PERGAMON

Journal of Geodynamics 32 (2001) 289–310

JOURNAL OF GEODYNAMICS

www.elsevier.com/locate/jgeodyn

A neural-network model for earthquake occurrence

Bertalan Bodri*

Research Group for Geophysics and Environment Physics, Hungarian Academy of Sciences, c/o Geophysics Department, Eötvös University, Budapest, Ludovika 2, Hungary 1083

Received 10 January 2000; received in revised form 21 June 2001; accepted 9 July 2001

Abstract

Changes in seismic activity patterns can occur during the process of preparation of large earthquakes, and such changes possibly are the most reliable long-term earthquake precursor examined to date. In the present work, seismicity rate variations in the Carpathian–Pannoman region, Hungary, and the Peloponnesos–Aegean area, Greece, have been used to develop neural network models for the prediction of the origin times of large ($M \geqslant 6.0$) earthquakes. Three-layer feed-forward neural network models were constructed to analyse earthquake occurrences. Numerical experiments have been performed with the aim to find the optimum input set configuration which provides the best performance of a neural network. It was possible to reach sufficient training tolerance for the constructed networks (correspondence between predicted by the model outputs and known from experience outputs within the limits of given error thresholds) only when the input set contained seismicity rate values for different magnitude bands (when such data appeared representative enough) and also for more than one time intervals between large earthquakes. The specific structure of the network input generates the question of whether this configuration has some relationship to the physics of the strain accumulation and/or release process. The remarkably satisfactory performance of the constructed neural networks suggests the usefulness of the application of this tool in earthquake prediction problems.

1. Introduction

Since quantitative prediction is the aim of every science, earthquake prediction has long been considered a most important, ultimate goal of seismology. The time scale of forecasting of earthquakes is generally classified as short-, intermediate-, and long-term, according to the expected time interval to an impending earthquake (Wallace et al., 1984). Attempts for short-term

* Tel./fax: +36-1-210-1089.
E-mail address: bober@ludens.elte.hu

0264-3707/01/$ - see front matter
PII: S0264-3707(01)00039-4

predictions try to identify precursors on time scales of hours to weeks before an earthquake, and are based on the measurement of rapidly changing anomalies, such as foreshocks, enhanced rates of ground deformation, changes of water chemistry, etc. (Wyss and Burford, 1987; Roeloffs, 1988; Fraser-Smith et al., 1990). Long-term predictions, the subject of primary interest in the present work, are made a few years to a few decades before the expected earthquakes. They are based mostly on analysis of earthquake recurrence times and changes of broad seismicity patterns (Habermann, 1988; Matthews and Reasenberg, 1988; Carlson, 1991; Shaw et al., 1992; Eneva and Ben-Zion, 1997a; Triep and Sykes, 1997). Changes in seismicity patterns are the most successful long-term precursors examined to date.

Seismicity databases (catalogs) are the most popular source of data for long-term prediction studies for a number of reasons (abundance, existence for almost all regions of the world, availability on a continuous basis, data directly reflecting processes in the regions of mainshock preparation, etc.). Generally, investigators search the catalogs for long-term changes in seismic activity, and look whether or not these changes are followed by large events. However, the anomalies that different authors interpret as precursors for large earthquakes may be quite different in different areas (Habermann, 1988; Shaw et al., 1992; Triep and Sykes, 1997), and in spite of attempts to support them with mechanical models (Sykes, 1983; Carlson, 1991; Shaw et al., 1992; Bufe and Varnes, 1994; Eneva and Ben-Zion, 1997b) and also regardless of some successful real-time predictions (e.g. Wyss and Burford, 1987), are more or less qualitative and preliminary.

Recently, efforts have been made to investigate the potential of artificial neural networks (ANNs) as a tool for simulation of the behaviour of systems that are governed by nonlinear multivariate and generally unknown interconnections within a noisy, poorly-controllable physical environment. The advantage of this framework is that the neural network provides a black-box approach and the user need not know much about the physics of the process simulated. In the present work, we make a preliminary and simple attempt to relate neural network ideas to seismic activity patterns in the Carpathian–Pannonian area, Hungary, and the Peloponnesos–Aegean region, Greece (Fig. 1), and develop ANN models focussing on the prediction of the origin times of large earthquakes. The choice of the ANN approach is motivated by the lack of clear causal relations between seismicity patterns and related crustal environments which could support the use of knowledge-based approaches, and also by previous successful applications of ANNs for the solution of different seismic problems (e.g. Chiaruttini et al., 1989; Chiaruttini and Salemi, 1993; Dai and MacBeth, 1997). Simple experiments are reported below in an attempt to construct neural networks and to demonstrate the performance and feasibility of ANN models using seismicity rates (number of earthquakes in a unit time) as input values to forecast occurrences of large earthquakes. Consideration of seismicity rates seems to us attractive because many seismologists share the view that changes in seismicity rates can occur as part of the process of preparation of large earthquakes.

2. Seismotectonics and earthquake data

Only a very short account of these characteristics wifi be presented below, for more details the reader is referred to Royden and Horváth (1988), and Papazachos et al. (1991) and Kiratzi and Papazachos (1995).

