

# On the Ability of the Benford's Law to Detect Earthquakes and Discriminate Seismic Signals

by J. Díaz, J. Gallart, and M. Ruiz

*Online Material:* Tables of hypocentral parameters; figure showing examples of teleseismic detections using the BL<sub>c</sub>.

## INTRODUCTION

Benford's law (BL), also known as the first-digit law, is an intriguing pattern in data sets, first discovered more than a century ago then forgotten and rediscovered in the twentieth century. This law establishes that the frequency of occurrence of the first digit in many real-world data sets is not uniform, but instead favors the first digits. In particular, the occurrence of numbers starting with 1 and 2 is close to 30% and 18%, respectively, whereas numbers starting with 8 or 9 are close to 5%. This law was first formulated in the late nineteenth century by the Canadian–American astronomer and mathematician Simon Newcomb, who noticed that library books of logarithms were more thumbed in the first pages, those containing numbers starting with 1. He proposed a law stating that the probability of a digit  $N$  being the first digit of a number was equal to  $\log(N + 1) - \log(N)$  (Newcomb, 1881).

This law remained unnoticed for the scientific community for more than 50 years. In 1938, it was rediscovered by the engineer Frank Benford, who presented data collections from up to 20 different sources, including financial data, population of cities, or averages of the American baseball league, each of them satisfying the expected proportionality (Benford, 1938). His work became popular, and the rule finally took his name. Benford extended the rule to any arbitrary base of the logarithm and confirmed that it applies further than the leading digit.

The law is commonly expressed as

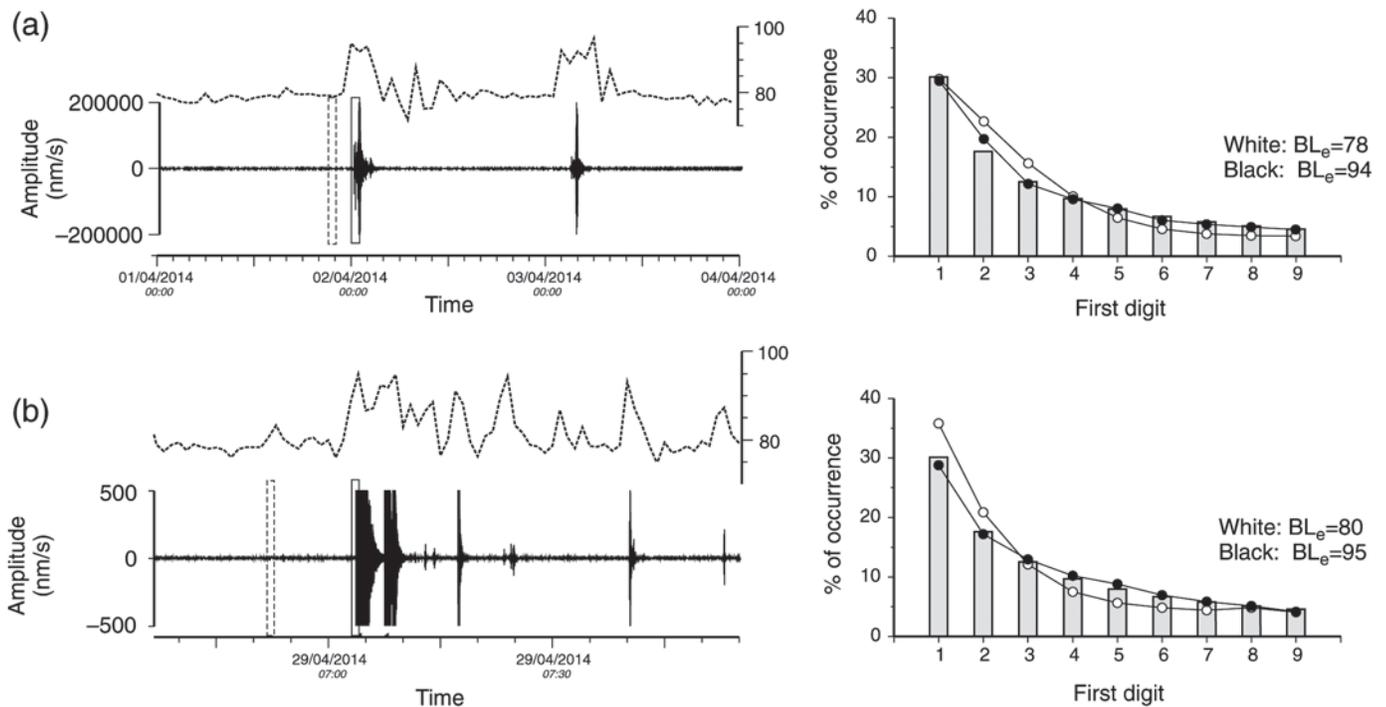
$$P(d) = \log_{10}(d + 1) - \log_{10}(d) = \log_{10}\left(1 + \frac{1}{d}\right), \quad (1)$$

in which  $P(d)$  is the probability of occurrence of the first significant digit  $d$ , considering this digit as the leftmost nonzero value and disregarding any negative sign or decimal point ( $d = 1, 2, \dots, 9$ ). This implies that if  $X$  is uniformly distributed, the variable  $Y = 10^X$  matches BL. Data from any distri-

bution will tend to conform to BL as long as the distribution spans several orders of magnitude and is reasonably smooth (Fewster, 2009). Even if the original BL applies only to data that span an integer number of base units, Sambridge *et al.* (2011) extended its validity to an arbitrary range of data for any base  $B$ . It is also relevant to note that BL is the only probability distribution that is both scale and base invariant, and therefore the conformity of a data set does not depend on its units (Pinkham, 1961).

Varian (1972) first suggested the goodness of fit to BL can be used to investigate the naturalness of certain data sets. This idea was later employed by Nigrini (1992, 1996), who proposed the use of BL in forensic accounting, for example, to validate unmanipulated data. Once it has been established that a certain data set follows BL under normal circumstances, any deviation can be attributed to data manipulation. This has been exploited with success to detect fraud in business accounts, tax returns, or stock markets, and the approach is now legally accepted as evidence in the United States legal system.

The applicability of BL to the natural sciences was explored by Nigrini and Miller (2007), who verified that hydrological streamflow statistics verify the law, while the distribution of sizes of lakes and wetlands does not conform. Sambridge *et al.* (2010) have analyzed the BL fitting of different data sets from the fields of physics, astronomy, geophysics, or chemistry, ranging from the rotation frequencies of pulsars to atmospheric temperature anomalies or the numbers of infectious diseases. Sen De and Sen (2011) have shown that BL can be used to detect cooperative quantum phenomena in many-body systems. Geyer and Martí (2012) have checked the conformity to this law of different volcanology-related data sets, such as the area and age of collapse calderas or the duration of volcanic eruptions. Moving closer to seismology, Sambridge *et al.* (2010) have shown the BL compliance of seismic wavespeeds below the southwest Pacific and of a very large database of global earthquake depths. They also showed the onset of the 2004 Boxing Day Sumatra–Andaman earthquake could be detected just by analyzing the first-digit distribution of displacement counts in the histogram of seismic waveform.



