Expert Systems with Applications 38 (2011) 15032–15039

Contents lists available at ScienceDirect

Expert Systems with Applications

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

Artificial neural networks for earthquake prediction using time series magnitude data or Seismic Electric Signals

Maria Moustra, Marios Avraamides, Chris Christodoulou *

Department of Computer Science, University of Cyprus, 75 Kallipoleos Avenue, P.O. Box 20537, 1678 Nicosia, Cyprus

ARTICLE INFO

Keywords:
Neural networks
Earthquake prediction
Seismic Electric Signals

ABSTRACT

The aim of this study is to evaluate the performance of artificial neural networks in predicting earthquakes occurring in the region of Greece with the use of different types of input data. More specifically, two different case studies are considered: the first concerns the prediction of the earthquake magnitude (M) of the following day and the second the prediction of the magnitude of the impending seismic event following the occurrence of pre-seismic signals, the so-called Seismic Electric Signals (SES), which are believed to occur prior to an earthquake, as well as the time lag between the SES and the seismic event itself. The neural network developed for the first case study used only time series magnitude data as input with the output being the magnitude of the following day. The resulting accuracy rate was 80.55% for all seismic events, but only 58.02% for the major seismic events ($M \geqslant 5.2$ on the Richter scale). Our second case study for earthquake prediction uses SES as input data to the neural networks developed. This case study is separated in two parts with the differentiating element being the way of constructing the missing SES. In the first part, where the missing SES were constructed randomly for all the seismic events, the resulting accuracy rates for the magnitude of upcoming seismic events were just over 60%. In the second part, where the missing SES were constructed for the major seismic events ($M \geqslant 5.0$ on the Richter scale) only by the use of neural networks reversely, the resulting accuracy rate by predicting only the magnitude was 84.01%, and by predicting both the magnitude and time lag was 83.56% for the magnitude and 92.96% for the time lag. Based on the results we conclude that, when the neural networks are trained by using the appropriate data they are able to generalise and predict unknown seismic events relatively accurately.

1. Introduction

Earthquakes are one of the most costly natural hazards faced by the nation in which they occur without an explicit warning and may cause serious injuries or loss of human lives as a result of damages to buildings or other rigid structures. During the last decades there has been an increasing interest and academic research on predicting seismic events. In the effort to predict earthquakes, people and researchers have tried to associate an impending earthquake with such varied phenomena as seismicity patterns, electromagnetic fields, weather conditions and unusual clouds, radon or hydrogen gas content of soil or ground water, water level in wells and animal behaviour (like for example the recently reported pre-seismic anticipatory behaviour in the common toad *Bufo bufo*, see Grant and Halliday (2010) and references therein for reports of seismic activity responses of other species). Earthquake prediction, which aims to specify three elements, namely when, where and how large the impending earthquake will be, constitutes the most important unsolved problem of seismology.

In the past, various efforts have been made to solve this particular problem. These efforts led to the construction of models, which attempted to comprehend the nature of seismic phenomena and predict high magnitude (M) seismic events based on different approaches. Some of the most important efforts which are related to this study are reviewed below. For a detailed survey of earthquake prediction efforts, see Adeli and Panakkat (2008).

One of the most debated methods for earthquake prediction is a method called "VAN", after the initials of three Greek scientists from the University of Athens, Varotsos, Alexopoulos and Nomicos (Uyeda, 1997). These scientists found that transient variations of the earth's electric field, known as Seismic Electric Signals (SES) are observed before an earthquake. The SES are used to determine the location and the magnitude of the impending earthquake. After a SES is recorded, an earthquake occurs within several days to several weeks based on the SES's type (Varotsos & Alexopoulos, 1984a, 1984b). To determine the epicentre of an impending earthquake, a process of elimination of the possible seismic areas is applied, including the selectivity effect, the polarity effect and the ratio of

* Corresponding author. Tel.: +357 22892752; fax: +357 22892701.
E-mail addresses: cs04mm2@cs.ucy.ac.cy (M. Moustra), cs03am1@cs.ucy.ac.cy (M. Avraamides), cchrist@cs.ucy.ac.cy (C. Christodoulou).

0957-4174/$ - see front matter
doi:10.1016/j.eswa.2011.05.043

the two components of the SES. The magnitude of the earthquake is estimated using the data from the specific station which was collected for the same seismic area in the past (Varotsos & Lazaridou, 1991). The VAN group scores their prediction "successful" when the actual earthquake occurred within several days to several weeks after the precursor SES is recorded, within ca. 100 km from the predicted epicentre and within ca. 0.7 units of the predicted magnitude on the Richter scale (RS). Based on these criteria, about 60% of their predictions are successful and about 60% of Greek earthquakes of $M > 5.3$ on the RS are successfully predicted (Uyeda, 1997). However, this success has not been widely recognised by the scientific community, including the Greek seismological community. Some argue for example, that the VAN's SES are all noise unrelated to earthquakes, and others persist that the success is attained by chance (Uyeda, 1997).

Bodri (2001) attempted to relate neural network ideas to seismic activity patterns in the Carpathian–Pannonian area of Hungary, and the Pelloponnesos–Aegean region of Greece. A three-layer feed-forward multilayer perceptron neural network model using error backpropagation as learning algorithm was developed for the prediction of the origin times of large earthquakes ($M \geqslant 6.0$ on the RS). The network used as input the mean seismicity rates (number of earthquakes per unit time) in selected time intervals, and more specifically within the time intervals between the recorded $M \geqslant 6.0$ (RS) earthquakes. The results of this effort were particularly satisfactory despite the fact that the training set was inadequate because of the infrequent occurrence of large earthquakes. The neural networks managed to predict the origin times of such events with a deviation of ±6 months. The impressive performance of the neural networks revealed the usefulness of such tools in the problem of earthquake prediction.

