Tectonophysics 503 (2011) 189–194

Contents lists available at ScienceDirect

Tectonophysics

journal homepage: www.elsevier.com/locate/tecto

Identifying long-range correlated signals upon significant periodic data loss

P.A. Varotsos *, N.V. Sarlis, E.S. Skordas

Solid State Section and Solid Earth Physics Institute, Physics Department, University of Athens, Panepistimiopolis, Zografos 157 84, Athens, Greece

ARTICLE INFO

Article history:
Received 27 May 2010
Received in revised form 17 January 2011
Accepted 27 February 2011
Available online 3 March 2011

Keywords:
Seismic electric signals
Detrended fluctuation analysis
Geoelectrical field measurements
Long-range correlated signals

ABSTRACT

When monitoring geophysical parameters, data from segments that are contaminated by noise may have to be abandoned. This is the case, for example, in the geoelectrical field measurements at some sites in Japan, where high noise – due mainly to leakage currents from DC driven trains – prevails almost during 70% of the 24 hour operational time. We show that even in such a case, the identification of seismic electric signals (SES), which are long-range correlated signals, may be possible, if the remaining noise free data are analyzed in natural time along with detrended fluctuation analysis (DFA).

1. Introduction

In many cases of geophysical and/or geological interest, it happens that for substantial parts of the time of data collection, high noise prevents any attempt for extracting a useful signal. Data for such time segments are removed from further analysis. The appearance of such a noise may be periodic as in the case treated in the present work. It is the objective of this paper to examine whether the remaining data allow the identification of long-range temporal correlations.

The present study was motivated from the results of geoelectrical measurements in Japan aiming at the detection of Seismic Electric Signals (SES), which are low-frequency (≤ 1 Hz) variations of the electric field of the earth that precede earthquakes (Varotsos and Alexopoulos, 1984a,b). SES sometime appears as a single signal lasting for minutes but often many SES (hereafter called “pulses” as needed) keep appearing during certain length of time, which may be as long as a few days or more. Such a case is called SES activity. It has been shown that the SESs in a SES activity have *long range* temporal correlations characteristic to *critical* phenomena (Varotsos et al., 2002). The measurements in Japan have detected clear SES either at noise-free measuring sites or at noisy stations when the SES happened to occur at midnight, i.e., when the noise level was low (Uyeda et al., 2000, 2002). The major difficulty at many sites is the contamination of records by high noise due to leakage currents from DC driven trains and other artificial sources, against which some countermeasure such as independent component analysis to extract signals has been attempted (e.g., see Orihara et al. 2009). The low noise time occurs from 00:00 to 06:00 and from 22:00 to 24:00 local time (LT) when nearby DC driven trains cease service, i.e., almost only 30% of the 24 h. Thus, the question arises whether it is still possible to identify SES upon removing the noisy data segments lasting for the period 06:00 to 22:00 every day. The answer to this question is attempted in this paper for the case when the duration of SES activity is much longer compared to those of individual pulses, i.e., a few days to a few weeks or even more, although admittedly long lasting SES activity is rather seldom, limiting the applicability of the results described below.

The key point in the present work is the use of the following two modern methods: The natural time analysis of the remaining data and the detrended fluctuation analysis (DFA). The present question differs from the one in which we investigated (Skordas et al., 2010) the effect of the random in time removal of data segments of fixed length on the scaling properties of SES activities. It also differs from the case in which the lengths of the lost or removed data segments are random and may follow a certain type of distribution (Ma et al., 2010).

We now briefly describe the time series analysis in natural time χ, which is a new time domain (Varotsos, 2005; Varotsos et al., 2002, 2003a,b;). In a time series comprising N events, the natural time $\chi_k = k/N$ serves as an index for the occurrence of the k-th event. The evolution of the pair (χ_k and Q_k) is studied, where Q_k is a quantity proportional to the energy released in the k-th event. For dichotomous signals, which is frequently the case of SES activities, the quantity Q_k can be replaced by the duration of the k-th pulse. By defining $p_k = Q_k / \sum_{n=1}^{N} Q_n$, we have found that the variance $\kappa_1 = \langle\chi^2\rangle - \langle\chi\rangle^2$, where $\langle f(\chi)\rangle = \sum_{n=1}^{N} p_k\, f(\chi_k)$, of the natural time χ with respect to the distribution p_k may be used for identifying *criticality*, and hence the SES activities. More specifically, the following relation should hold for SES activities

$$\kappa_1 \approx 0.070 \tag{1}$$

* Corresponding author. Tel.: +30 210 9617573; fax: +30 210 9601721.
E-mail address: pvaro@otenet.gr (P.A. Varotsos).