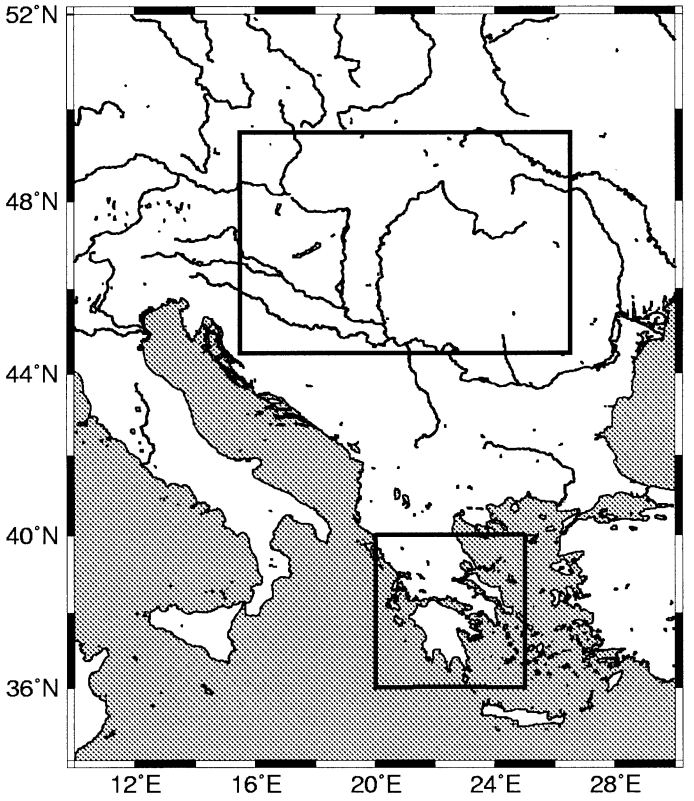

Fig. 1. Part of Europe showing geographical location of the study areas (framed with solid lines) in the present work. Northern rectangle: Carpathian–Pannonian area. Southern rectangle: Peloponnesos–Aegean region.

The Carpathian–Pannonian system includes the back-arc extensional Carpathian basin (with its central part known as the Pannonian basin) surrounded by mountain ranges of the Alps, Carpathians and Dinarides. Although the area represents a part of the coffision zone between Eurasia and Africa, it exhibits definitely low seismic activity with respect to other parts of this zone. Because of a scattered distribution of the epicenters and a complicated regional geology, it is difficult to reveal relations between epicentral distribution and tectonic setting. The area is broken up into a number of crustal blocks of different size. Internal deformation due to relative motions of these blocks is likely to determine the seismic activity. Because of this structure, simple plate tectonic models can be applied to the region with great caution only (Gutdeutsch and Aric, 1988). Data on seismicity were taken from the Hungarian Earthquake Catalog (456–1986) (Zsíros et al., 1988) completed by the list of shocks for the subsequent period 1987–1994 (Zsíros, 1997, personal communication). The minimum level of completeness was estimated as the magnitude below which the log cumulative number of events with known magnitudes in the whole catalog started to depart from a linear magnitude dependence, and a threshold value of $M=2.6$ was found. From 1880 A.D., the set of $M \geqslant 2.6$ events appears stationary in the sense that the series of the yearly number of such earthquakes does not show any linear trend over the period of 115 years (regression line to data in Fig. 2 has the slope of 0.05 ± 0.07). Subsequent considerations will thus