Lakshmi and Tiwari (2007) examined the temporal evolution of seismicity of the Central Himalaya (CH), Western Himalaya (WH) and Northeast Himalaya (NEH). A multilayer feed-forward artificial neural network (ANN) model was developed to simulate monthly resolution earthquake frequency time series for the three regions. The learning algorithm used was the backpropagation with gradient descent optimisation technique. Cross-validation was also utilised to test the networks generalisation ability. The data used concerned the seismic events which occurred in the period of 1960–2003 and for magnitude of $M \geqslant 4$ (RS). The earthquake monthly frequency data was used as input values to the neural network. More specifically, a temporal sequence of the previous five monthly frequency data was selected as input. The frequency value of the next month was used as output of the network. According to the sum-squared error that was calculated for each region, in order to measure the differences between the actual and predicted values, the results obtained by the ANN model were described as reasonably good. Furthermore, the results showed that the earthquake dynamics in the regions of WH and NEH are better "organised" than in the CH, since the earthquake processes of WH and NEH have a higher predictive correlation coefficient, at 50–55%, in contrast to the CH which has 30%.

Lakkos, Hadjiprocopis, Comley, and Smith (1994) used a feed-forward neural network which was simulated using the XERION software package (van Camp, 1993) and the Delta-Bar-Delta as training algorithm (van Camp, 1993) to predict the geographical location (longitude and latitude) and the magnitude of an impending earthquake. The input data presented to the network consisted of the two components of the SES of the VAN method (East–West, North–South). The data used for training was collected by the monitoring station of Ioannina in North Western Greece, but the data quantity was not sufficient. Thus they expanded the original data set by a factor of five by adding a small amount of Gaussian noise. After testing the neural network using data that was not part of the training data set, the results showed that the network was able to give more accurate predictions for the geographical area of 20.0°E–21.5°E and 37.5°N–40.0°N since the majority of training data was associated with this area. In addition, the epicentre location can be predicted with an error of less than 0.3° and the magnitude with an error of less than 0.5 on the RS (Lakkos et al., 1994). However, the authors do not clarify the magnitude range of the data used for training and testing.

Another example is the probabilistic neural network (PNN) that was implemented for the magnitude prediction of the largest earthquake in a pre-defined future time period (Adeli & Panakkat, 2009). More specifically, the future time period that they considered was the following fifteen days and the region examined was Southern California. The PNN takes as input eight mathematical earthquake parameters called seismicity indicators (Gutenberg & Richter, 1956) and classifies the predicted magnitude in one of the several output classes. The indicators are: the time elapsed during a particular number (n) of significant seismic events before the month in question, the slope of the Gutenberg-Richter inverse power law curve for the n events, the mean square deviation about the regression line based on the Gutenberg-Richter inverse power law for the n events, the average magnitude of the last n events, the difference between the observed maximum magnitude among the last n events and the expected ones using the Gutenberg–Richter relationship (Gutenberg & Richter, 1956) known as the magnitude deficit, the rate of square root of seismic energy released during the n events, the mean time or period between characteristic events, and the coefficient of variation of the mean time. Three different statistical measures have been used for the model's evaluation: the probability of detection, the false alarm ratio and the true skill score. According to the results based on these three metrics, the PNN gave good prediction accuracies for magnitudes between 4.5 and 6.0 (RS), but not for magnitudes greater than 6.0 (RS).

Artificial neural networks (ANNs) have been used by many researchers to investigate their potential as a tool for simulation of the behaviour of systems that are governed by nonlinear multivariate data and generally unknown interconnections within a noisy, poorly-controllable physical environment. The advantage of this framework is that the ANN provides a black-box approach and the user does not need to know much about the nature of the process being simulated. Considering this advantage, in the present paper we describe two test cases for earthquake prediction for the region of Greece by applying ANNs. A number of inherent features of ANNs make them suitable for such a task and for a huge number of other applications (see for example Velido, Lisboa, and Vaughan (1999) and Paliwal and Kumar (2009), for surveys). More specifically, in contrast to model-based techniques, ANNs are data-driven self-adaptive (through learning) systems, as they do not need any *a priori* assumptions with regards to the models of the scenarios being investigated, or if they do, they are minimal. ANNs can usually generalise pretty well after they are trained with a sample of the data, which could even be noisy (Haykin, 2009).

The remainder of the paper is structured as follows: Section 2 describes the first case study which concerns earthquake prediction using only time series magnitude data, Section 3 describes the second case study which concerns earthquake prediction using the SES of the VAN method discussed above, and finally Section 4 gives a discussion and conclusions.

2. Case study I: earthquake prediction using only time series magnitude data

2.1. Methodology and data preprocessing

The first case study outlines a methodology for predicting the magnitude of the most important seismic event for the following day in Greece, using only time series earthquake magnitude data. A three-layer perceptron neural network with the backpropagation learning algorithm (Rumelhart, Hinton, & Williams, 1986) was implemented. The backpropagation algorithm is the most commonly used one and has been applied successfully to a broad range of problem types and fields, such as pattern recognition, regression as well as prediction problems. As mentioned above, the data presented to the network form a time series of a single input, which is the maximum magnitude of seismic events for each day in a given period. Rather than present a single input to the network, the data is presented as a number of days at a time (that number is referred to as the depth of the network). This is in essence a moving time window across the data set, with width equalling the depth, so each day of the data is presented multiple times with a varying set of neighbouring days' data.