▲ **Figure 1.** Identification of teleseismic and local events using Benford's law (BL). (Left) The vertical seismic components are overlaid by the Benford's law estimator ( $BL_e$ ) values. Solid- and dashed-line boxes, respectively, show examples of time windows with and without positive detections. (Right) BL predictions for the distributions of first digits (gray bars) and the results for time windows on the seismic waveform with or without detections (black and white circles, respectively). (a) Two teleseismic events with epicenters in northern Chile. The seismic trace is unfiltered, and the time-window length used for the analysis is 1 hr. (b) A local swarm near Lourdes (southern France), at about 50 km of the recording site. The trace is band-pass filtered between 5 and 40 Hz. The amplitudes have been clipped to show minor events. The time window used in this case is 60 s.

Therefore, it now seems well established that BL is a common feature across a wide set of physical sciences, among which are the Earth sciences, even if the explanation of the law is unclear. Benford suggested it must be a broad law related to the general logarithmic character of the natural phenomena. BL can be considered either a fundamental law of nature, somehow similar to the second thermodynamics law of entropy, or just a curious empirical phenomenon. Following [Berger and Hill \(2010\)](#), it may be stated that, although BL now rests on solid mathematical ground, its ubiquity in real-life data remains mysterious.

There are few examples thus far of the application of the BL to discern between signals and background noise in continuous time series. Following the [Sambridge \*et al.\* \(2010\)](#) approach, our purpose here is to explore the use of the conformity to BL as a detection criterion applied to amplitude values of the continuous stream of seismic data and hence to allow detection of natural transient events and, in particular, to identify seismic-phase arrivals. We have developed a simple code that can be applied directly to seismic data. [Sambridge \*et al.\* \(2010\)](#) suggested large earthquakes may obey BL, whereas small local seismicity does not. As illustrated in [Figure 1](#), our first-order results show positive detections are obtained for both local and distant events once the appropriate parameters are chosen. We also present evidence that the criterion can be used in a reverse

way (i.e., looking for segments of the data set with a very poor fitting to BL) to identify specific seismic waves. The observations derived from this procedure can also contribute to a better understand of why the BL holds true.

## METHODOLOGY

As stated previously, a main goal of this contribution is to assess whether the conformity of the seismic data to BL can be used as a criterion to identify the arrival of earthquake signals. The detection of seismic arrivals is routinely performed using different methods, from simple amplitude thresholds to methods based in artificial neural network approaches, in which parameters such as variance, skewness, and kurtosis are taken into consideration ([Bormann, 2012](#)). The most popular method is probably the study of the short-time average/long-time average (STA/LTA) ratios, in which short and long averages are calculated over the data stream and detection is established if this ratio exceeds a fixed threshold. At least in this early stage, our study is not focused on determining a procedure that improves the performance of the classical detection methods, but rather to investigate the ability of BL for detection purposes.

For our test, we considered the vertical component of a Trillium 240 s sensor installed at the Canfranc Underground Laboratory (LSC; station code CANF; network code LC)

located in the central Pyrenees, which records continuously with a sampling rate of 100 samples/s. The response of the instrument has been removed, and the trace is represented in velocity values (nm/s). This station, located inside a tunnel, has a remarkably small background noise, favoring the identification of low-energy seismic arrivals. We focused here on two time periods for our study, even though additional tests were carried out to confirm the results. The first data set includes the arrivals of two large teleseismic events, whereas the second holds the records generated by a seismic swarm, the epicenters of which were located roughly 50 km northeast of the recording site.

Our procedure is developed in the Seismic Analysis Code (SAC) environment (Goldstein *et al.*, 2003), but it can be easily adapted to other formats. Signal preprocessing is simple, as it only involves the removal of possible linear trends and zero offsets, which will perturb the estimations. A classical Butterworth band-pass filter has been applied when dealing with local seismicity. The method itself starts with splitting the time series into segments of a fixed length, which may eventually overlap. For each of those segments, we calculate the frequency of occurrence of the first digits of the amplitude values and evaluate the conformity to BL by using statistical criteria. In a first step, we follow the procedure of Sambridge *et al.* (2010), who defined a Benford's law estimator ( $BL_e$ ) as

$$BL_e = 100 \times \left\{ 1 - \sqrt{\frac{\sum_{d=1}^9 [P_{\text{obs}}(d) - P_{\text{BL}}(d)]^2}{P_{\text{BL}}(d)}} \right\}, \quad (2)$$

in which  $P_{\text{obs}}(d)$  stands for the observed proportion of digit  $d$  and  $P_{\text{BL}}(d)$  is the proportion expected according to BL. Hence,  $BL_e$  values close to 100 will correspond to signals matching the BL. If seismic signals and background noise have a different degree of adjustment to BL, fixing a threshold value for the  $BL_e$  will allow identifying seismic arrivals.

The  $BL_e$  values calculated for each segment can be plotted as a time series against the seismic signal to compare their evolution. We have arbitrarily chosen to plot each  $BL_e$  value at the beginning of the sampled signal interval. There is a clear correlation in Figure 1 between the seismic data and the BL evaluator, for both large teleseismic events and small local signals.

Only two parameters are needed to implement this method, the segment length and the threshold value. This is in contrast to most of the usual methods, which typically require further parameterization; as an example, STA/LTA procedures need to fit the STA and the LTA lengths and the threshold STA/LTA ratios for triggering and detragging. The basic parameter in our approach is the length of the signal segments for which BL is estimated. This acts as a kind of band-pass filtering, with larger intervals displaying low-frequency features and short intervals enhancing the high-frequency details. Because of the statistical character of the BL, we suspect that if the signal interval is too short, the conformity to BL will fail. The threshold value of the  $BL_e$  must be fixed empirically, as the formal chi-square test is very sensitive to small deviations from

BL when large data sets are considered, making it very difficult to reach a conformity conclusion (Nigrini, 1996). Nigrini suggested that a test that ignores the number of records will be more useful for forensic purposes and proposed using the mean absolute deviation (MAD), defined as

$$MAD = \sum_{d=1}^9 \frac{|P_{\text{obs}}(d) - P_{\text{BL}}(d)|}{9}. \quad (3)$$

MAD values from 0 to 0.006 denote close conformity, values from 0.006 to 0.012 show acceptable conformity, and values in the 0.012–0.05 range denote marginally acceptable conformity to BL. In the following sections, we explore the use of these parameters and evaluators in the case of distant and local seismic events.