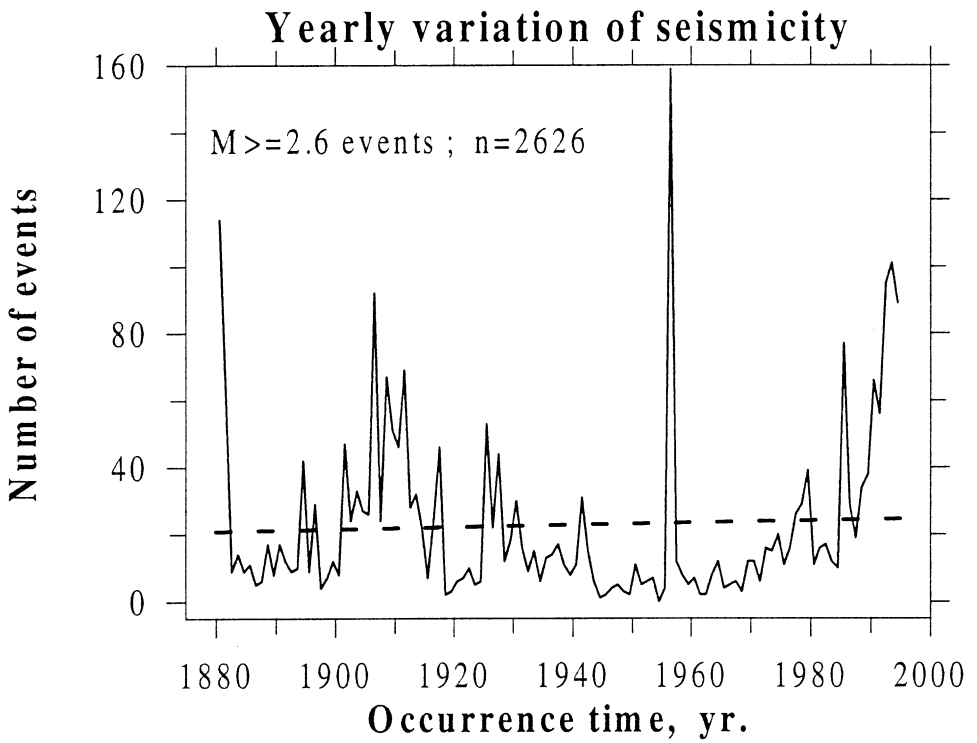

Fig. 2. Yearly variation of the number of $M \geqslant 2.6$ earthquakes in the Carpathian–Pannonian area from A.D. 1880 to 1994. Dashed line represents regression line fitted to data (slope of regression line: 0.05 ± 0.07).

be restricted to $M \geqslant 2.6$ earthquakes occurring from 1880 A.D. Their number in the study area is $n = 2626$. The above result, however, may be misleading since the apparent stationarity might be attributable to the relatively large number of aftershocks following a few strong earthquakes in the initial section of the series (e.g. an $M6.2$ event in 1880 at Zagreb, Croatia, the $M5.7$ earthquake in 1906 in Little Carpathians, NW Hungary, etc.). A more rigorous consideration of the problem of homogeneity of data in the following section shows that a significantly higher threshold magnitude of $M = 4.0$ should be introduced to get rid of "hidden problems" in data.

The Aegean area and its surroundings are considered one of the seismically most active region of the world. On a large scale, this area represents an edge zone of the African-Eurasian collision, where the African plate approaches Eurasia and subducts beneath it along the Hellenic arc outlined by earthquake hypocenters (Kiratzi and Papazachos, 1995). Because of a complex plate tectonic situation, seismicity in the area selected for study is increased even on the generally high background of the larger Aegean region. Our choice was motivated also by the fact that this is approximately the test area of the VAN group's (Varotsos and colleagues in Greece) predictions, and has recently been the target of considerable research interest (see, e.g. Varotsos and Kulhanek, 1993; Geller, 1996a). As data source on seismicity, we used the Greek SI-NOA (Seismological Institute, National Observatory of Athens) catalog for 15 years from 1982 to 1996. The space window 20–25° E and 36–40° N was applied, and, because some 98% of earthquakes were of shallow depth of less than 50 km, no lower limit on hypocenter depths was set on. To use a

surface wave magnitude (M_s) as usually defined, we added 0.5 to the local magnitudes M_L reported by SI-NOA (Geller, 1996b). Altogether 11,092 earthquakes fall in the considered time-space window. The series of the monthly number of all events with known magnitudes (Fig. 3) does show a linear trend over the observation period. However, from a threshold magnitude of $M=3.5$ (with $n_{M\geqslant 3.5}=7951$) introduced in subsequent considerations, this trend is already not significant.

3. Testing artificial changes of seismicity rate

The catalog parameters necessary for seismicity rate analysis are earthquake origin times and magnitudes. Seismicity rate variations can be well illustrated by the cumulative number curves. Generally, significant rate changes can be observed in most seismicity data sets. Fig. 4 shows variations of the cumulative number of events in the Carpathian region between 1880 and 1995, for different magnitude bands. The diagram clearly demonstrates not only variations of the seismicity rate with time, but also the dependence of the rate changes on the size of the events considered. The rates for small events ($M=2.6$–3.9), for example, can be divided roughly in three intervals: high rate between 1880–1920, low rate between 1920 and 1980 and high rate since 1980, with slopes of the cumulative number curves of 23.5 ± 0.9, 10.8 ± 0.3, and 40.3 ± 3.1 events a year for the particular intervals, respectively. On the other hand, the rate of occurrence of events with