The data was derived from the Seismological Institute, National Observatory of Athens (SINOA) and covered a period of 21 years, and more specifically from 01/01/1980 to 01/01/2001. An input file for the neural network was created containing the normalised maximum magnitude value for each day. For any day where there was no seismic activity a zero was recorded in the file for the specific day as the value of the magnitude. In addition, we assume that the location of occurrence of the seismic event is "known" since the network's training concerns data regarding the area of Greece only (i.e., we assume that the "where" independent variable is constant).

For the evaluation of the model, the accuracy rate was calculated based on the Mean Absolute Error (MAE), which resulted from the prediction of all the seismic events that were included in the dataset. Supposing that the maximum deviation between the actual (target) and the predicted value was related to the maximum percentage error (100%), we calculated the percentage error of the corresponding MAE. The difference between the maximum success rate (100%) and the percentage error is equal to the accuracy rate of the prediction.

Before proceeding to the analysis of the results, we tested the effect of a variety of values for the different parameters of the backpropagation algorithm (learning rate, momentum), as well as the width of the input moving time window and the architecture of the network, aiming for the best possible performance of the neural network. Based on this experimentation, the optimal architecture of the network was found to be a network with one hidden layer of 50 hidden nodes, while the optimal width of the time moving window, which corresponds to the number of inputs of the network, was found to be 30 (corresponding to 30 consecutive days of data prior to the day for which a prediction was made). In addition, the network appeared to be more efficient with the learning rate at 0.4 and the momentum at 0.8. All neurons in the network use a sigmoidal activation function.

Certainly, it has to be noted that the big challenge in earthquake prediction and the true test of the capability of the neural network is to be able to predict the next major seismic event. Those events constitute the outliers for the particular application and they have devastating consequences in the regions they occur. If an accurate forecast is achieved for such seismic events, then the results would be beneficial for the population since the danger of loss of human lives would be minimised. For this reason, we decided to examine also the detection of the seismic events that are considered outliers and find the accuracy rate that the network achieves for their prediction. The description of the procedure for detecting and predicting the major seismic events is explained in the following paragraphs.

After considering some methods for detecting outliers, we decided that for the particular application the Boxplot method would be more useful (Hoaglin, 2003). This decision was taken because the Boxplot method would detect the possible outliers in the entire dataset, while considering the seismic activity that was recorded in the entire time period which is examined. Moreover, the Boxplot method derives a graphic result and facilitates the examination of the outliers.

By observing Fig. 1, where the graphical result of the Boxplot method is presented, we note that the minimum and maximum values of all the data are 2.10 and 5.10, respectively. Given that these two values are the minimum and maximum threshold for the magnitude, we conclude that any value less than or equal to 2.00 and greater than or equal to 5.20 on the RS is considered as an outlier. It is remarkable that the results show that for the period that we examine, 50% of the recorded magnitudes are greater than 3.70 on the RS.

2.2. Results

The accuracy rate that resulted from the prediction of *all* the seismic events that occurred in Greece for the above-mentioned period is 80.55%. As a general observation, we can say that the achieved accuracy rate is relatively high and consequently we conclude that once the ANN is trained by using the appropriate data, it is able to generalise and predict unknown seismic events accurately.

Moreover, as we already mentioned in Section 2.1, we also examined the efficiency of the model to predict the outliers ($M \geqslant 5.2$, RS). The accuracy rate that resulted from the prediction of the outliers by training the neural network model is 52.81%. The accuracy rate was calculated based on the MAE as described in Section 2.1, but the MAE resulted from the prediction of the major seismic events instead of all the seismic events.

In order to improve the performance and the accuracy of the ANN in predicting major seismic events (outliers) we attempted to train the network in two phases. The first phase involved the training of the network using only the outliers as training data. More specifically, the magnitudes of the major seismic events ($M \geqslant 5.2$, RS) were presented to the network as inputs one by one. Afterwards, in second phase the network was trained using the time series of all the seismic events that we had available, as described before. The weights of the network which resulted from the last iteration of the first training phase were saved to be used at

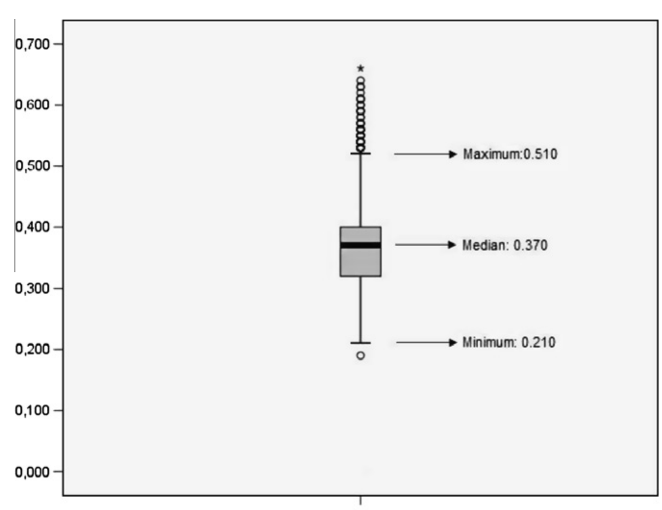

Fig. 1. The graphical result derived by applying the Boxplot method for the historical magnitude seismic data of Greece. It shows that any value less than or equal to 2.00 and greater than or equal to 5.20 on the RS is considered as an outlier, given that the minimum and maximum magnitude threshold values of all the data are 2.10 and 5.10, respectively.

the second training phase. Thus, instead of the random initialisation of the weights before the second training phase was initiated, the weights from the first phase were loaded. In this way, when the network starts the second training phase, it already has a reduced error and consequently we would expect better predictions. The accuracy rate resulted by training in two phases is 58.02%.

