Cross-correlation analysis for geoelectric time series associated with an earthquake by means of mutual information theory

A. Ramírez-Rojas, R. T. Páez-Hernández, and J. Rubén Luévano
Departamento de Ciencias Básicas, Universidad Autónoma Metropolitana Azcapotzalco, Avenida San Pablo 180 Col. Reynosa, México, 02200 D.F., México,
E-mail: arr@correo.azc.uam.mx; alexramno@gmail.com

Received 30 de junio de 2011; accepted 25 de agosto de 2011

We apply information theory to characterize electric self-potential time series monitored in the most important seismic region in México. We compute the level of global cross-correlation between two perpendicular components of such time series. Both time series were monitored, for a period of several months, before, during and after of a Ms = 7.4 earthquake occurred on September 14, 1995. Our calculations are based on the re-scaled average mutual information index \( \lambda = \lambda(I) \), where \( I \) is the mutual information function and \( \lambda \) is a measure of strength of nonlinear correlations between two time series. Our findings suggest that the high level of global correlation observed between both electric field components during several months before the event, is associated with the preparation mechanism of the occurred earthquake.

Keywords: Mutual Information; Cross-correlation; electric self-potential.

Aplicamos la teoría de la información a series de tiempo del autopotencial eléctrico monitoreado en la región sísmica más importante de México. Calculamos el nivel global de la correlación cruzada entre dos componentes perpendiculares del campo eléctrico medido. Ambas series se monitorearon algunos meses antes, durante y después del sismo de Ms = 7.4 ocurrido el 14 de septiembre del 1995. Nuestras calculaciones están basadas en el índice re-escalado promedio de información mutua \( \lambda = \lambda(I) \) (donde \( I \) es la función de información mutua y \( \lambda \) es una medida de la intensidad de las correlaciones no lineales entre dos series de tiempo. Nuestros resultados sugieren que el alto nivel de correlación global observado entre las dos componentes del campo eléctrico algunos meses antes del evento, está asociada con el mecanismo de preparación del sismo ocurrido.

Descriptores: Información mutua; correlación cruzada; auto-potencial eléctrico.

PACS: 05.45Tp; 05.10.a

1. Introduction

The characterization of nonlinear features of dynamical systems from their time series plays a central role in the study of complex systems. Hence, the introduction of indicators of nonlinearities is an important task. Some ideas of information theory related to the nonlinear predictability that were introduced by Tsonis [1], who applied methods based on the entropy concept, have provided insights into the nature of linearity and nonlinearity of a system. For linear systems, the standard measure of dependence between two random variables is the linear cross-correlation function. However, this indicator does not work well when a nonlinear relationship is present. A measure of global dependence between two time series is the so-called mutual information function \( I \). By definition [4], \( I \) is an unbounded function \((0 \leq I < \infty)\), so a direct analysis of it is not always recommendable. Hence, it is more convenient to consider the re-scaled average mutual information index, \( \lambda = \lambda(I) \), because it is bounded: \( 0 < \lambda(I) < 1 \). When \( \lambda = 0 \) (\( I = 0 \)), both signals are uncorrelated and there is a significant correlation when \( \lambda = 1 \) (\( I \to \infty \)). Just like \( I \), the index \( \lambda \) identifies only global correlations without any kind of specification of linear or non-linear dependence [1]. The information theory concepts [2,3,4] have been applied in studies of financial time series. On the other hand, earthquake prediction is considered one of the most challenging and debated questions in the scientific community. One experimental methodology used in the search of seismic precursory signals consists in monitoring the geoelectric field in active seismic regions. This technique was implemented by Varotsos et al. [5,6] and consists in monitoring the electric self-potential of the ground by means of electrode pairs buried in the ground and separated \( L \) meters. Two perpendicular components of the local electric field \( (E=\Delta V/L) \) are obtained. Each component of \( E \) is monitored in two perpendicular lines: north-south (NS) and east-west (EW). The short-term earthquake prediction based on electromagnetic phenomena is considered a promising candidate to obtain information of imminent EQs [7,8]. In this context, it was previously reported in Ref. 9 the cross-correlation linear coefficient between both perpendicular components, NS and EW of the geoelectric field monitored in the South Pacific Mexican coast. The reported results show complex behaviors suggesting that non-linear dependences were present previous to the EQ (M7.4) which occurred in September 14, 1995. The aim of this work is to analyze, by means of the re-scaled parameter \( \lambda \), the global correlation between two geoelectric time series (NS and EW channels) monitored simultaneously and associated with the mentioned EQ. Our results show an important cross-correlation, in terms of \( \Box(I) \) suggesting that the preparation mechanism started approximately five months before the main shock. The work is organized as follows: In Sec. 2 a brief review of re-scaled parameter and mutual
information is given. Section 3 is devoted to the monitoring time series. The results are described in Sec. 4.