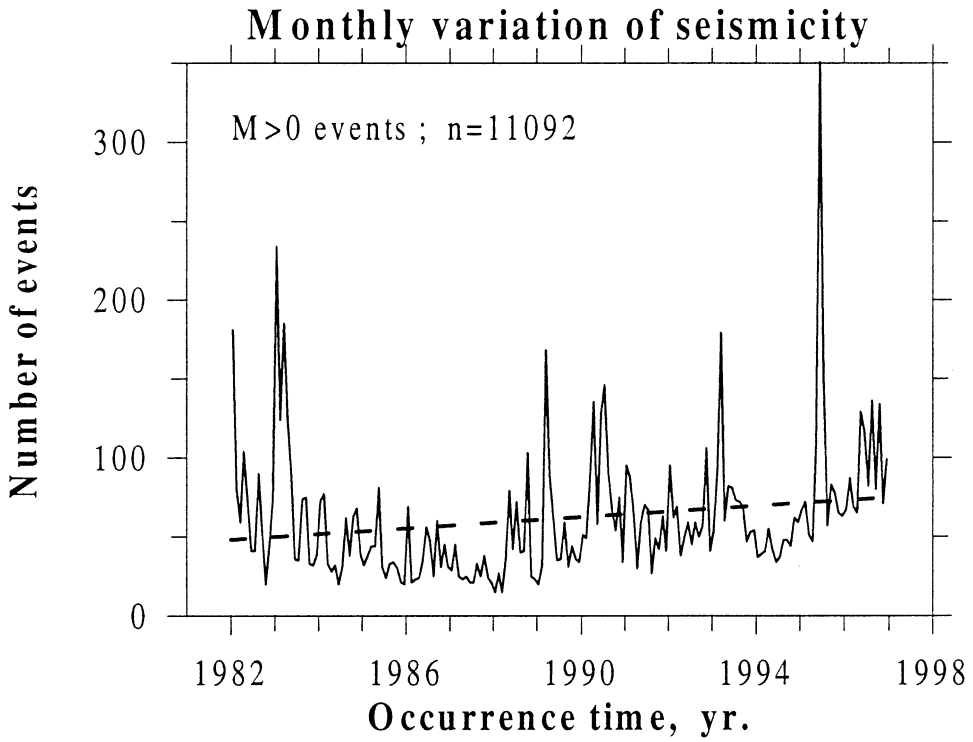

Fig. 3. Variation of the monthly number of all reported earthquakes with known magnitudes in the Peloponnesos–Aegean region for the interval of 15 years of 1982–1996. Dashed line: regression line fitted to data, its slope: 1.8 ± 0.7.

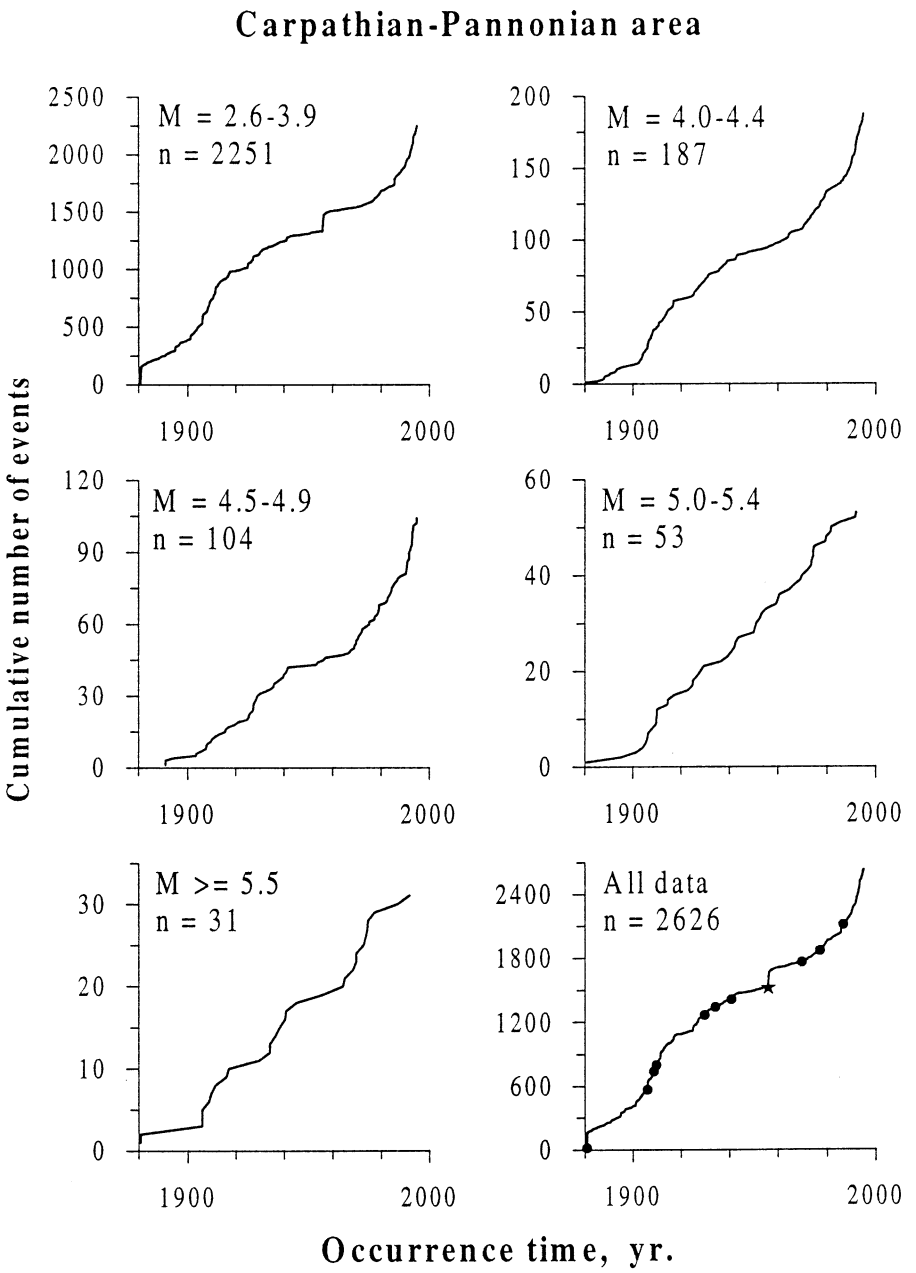

Fig 4. Variations of the cumulative number of events versus time, in six magnitude bands in the Carpathian–Pannonian area. Dots indicate occurrences of large ($M \geqslant 6.0$) earthquakes, asterisk shows the $M5.6$ Dunaharaszti (Central Hungary) earthquake in 1956, followed by a large number of reported aftershocks.

