Proceedings of the 6th WSEAS International Conference on Multimedia Systems & Signal Processing, Hangzhou, China, April 16-18, 2006 (pp208-214)

An Application of Pattern Recognition techniques for the Analysis of Geoelectrical Signals in relation to the earthquake activity of Western Greece

APOSTOLOS IFANTIS, VASILIS NIKOLAIDIS, GEORGE ECONOMOU*
Department of Electrical Engineering, Control System and Signal Processing Laboratory
Technological Educational Institute of Patras
M. Alexandrou 1 Koukouli, Patras
GREECE
*Department of Physics, Electronics Laboratory, University of Patras, Patras 26500, Greece.

Abstract: - A pattern recognition based approach is used in this study to examine single-channel Long Term Geoelectric Potential difference (LTGP) data recorded during the 1998-2003 period in Western Greece. Seeking a schema that automatically discovers data features containing information possibly related to seismic activity, patterns in LTGP data recorded during small (72-hour) consecutive time segments are examined, and pattern recognition methods are used to link them to seismic information of co-occurring seismic events, such as epicenter and depth.

Key-Words: - Geoelectrical signals, Pattern recognition, Similarity, Earthquake

1 Introduction

The investigation of electromagnetic phenomena as means for short-term earthquake prediction is an on-going process. Many cases of electromagnetic frequencies associated with earthquakes have been reported over the years [1][2][3] and investigations where such measurements are analyzed using sophisticated signal analysis techniques have taken place [4][5][6]. Such studies indicate that preseismic and postseismic geotectonic activity accompaning an earthquake may affect the characteristics of the geoelectrical signals, suggesting the possibility of a prediction schema for oncoming earthquakes [5][6]. However, given that such signals are affected by many factors, a clear correlation between Long Term Geoelectric Potential difference (LTGP) signals and earthquakes is still under investigation. This paper attempts to determine further such possible links between LTGP data and seismic activity information. In particular, we investigate whether there are differences between LTGP data recorded during time periods with significant seismic events occurring in the mainland area of Southwest Peloponnese and data recorded during periods with events in the Ionian Sea region. Both are areas of high seismic activity in Western Greece, a territory with the highest seismic activity in Europe. Furthermore, we also study a possible relation between LTGP data and epicenter depth of the aforementioned seismic events. This work focuses on data recorded during a six year time period (1998-2003), when several major earthquakes occurred in the examined area (Table 1). Recordings have been collected during a independent experimental investigation at the University of Patras Seismological Laboratory (UPSL).

1.1 Data Acquisition System

LTGP data is measured between pairs of electrodes placed in the ground at specific locations. The electric field is continuously monitored, usually in two perpendicular directions (N-S and E-W), by an appropriate number of electrodes. In this system the monitoring of the LTGP potential difference is achieved by four sets of dipoles arranged in short as well as long distances (Fig.1). These dipoles make use of Pb-PbCl2 electrodes. Two sets have an electrode separation of 100m and are adjusted towards the N-S (ch0) and E-W (ch1) one perpendicular to each other. The third set (ch2) has an electrode separation of 300m and is directed towards the NE-SW. The fourth set has an electrode separation of 3000m in the direction NW-SE (ch4). The dipoles are set on the outskirts of the University of Patras, in Rio, in a rather quiet countryside and are based in Pleistocene compact conglomerates. For the purpose of this analysis, only N-S (ch0) data was used. The obtained LTGP signal appears in Figure 2, with the vertical axis showing the potential in mV and the horizontal axis the time interval in hours. Figure 2b illustrates the major seismic events (with magnitude >= 4.8 Ms) in the same period, the vertical axis being the magnitude of earthquakes in Richter scale and horizontal axis the time interval in hours. Details of these significant seismic events are listed in Table 1, while their epicenters appear in Figure 3.

