENDIX A

EPS

The EPS is a statistically based filtering process devised for reducing noise while preserving the most noticeable changes (edges) in the data (Luo et al., 2002). Though we are going to describe the 1D EPS algorithm it can be generalized for two dimensions or three dimensions (AlBinHassan et al., 2006).

The EPS can be viewed as a simple modification of the running-average smoothing method. As in the running average, the only parameter of the EPS algorithm is the window length. We will show how it works by means of an example.

Let us consider a five-point EPS operator. Given any arbitrary sample of the data s_i, there are five shifted windows that include it:

Window 1: $(s_{i-4},s_{i-3},s_{i-2},s_{i-1},s_i)$
Window 2: $(s_{i-3},s_{i-2},s_{i-1},s_i,s_{i+1})$

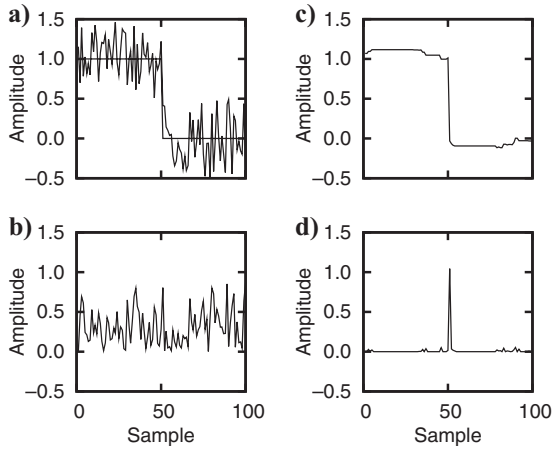

Figure A-1. (a) Step function contaminated with noise. (b) Absolute value of derivative of data in (a). (c) Data in (b) filtered using a 25-point EPS operator. (d) Absolute value of filtered data derivative.

Window 3: $(s_{i-2},s_{i-1},s_i,s_{i+1},s_{i+2})$
Window 4: $(s_{i-1},s_i,s_{i+1},s_{i+2},s_{i+3})$
Window 5: $(s_i,s_{i+1},s_{i+2},s_{i+3},s_{i+4})$

On output, sample s_i is replaced by the mean of the window with the smallest standard deviation. This process is repeated for all data samples. As a result, the filtering process assigns the mean of the most homogeneous data window around the *i*th sample to the *i*th point location. Thus, if a window contains an abrupt change, the standard deviation will be high and this window will not be selected for the EPS output. Consequently, noise is filtered by the averaging and edges are preserved.

The EPS is illustrated in Figure A-1. A step function contaminated with random noise is shown in Figure A-1a. Figure A-1b shows the absolute value of its derivative. Clearly, the step location is masked by the noise and cannot be easily identified. After the application of a 25-point EPS operator, the step location is clearly visible (Figure A-1c) and its exact location can be easily identified by observing the derivative (Figure A-1d).

REFERENCES

AlBinHassan, N. M., Y. Luo, and M. N. Al-Faraj, 2006, 3D edge-preserving smoothing and applications: Geophysics, **71**, no. 4, P5–P11.
Boschetti, F., M. Dentith, and R. List, 1996, A fractal-based algorithm for detecting first arrivals on seismic traces: Geophysics, **61**, 1095–1102.
Coppens, F., 1985, First arrivals picking on common-offset trace collections for automatic estimation of static corrections: Geophysical Prospecting, **33**, 1212–1231.
Denis, A., and F. Crémoux, 2002, Using the entropy of curves to segment a time or spatial series: Mathematical Geology, **34**, 899–914.
Feder, J., 1988, Fractals: Plenum.
Gelchinsky, B., and V. Shtivelman, 1983, Automatic picking of first arrivals and parameterization of traveltime curves: Geophysical Prospecting, **31**, 915–928.
Hatherly, P., 1982, A computer method for determining seismic first arrival times: Geophysics, **47**, 1431–1436.
Jiao, L., and W. M. Moon, 2000, Detection of seismic refraction signals using a variance fractal dimension technique: Geophysics, **65**, 286–292.
Klinkenberg, B., 1994, A review of methods used to determine the fractal dimension of linear features: Mathematical Geology, **26**, 23–46.
Korvin, G., 1992, Fractal models in the earth sciences: Elsevier.
Luo, Y., M. Marhoon, S. A. Dossary, and M. Alfaraj, 2002, Edge-preserving smoothing and applications: The Leading Edge, **21**, 136–158.
Mandelbrot, B. B., 1983, The fractal geometry of nature: W. H. Freeman and Company.
Murat, M., and A. Rudman, 1992, Automated first arrival picking: A neural network approach: Geophysical Prospecting, **40**, 587–604.
Peitgen, H. O., H. Jurgens, and D. Saupe, 1992, Fractals for the classroom, part one: Introduction to fractals and chaos: Springer-Verlag.
Peraldi, R., and A. Clement, 1972, Digital processing of refraction data: Study of first arrivals: Geophysical Prospecting, **20**, 529–548.
Sabbione, J. I., and D. R. Velis, 2008, First arrivals picking using fractal dimension analysis: VII Congreso de Exploración y Desarrollo de Hidrocarburos, La Geofísica: Integradora del conocimiento del subsuelo, 187–200.
Spagnolini, U., 1991, Adaptive picking of refracted first arrivals: Geophysical Prospecting, **39**, 293–312.
Tibuleac, I. M., E. T. Herrin, J. M. Britton, R. Shumway, and A. C. Rosca, 2003, Automatic secondary seismic phase picking using wavelet transform: Proceedings of the 25th Seismic Research Review (SRR-03), Nuclear Explosion Monitoring: Building the Knowledge Base, Session 3–04, 352–359.
Tosi, P., S. Barba, V. D. Rubeis, and F. D. Luccio, 1999, Seismic signal detection by fractal dimension analysis: Bulletin of the Seismological Society of America, **89**, 970–977.
Turcotte, D. L., 1997, Fractals and chaos in geology and geophysics, 2nd ed.: Cambridge University Press.
Yilmaz, O., 2001, Seismic data analysis: Processing, inversion, and interpretation of seismic data: SEG Investigations in Geophysics Series No. 10.
Yung, S. K., and L. T. Ikelle, 1997, An example of seismic time picking by third-order bicoherence: Geophysics, **62**, 1947–1951.