3. Case study II: earthquake prediction using Seismic Electric Signals

The second case study concerns the use of Seismic Electric Signals (SES) which were recorded and utilised by the VAN team in Greece. After a major earthquake hit Athens in 1981 causing serious injuries, the VAN team proposed that some electric current may be generated in the earth region of the earthquake before the event, so they started measuring the electric field of the earth (Uyeda, 1997). Indeed, the VAN team found that these electric field variations of the earth which they measured and studied, called SES, precede earthquakes (Varotsos & Alexopoulos, 1984a). More specifically, the SES are defined as low frequency ($\leqslant$1 Hz) electric signals and have been first observed in Greece (Varotsos, Sarlis, Skordas, & Lazaridou, 2006).

The measurement of SES is relatively simple. The components of the electric field are measured between two electrodes located into the ground at desired distances in two perpendicular directions, East–West (E–W) and North–South (N–S). Because of the geomagnetic field variations, rain fall, man-made noise, electro-chemical instabilities of electrodes, etc., the geoelectric potential is constantly changing (Uyeda, 1997). Thus, the precursor signals must be distinguished from this noise. This difficult task was only achieved by the VAN group who discovered that these pre-seismic signals exist only if one chooses the correct locations (sensitive stations) (Uyeda, 1997).

In the current case study we attempted to predict the impending earthquakes using these precursory SES as input data to the neural network. Unfortunately, the number of the SES that the VAN team published was not sufficient for training and testing the ANNs. As a result, we decided to use additional data from the seismicity catalogues of the Seismological Institute, National Observatory of Athens (SINOA) and construct the corresponding SES.

The study is mainly based on 29 original SES which were recorded by the VAN team before the corresponding seismic events. These signals occurred between 17 January 1983 and 18 November 1997 and were published in Varotsos (2005) and Uyeda, Al-Damegh, Dologlou, and Nagao (1999). This case study is separated in two parts with the differentiating element being the way of constructing the additional SES to be used as input data to the neural networks. In the first part the SES are constructed randomly for all seismic events, whereas in the second part the SES are constructed for major earthquakes only by using ANNs reversely.

3.1. Constructing additional SES randomly

3.1.1. Methodology and data preprocessing

The data for the seismic events that occurred in Greece which was derived from the SINOA, covered the period between 01/01/1980 and 01/01/2001. Given that we had the original SES for the 29 seismic events, we needed to construct the SES for the remaining seismic events.

Based on experiments, Varotsos and Alexopoulos (1984a, 1984b) showed that the interesting quantity of each SES is the maximum value of the potential change ΔV, in other words the signal measured between the two electrodes. Thus, according to this conclusion and Eq. (1) which gives the relationship between the SES amplitude expressed as $\Delta V/L$ and the magnitude M (where L is the distance between the two buried in the earth electrodes used for the signal's measurement), that results in a linear dependency between them with an intercept b (Varotsos, Alexopoulos, & Lazaridou, 1993), an attempt was made to construct the remaining SES.

$$\log\frac{\Delta V}{L} = (0.34 - 0.37) * M + b \tag{1}$$

The plots for dipoles of the two different orientations have the same slope (0.34–0.37), but different intercepts (Varotsos, 2005). Since the earthquake magnitude and $\Delta V/L$ are given for the 29 seismic events which we had access to, we only needed to find the constant b for each station. The SES for the 29 seismic events were recorded at two different stations, IOA (Ioannina in north western Greece) and PIR (Pirgos in the Pelloponese). Consequently, we needed to find the constant value of b, for each one of the two stations. The b values for the two stations were found to be $b_{IOA} \approx 1.2$ and $b_{PIR} \approx 1.2$ by linear approximation using Eq. (1). Then, verification was carried out based on the original 29 seismic events and we discovered that the SES values that were calculated according to the equation did not agree with the original SES. As a result, we decided initially to construct the SES randomly for all seismic events. More specifically, for each seismic event that we did not have the corresponding SES, we applied the random construction of SES. According to the 29 seismic events, the SES of the N–S direction vary from -13 to 12 mV/km, and the SES of the E–W direction from -26 to 58 mV/km. Unfortunately, Varotsos and his group do not refer at all to the ranges of SES, therefore we based our construction on those 29 original SES. Thus, random values for each direction of each seismic event were selected based on their corresponding range.

The data was normalised to values from 0 to 1 or from -1 to 1, according to the following cases:

(i) In the case of the magnitude and the time lag, where the values are only non-negative, the formula that was used can be seen in Eq. (2) below:

$$X_i = \frac{S_i - Min}{Max - Min} \tag{2}$$

where S_i is the non-negative value of the ith pattern of the specific attribute that we wish to normalise, Min and Max are the minimum and maximum values of the corresponding attribute, respectively and X_i is the normalised value of the ith pattern of the specific attribute.