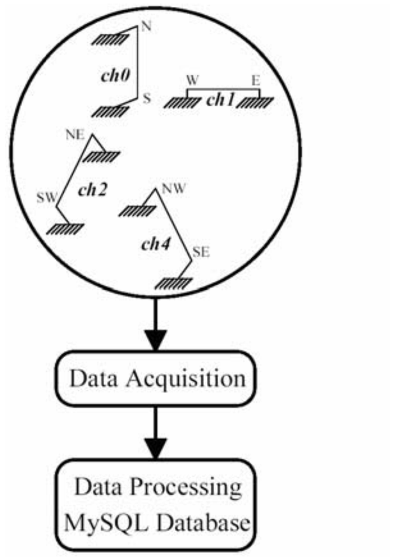
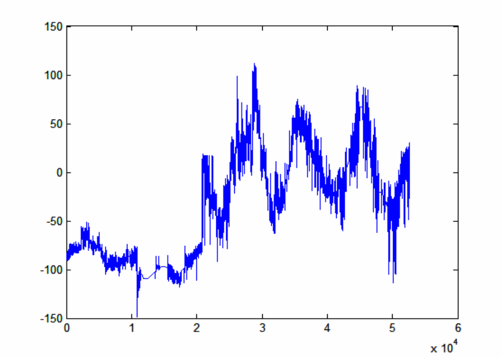
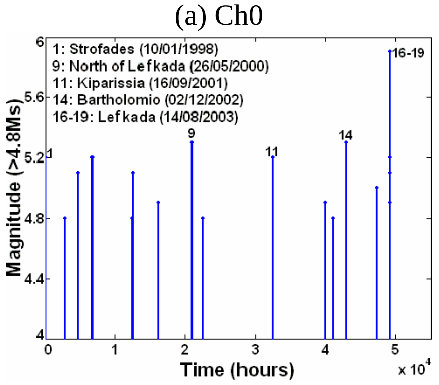


Fig.1 Data Acquisition System



(a) Ch0



(b) Seismic Events

Fig.2 Geoelectic and Seismic signals

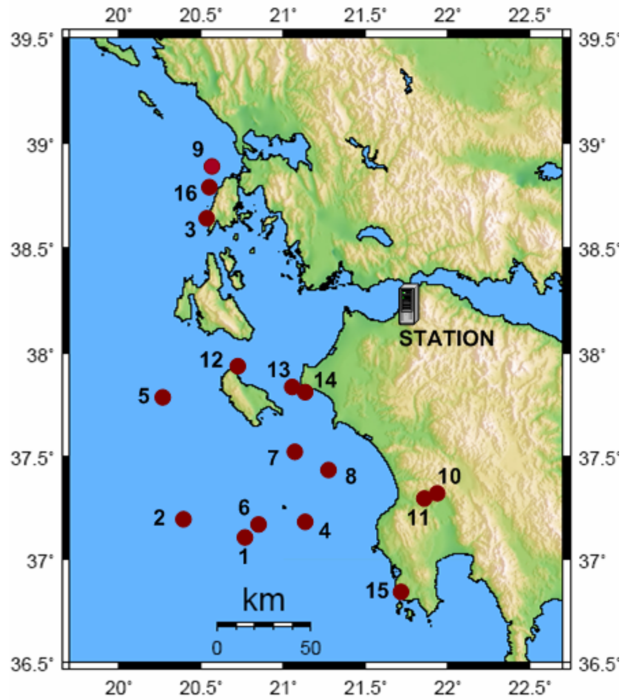


Fig. 3. Epicenters of the major earthquakes in Patras area in Western Greece during the period 1998-2003.

2 Problem Formulation

The six-year original LTGP data consists of 52560 samples per channel, each value being the average of 180 actual samples taken hourly and corresponding to an hour of data acquisition. Although the original data was recorded in four channels, this analysis only uses data from channel 0 (Figure 2a). Correlation of this data with seismic activity information is to be investigated.

3 Problem Solution

3.1 Data Preprocessing

In order to create coordinate vectors which would be used in pattern recognition analysis, preprocessing was applied to the channel 0 LTGP data. Initially a low frequency filtering step was applied, with each sample being replaced by its difference to the channel's average signal level during the past five days (i.e. the mean value of the previous 120 samples).

