Geoelectric Field Signal Investigation using Multidimensional Techniques and its Possible Relation to Earthquakes in Western Greece

Apostolos Ifantis, George Economou, Christos Theoharatos, *Member, IEEE*, Dimitris Petropoulos, and Gerasimos-Akis Tselentis

Abstract—**In this study, time series similarity and multi-dimensional scaling (MDS) techniques are used to examine recordings of the geoelectric potential difference, 10-days prior to significant seismic events in the area of Western Greece. The data have been collected during a five-year experimental investigation, at the earthquake prediction section of the University of Patras Seismological Laboratory (UPSL). The presence of similar patterns in the recorded time series data of the geoelectric field, ten days before the occurrence of seven significant earthquakes in the seismic area of Patras city, is investigated. Certain interesting clustering tendencies are detected in the MDS-plot of the projected time sequences data, indicating the existence of a possible correlation between the geoelectric signal structure and the impeding seismic activity. An explanation of the above results is also provided.**

I. INTRODUCTION

AMONG all earthquake precursors, earthquake-related electromagnetic phenomena are considered as the most promising candidates for short-term earthquake prediction. Over the years, numerous electromagnetic signals at a broad range of frequencies associated with earthquakes have been reported [1]-[3]. However, due to the many possible sources of noise, correlation between these signals and earthquakes is still controversial. Detected signals vary in duration and pattern, having specific features and spectral characteristics. Nevertheless, there is strong evidence that anomalous changes of the geoelectromagnetic field do occur prior to strong earthquakes. Currently, a plethora of geoelectromagnetic signal-change investigations take place for the purpose of earthquake prediction. Long-Term Geoelectric Potential (LTGP) measurements are carried out and the abundance of the generated time series data recordings are searched using sophisticated signal analysis techniques [4]-[6]. An earthquake is not an instantaneous phenomenon; it is accompanied with preseismic geotectonic preparations, thus correlation of the behavior of the LTGP and an oncoming earthquake is possible [5], [6].

In this paper, the existence of similarities in time sequences of the geoelectric potential data is investigated, over a 10-days time period, prior to major earthquakes that took place in the area of Western Greece. For this purpose, time series similarities search is studied using exploratory data analysis (EDA) for data projection and visualization [7], applied to sequences of the data. The reduction of dimensionality can lead to an increased capability of extracting knowledge from the underlying data by means of visualization, and to novel possibilities in designing efficient and possibly more effective classification schemes. Recordings have been collected during a five-year (1993-1997) independent experimental investigation in cooperation with the Varotsos-Alexopoulos-Nomikos (VAN) group, at the UPSL. During this period, several destructive earthquakes have occurred in the area of Western Greece, a territory with the highest seismic activity in Europe.

II. LTGP ACQUISITION SYSTEM

The monitoring of the geoelectric potential difference in the system is accomplished using a dipole set of Pb-PbCl$_2$ electrodes located 100m apart, having an E-W direction. Recorded data are referred as Ô*Channel 0'*. The exact geographical position of the station can be found in Fig. 1.

The signal produced by this dipole is initially directed to an electronic VAN device. Afterwards, it is directed to an A/D converter, which is set to digitize at a rate of 3 samples per min. The converter is connected to an ordinary PC where monitoring and processing of the signals is taking place. This channel is also connected to a pen-recorder and a graph paper illustrates continuously the field-changes in the area. The obtained electro-telluric signal is transmitted via a dedicated line to the control room. The channel signal is anti-alias filtered with a 30Hz Butterworth low-pass filter sampled at 100Hz and converted to digital form with a 32-bit resolution.

Manuscript received October 1, 2005. This work was supported by the by the National Ministry of Education under the project entitled “Soft computing techniques for multichannel processingÓ in the context of the program “ArchimidisÓ

A. Ifantis is with the Department of Electrical Engineering, Control System and Signal Laboratory, Tech. Inst. of Patras, Greece (corresponding author, phone: +30 2610 369238; e-mail: ifantisa@teipat.gr).

G. Economou is with the Department of Physics, Electronics Laboratory, University of Patras, Patras 26500, Greece (e-mail: economou@physics.upatras.gr).