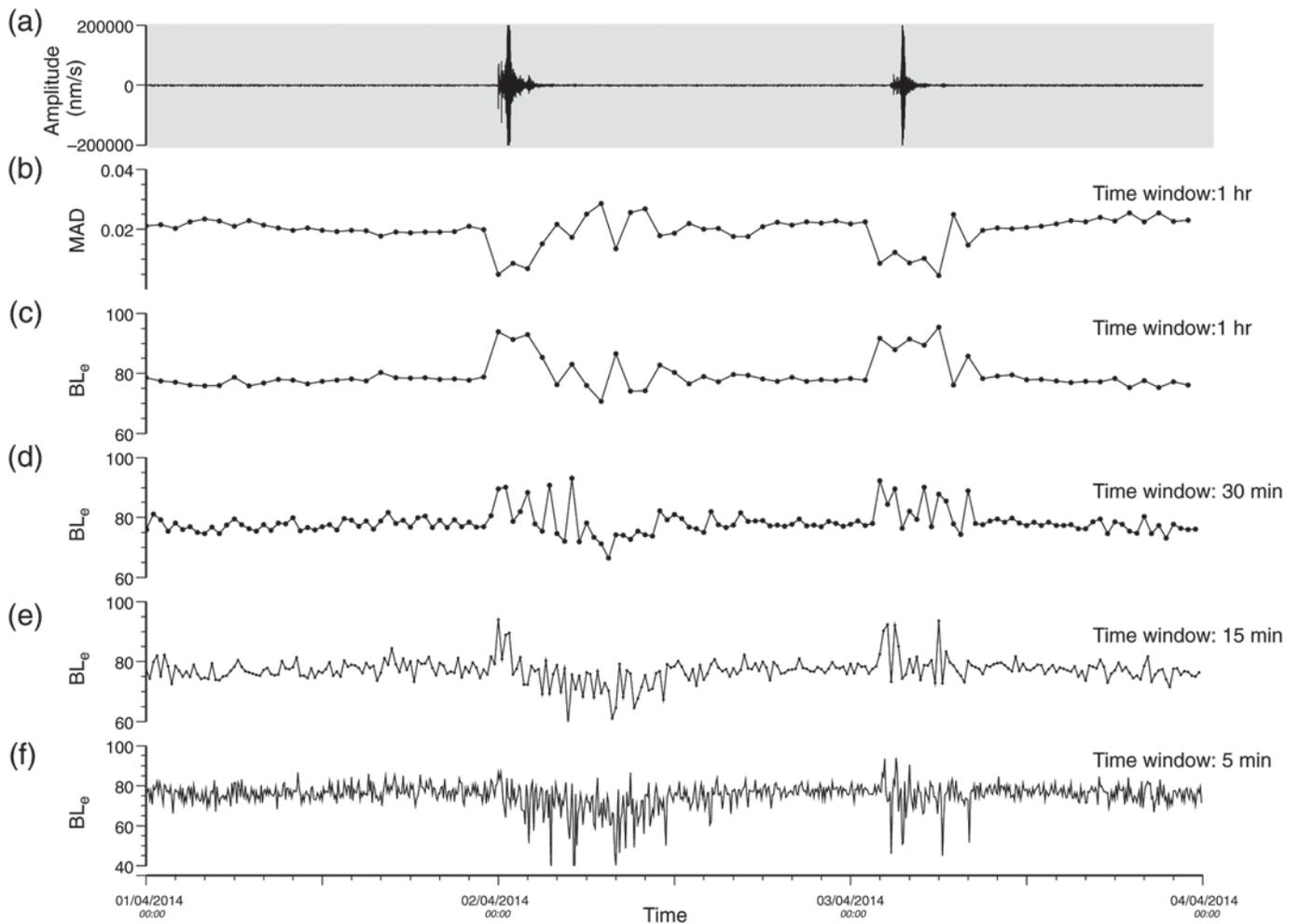
## APPLICATION TO DISTANT AND LOCAL SEISMIC EVENTS

### Teleseismic Events

To explore the ability of BL to detect the arrival of teleseismic waves, we merged three days of data starting on 01 April 2014 and ending on 04 April 2014. This includes the recording of two  $M_w$  8.2 and  $M_w$  7.7 large events with epicenters in northern Chile. Three additional events with magnitudes larger than 6 have been reported within this time period in the same epicentral zone (⊕ Table S1, available in the electronic supplement to this article).

For large teleseismic events, different seismic phases can be recorded over several hours with amplitudes clearly higher than background noise. Therefore, testing segment time lengths that range from one hour to a few minutes seems appropriate in this case. We have prospected in the range from 1 hr to 5 min. In Figure 2, length intervals of 30 and 60 min, involving 180,000 and 360,000 data samples per segment, respectively, show  $BL_e$  values around 95 during the arrival of the main seismic phases, in contrast with background values below 80. Intervals of 15 min still succeed in identifying the main teleseismic events, but the image starts to blur. Shorter time segments (5 min) still react to the teleseismic arrivals but seem clearly less appropriate. If we take the MAD estimator into consideration, values around 0.01 are reached during the seismic arrivals, thus entering into the acceptable conformity range proposed by Nigrini (1996). Figures 2b and 2c show the two estimators ( $BL_e$  and MAD) have a very close similarity, even if the definition used for  $BL_e$  results in the opposite sign for both estimators.

A general test of the method is beyond the objectives of this study, but we have verified successfully that the procedure works for seismic stations with typical levels of background seismic noise. ⊕ Figure S1 shows the application of our procedure to three events with magnitudes  $M_w$  6.0, 7.5, and 8.3, recorded at several stations of the temporary TopoIberia-Iberarray network deployed over Iberia and northern Morocco. The arrival of the  $P$  waves is clearly identified in all the cases, using time windows adapted to each event and ranging between 1 and 5 min.



▲ **Figure 2.** Effect of the length of the time window used in the calculation of the  $BL_e$  for teleseismic events: (a) the raw vertical seismic signal, (b) the mean absolute deviation (MAD) estimator using a 1 hr long window, and (c–f)  $BL_e$  results for 60, 30, 15, and 5 min. The limits between time windows are marked by dots.