#	Date	Time (GMT)		LAT (N)	LON (E)	Depth	Mag
1	10/01/1998	19	21	37.12	20.73	5	5.2
2	01/05/1998	04	00	37.20	20.43	5	4.8
3	16/07/1998	17	29	38.66	20.56	5	5.1
4	06/10/1998	12	27	37.19	21.13	5	5.2
5	08/10/1998	03	50	37.79	20.27	5	5.2
6	07/06/1999	23	55	37.18	20.83	5	4.8
7	11/06/1999	07	50	37.56	21.11	58	5.1
8	04/11/1999	02	08	37.45	21.32	12	4.9
9	26/05/2000	01	28	38.91	20.58	5	5.3
10	25/07/2000	19	33	37.31	21.91	5	4.8
11	16/09/2001	02	00	37.29	21.83	5	5.2
12	28/7/2002	17	16	37.95	20.73	19	4.9
13	14/09/2002	19	50	37.81	21.06	8	4.8
14	02/12/2002	04	58	37.80	21.15	17	5.3
15	29/04/2003	01	51	36.83	21.72	67	5.0
16	14/08/2003	05	14	38.79	20.56	12	5.9
17	14/08/2003	08	41	38.81	20.56	14	4.9
18	14/08/2003	12	18	38.76	20.67	8	5.1
19	14/08/2003	16	18	38.76	20.67	9	5.2

Table 1: List of the major earthquake events occurring in Western Greece during the 1998-2003 period.

Since no average signal level could be computed for the first 120 samples, these were removed from further analysis. The remaining 52440 samples were then sliced into 728 intervals each corresponding consecutive 72-hour periods (3-day segments). Data corresponding to the last 24 hours of the six year period was removed at this point, since it could not form a complete 72 hour section. Next, the Discrete Fourier Transform (DFT) of the data in each 72-hour segment was recorded, resulting in a 728 x 72 dataset spanning more than 99% of the original six year period. Finally, Principal Component correlation Analysis (PCA) was performed on the dataset, to expose a set of new orthogonal axes on which the data reaches maximum variance [7][8]. On the less significant PCA axes the projected data had very low variance, allowing the removal of these axes from further consideration i.e. the reduction of the dimensionality of the problem from 72 to 30 coordinates with little data loss. A 3-d projection of the resulted 72 x 30 dataset is shown in Figure 4.

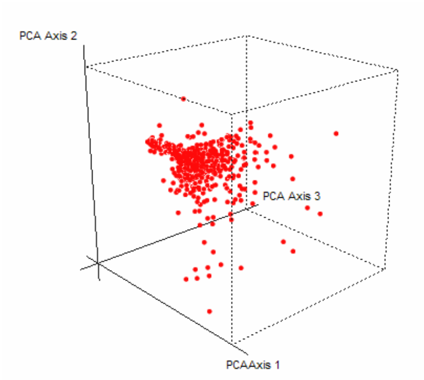


Fig.4. A 3-d projection of the entire data.

3.2 Event Epicenter Location (Unsupervised)

The next step taken involves taking into account the recorded earthquake data and link it to possible patterns in the data. As mentioned earlier, the 728 data vectors were created using an automated process in which time was partitioned into consecutive segments. These segments were flagged according to the co-occurring recording seismic activity. However, segments corresponding to periods of major activity (containing a major earthquake) could contain any ratio of preseismic or postseismic time, since earthquakes could have occurred at any point in the segment. Also, some of these periods happen to contain more than one major seismic event. The majority of the segments do not contain periods of major seismic activity, but do coincide with low magnitude seismic events which are very common in the examined geographical region (as discussed later). Focusing on the analysis of exceptionally active seismic periods, periods of minor or no earthquake activity were removed from further pattern recognition analysis. This resulted in a 15 x 30 dataset. The 15 records correspond to active time segments (periods), but since some of the 19 major earthquakes (Table 1) occurred during time assigned to a single segment there are less active periods than earthquakes. As seen in Table 2, events 4 and 5 were both found in period 92, while events 16, 17, 18 and 19 are in period 682. In fact, events 16, 17, 18 and 19 occurred within hours from each other and have the same epicenter. Each period is then assigned to a class. The classes are created based on some seismic activity information related to the period. Here, information relating to the epicenter location of the seismic activity was used; in particular, two cases were defined:

(a) Periods with major earthquake activity in the Ionian Sea region (class A, 12 periods).