0040-1951/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.tecto.2011.02.011

Beyond the condition of Eq. (1), we have shown that the SES activities, when analyzed in natural time, exhibit *infinitely* ranged temporal correlations and obey the conditions (Varotsos et al., 2005, 2006a,b):

$$S, S_{-} < S_u, \tag{2}$$

where S is the entropy S in natural time defined as: $S \equiv \langle \chi \ln \chi \rangle - \langle \chi \rangle \ln \langle \chi \rangle$ (Varotsos et al., 2003a) and S_{-} is the entropy obtained upon time reversal. Eq. (2) states that both S and S_{-} are smaller than the value S_u ($=\ln 2/2 - 1/4 \approx 0.0966$) of a "uniform" (u) distribution, e.g. when all p_k are equal.

The fact that SES activities exhibit *critical* dynamics, is believed to be related to their generation mechanism (see Varotsos et al., 1993, and references therein). In the focal area of an impending earthquake (EQ hereafter), which contains ionic materials, the stress gradually increases. In ionic solids a number of extrinsic defects are always formed because they contain aliovalent impurities. These extrinsic defects are attracted by the nearby impurities and hence form electric dipoles the orientation of which can change through defect migration. When the stress (pressure) σ reaches a critical value σ_{cr}, a *cooperative* orientation of these dipoles occurs generating SES.

We now summarize the detrended fluctuation analysis DFA (Peng et al., 1994; Taqqu et al., 1995) which is a novel method that has been developed to address the problem of accurately quantifying long range correlations in non-stationary fluctuating signals. It has been applied to diverse fields ranging from DNA (Peng et al., 1993; Stanley et al., 1999), to meteorology (Ivanova and Ausloos, 1999), and economics (Vandewalle and Ausloos, 1997; Ivanov et al., 2004). DFA is, in short, a modified root-mean-square (rms) analysis of a random walk. In principle, it estimates the deviations from the local trends $y_s(n)$ of a non-stationary long time series of length N piecewise by dividing it into small segments with length s and compute the Fluctuation function F(s), which is the variance of $y_s(n)$:

$$F(s) = \sqrt{\frac{1}{N}\sum_{n=1}^{N} \left[y_s(n)\right]^2} \tag{3}$$

F(s) corresponds to the trend-eliminated root mean square displacement of the random walker. Then, the above computation is repeated for a broad number of scales s to provide a relationship between F(s) and s.

When a power–law relation between F(s) and s, i.e.

$$F(s) \propto s^{\alpha} \tag{4}$$

is found, it indicates the presence of scaling-invariant (fractal) behavior embedded in the fluctuations of the signal (Peng et al., 1994; Taqqu et al., 1995). The fluctuations can be characterized by the scaling exponent α, a self-similarity parameter: If $\alpha = 0.5$, there are no correlations in the data and the signal is uncorrelated (white noise); the case $\alpha < 0.5$ corresponds to anti-correlations, meaning that large values are most likely to be followed by small values and vice versa. If $\alpha > 0.5$, there are long-range correlations, which are stronger for higher α (Bashan et al., 2008). Note that $\alpha > 1$ indicates a non-stationary local average of the data and the value $\alpha = 1.5$ indicates Brownian motion (integrated white noise).