### Local Seismicity

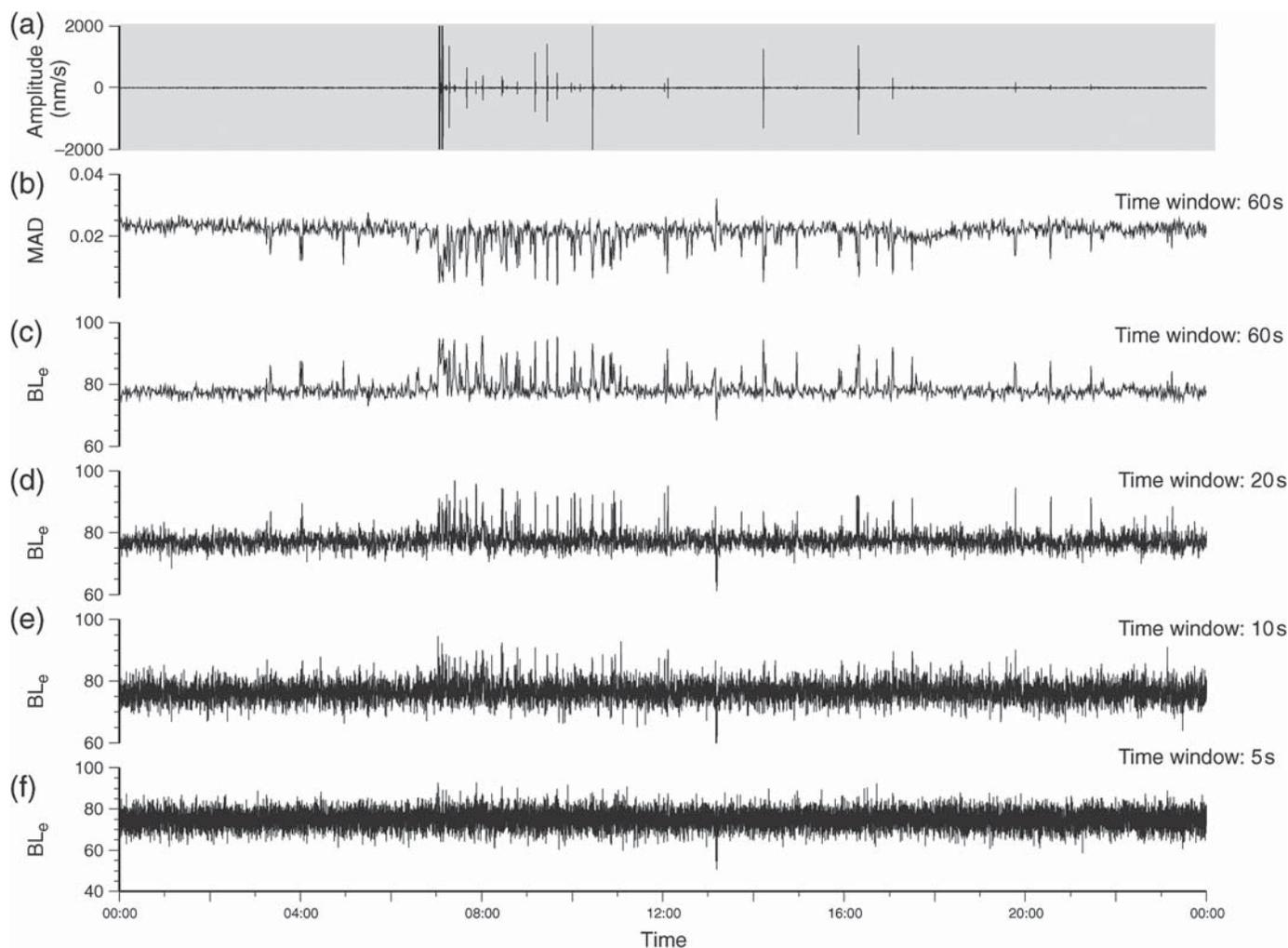
A local seismic burst occurred on 29 April 2014 in the Lourdes area of the French Pyrenees, at about 50 km of the recording site. The main event during this episode reached a local magnitude ( $M_{LV}$ ) of 4.5 and was followed 4 min later by an  $M_{LV}$  3.4 event. Up to 11 additional aftershocks are reported in the RéNaSS (French national short-period network) catalog, with magnitudes ranging between 1.4 and 3.2 (Table S2), although in our seismic record several additional events can be recognized. The small magnitude of the expected events makes it necessary to apply a band-pass filter prior to any detection analysis. We used a 5–40 Hz band-pass filter, although other values around the same frequencies are also useful. The whole seismic record of day 29 April 2014 was analyzed to explore the ability of our BL tests to detect seismic events.

The amplitude of the signals shows a large variability, extending from  $\pm 1.5 \times 10^5$  nm/s for the largest event to  $\pm 50$  nm/s for the smallest ones. Figure 3 shows the results obtained for length intervals ranging from 60 to 5 s. Using an interval of 1 min, the  $BL_e$  allows discrimination of most event

arrivals, but it does not provide a good resolution for events with close origin times.  $BL_e$  values exceeding 90 are obtained for the major events. A time interval of 20 s provides a better resolution, even if the absolute values of  $BL_e$  are slightly smaller. Intervals as short as 10 or 5 s, involving just 1000 and 500 data samples, respectively, still react to BL for the main events, but the discrimination between signal and noise becomes more difficult. As with the case of teleseismic signals, the MAD estimator has a very similar evolution. The larger events result in MAD values below 0.005, within the close conformity range proposed by Nigrini (1996). Small events present MAD values ranging between 0.008 and 0.014, still inside the acceptable conformity interval.

### Additional Skills of the Method

The  $BL_e$  not only allows the identification of large seismic events, but it is also able to discern more subtle effects. In Figure 2c, looking at the second major Chilean event, the  $BL_e$  shows high values nearly 1 h before the major event. A close inspection of the seismic waveform (enlarged view in Fig. 4a)



▲ **Figure 3.** Effect of the length of the time window used in the calculation of the  $BL_e$  for local events: (a) 5–40 Hz band-pass filtered seismic signal (clipped to display minor events), (b) MAD estimator using a 60 s window, and (c–f)  $BL_e$  results for 60, 20, 10, and 5 s. The limits between time windows are marked by dots.

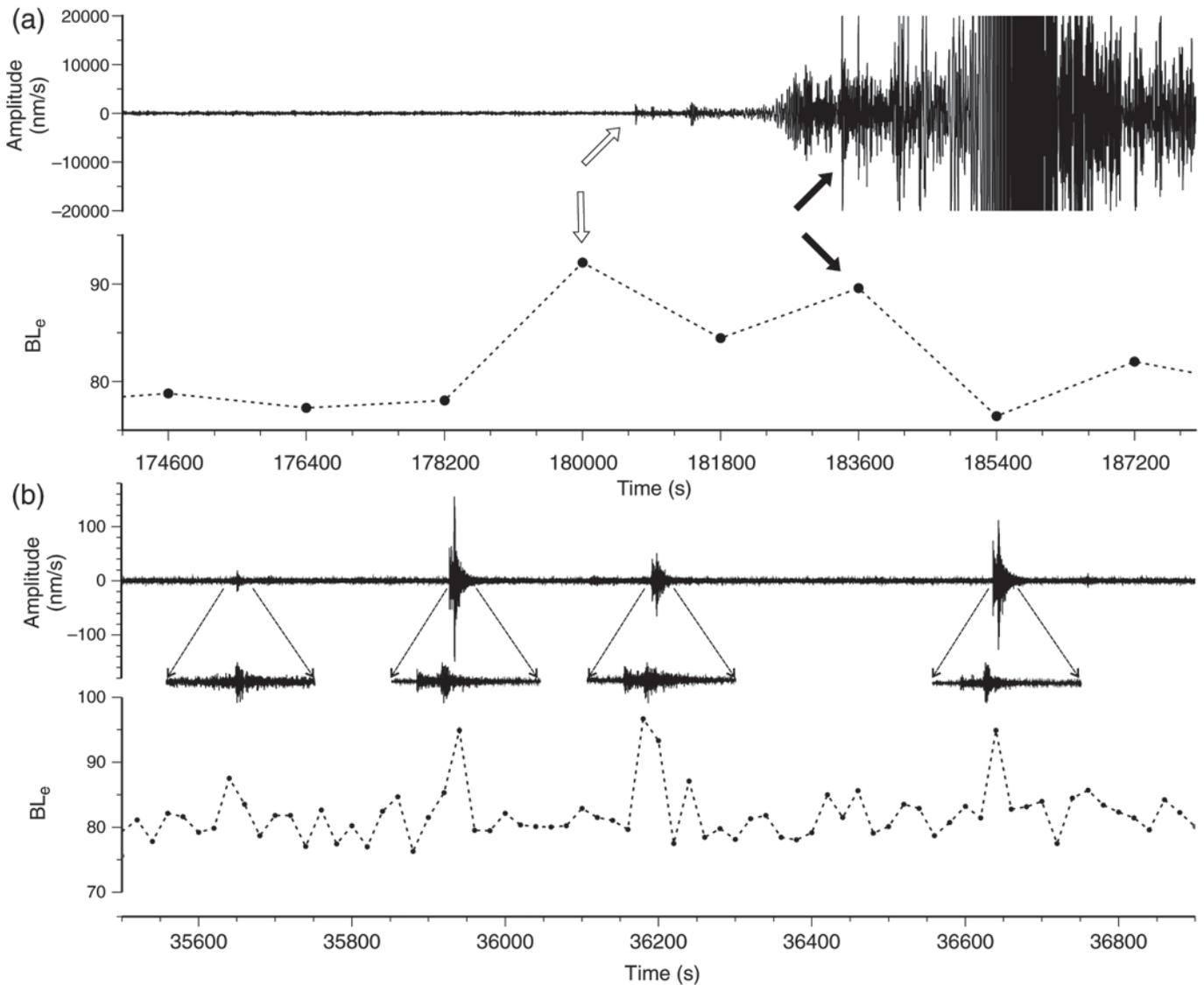
shows a seismic peak (open arrow) previous to the arrival of the large  $M_w$  7.7 event (solid arrow). This earlier peak is not spurious, but matches the arrival of the  $P$  phase of a much less energetic  $M_w$  6.5 event triggered 45 min before the major event (Ⓜ Table S1). It is important to note that even if the amplitudes of the two events differ by an order of magnitude, the  $BL_e$  reaches similar values for both events. Therefore, we can conclude the conformity to BL does not depend primarily on the amplitude of the incoming signal.