(b) Periods with major earthquake activity in the Southwest Peloponnese (mainland) region (class B, 3 periods).

The resulting partitioning of time segments with major seismic events is listed in Table 2. Automated analysis or classification of the LTGP data is then performed, and the two partitioning results are compared. In this case, however, visual inspection of the 2-d scatter plot of the LTGP data projected on the first and second principal component axes reveals strong separation of the data attributed to the two cases (Figure 5). On the most significant PCA axis (the axis where data reaches maximum variance, horizontal axis in Figure 5) all class A data was distributed in the lower value range away from class B data, and suggesting little need for further pattern recognition analysis.

Period #	Period Start Time (Hours)	Period End Time (Hours)	Major Seismic Events in Period	Class Assign.
2	192	264	1	A
39	2856	2928	2	A
64	4656	4728	3	A
92	6672	6744	4,5	A
173	12504	12576	6	A
174	12576	12648	7	A
223	16104	16176	8	A
291	21000	21072	9	A
311	22440	22512	10	B
450	32448	32520	11	B
555	40008	40080	12	A
571	41160	41232	13	A
597	43032	43104	14	A
657	47352	47424	15	B
682	49152	49224	16,17,18,19	A

Table 2: Periods with major seismic events and related class assignments.

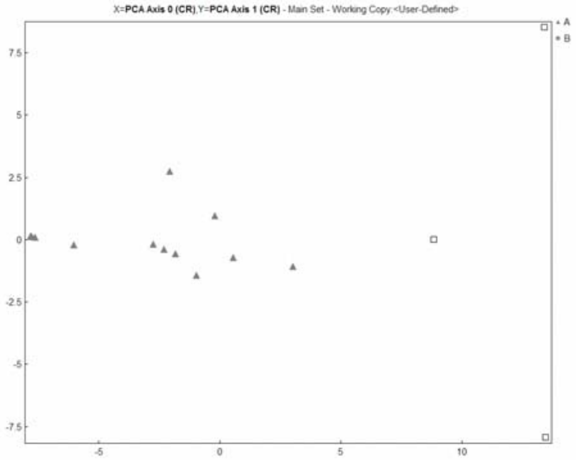


Fig.5 Projection of the data on the two most significant principal component axes showing the two cases of earthquake activity.

Still, to verify that the partitioning still holds when the full 30-d coordinate space is considered, three classic unsupervised clustering methods were applied to the LTGP data:

(a) *k-means* [10] The method was set to create 2 classes and run for a limit of 1000 iterations. Euclidean distance was used as the similarity metric. The method was initialized with cluster centers which were randomly picked vectors from the data. To eliminate output instabilities caused by the random initialization, the method was executed 20 times, 17 of which produced a clustering identical to the original two-case partitioning of Table 2 (Figure 6).

(b) *LVQ neural network* [11][12] Set to create 2 classes and running for 501 iterations, the method also resulted in clusters identical to those of the original two-case partitioning. This was expected since the method is similar, in principle, to k-means.

(c) *Single link hierarchical clustering* [7] with Euclidean distance used as the similarity metric. Seeking links between nearest neighbors the 4-class partitioning resulted in a single class containing all elements of original class A, while the records of class B formed single one-element classes (Figure 7).

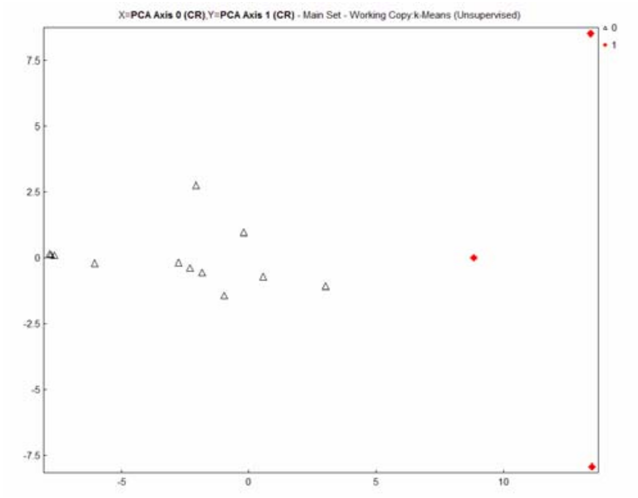


Fig.6. Projection of the data on the two most significant principal component axes, results of k-means clustering shown.