For stationary signals with long-range power–law correlations the value of the scaling exponent α is interconnected with the exponent β characterizing the power spectrum $S(f) \sim f^{-\beta}$ (f = frequency) through (Peng et al., 1993)

$$\beta = 2\alpha - 1 \tag{5}$$

When employing natural time, DFA seems to distinguish (Varotsos et al., 2003b) SES activities from artificial noise because, for the SES activities the α-values lie approximately in the range

$$0.9 \le \alpha \le 1.0, \tag{6}$$

while for the artificial noise (caused by man-made sources) investigated in Greece (Varotsos et al., 2003a,b) the α-values are markedly smaller, i.e., $\alpha = 0.65$–0.8. In other words, the artificial noise recorded in Greece, which at the most lasts for 24 h, may have long-range correlations, e.g. $\alpha \approx 0.75$ (see Fig. 9 of Varotsos et al. (2003a)), but none of several artificial noises studied was found to exhibit infinitely ranged long-range correlations (i.e., having α-value close to unity).

2. Data analysis and results

Let us suppose that we have a long time series of data s(i) (shown in red in the example of Fig. 1), with a duration appreciably larger than 24 h for instance, and we are forced to remove the same segment of these daily data. The portion of the 24 hour data that remain will be hereafter labeled p_r and the number of data corresponding to one period, say 24 h, T. Thus, every T samples, $(1-p_r)T$ of them (belonging to the shaded parts of Fig. 1) are removed. The remaining segments (blue in Fig. 1) are concatenated to form the new time series c(i) which is subsequently read in natural time. We now impose the following conditions (7) and (8) on c(i) for classifying the signal as SES activity. The condition (7) comes from the relation (6) after considering the reasonable experimental error:

$$0.85 \le \alpha \le 1.10 \tag{7}$$

The condition (8) comes from Eqs. (1) and (2) also by considering the reasonable experimental error in κ_1:

$$|\kappa_1 - 0.07| \le 0.01, S \le S_u, S_{-} \le S_u \tag{8}$$

In the following subsections, in order to solve our problem, synthetic signals will be produced and analyzed whether they obey conditions (7) and/or (8) using a Monte Carlo comprising 10^3 realizations. The Monte Carlo procedure has been used to "average" over the possible realizations of the synthetic SES activities and noises that will be discussed later in Sections 2.1 and 2.2 as well as the fact that both types of electric signals may start any time of the day. Thus, one should randomly select an integer i_{init} from 1 up to T, and keep in c(i) the samples $s(i_{init})$ to $s(i_{init}+p_rT-1)$ of s(i), i.e., we keep p_rT samples in total. The next segment to be kept in c(i) is $(1-p_r)T$ samples after $s(i_{init}+p_rT-1)$, starting from $s[i_{init}+p_rT-1+(1-p_r)T+1=i_{init}+T]$ up to $s(i_{init}+T+p_rT-1)$ and so on (see the blue lines in Fig. 1). This way we periodically remove $(1-p_r)T$ samples and keep p_rT every T samples from the original signal s(i). This Monte Carlo simulation allows us to evaluate the probability to identify the original signal as a SES activity.

The probability that the condition (7) is satisfied will be hereafter labeled p_1. By the same token, the probability to satisfy the condition (8) is designated by p_2. Finally, the probability to obey either condition (7) or condition (8) will be labeled p_3. Upon considering the number of the Monte Carlo realizations ($M = 10^3$), a plausible estimation error

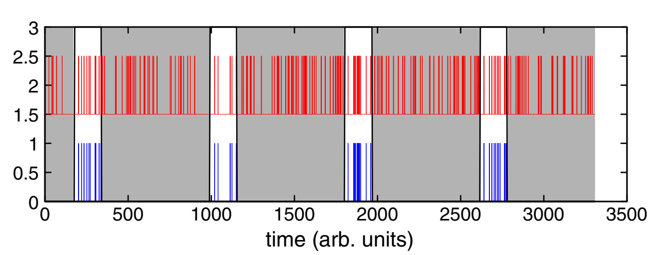

Fig. 1. An example of the procedure described in Section 2.1, showing data segments that are periodically removed every $T \approx 813$ arbitrary units from the original dichotomous time-series (red). The gray shaded areas correspond to the high noise periods (e.g. 06:00–22:00 LT in Japan) which have to be discarded daily.