2. Mutual information and methodology

The mutual information function is defined as follows: Consider two real valued random variables X and Y. Let \( p_x(x) \) and \( p_y(y) \) the probability density functions of the X and Y respectively, and \( p_{X,Y}(x,y) \) their joint probability density function. The mutual information \( I(X,Y) \) is defined as:

\[
I(X,Y) = \int \int p_{X,Y}(x,y) \log \frac{p_{X,Y}(x,y)}{p_x(x)p_y(y)} \, dx \, dy
\]

where \( p_x(x) \) and \( p_y(y) \) are the marginal probabilities and \( p_{X,Y}(x,y) \) is the joint probability. The mutual information function satisfies the following properties: a) \( I = 0 \), indicates that X and Y are statistically independent and b) \( 0 \leq I < \infty \), means that the mutual information is an unbounded function, this last becomes problematic for comparisons between both data sets. The re-scaled average mutual information index \( \lambda(I) \) is:

\[
\lambda(I) = \sqrt{1 - e^{-2I}}
\]

As it has been described by Tsonis in Ref. 1, the quantity \( \lambda(I) \) capture both linear and nonlinear dependence between X and Y and, as \( 0 \leq \lambda \leq 1 \) so that it is more convenient in order to quantify the level of cross-correlation. To compute \( I \), and then \( \lambda(I) \), the time series is partitioned in windows. We select a set of N non-overlapping windows of size s, and I is computed on each window. Hence, the marginal probabilities, \( p_x(x) \) and \( p_y(y) \), and the joint probabilities \( p_{X,Y}(x,y) \) are obtained by a frequency analysis from the sample space.

In our analysis we considered windows of 2048 data points corresponding to 1 hr 8 min. The whole monitored period is around a year, from January 04 to December 19, 1995. In both cases, each data point is a measurement of the electric self-potential fluctuations of the ground with a sampled frequency of 0.5 Hz.

3. Data set

The geoelectric time series analyzed are depicted in Fig. 1, and consist of the measurements of the electric self-potential fluctuations, \( \Delta V \), between two pairs of non-polarized electrodes, buried 2 m of depth into the ground and separated 50 m. A couple of electrodes were oriented towards the North-South (NS channel), and the other one towards the East-West (EW channel) direction [10], as it is indicated by the so-called VAN methodology [5,6]. The data set analyzed in this paper was monitored during the year 1995 at the station located in Acapulco (16.85°N, 99.9°W). Two time series were simultaneously recorded (NS and EW channels) with sampling rate of \( \Delta t = 2 \) s.

4. Results

Figure 2 shows the sequence of scatter plots corresponding to twelve segments. Each graph is a short segment of signals of the consecutive months, from January to December of 1995. The cross-correlation of short segments of NS and EW suggests nonlinear dependence between both components of the electric field.

It is noted that the cross-correlation behavior gives a qualitative description of the dependence between two time series. Each scatter plot suggests that different kinds of underlying dynamical behavior follow when the time goes on. In Fig. 2(a) a quasi-linear dependence between NS and EW components is observed previous to the preparation process of the EQ. In Fig. 2(b) and Fig. 2(c) complex underlying dynamics are dominant. The sequence of Fig. 2(d) to Fig. 2(h) shows that linear dependences are lost. The segment where the EQ occurred (Sept. 14, 1995) is shown in Fig. 2(i). After that, the system goes back to its equilibrium dynamics and the local electric field fluctuate randomly as it is shown in Figs. 2(k) and 2(l) and apparently the seismic activity has finished showing an uncorrelated dependence.

In Fig. 3(a) the \( \lambda \) behavior is depicted. An important level of correlation can be noticed, except along April when, according to Mexican catalogues the seismic activity was very low. After that month, \( \lambda \) describes a strong dependence, ranging in the interval \( 0.8 < \lambda < 0.99 \). This behavior is maintained until September and it is associated possibly to the preparation process. A few days before the EQ and some weeks after it, there are correlation values fluctuating between: \( 0.35 < \lambda < 0.85 \), and are associated with the relaxation process and aftershocks.

Comparatively the Fig. 3(b) depicts the linear cross correlation coefficient, displaying large fluctuations. Such values are associated with the linear fitting of points like it is shown in Fig. 2.
Figure 2. Sequences of scatter plot maps of excerpts showing the evolution of cross-correlation between both time series. Each window represents one excerpt of each month of 1995 year. Different features in their dependences can be observed.

Figure 3. (a) Rescaled average mutual information index $\lambda(I)$. A clear structure showing a significant nonlinear correlation previous to EQ is noted. (b) The linear cross correlation index is given to comparison.
As expected, we found a strong non linear dependence between both components, particularly in a period previous the big earthquake occurrence. Also, we observe that $\lambda$ shows a specific structure for a given selection of window’s size $s = 2048$, or 1 hr 8 min. It should be noted that, in particular, for different selections of window’s size, $s = 512, 1024,$ and 4096 points, we observed that the remarkable structure shown in Fig. 3 is lost, to become more uniform. A thorough analysis of this phenomenon will be given elsewhere. Here we want to emphasize that, for this purpose, we need to adapt our methodology to be compared with the spectral power of the signals.

5. Concluding remarks

We have computed the index $\lambda$ for geoelectric signals. These signals are recorded simultaneously in two orthogonal components and in a time interval containing an earthquake of $M = 7.4$. That index is compared with the cross correlation coefficient, which only detects the linear dependencies between both components. For such purpose, we make a partition of the time series in a sequence of non overlapping windows, having size $s$, and compute $\lambda$ for each them obtaining a local value of $\lambda$. Despite the fact that I display an irregular behavior, the values of $\lambda$ are more regularly distributed. We found that there are many windows preceding the earthquake which display higher values of $\lambda$ indicating strong cross-correlation. However, there exist some windows with relatively lower values of $\lambda$, particularly in the time windows associated to lower seismic activity. Also, we observe that $\lambda$ is sensitive to the selection of $s$. Hence, we conclude that the statistical properties of the mutual information $I$, considered as an observable of a time series, provide us with valuable qualitative information about the preparation process preceding a big event. We conclude that $\lambda$ can be considered as a qualitative indicator of a precursory behavior of the seismicity.

Acknowledgements

This work was supported by the Irreversible Processes Physics Area and Basic Sciences Department of the Universidad Autónoma Metropolitana.