$M \geqslant 5.0$ has remained approximately constant (with a slope of 0.84 ± 0.01 events/yr) since 1900. Three rate intervals can be observed also in the set of all events because most of the shocks are small. Variations in seismic rates appear similarly noticeable also in Fig. 5, constructed for the

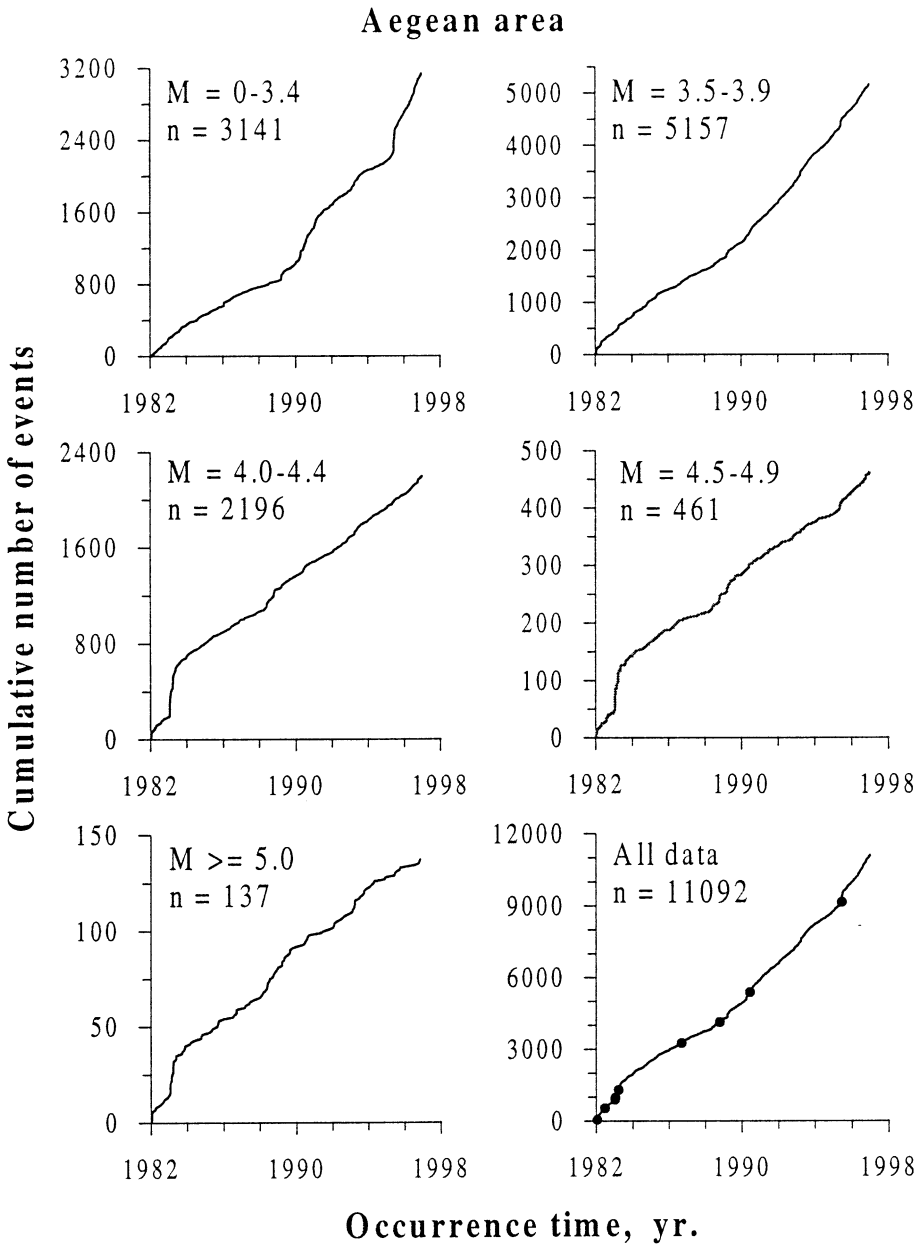

Fig. 5. Same as Fig. 4, but for the Peloponnesos–Aegean region.

Greek data set. For brevity, we do not list here numerical values of changes of the slope of the seismic rate curves in particular intervals of time and magnitude ranges.