Ch. Theoharatos is with the Department of Physics, Electronics Laboratory, University of Patras, Patras 26500, Greece (e-mail: htheohar@upatras.gr).

D. Petropoulos is with the Department of Physics, Electronics Laboratory, University of Patras, Patras 26500, Greece (e-mail: dpetrop@upatras.gr).

G.-A. Tselentis is with the Department of Geology, Seismological Laboratory, University of Patras, Patras 26500, Greece (e-mail: tselenti@upatras.gr).

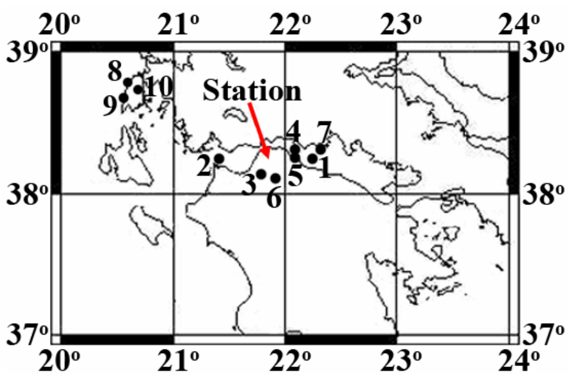

Fig. 1. Epicenters of the major earthquakes in Patras area in Western Greece during the period 1993-1997.

The digitization rate for observing long time variations in our station is set at 1sample per 20 sec. For the purposes of this work a time representation of 1 sample per hour is selected. That is, approximately 43600 data (points) describe the geoelectric potential-difference signal during the period from 1993 to 1997. This signal is illustrated in Fig. 2.

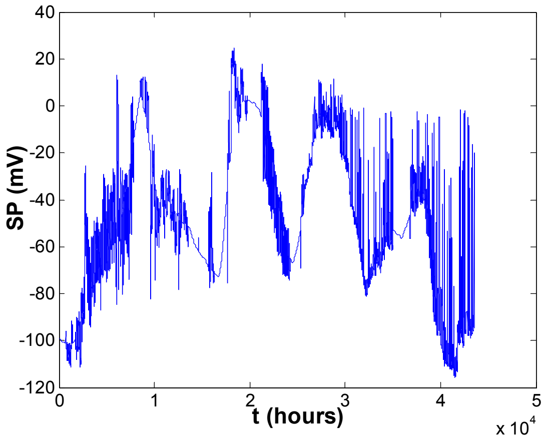

Fig. 2. Geoelectric signal (*Channel 0*).

III. TIME SERIES SIMILARITY SEARCH AND VISUALIZATION

Time-series are the principal data format of signals in many applications. Usually, a time-series is an ordered sequence of real numbers collected at uniformly spaced time instances (in our case representing geoelectric field measurements), each number representing a value at a time point. Searching for similar patterns among time-series intervals is an open research field, extensively investigated by the research community as a tool for exploring classification, clustering and mining of association rules [8]-[10]. The most common notion for time series similarity refers to time-series shape similarity.

Analysis of time-series data is rooted in the ability to find similar patterns. The definition of similarity / dissimilarity between time sequences is of crucial importance for time-series searches. Among the plethora of proposed (dis)similarity measures, the Euclidean distance, which is a special case of the L_p-norm, is the most popular one. Given two data sequences $X=(x_1,x_2,...,x_n)$ and $Y=(y_1,y_2,...,y_m)$ with $n=m$, their Euclidean distance is defined as:

$$L_2(X,Y)=\sqrt{\sum_{i=1}^{n}(x_i-y_i)^2} \qquad (1)$$

It is generally recognized that the Euclidean distance is affected by offsets of the sequences. In the case where shape similarity is sought, offset shifting and amplitude scaling transformations alleviate this problem [10]. Given a time series, $S(s_1,s_2,...,s_n)$ each element is transformed by the form of $\frac{s_i-mean(S)}{std(S)}$.