This point is confirmed by the analysis of the local seismic series. Figure 4b presents a segment of about 25 min of signal, including an  $M_{L_V}$  1.4 event and three uncataloged earthquakes. The  $BL_e$  reacts similarly for the three larger events and even successfully allowed the identification of the first one, with a signal-to-noise ratio (SNR) close to 2 and amplitude of a few tens of nanometers per second.

To check the efficiency of the BL method with respect to the classical STA/LTA algorithm, we present the comparison between both procedures at different time scales (Fig. 5).

We used 1 s for STA, 50 s for LTA, a ratio of 6 for triggering, and a ratio of 5.5 for detragging. Those values are common when dealing with local seismicity. A threshold value of 85 in the  $BL_e$  was fixed to identify time intervals matching BL and hence corresponding to seismic arrivals. Figure 5a presents nearly 14 hours of signal starting before the main event. A clear overall similarity is observed between the two methods. To further analyze them, Figure 5b shows an enlarged view, including an  $M_{L_V}$  2.0 event and up to five additional minor events. The  $BL_e$  is able to detect all the signals, whereas the STA/LTA fails to recognize the first two events. We can conclude the new method presented here has a similar performance to STA/LTA, even if a formal assessment on the relative performance of both methods would require a more extensive exploration of the parametric space, as the number of detections of each method is strongly dependent on the chosen threshold values.

Although BL clearly succeeds to identify seismic events, it must be pointed out that its ability to precisely detect the onset of the first arrivals appears more limited. As the time