(ii) In the case of SES, which contain negative values as well, the normalisation formula of Eq. (2) was modified according to the sign of the SES value. For the negative SES, we used the sub-formula given by Eq. (3):

$$X_i = \frac{S_i}{|Min|} \tag{3}$$

where S_i is the ith pattern of the negative SES (either N–S or E–W), Min is the minimum value of the corresponding signal, and X_i is the normalised value of the ith pattern of the corresponding signal.

(iii) For the positive SES, we used the sub-formula given by Eq. (4):

$$X_i = \frac{S_i}{Max} \tag{4}$$

where S_i is the ith pattern of the positive SES (either N–S or E–W), Max is the maximum value of the corresponding signal, and X_i is the normalised value of the ith pattern of the corresponding signal.

Table 1
Maximum and minimum value for each direction (North–South, East–West) of the SES.

	MIN (mV/km)	*MAX* (mV/km)
N–S	−13	12
E–W	−26	58

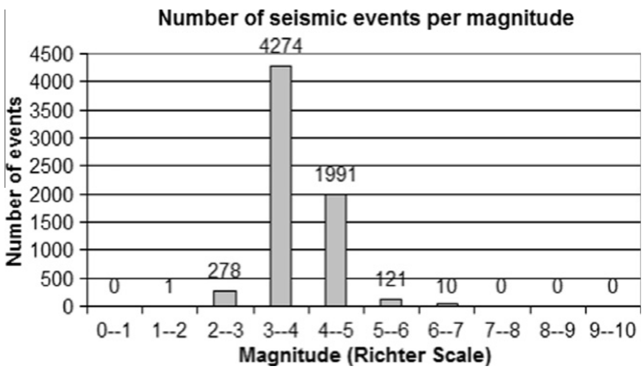

Fig. 2. The distribution of seismic events based on their magnitude.

Based on the above formulas, we needed the minimum or the maximum signal value of the corresponding direction (N–S or E–W) in order to calculate the normalised value of each SES. As mentioned before, Varotsos and his group do not refer at all to these quantities. As a result, we had to assume that the *Min* and *Max* values of each direction's signal are as given in Table 1, which were derived from the original dataset of the 29 seismic events.

Fig. 2 presents the exact number of seismic events that occurred in Greece for the period of 01/01/1980–01/01/2001 distributed in magnitude bins of 1 on the RS. It is remarkable that the majority of seismic events had a magnitude between 3 and 4 on the RS, whereas the seismic events of 5–6 and 6–7 on the RS are 121 and 10, respectively. The very low number of the major seismic events might affect the efficiency of the neural network in predicting major seismic events.

3.1.2. Design

Three variations of ANN models have been developed for earthquake prediction. In all cases, the networks take as input the SES of the North–South and East–West directions. The part that makes each network different from each other is the input or output layer. The first model, which we will refer to as the basic one, has only one output node which is the magnitude of the predicted upcoming seismic event (see Fig. 3). The second model apart from the SES, takes the average magnitude for the previous 30 days as an extra input and gives the magnitude of predicted upcoming seismic event as output (see Fig. 4). Finally the third model is the same as the basic model, but apart from the output node for the magnitude, it has another output which is trained to predict the time lag

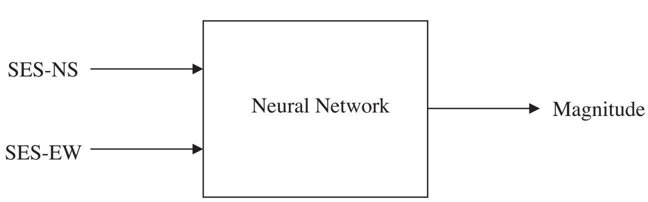

Fig. 3. Schematic of the 1st (basic) neural network model: takes as inputs the SES and predicts the magnitude of the upcoming earthquake.

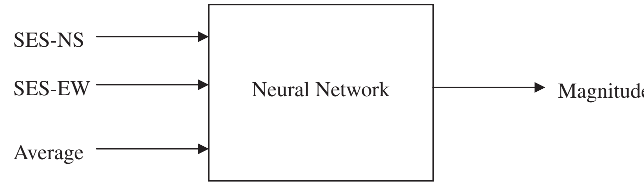

Fig. 4. Schematic of the 2nd neural network model: takes as inputs the SES and the average magnitude of the 30 previous days and predicts the magnitude of the upcoming earthquake.

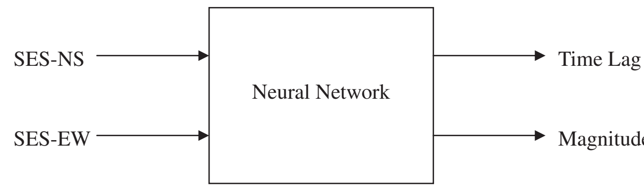

Fig. 5. Schematic of the 3rd neural network model: takes as inputs the SES and predicts the magnitude of the upcoming seismic event and the time lag till its occurrence.

between the date on which the SES was recorded and the date of the upcoming major seismic event (see Fig. 5). The aim of the third neural network is to check whether there is correlation between the SES and the time lag. In other words, we want to check whether it is possible to predict the exact occurrence time of a major seismic event by using the SES recorded before that event.