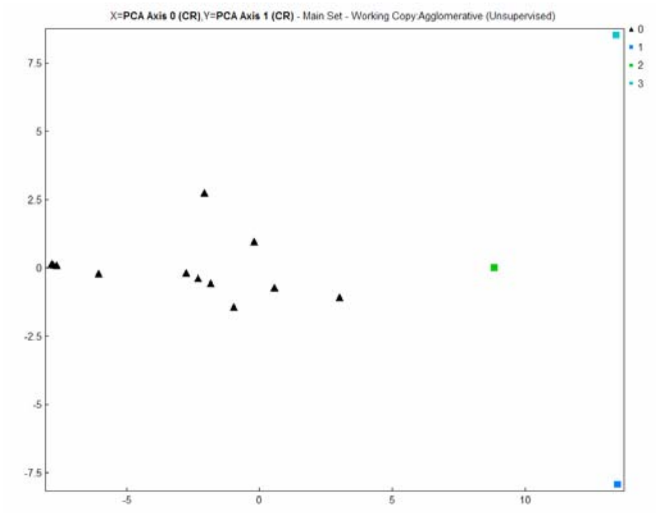
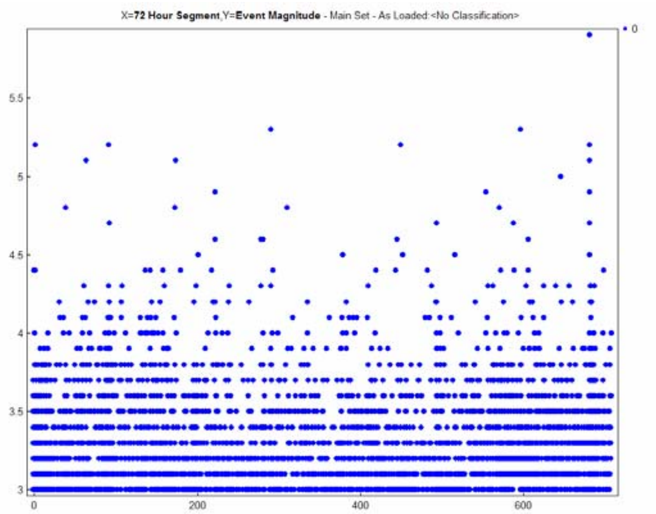


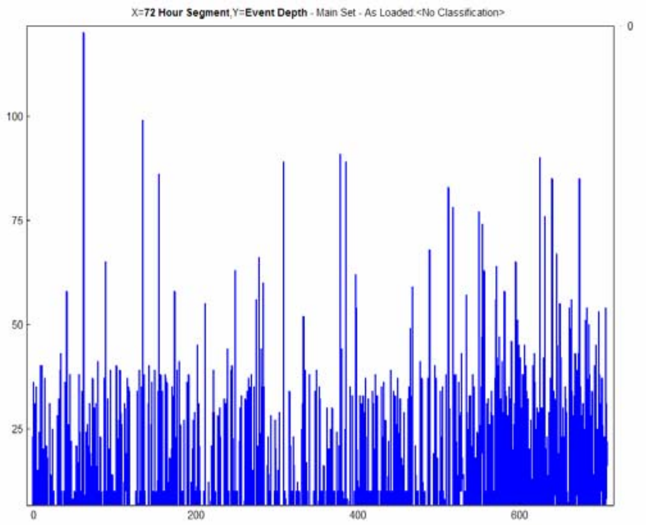
Fig.7. Projection of the data on the two most significant principal component axes, results of hierarchical clustering shown.