Occasional abrupt steps in the cumulative number curves (e.g. sharp increase in Fig. 4 in the year 1956, or in Fig. 5 at around 1983) are attributable to the large number of aftershocks accompanying some relatively large events. Generally, the tendency for the seismic events to

occur groupped in space and time is a phenomenon widely observed at different space–time–magnitude scales (e.g. Shlien and Toksöz, 1970). Since the pioneering work by Omori (1894), a variety of statistical methods have been applied to quantify the characteristics of the clustering of seismicity in time and space and/or different seismo-tectonic environments (e.g. Prozorov and Dziewonski, 1982; Reasenberg, 1985; Ogata, 1988). Earlier approaches relied on the model of a Poisson cluster process, where usually the strongest event in an earthquake sequence is considered a cluster centre or main event, and the mainshock occurrence is taken as a stationary process with a uniform rate over time. However, it has become a common view that the behaviour of mainshocks does not seem to be Poissonian either (e.g. Smalley et al., 1987; Kagan and Jackson, 1991; Main, 1995). To illustrate this, the results of a numerical experiment of ours worth mentioning. Although there is no standard method for aftershock removal, following Kagan's (1996) procedure we have carried out declustering of the Greek data set ($n=7951$ events with $M \geqslant 3.5$) considered in the present study. Having removed some 40% of the earthquakes, we counted the frequency distribution of 548 time intervals of equal length (some 10 days) with 1, 2, 3, ..., etc., declustered earthquakes within each interval. A Chi-square test of the resulting frequency histogram has shown that the zero hypothesis (i.e. that the distribution is Poissonian) might have had the probability of only as small as 340 to a million.

Kagan and Jackson (1991) have analysed statistically the long-term properties of several instrumental earthquake catalogs. They found that both short-term clustering and the long-term variation of seismicity were scale invariant governed by power-law temporal distribution, though the fractal dimensions (exponents of the power law) in the two scale ranges might be different. Understanding earthquakes as critical phenomena with fractal spatial and temporal correlations, Main (1995) argues that "it would appear dangerous to "clean" catalogs by removing foreshocks and aftershocks to artificially produce a more Poissonian distribution, since these particular events are the strongest evidence for criticality as the underlying statistical physical process". As already mentioned previously, in the pressent study we apply ANN modelling as a black-box approach to earthquake phenomenology, and in such an approach prefer to use naturally occurring, "clustered as registered" input data.

Although earthquake catalogs seem to be simple lists of events, they are, in fact, complex data sets. Seismicity rate variations considered in any attempt of prediction should be natural, i.e. associated with temporal variations in the seismic process itself. However, the observed seismicity rates usually contain significant man-made impacts. Such artificial changes should be identified and corrected for. The importance of delineating man-made changes in studies of seismicity rates was demonstrated by, e.g. Habermann (1987, 1991). Artificial changes arise mainly from progressive development in the seismic monitoring networks (closure and opening of stations, changes in instrumentation) and in processing techniques. They can generally be classified into three categories: (1) detection changes arising from increased capability of a network to recognize and locate events due to installation of new stations, (2) systematic changes in magnitudes of events (magnitude shifts), caused by changes in the station distribution used to determine magnitudes, and (3) reporting changes which are related to lack of reporting magnitudes for detected events. While detection changes can be accounted for easily enough, the last two effects cannot be simply characterized as they are complex and highly variable.

Even when artificial changes in seismicity patterns have been eliminated to some degree at earlier stages of processing of data, further analysis is indispensable. Man-made changes generally

affect smaller events, i.e. naturally divide seismicity data into a subset of smaller events which are affected by the change and a subset of larger events which remain unaffected (Habermann, 1987). To reveal the impact of changes in monitoring networks and procedures on seismicity catalogs and identify a magnitude cutoff which eliminates affected events from consideration, Habermann (1987; see also Habermann, 1991; Eneva et al., 1994) proposed a quantitative technique based on z-statistics, the general parametric statistical test used for evaluating the difference between two means. The z-value is calculated as

$$z = \frac{v_1 - v_2}{\sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}}, \tag{1}$$

where v_1 and v_2 are mean seismic rates for some two observation periods, s_1 and s_2 their standard deviations, and N_1, N_2 the number of events in each period. The z-value is an indication of the strength of a change, where the amount of change of a variable as well as the corresponding standard deviations and also durations of the periods used to define the change are considered. Negative z-values indicate rate increases while a positive z corresponds to rate decrease. The magnitude cutoff is determined by examining the distribution of an observed change in the magnitude domain. To find the correct cutoff one has to examine a wide range of possible cutoffs using a plot called a magnitude signature. The motivation for magnitude signature plots is the observation that seismicity rate changes are often a strong function of magnitude. Concerning the question of how to divide the events into subgroups for examination, Habermann (1987) suggests to group the events into bands with simple lower or upper cutoffs. This approach has the advantage that it divides the data into subsets corresponding to the manner that man-made changes effect seismicity (as discussed above). Besides, this division smooths the data and makes interpretation of the changes easier.