For each variation of the neural networks we had tested a range of values for the parameters of the backpropagation algorithm (learning rate, momentum) and the structure of the network in order to find the optimal combination that would provide the best possible performance of each neural network model. Based on this experimentation the optimal structure for the first ANN model, which is used in the prediction of the magnitude, consists of one hidden layer with 10 hidden neurons. The values of the learning rate and the momentum term are 0.6 and 0.8, respectively. The structure of the second ANN model consists of one hidden layer with 5 hidden neurons. The optimal values for learning rate and momentum are 0.5 and 0.8, respectively. Finally, the optimal structure for the third ANN model consists of one hidden layer with 10 hidden neurons (as the basic model). The values of the learning rate and the momentum term are 0.6 and 0.8, respectively (as in the basic model). All neurons in all three ANN models use a sigmoidal activation function.

3.1.3. Results

The accuracy rate on all seismic events resulted by the basic model is 60.54% and by the second model 60.66%. We expected that the second model which has the average value of magnitude for the previous thirty seismic events as an additional input would give better results than the basic model since we had given more information to the network concerning the magnitude. However, the increase of 0.12% in the accuracy rate of the second model from the basic model does not seem to be statistically significant.

The networks are not capable to predict any major seismic events ($M \geqslant 5.0$, RS), which is the main problem of earthquake prediction. This is due to the fact that the number of major seismic events is very low (see Fig. 2) making the forecasting of such events more difficult.

Finally, extremely weak predictions resulted from the third network (Fig. 5), which led us to the conclusion that there is no correlation between SES and time lag in the first part of case study II. In other words, when the missing SES are constructed randomly for all seismic events, the SES cannot predict the time lag between the recorded SES and the impending seismic event.

3.2. Constructing additional SES using neural networks reversely

3.2.1. Methodology and data preprocessing

Due to the inability of the neural network models of the first part of the current case study (Section 3.1) to predict major seismic events ($M \geqslant 5.0$, RS), we were led to an amendment of our previous methodology. The most important issue is that we did not have access to any SES that corresponded to seismic events with $M < 5.0$ (RS) (which according to Varotsos and Alexopoulos (1984a) did exist), and consequently the constructed SES for these seismic events (with $M < 5.0$, RS) with data having $M \geqslant 5.0$ (RS) were most probably not realistic. Thus, we decided to isolate the major seismic events with $M \geqslant 5.0$ (RS) and train the neural network models with a dataset including only these seismic events. However, we still had SES only for the 29 major seismic events which are insufficient for training and testing the neural network models. Therefore, we filtered the seismicity catalogues of the SINOA to find the major seismic events ($M \geqslant 5.0$, RS) for the period between 1980 and 2007, and we found 157 new major seismic events on top of the 29 we already had. The next step was to construct the corresponding SES for the 157 seismic events.

In order to construct SES which would correspond to these 157 major seismic events, we implemented two neural network models which generated appropriate data for each of the networks used in the prediction of the magnitude only (corresponding to the network of Fig. 3) and the magnitude and time lag (corresponding to the network of Fig. 5). In particular, the magnitude of each seismic event was presented to the neural networks as input in order to give as output the corresponding SES, and SES and time lag. In other words, we used the neural networks of Section 3.1.2 (Figs. 3 and 5) reversely. Schematic diagrams of the neural network models that were implemented for the construction of the data can be seen in Figs. 6 and 7.

Both networks were trained with the backpropagation learning algorithm and the leave-one-out cross validation method (Vapnik, 1998). It is important to note that, in this training phase we only used the data of the 29 original events for which both the magnitude and the SES are known. Once the training phase had been completed successfully, we performed the testing phase. In this phase, we presented the other 157 major seismic events and the networks generated information about the SES (and the time lag)

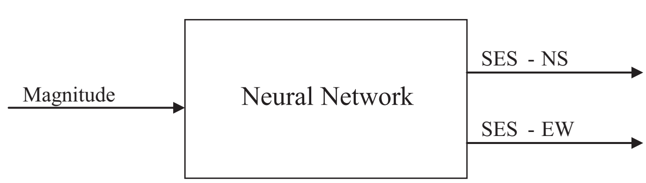

Fig. 6. Schematic of the neural network model which generates data that is used for the prediction of the magnitude of an upcoming earthquake.

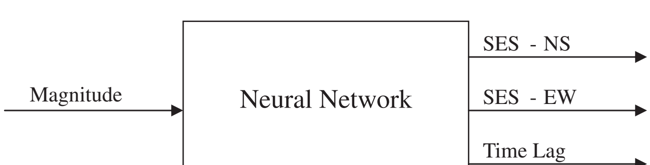

Fig. 7. Schematic of the neural network model which generates data that is used for the prediction of the magnitude of an upcoming earthquake and its time lag from SES occurrence.

for each one of these events in the testing dataset. The SES were normalised in the same way as described in Section 3.1.1.

3.2.2. Design

As mentioned above, the goal of this part of the study is to design and implement ANN models that can predict major seismic events ($M \geqslant 5.0$, RS) with practical accuracy. In order to fulfil this goal successfully, two neural networks from Section 3.1 were used. More specifically, the basic model (Fig. 3) for the prediction of magnitude and the model predicting the magnitude and the time lag (Fig. 5) were used. The main difference between the neural networks used in the current section and the neural networks of Section 3.1 is the data taken as input, since they concern the SES of only major seismic events. In particular, for the basic model (Fig. 3) the data constructed by the network of Fig. 6 was used as input data, while for the model predicting the magnitude and the time lag (Fig. 5), the data constructed by the network of Fig. 7 was used as input data. For each neural network model we had tested a range of values for the parameters of the backpropagation algorithm (learning rate, momentum) and the structure of the network in order to find the optimum configuration which provides the best possible performance of the ANN models and determined the structure and the parameters that characterise the specific networks. Based on this experimentation, both networks consist of one hidden layer with 6 hidden neurons and the optimal values of the learning rate and the momentum term are 0.01 and 0.2, respectively. Both networks used the backpropagation learning algorithm with a sigmoidal activation function for each of their neurons and the leave-one-out cross validation procedure for training and testing, respectively.