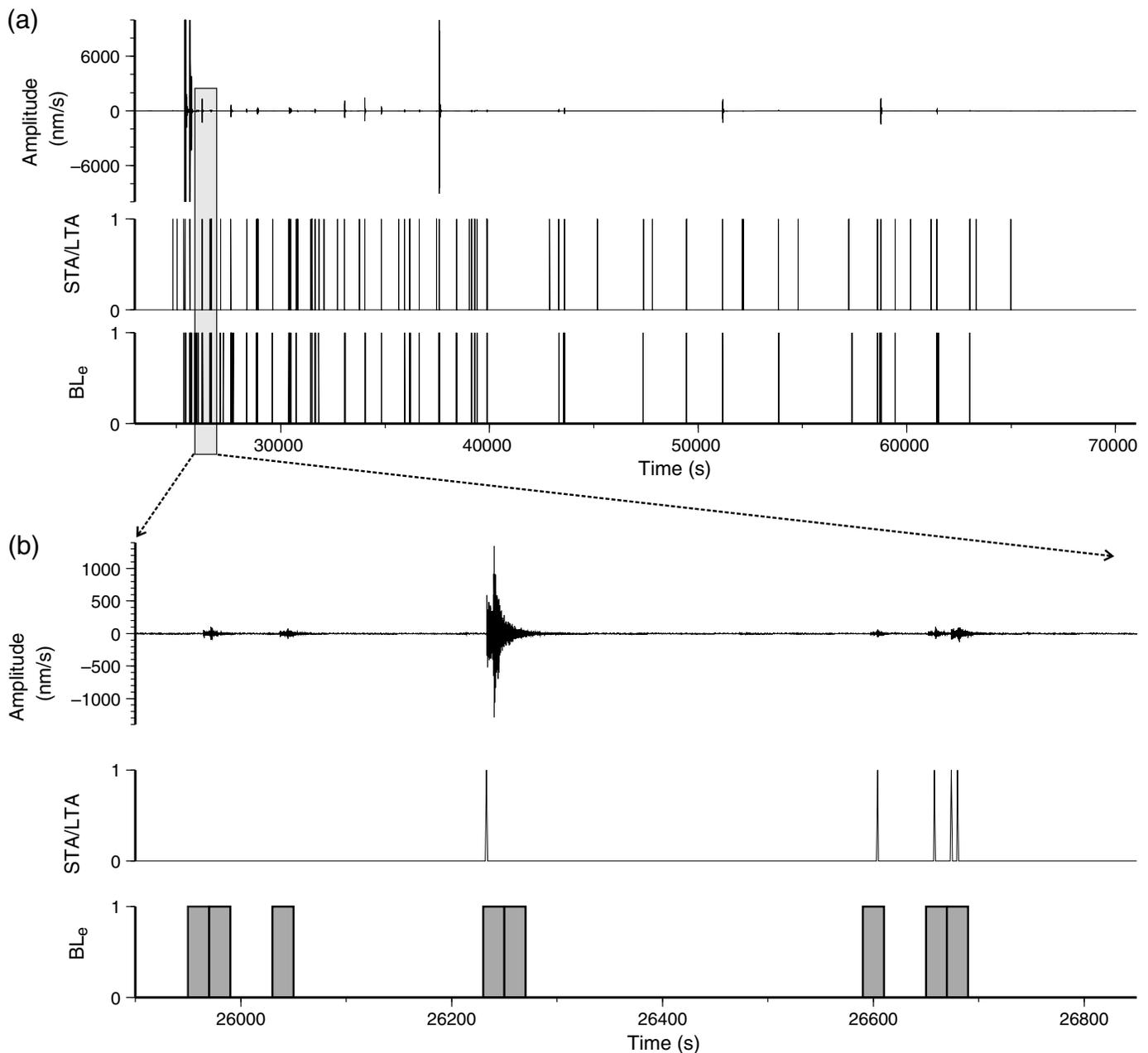


▲ **Figure 4.** (a) Seismic trace and  $BL_e$  during the arrival of two successive northern Chile events, with magnitudes 6.5 and 7.7 and separated approximately by 45 min. Open and solid arrows show the  $P$ -wave arrivals and the effect on  $BL_e$ , respectively. Dots denote the beginning of the 30-min-long time windows used in this case. (b) Seismic trace and corresponding  $BL_e$  (time windows: 20 s) for a portion of the local seismic swarm. An event arriving at 36,630 s was reported with  $M_{LV}$  1.4; the rest are uncataloged. The insets present the events expanded in time and with amplitude normalized. The  $BL_e$  values are very similar in spite of the large amplitude differences.

intervals used to calculate the conformity of the signals must be relatively large to preserve its statistical significance, the resolution on the arrival time is small. However, the accuracy of the results can be improved significantly by using overlapping intervals.

When inspecting the  $BL_e$  series calculated using short-time intervals for teleseismic events, we noticed that some low  $BL_e$  values (minima) consistently appear in correlation with the arrival of surface waves. Figure 6 shows the signal corresponding to the larger Chilean event once a 50 mHz low-pass filter is applied. This enhances the arrivals of the so-called global-circling waves (GCW), surface waves which travel

around the globe several times and can be detected at each passage over the recording site, roughly every 3 hr and 20 min. Full and dashed boxes highlight the GCW travelling along the short and long great circles between the earthquake and the recording site. It seems clear that a correlation does exist between the identified GCW and the low values in the  $BL_e$  calculated using a 2 min time window. Although the base  $BL_e$  value is around 75–80, the extrema range between 20 and 45. Therefore, this severe lack of conformity to BL can provide a new tool to identify those particular seismic phases. This behavior can provide new clues to understand the relationship between BL and seismic signals, as will be developed in the next section.



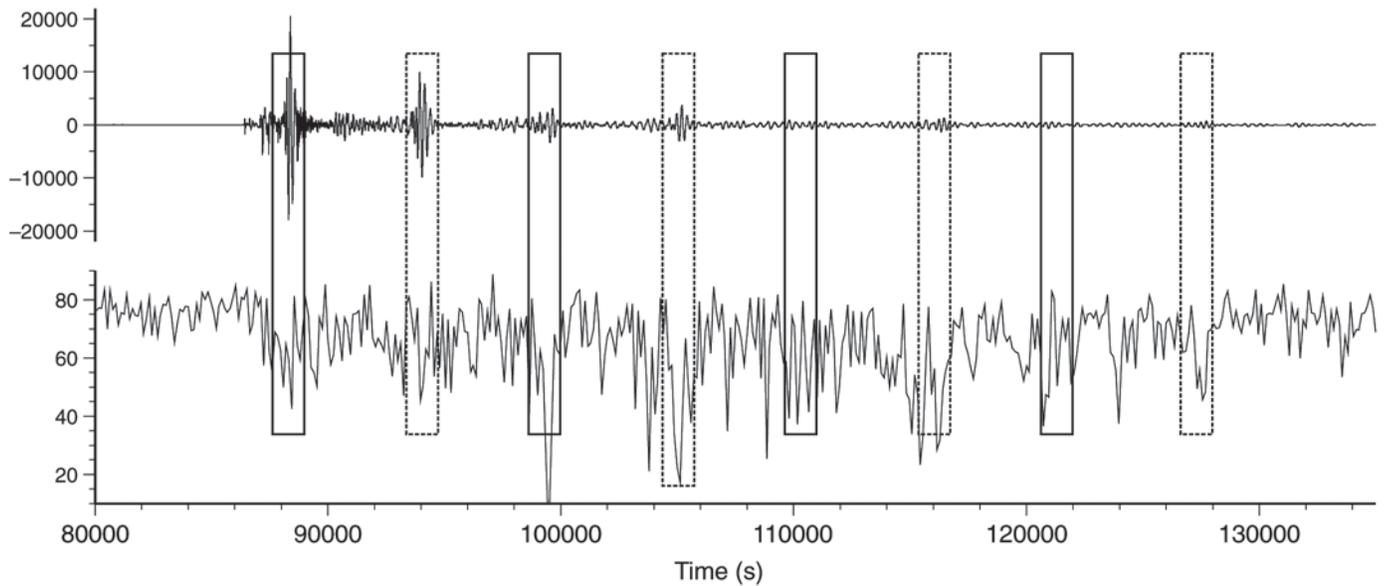
▲ **Figure 5.** Detections during the local seismic swarm using, respectively, a short-time average/long-time average (STA/LTA) algorithm (STA 1 s, LTA 50 s, triggering ratio 6, detragging ratio 5.5) and the  $BL_e$  (time-window length 20 s, threshold value 85). (a) and (b) show the results at different time scales.

## BL CONFORMITY AND FREQUENCY CONTENT

Seismic waveforms carry complex information both in amplitude and phase properties. Most of this information will be discarded when reducing the signal to only its first-digit frequencies, and thus the ability of the BL method to detect seismic events seems surprising. However, a similar drastic reduction of the signal information has been used successfully in other seismological fields, as for example in the calculation of ambient noise tomography. In this case, the one-bit normalization (Cupillard *et al.*, 2011), a procedure originally intro-

duced in communication theory in which only the sign of the signal is preserved, is widely used to calculate the correlation between data from different seismic stations.

Sambridge *et al.* (2010) stated that major earthquakes follow BL better than noise due to their larger dynamic range. We have shown here that this pattern is true not only for significant teleseismic events, but also for very small signals, with dynamic ranges of just a couple of decades and very low SNR, which still react positively to the BL test. Hence, we have documented that  $BL_e$  values do not depend primarily on the amplitude of the incoming seismic signal. In addition, we have



▲ **Figure 6.** Observation of global surface waves after the larger Chilean earthquake. The seismic trace has been low-pass filtered with a corner frequency of 5 mHz.  $BL_c$  has been calculated on the unfiltered trace with a time-window length of 120 s. The interval including the global-circling waves (GCW) traveling in the opposite sense around the globe is displayed using solid- and dashed-line boxes. In most cases,  $BL_c$  minima correlate with GCW arrivals.