Following a geometric approach, a time series of length n can be considered as a point (*feature-vector*) in a multidimensional space $\mathbf{R}^{\mathbf{N}}$ resulting in a much-simplified representation, and thus can be treated as a collection of points in $\mathbf{R}^{\mathbf{N}}$. In this way, shifting and amplitude scaling of a time series correspond to the movement of the point in straight lines. The definition of the pairwise time series similarity provides a mean of indexing time series data. Given a query sequence $Q=(q_1,q_2,...,q_n)$, then a set of stored data sequences $S=(S_1,S_2,...,S_n)$ can be ordered according to their similarity to Q as defined by L_2, performing a nearest neighbor search in the multidimensional space.

When all of the stored data sequences S need to be examined as a whole and in order to investigate possible relations among them, their direct indexing is prohibitive due the space's high dimensionality. The core idea is to reduce the dimensionality of the input space. This is accomplished by projecting the N-dimensional blobs that represent the time-series into a k-dimensional space so that (i) $k<<N$ and (ii) the resulted distances are preserved as well as possible. Thus, dimensionality reduction is the procedure of finding a suitable lower-dimensional space in which to represent the original data. That alternative representation of the data can be helpful to explore high-dimensional data with the goal to discovering structures or patterns, visualize the data using scatter-plots when dimensionality is reduced to 2-D or 3-D and finally analyze the data using statistical methods, such as clustering, smoothing, probability density estimation, or classification [7].

In this case, *dimensionality reduction* plays a significant role. Although the data may have very high dimensions, these are typically embedded into manifolds (or subspaces) that are of substantially lower dimensions. Identifying these manifolds (or subspaces) is critical in understanding data relations. The general process of finding a configuration of points whose inter-point distances correspond to similarities or dissimilarities is often c*alled multidimensional scaling* (MDS) [7] and it is used to discover the underlying distribution of the original feature space. In general, MDS finds an embedding of the data points in a low-dimensional space such that the proximity between objects in the multi-dimensional space is represented with some degree of fidelity by the distances between points in the resulted feature space. This means that observations that are close together in a high-dimensional space should be close in the low-dimensional one. Algorithmically, given the *M* time-series, their pairwise distances and the desired dimensionality of the produced space k, the optimization criterion is given by:

$$Stress = \sqrt{\frac{\sum_{ij}\left(D\left(S_i,S_j\right)-D\left(S_{ki},S_{kj}\right)\right)^2}{\sum_{ij}D\left(S_i,S_j\right)^2}}, \qquad (2)$$

where $D\left(S_i,S_j\right)$ is the distance between time-series S_i and S_j and $D\left(S_{ki},S_{kj}\right)$ is the Euclidean distance at the k-dimensions. Actually, the *stress* is a number indicating how well the distances are preserved in the newly formed feature space. The steepest descent algorithm starts with an assignment of the time-series to the k-dimensions and proceeds by minimizing the *stress* in (2) by moving points. The mapping generally is nonlinear and can reveal the overall hidden structure of the original data. However, it is inevitable that there will be some stress in the resultant display.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

The major seven earthquakes (>4.0M_S) that occurred during the period under investigation together with some additional details are depicted in Table 1. In order to identify possible correlations of the signal prior to these events, a set of twelve data sequences S, each one of length *N=240* corresponding to a point in a 240-dimension space, were used as input to our data mining system. These 240-point sequences corresponded to 10-days duration (one point per hour), where the sequences end-point coincided with the earthquake occurrence time. For comparison purposes, five additional time sequences of the same time duration were also constructed. These were located at periods of relative seismic quiescence (<3.5 M_S), in the area and their exact position in time is shown in Table 2. Finding such periods was not easy due to the area's high seismicity.

TABLE 1
THE SEVEN MAJOR EARTHQUAKES THAT OCCURRED IN PATRAS AREA IN WESTERN GREECE DURING THE PERIOD 1993-1997

No.	Point	Magnitude	Distance (km)	Date
1	1863	4.9	44	18/03/93
2	3575	4.0	60	29/05/93
3	4700	5.1	15	14/07/93
4	21505	5.6	42	15/06/95
5	21755	4.0	42	25/06/95
6	24030	4.8	29	28/09/95
7	42453	4.9	55	05/11/97

TABLE 2
FIVE SEGMENTS OF THE GEOELECTRIC POTENTIAL SIGNAL DURING PERIODS OF SEISMIC QUIESCENCE.