3.2.3. Results

The accuracy rate that resulted from the prediction of the magnitude is 84.01% and the corresponding rates from the prediction of both magnitude and time lag are 83.56% for magnitude and 92.96% for time lag. The accuracy rates were calculated based on the MAE, as described in Section 2.1.

4. Discussion and conclusions

In the present paper, two case studies were described for the prediction of magnitude (and time lag) of earthquakes in the region of Greece. Feed-forward multilayer perceptron type neural networks were implemented using the backpropagation learning algorithm for training. For the first case study (Section 2) the input data presented to the network is only a time series of earthquake magnitude data, whereas for the second case study (Section 3) the input data consist of the Seismic Electric Signals of the North–South and East–West directions.

Regarding the first case study, we observed that the accuracy rate for predicting the outliers (major seismic events with $M \geqslant 5.2$ on the RS) is reduced considerably at 52.81%, in contrast to the accuracy rate which concerns all the seismic events which reaches 80.55%. We expected that since the number of the seismic events that are considered as outliers is very low. More specifically, only 82 out of 7671 seismic events are considered as outliers ($M \geqslant 5.2$, RS), so because of the very low number of the outliers in the historical data, the training of the neural network is not efficient. Consequently, the network is not able to predict major seismic events ($M \geqslant 5.2$, RS) with high accuracy.

Unfortunately, our attempt to improve the accuracy of the prediction of the major seismic events ($M \geqslant 5.2$, RS) by training the network in two phases cannot be considered as successful, since the accuracy rate only had a slight increase from 52.81% to 58.02%. Even though the neural network can be considered more familiar with the outliers because of the initial training of the first phase (with major seismic events) where the resulting weights are kept and used for the second phase, the unsuccessful prediction of outliers was expected, given that the outliers are the minority of the seismic events of a specific area and furthermore they occur at irregular time intervals for a specific period of time.

The lack of enough data for Seismic Electric Signals (SES) was the most important obstacle in our second case study since without a substantial amount of data the study would have been meaningless. Despite this, we attempted a solution by constructing the SES. The first attempt was by constructing the missing SES of all the seismic events randomly according to the range of the amplitudes of the 29 original SES. The relatively low accuracy rates of 60.54% and 60.66% resulted from the networks of Figs. 3 and 4, respectively, and the weak prediction of the network of Fig. 5, were inevitable since the range of the amplitudes of the original SES corresponded only to seismic events with $M \geqslant 5.0$ (RS). Consequently, the artificial SES for earthquakes of $M < 5.0$ (RS) were not realistic justifying the low accuracy rates obtained. This led us to the second attempt where the missing SES for major seismic events only ($M \geqslant 5.0$, RS), were constructed by using neural networks reversely. This was done since there are no published SES for seismic events with $M < 5.0$ (RS) and for their construction, the 29 original SES (corresponding to seismic events of $M \geqslant 5.0$, RS) were used again. The accuracy rate increased to 84.01%, leading us to the conclusion that if we had the real SES for all the seismic events, the results would have been more accurate both for earthquakes of $M < 5.0$ (RS) and $M \geqslant 5.0$ (RS).

The aim of the second neural network of Section 3.2.2 corresponding to Fig. 5 was to check whether there is correlation between the SES and the time lag when the missing SES were constructed from major seismic events using ANNs reversely. The increased accuracy rate of 92.96% for predicting the time lag between the SES and the origin time of the seismic event in contrast to the accuracy rate for predicting magnitude (83.56%) implies that it is possible to predict the occurrence time of a major seismic event by using the SES recorded before that event. Thus in this case there is correlation between the SES and the time lag.

Regardless of certain inadequacies of the dataset due to scarcity of available SES, we found the performance of the constructed ANN models quite satisfactory, and this supports the usefulness of the application of this tool to similar problems.

As mentioned above, the VAN group scores their prediction "successful" when the actual earthquake occurred within several days to several weeks, within ca. 100 km from the predicted epicentre and within ca. 0.7 units of the predicted magnitude (RS). For that reason, we were not able to compare the results of the VAN group with ours, as we are calculating the accuracy rates by determining the exact difference between target and predicted values.

Moreover, the main difference between our methodology and that of the VAN group is that we attempted to find a correlation between the SES and the magnitude using neural networks, whereas the VAN team found a linear correlation empirically (see Eq. (1)) and based on this they calculate the magnitude of the impending earthquake. More specifically, the two electrodes which are used to determine the electric field constitute what the VAN team calls a line (E–W line and N–S line). Therefore, for a given line of a given station, the ΔV-values of the SES emitted from a given seismic region increase with the magnitude (Varotsos & Alexopoulos, 1984a, 1984b). In addition, we attempted by using ANNs to find a correlation between the SES and the time lag from the SES and the exact occurrence of the impending earthquake. This is in contrast to the VAN group who claimed that after a SES occurs, the seismic event follows within several days to several weeks based on the SES's type (Varotsos & Alexopoulos, 1984a, 1984b), and have not attempted to predict the exact date of occurrence of the impending earthquake.