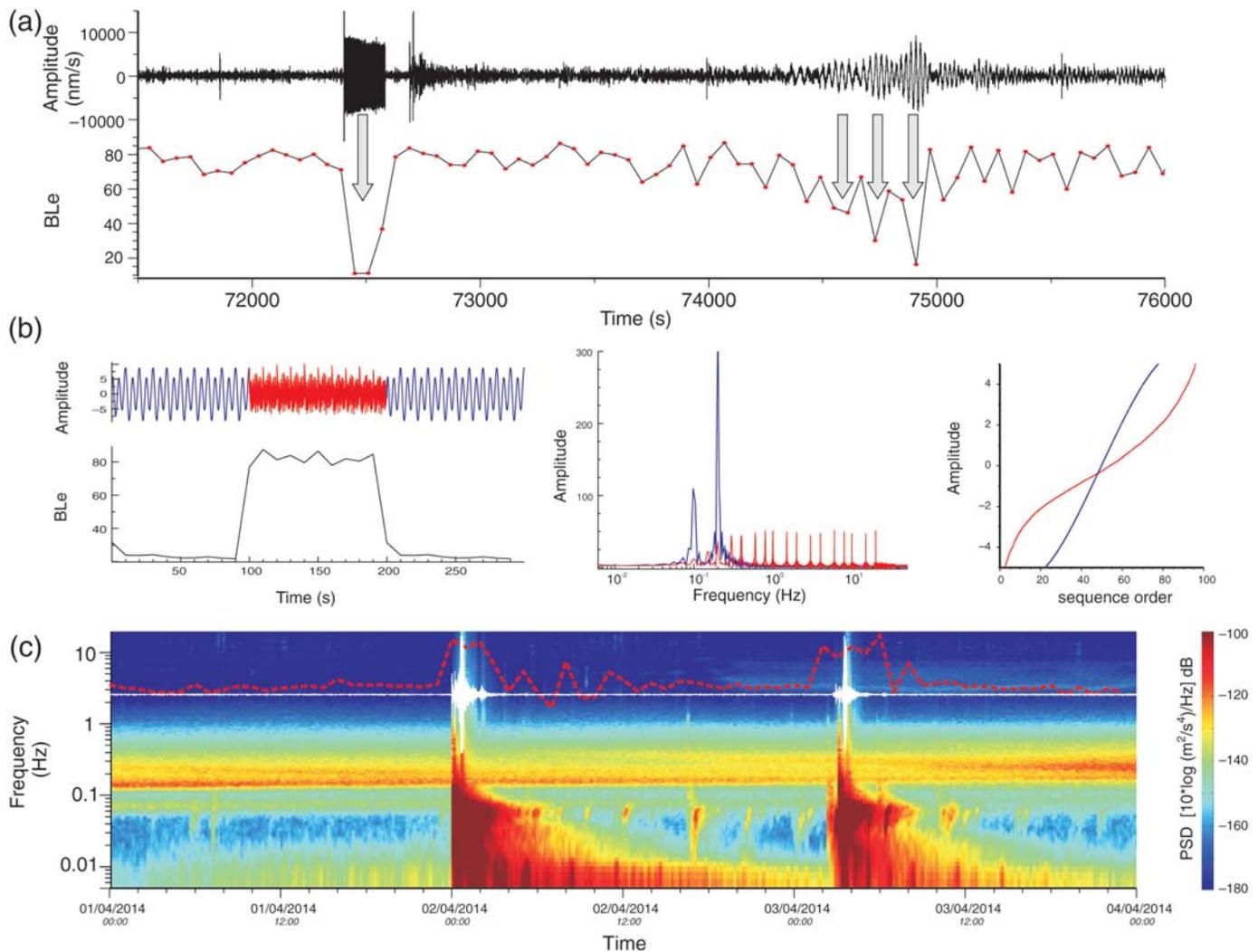
seen that if short-time windows are used to investigate teleseismic events, low  $BL_c$  values seem to be associated to the arrival of surface waves. This raised the hypothesis that the BL conformity of the seismic record is not related to its dynamic range, but rather to the changes in the frequency content of the incoming signal.

To document this issue, we investigated a record on a broadband seismic station deployed in northern Iberia during the TopoIberia experiment (Díaz *et al.*, 2010) that turned out to be installed close to a water pump that only operated occasionally. When the pump was activated, a large amplitude nearly monochromatic signal (close to 50 Hz) was clearly recorded by the seismometer. Figure 7a shows an example of such signal recorded shortly before the arrival of an  $M_w$  6.1 earthquake from the Alaska Peninsula. When applying our BL test with a short time window of 60 s, the monochromatic signal clearly correlates with  $BL_c$  values below 20. During the arrival of the surface waves generated by the seismic event, which appear in packets with different, monochromatic contents, the  $BL_c$  shows similar low values in the 40–20 range associated with the arrival of each surface-wave packet.

To further assess this relation between BL conformity and frequency content, we performed a synthetic test involving the generation of a composite signal. This trace includes two segments of nearly monochromatic sinusoidal signal separated by a segment with richer frequency content (Fig. 7b). When analyzing the conformity to BL of this synthetic signal, a clear difference appears between the different segments, with  $BL_c$  values reaching 80 during the rich frequency content interval and lying below 30 during the nearly monochromatic signal segments.

Looking for an ultimate explanation for this behavior, one may remember that if the values of a data set, ranked from small to large, approximately follow a geometric sequence  $S_N = Ar^{N-1}$ , the data set will match the BL (Raimi, 1976). The right panel in Figure 7b clearly illustrates that signals with a rich frequency content are closer to following a geometric sequence and hence to follow BL. The synthetic signal with poor spectral content (blue line) results in a near-linear sequence once ordered according to its amplitude, whereas the signal with richer frequency content (red line) results in a geometric (exponential) one.

In the absence of seismic events, the seismic record is dominated by the microseismic peak, the signal generated by ocean waves striking the coastline and interacting in open waters. The frequency content of this signal is restricted to the 0.05–0.8 Hz band, with most of the energy focused between 0.1 and 0.3 Hz (the double-frequency peak). This means the background seismic energy has relatively poor frequency content. In contrast, the signals generated by seismic events encompass a much wider frequency range, which may extend from a few millihertz to several tens of hertz in the case of large events. Figure 7c presents the spectrogram calculated for a three-day period including the two large Chilean events, together with the seismic record and the corresponding  $BL_c$  calculated using a one-hour time window. The Figure clearly documents the relationship between the spectral content of the signal and its conformity to BL. Therefore, the different degree of conformity to BL that exhibits the seismic waves and the background noise should be related primarily to their corresponding spectral contents.



▲ **Figure 7.** (a) A seismic trace for station IB E131, located in northwest Iberia, including the recording of a monochromatic signal related to a water pump engine just preceding the record of an  $M_w$  6.1 earthquake from the Alaska Peninsula (16 July 2011 at 19:59:12). The  $BL_e$  reacts with very low values, both to the artificial monochromatic signal and to the arrival of the surface wavetrains. (b) A synthetic test to illustrate the dependence of  $BL_e$  on the frequency contents: (left) the synthetic signal and the associated  $BL_e$ , which clearly changes from values around 25 for the first signal to values close to 80 for the second section; (center) the spectra of the two portions of the synthetic signal; and (right) the distribution of each signal, after it has been ordered according to its amplitude. The poor frequency signal has a near linear progression, whereas the richer signal shows a geometric progression. (c) The spectrogram calculated during the arrival of the Chilean events, seismic trace, and  $BL_e$  using a time-window length of 1 hr. Reddish colors in the spectrogram are for large energy. During the arrival of the seismic signals, energy spreads over a large portion of the spectra, whereas in the absence of events, energy is mostly concentrated around the microseismic peak.