No.	Start Point	End Point
A	10100	10339
B	21440	21679
C	32240	32479
D	37000	37239
E	38561	38800

In order to visualize inter time-series similarities, their MDS plot was constructed using pairwise Euclidean distances. In Fig. 4 the 2-D MDS plot is illustrated, where the seven time-blobs recorded prior to earthquakes are displayed using a red-star symbol and labeled according to Table 1 data from 1 to 7, while the rest five sequences that correspond to seismic quiescence are presented by a green-circle, also labeled in accordance to Table 2 data format. Results are interesting, indicating a certain correlation between the temporal patterns of the time series preceding the seismic event and the earthquake's focal area.

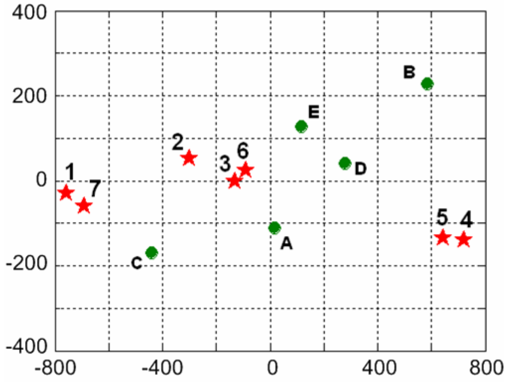

Fig. 4. 2-D MDS clustering results of the 7 earthquakes (red stars) and the 5 random blobs (green circles) of seismic quiescence, in Patras area.

In order to further exploit the underlying field pattern of the broader territory, three extra time sequences of the same time duration were also consider. These were related to the three major earthquakes that occurred during the same time-period in the area of North Ionian, far away from the recording station and presented using the same details in Table 3. Following the same experimental methodology, the MDS plot was produced for studying terrain-pattern clustering tendencies from different located areas and illustrated in Fig. 5.

TABLE 3
THE THREE MAJOR EARTHQUAKES THAT OCCURRED IN SOUTH IONIAN AREA IN WESTERN GREECE DURING THE PERIOD 1993-1997

No.	Point	Magnitude	Distance (km)	Date
8	10082	5.3	147	25/02/94
9	16744	4.9	159	29/11/94
10	16783	4.8	148	01/12/94

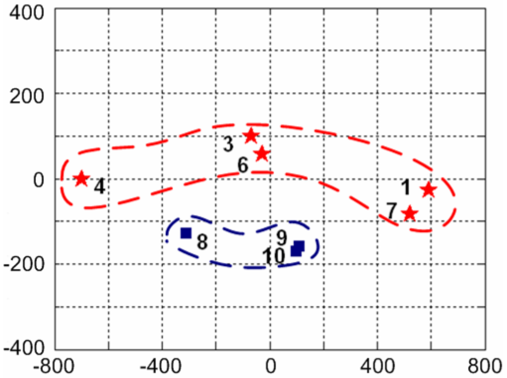

Fig. 5. 2-D MDS plot of the five-strogest events (red stars) in Patras area and the three events (blue squares) in North Ionian area.

A more detailed analysis is provided in the sequel, related to the history of the previously mentioned seismic events. During the recording period, seven significant earthquakes with magnitude Ms ≥4.0R were recorded. The earthquake No. 3 was very strong and caused damages in the city of Patras having its epicenter point very close to our geoelectric monitoring station, as depicted in Fig. 1. The earthquake No. 6 had its epicenter point a few *km* away from the earthquake No. 3. Thus the geometric proximity of the corresponding points in the 2-D graph of Fig. 4 & 5 indicate a similarity in the signal structure prior to significant seismic events that originate from the same area. The same general signal behavior appears at the time interval in the case of earthquake No. 4 and earthquake No. 5, as it is also shown in Fig. 4 & 5. The same clustering tendencies are also apparent in the other pair of earthquakes, No. 1 and No. 7, having their epicenters close enough, as plotted in the epicenter map of Fig. 1. Earthquake No. 2 that has its epicenter point located at another geographic area reveals a different time signature and does not cluster with any of the other earthquakes. Thus the geoelectric recordings display sensitivity to the geotectonic area of the earthquake disturbance. Moreover, the different time signatures of time-series data taken during seismic quiescence are distributed in a more uniform way of visualization, without the creation of any obvious clusters. This can be explained by the absence of a characteristic and prevailing disturbance that would have otherwise modified their temporal behavior. Finally, the three major earthquakes that occurred in the area of North Ionian in Western Greece tend to cluster together as presented in Fig. 5, revealing the different underlying territory pattern of the underwater terrain. By applying dimensionality reduction techniques, the different manifolds of the underlying field distributions are formed, resulting in a clear geographical classification scheme.