Fig. 6 shows the magnitude signature plot for the Carpathian area. Statistics has been calculated from a comparison of seismicity rates for the $M \geqslant 2.6$ catalog during two periods (October 1908–March 1934 and March 1934–March 1977) between the three largest events in the area for the whole observation time. Magnitude band here means all events below or above a given magnitude. The characteristic features of typical magnitude signature plots have been described in detail by Habermann (1987). Fig. 6 shows the typical features of a magnitude signature for a detection decrease (positive z throughout all magnitudes, lack of change in data sets containing larger events on the right side of plot, etc.; see Habermann, 1987). The magnitude cutoff which eliminates this detection decrease is determined by the lowest limit of the nearly-horizontal section of the curve on the right side, or by the strongest change on the left; both inferring in the present case a threshold magnitude of about 4.0. Application of this cutoff magnitude sharply decreases the number of events ($n_{M \geqslant 4.0} = 375$) in the data set available for further analysis, but at the expense of data loss homogeneity is satisfied. Detection decreases are usually related to the closure of stations. In local catalogs, detection decreases in one region can in principle be related also to installation of stations in another region and the associated increase in workload. In the present case, however, especially in the light of historical-political events of the 20th century in the region (wars, changes of state borders, etc.), the first possibility seems to us more probable.

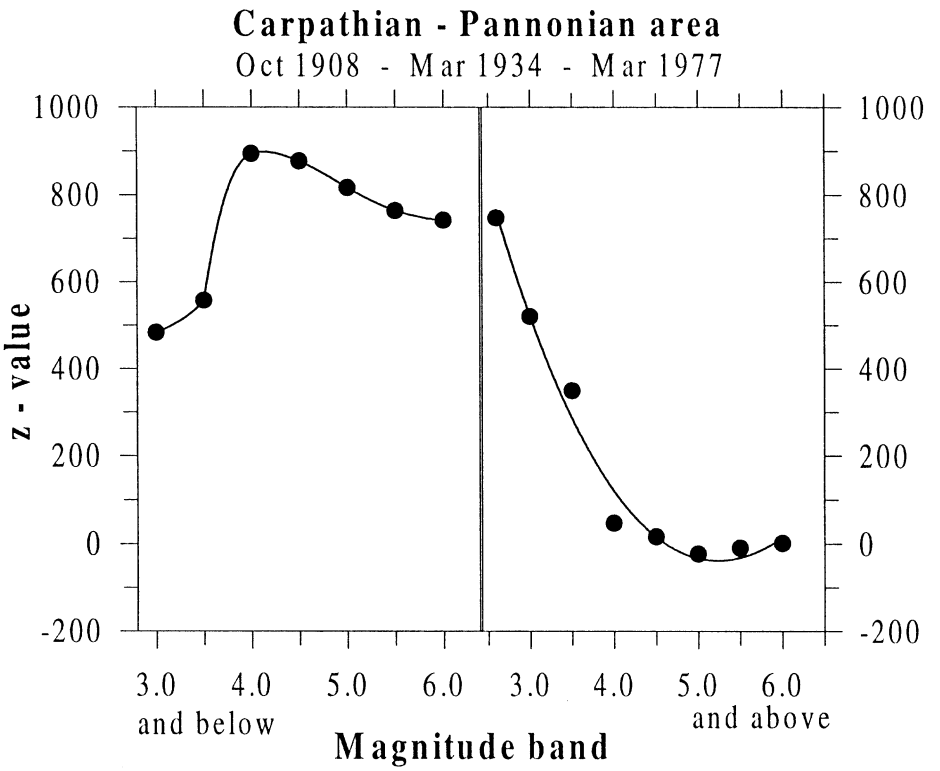

Fig. 6. Magnitude signature plot (distribution of seismicity rate changes in the magnitude domain) for the Carpathian–Pannonian area. Seismicity rates during the two time intervals are calculated for the $M \geqslant 2.6$ catalog. Magnitude band here along the horizontal axis includes all events below or above a given magnitude (e.g. the point above 5.0 on the left side of the plot shows the z-value for the events with $M < 5.0$, and on the right side of the plot for those with $M \geqslant 5.0$). The clearest cutoff is given by the leftmost boundary (in the present case $M \approx 4.0$) of the curve of z-values around zero on the right side of the plot. This is the lowest level from which detection is unaffected. Variations of z-value show characteristics of a detection decrease below magnitude $M = 4.0$ (see, e.g. Habermann, 1987).

An apparent decrease of the general seismicity level in the Carpathian basin from the 1st to the 2nd third of the 1900s is noted also by Zsíros (2000), without an attempt for interpretation. Although other types of man-made changes in seismicity rates have not been analysed in the present work, we believe that application of the relatively high cutoff magnitude found above can reasonably account also for possible reporting changes and magnitude shifts in the data set. Fig. 7 shows the analogous magnitude signature curves for the Greek data set. Comparison of seismic rates here is made for all events within two time intervals from October 1988 to June 1990 and June 1990 to June 1995 between three large ($M \geqslant 6.0$) earthquakes. The magnitude signature plot in this case shows signs of a detection increase, which, however, ceases from a cutoff magnitude of 3.5. As already mentioned in the previous section, the series of $M \geqslant 3.5$ events (with $n = 7951$) has been found stationary over the whole observation period. The present examples illustrate that the homogeneity test is necessary, since it helps avoiding problems associated with possibly spurious seismicity rate changes.