## CONCLUSIONS

The first-order result of our study is that seismic signals, both from distant or local events, can be easily detected using an estimator based on the conformity to BL of the incoming signal. The sensitivity of this method is similar to that of the classical STA/LTA method for local events, even if a careful comparison is beyond the objectives of this study. The detection of the seismic signals does not depend basically on their amplitude, but on their frequency content. Nearly monochromatic signals result in very poor conformity to BL, whereas

frequency-rich signals provide high degrees of conformity. This feature can be exploited to detect the late arrivals of seismic wavetrains associated with large events (their low-frequency content results in lows on the  $BL_e$ ) or to discriminate specific natural or anthropogenic seismic signals from the ambient noise.

Our study provides a first evaluation of the proper parameterizations to identify the seismic arrivals using BL. Time-window lengths around 30 min provide good detections of teleseismic signals, whereas time lengths around 20 s are appropriate to identify local events. The conformity estimators can

be used to automatically detect arrivals when its value overpasses a threshold value, which seems to be around 85 for the defined  $BL_c$  or in the 0.010–0.015 range for the MAD estimator. However, as in other detection methods, those threshold values need to be adapted to each specific case, as the final number of detections is very sensitive to the chosen value.

From a practical point of view, this study presents a new way to discern between seismic signals and background noise based on a quite intriguing feature, Benford's Law. This method has some advantages with respect to the existing ones, such as its easy parameterization or its ability to recognize changes in the frequency content, but also some pitfalls, as the difficulty to provide accurate picks on the phase arrival times. In any case, it can be considered and used as an additional tool of the seismological toolbox. ✉

## ACKNOWLEDGMENTS

We want to acknowledge all staff of the Laboratorio Subterráneo de Canfranc (LSC) for their help in the installation and maintenance of the Geodyn facility. This is a contribution of the Geodyn project (MICINN ICTS2009-33). Additional support is provided by the TopoIberia (CSD2006-00041) and Misterios (CGL2013-48601-C2-1-R) projects. Many of the figures were produced using Generic Mapping Tools software (Wessel and Smith, 1998). Seismic data are curated at Institut de Ciències de la Terra Jaume Almera–Consejo Superior de Investigaciones Científicas (ICTJA-CSIC) and may be obtained upon request.

## REFERENCES

Benford, F. (1938). The law of anomalous numbers, *Proc. Am. Philos. Soc.* **78**, 551–572.

Berger, A., and T. Hill (2010). *Fundamental Flaws in Feller's Classical Derivation of Benford's Law*, University of Alberta, Edmonton, Canada.

Bormann, P. (Editor) (2012). *New Manual of Seismological Observatory Practice (NMSOP-2)*, International Association of Seismology and Physics of the Earth's Interior, GeoForschungsZentrums German Research Centre for Geosciences, Potsdam, Germany (<http://nmsop.gfz-potsdam.de>, last accessed June 2014), doi: [10.2312/GFZ.NMSOP-2](https://doi.org/10.2312/GFZ.NMSOP-2).

Cupillard, P., L. Stehly, and B. Romanowicz (2011). The one-bit noise correlation: A theory based on the concepts of coherent and incoherent noise, *Geophys. J. Int.* **184**, 1397–1414, doi: [10.1111/j.1365-246X.2010.04923.x](https://doi.org/10.1111/j.1365-246X.2010.04923.x).

Díaz, J., A. Villasenor, J. Morales, A. Pazos, D. Cordoba, J. Pulgar, J. L. Garcia-Lobon, M. Harnafi, R. Carbonell, J. Gallart, and the

TopoIberia Seismic Working Group (2010). Background noise characteristics at the IberArray broadband seismic network, *Bull. Seismol. Soc. Am.* **100**, no. 2, 618–628, doi: [10.1785/0120090085](https://doi.org/10.1785/0120090085).

Fewster, R. M. (2009). A simple explanation of Benford's law, *Am. Stat.* **63**, no. 1, doi: [10.1198/tast.2009.0005](https://doi.org/10.1198/tast.2009.0005).

Geyer, A., and J. Martí (2012). Applying Benford's law to volcanology, *Geology* **40**, no. 4, 327–330.

Goldstein, P., D. Dodge, M. Firpo, and Lee Minner (2003). SAC2000: Signal processing and analysis tools for seismologists and engineers, Invited Contribution to *The LASPEI International Handbook of Earthquake and Engineering Seismology*, W. H. K. Lee, H. Kanamori, P. C. Jennings, and C. Kisslinger (Editors), Academic Press, London, United Kingdom.

Newcomb, S. (1881). Note on the frequency of use of different digits in natural numbers, *Am. J. Math.* **4**, 39–40.

Nigrini, M. J. (1992). The detection of income tax evasion through an analysis of digital frequencies, *Ph.D. Thesis*, University of Cincinnati, Ohio.

Nigrini, M. J. (1996). A taxpayer compliance application of Benford's law, *J. Am. Tax. Assoc.* **18**, 72–91.

Nigrini, M. J., and S. J. Miller (2007). Benford's law applied to hydrology data: Results and relevance to other geophysical data, *Math. Geol.* **39**, no. 5, 469–490.

Pinkham, R. S. (1961). On the distribution of first significant digits, *Ann. Math. Stat.* **32**, no. 4, 1223–1230, ISSN: 0003-4851.

Raimi, R. A. (1976). The first digit problem, *Am. Math. Mon.* **83**, no. 7, 521–538, ISSN: 0002-9890.

Sambridge, M., H. Tkalcic, and P. Arroucau (2011). Benford's law of first digits: From mathematical curiosity to change detector, *Asia Pacific Mathematics Newsletter* **1**, no 4, 1–5.

Sambridge, M., H. Tkalcic, and A. Jackson (2010). Benford's law in the natural sciences, *Geophys. Res. Lett.* **37**, L22301, doi: [10.1029/2010GL044830](https://doi.org/10.1029/2010GL044830).

Sen De, A., and U. Sen (2011). Benford's law: Detection of quantum phase transitions similarly as earthquakes, available at <http://arxiv.org/abs/1103.5398v1> (last accessed August 2014) [quant-ph].

Varian, H. (1972). Benford's law (Letters to the Editor), *Am. Stat.* **26**, no. 3, 65, doi: [10.1080/00031305.1972.10478934](https://doi.org/10.1080/00031305.1972.10478934).

Wessel, P., and W. H. F. Smith (1998). New, improved version of the Generic Mapping Tool released, *Eos Trans. AGU* **47**, 576.

J. Díaz  
 J. Gallart  
 M. Ruiz  
*Institute of Earth Sciences "Jaume Almera"*  
*Consejo Superior de Investigaciones Científicas*  
*(ICTJA-CSIC)*  
 c/ Solé Sabarís sn  
 08028 Barcelona, Spain  
 jdiaz@ictja.csic.es

Published Online 26 November 2014