3.3 Event Depth (Supervised)

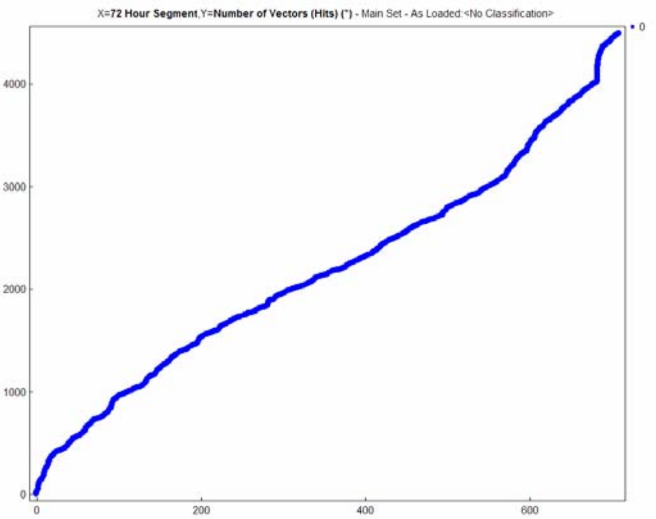
As mentioned earlier, the geographic area studied in this work is one of high seismic activity. By decision, this study focuses on the 19 major seismic events with magnitude of at least 4.8 Ms and listed in Table 1. This creates a manageable dataset; should this threshold be lowered to a magnitude of 3 Ms, the number of events to be examined would increase to about 4500, as shown in Figure 8.



(a)



(b)



(c)

Fig.8. Magnitude (a), depth (b) and rate (c) of seismic events with magnitude >= 3 occurring in Western Greece during the 1998-2003 period. Time units (in horizontal axis) are 72-hour periods.

Figure 8a shows the magnitude of the events, 8b is the depth of their epicenter, and 8c shows the cumulative number of these events (indicating the rate in which the events occurred. In all plots of Figure 8, the horizontal axis is time measured in 72-hour units, starting 120 hours past 1/1/1998, 00:00 GMT and consistent with the time segments resulting from the preprocessing process earlier. Examining the plots, a change in overall seismic behavior is evident, starting around time segment 550 and onwards. During segments 550 to 728 there is an increase in the rate of seismic events (Figure 8c), while the average depth of the events increases (Figure 8b). A significant number of major seismic events (events 12 to 19 of Table 1) also occur during this time.

Returning to the LTGP data, Figure 9 shows the values of the most significant principal component axis for each of the 728 time segments (with time also measured in 72-hour periods). Examining this plot, one notes that: (a) in segments 0 to 300 values are mostly negative and expose little variance besides few extreme values, (b) in segments 300 to 550 values become mostly positive, while (c) in segments 650 to 728 remain mostly positive but also display most variance often reaching negative values.

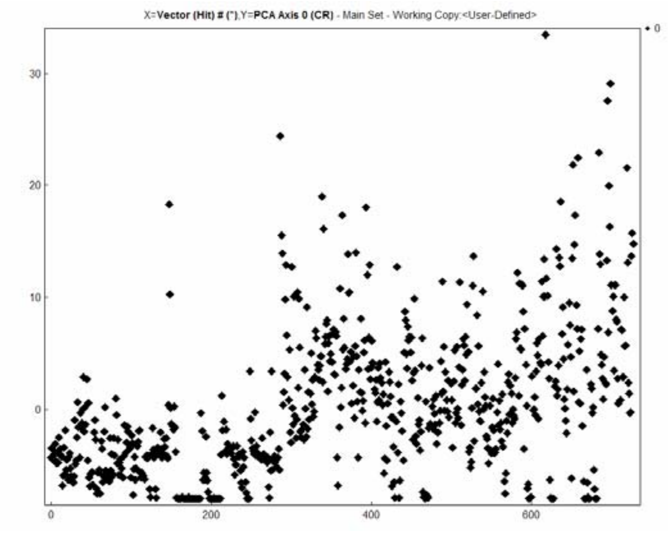


Fig.9. Values of LTGP data projected on the most significant principal component axis vs. time (in x-axis, expressed as 72-hour periods).