V. CONCLUSIONS

In this work we investigated the existence of possible temporal similarities in geoelectric signal recordings, for a period of ten days before significant earthquakes. Experimental results indicate the existence of significant correlation in the patterns of the recorded time sequences for earthquakes originating from the same region. In addition, the appearance of temporally similar precursor signals is consistent with increased seismic activity in the focal zone. This is possibly due to the evolution of the earth's crust towards self-organization at the critical point and involves the formation of structures in the fault zone.

The presented results indicate that the degree of correlation between seismic sequences from the same epicenter region and the similar geoelectric fluctuations could be used to a form a query time-series database system and provide evidence for upcoming earthquake events. The prediction error seems to increase for those earthquakes having either small magnitude or epicenter far away from the geoelectric signal monitoring station. This means that the earthquake generation process affects the LTGP dynamics in a manner dependant upon several parameters (e.g. geological composition, propagation mechanisms, etc).

The proposed method based on a dimensionality reduction technique, provides promising results that can be used jointly with other seismic precursor techniques to solve the difficult problem of short time earthquake prediction**.** However, further investigations are needed to better quantify these findings. The consideration of more LTGP channels into the training algorithm is pursued at present, expected to provide us with extra information and improve both the accuracy and the reliability of this method.

REFERENCES

[1] P. Varotsos and K. Alexopoulos, "Physical properties of the variations of the electric field of the earth preceding earthquakes I & IIQ́ *Tectonophysics*, vol. 110, pp. 73-125, Dec. 1984.

[2] K. Meyer and R. Pirjola, "Anomalous electrotelluric residuals prior to a large imminent earthquake in GreeceQ́ *Tectonophysics*, vol. 125, pp. 371-378, May 1986.

[3] G-A. Tselentis and A. Ifantis, "Geoelectric variations related to earthquakes observed during a 3-year independent investigationQ́ *Geophysical Research Letters*, vol. 23, no. 11, pp. 1445-1448, June 1996.

[4] M. Hayakawa, T. Ito and N. Smirnova, "Fractal analysis of ULF geomagnetic data associated with the Guam earthquake on August 8, 1993Q́ *Geophysical Research Letters*, vol. 26, no. 18, pp. 2797-2800, Sept. 1999.

[5] L. Telesca, V. Cuomo, V. Lapenna, "A new approach to investigate the correlation between geoelectrical time fluctuations and earthquakes in a seismic area of southern ItalyQ́ *Geophysical Research Letters*, vol. 28, no. 23, pp. 4375-4378, Dec. 2001.

[6] A. Ifantis, G. Economou and G.A.Tselentis, "Experimental Investigation of Electrotelluric Field Periodic Anomalies in Western Greece and their Possible Relation to Seismicity during a Five – Year PeriodQ́ *Acta Geophysica Polonica*, vol. 51, no.3, pp. 291-305, 2003.

[7] W.L. Martinez and A.R. Martinez, *Exploratory data analysis with Matlab*, Chapman & Hall/CRC, 2005, ch. 3.

[8] G. Das, D. Gunopulos and H. Mannila, "Finding Similar Time SeriesQ́ in *Proc. of 1st European Conf. on Principles and Practice of Knowledge Discovery in Databases*, Trondheim, Norway, 1997, pp. 88-100.

[9] R. Agrawal, C. Faloutsos and A.N. Swami, "Efficient similarity search in sequence databasesQ́ in *Proc. of 4th Int. Conf. Foundations of Data Organ. and Algorithms*, Chicago, USA, 1993, pp. 69-84.

[10] K.K.W. Chu and M.H. Wong, "Fast Time-Series Searching with Scaling and ShiftingQ́ in *Proc. of 18th ACM Symp. on Principles of Database Systems*, Philadelphia, USA, 1999, pp. 237-248.