As we stated above, our prediction concerns the following day (in the first case study) or the next major upcoming seismic event (in the 2nd case study), which is more difficult than predicting the impending earthquakes within a predefined future time period but not the exact date of occurrence. Therefore, we are confident that, should we were to predict the earthquakes for a specific future time period, like for the following fifteen days, as Adeli and Panakkat (2009) did, we would have had higher accuracy rates.

As a general observation, we can argue that the neural network models used in this study are able to predict the magnitude and the time lag of a major seismic event relatively accurately. This fact confirms that once the neural networks are trained by using the appropriate data, they are able to generalise and predict unknown seismic events accurately.

The accuracy rates presented in the current paper are all based on the out-of-sample performance for each model. In other words, the data used for testing the networks are different from data used for training. According to Tashman (2000), for a good evaluation of a forecasting study, the method should be assessed based on the out-of-sample performance. In this way the predictive capability of the model will match the conditions of the real-world (Adya & Collopy, 1998).

Another two validation criteria which are met by our study concern the network's architecture and the cross-validation. As it was mentioned, an experimentation was conducted for each model in order to find the optimal network architecture. Moreover, the leave-one-out cross validation method which was applied in Section 3.2 is recommended to be used since it facilitates the termination of learning and controls overfitting (Adya & Collopy, 1998).

References

Adeli, H., & Panakkat, A. (2008). Recent efforts in earthquake prediction 1990–2007. *Natural Hazards Review, 9*(2), 70–80.

Adeli, H., & Panakkat, A. (2009). A probabilistic neural network for earthquake magnitude prediction. *Neural Networks, 22*, 1018–1024.

Adya, M., & Collopy, F. (1998). How effective are neural networks at forecasting and prediction? A review and evaluation. *Journal of Forecasting, 17*, 481–495.

Bodri, B. (2001). A neural-network model for earthquake occurrence. *Journal of Geodynamics, 32*, 289–310.

Grant, R. A., & Halliday, T. (2010). Predicting the unpredictable; evidence of pre-seismic anticipatory behaviour in the common toad. *Journal of Zoology, 281*(4), 263–271.

Gutenberg, B., & Richter, C. F. (1956). Earthquake magnitude, intensity, energy and acceleration. *Bulletin of the Seismological Society of America, 46*(1), 105–146.

Haykin, S. (2009). *Neural networks and learning machines* (3rd ed.). New Jersey: Pearson Education.

Hoaglin, D. (2003). John W. Tukey and data analysis. *Statistical Science, 18*(3), 311–318.

Lakkos, S., Hadjiprocopis, A., Comley, R., & Smith, P. (1994). A neural network scheme for earthquake prediction based on the Seismic Electric Signals. In *Neural networks for signal processing IV, proceedings of the IEEE conference on neural networks and signal processing, Ermioni, Greece* (pp. 681–689).

Lakshmi, S., & Tiwari, R. (2007). Are northeast and western Himalayas earthquake dynamics better "organized" than central Himalayas: An artificial neural network approach. *Geofisica Internacional, 46*(1), 65–75.

Paliwal, M., & Kumar, U. A. (2009). Neural networks and statistical techniques: A review of applications. *Expert Systems with Applications, 36*, 2–17.

Rumelhart, D. E., Hinton, G. E., & Williams, R. S. (1986). Learning representations by back-propagating errors. *Nature, 323*, 533–536.

Tashman, L. (2000). Out-of-sample tests of forecasting accuracy: An analysis and review. *International Journal of Forecasting, 16*, 437–450.

Uyeda, S. (1997). The VAN method of short-term earthquake prediction. RIKEN international frontier program on earthquake research. *INCEDE newsletter, 5* (4), 2–4. Available from http://sems-tokaiuniv.jp/old/eprc/res/incede/incede2.html Accessed 31.03.10.

Uyeda, S., Al-Damegh, K. S., Dologlou, E., & Nagao, T. (1999). Some relationship between VAN seismic electric signals (SES) and earthquake parameters. *Tectonophysics, 304*, 41–55.

van Camp, D. (1993). *The XERION neural network simulator users guide*. University of Toronto.

Vapnik, V. (1998). *Statistical learning theory*. New York: Wiley.

Varotsos, P. (2005). *The physics of seismic electric signals*. Tokyo: Terra Scientific Publishing Company (Terrapub).

Varotsos, P., & Alexopoulos, K. (1984a). Physical properties of the variations of the electric field of the earth preceding earthquakes, I. *Tectonophysics, 110*, 73–98.
Varotsos, P., & Alexopoulos, K. (1984b). Physical properties of the variations of the electric field of the earth preceding earthquakes, II. Determination of epicenter and magnitude. *Tectonophysics, 110*, 99–125.
Varotsos, P., Alexopoulos, K., & Lazaridou, M. (1993). Latest aspects on earthquake prediction in Greece based on seismic electric signals, II. *Tectonophysics, 224*, 1–37.
Varotsos, P., & Lazaridou, M. (1991). Latest aspects on earthquake prediction in Greece based on seismic electric signals. *Tectonophysics, 188*, 321–347.
Varotsos, P., Sarlis, N., Skordas, E., & Lazaridou, M. (2006). Additional evidence on some relationship between Seismic Electric Signals (SES) and earthquake focal mechanism. *Tectonophysics, 412*, 279–288.
Velido, A., Lisboa, P. J. G., & Vaughan, J. (1999). Neural networks in business: a survey of applications (1992–1998). *Expert Systems with Applications, 17*, 51–70.