Prompted by such observations, and refocusing on the 19 major seismic events, a supervised method was trained to correlate the LTGP data of a time segment with the depth of co-occurring seismic events. The 15 x 30 dataset originally used in the unsupervised analysis earlier (consisting of 15 periods which in turn correspond to the 19 major events of Table 1) was split into two subset datasets, one set (of size 7) to be used for training the methods and one (of size 8) for testing and evaluating the trained method. The strategy used to create these sets is "interlaced half", i.e. every other record of the original dataset is placed in testing set, while the remaining data is considered known data and is used for training. In particular, a 7 x 30 training set containing data from time segments 39, 92, 174, 291, 450, 571, and 657, and a 8 x 30 testing set containing data from time segments 2, 64, 173, 223, 311, 555, 597, and 682 were created. All segments were assigned a class, according to the depth of co-occurring seismic events (or maximum depth if the segment contains more than one seismic event). Two classes were created, one for events of reported depth less that 10 km, a second for the remaining items. The training set combined with the class assignments were then used to train two supervised methods:

(a) *1-NNC* [9] The nearest neighbor classifier simply classifies each data point of the testing set to the class of the nearest item in the training set. Euclidean distance is used as the metric of similarity. Comparing the resulting classification with the known (correct) class, the method achieves 75% classification success, correctly classifying 6 out of the 8 records in the testing set.

(b) *Back-propagation Perceptron* [11] A well known multilayer neural network which uses the Generalized Delta Rule [13][14] to adjust connection weights and processing element biases. The encoding process is an iterative one, repeated until a satisfactory low error level (here 0.005) is attained or a maximum number of iterations (here 1000) is reached; global learning rate parameter was set to 0.6. A 30 x 30 x 2 topology was used, i.e. a single hidden layer of 30 elements, and one output element for each of the two classes. While training, the value of 1 was considered to be correct output of an element in the output layer if the input vector was of the class to which the element corresponds, 0 otherwise. The method reached the acceptable error level and stopped training after 80 iterations. Recall (or classification) is done in a single step where input values are placed in the input layer and the values towards the output layer are computed. When this process was applied to each vector of the testing set, the method achieved 87.5% classification success, correctly classifying 7 out of the 8 records in the set.

4 Conclusion

Unsupervised analysis of single-channel LTGP data recorded during 1998-2003 in Western Greece indicates that the data which co-occurred with the three mainland earthquakes (events 10, 11 and 15) is well separated from the data recorded during periods with earthquakes in the Ionian Sea region. Should the recorded LTGP be indeed affected by such activity, the partitioning could be attributed to differences of the geotectonic structures producing the two types of events, probable different earthquake generation mechanisms involved, as well as different signal propagation paths. Furthermore, the relatively high classification success of supervised methods trained to classify seismic activity depth using the same LTGP data is another indication that LTGP data may be linkable to seismic activity.

Acknowledgements

This work was supported by the by the National Ministry of Education under the project entitled “Soft computing techniques for multichannel processing”, in the context of the program “Archimidis

References:

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

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

[3] G-A. Tselentis and A. Ifantis, Geoelectric variations related to earthquakes observed during a 3-year independent investigation, *Geophysical Research Letters*, Vol.23, No.11, 1996, pp. 1445-1448.

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

[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 Italy, *Geophysical Research Letters*, Vol.28, No.23, 2001, pp. 4375-4378.

[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 Period, *Acta Geophysica Polonica*, Vol.51, No.3, 2003, pp. 291-305.

[7]M. R. Anderberg, *Cluster Analysis For Applications*, Academic Press, 1973.

[8]R. A. Johnson and D.W.Wichern, *Applied Multivariate Statistical Analysis*, Prentice-Hall International, 1992.

[9]Batchelor, B. G. *Practical Approach to Pattern Classification*, Plenum Press, 1974.

[10]Tou, J.T. and Gonzalez, R. C. *Pattern Recognition Principles*, Addison-Wesley, 1974.

[11]Simpson, P. K. *Artificial Neural Systems*, Pergamon Press, 1990.

[12] Kohonen, T. *Self-Organization and Associative Memory*, Springer-Verlag, 1988.

[13] Pao Y. *Adaptive Pattern Recognition and Neural Networks*, Addison-Wesley, 1989.

[14] Lippmann, R. P. *An Introduction to Computing with Neural Nets*, IEEE ASSP Magazine 14, 1987, pp 4-22.