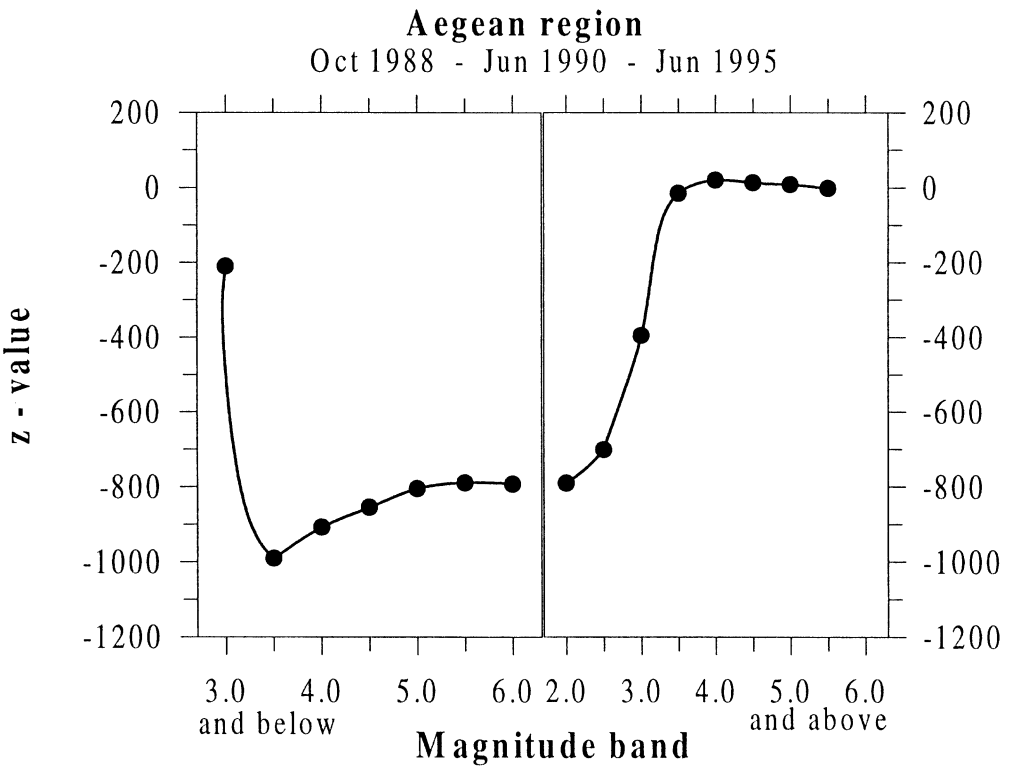

Fig. 7. Magnitude signature plot (comparison of seismicity rates within two time intervals as a function of magnitude bands) for the considered Pelopponesos–Aegean data set. The plot shows variations characteristic of a detection increase: negative z-values throughout the magnitude signature, lack of change ($z \sim 0$) in the data sets with larger events (on the right side of the plot) at $M \geqslant 3.5$, strongest change on the left side at magnitudes smaller than 3.5.

4. Neural-network modelling

For detailed information about the pattern recognition technique known as artificial neural-network modelling, the reader is referred to, e.g. Anderson (1995), only some basic features of the method and of the principles of its operation will be discussed below. Concerning applications to various geophysical problems, Raiche's (1991) review can be mentioned.

An artificial neural network (ANN) is a computing system made up of interconnected set of simple information processing elements, which, by analogy with biological nervous systems, are usually called neurons. The neuron collects inputs from single or multiple sources and produces output in accordance with a predetermined nonlinear function. A neural network model is created by interconnecting many of these neurons in a known configuration. Individually, the neurons perform trivial functions, but collectively, in the form of network, they are capable of solving complicated problems. Current ANNs are nevertheless much less complex than their biological counterparts, comprising much fewer components and operating in a way that is greatly simplified.

The type of network used in the present work appears the most popular one; the layered feed-forward network with error back propagation. Layered structure means that the network

includes an input layer, one or more hidden layers, and an output layer. In such a network, no communication is permitted between the neurons within a layer, and the processing elements do not take inputs from succeeding layers or from those before just the previous one. Fig. 8 illustrates a three-layer neural network scheme connected in feed-forward circuit. The propagation of the input signal through the system occurs in the following way. The input signal x_i, (an array of real values, each corresponding to an element of the input vector) is entered to each ith neuron in the input layer. The input neurons transmit these values across their links to the next (hidden) layer of neurons. Each link is associated with a weight w_{ij} which modifies the transmitted value. Every jth neuron of a hidden layer receives the summed transmitted values from all units of the previous layer and generates a bias b_j associated with that neuron. The result is then put through a transfer function f to generate a level activity h_j for this neuron. The activation levels of the hidden neurons are then transmitted across their outgoing links to the neurons of the output layer. As before, these values are weighted during transmission across the links, then summed at the output neuron and put again through an activation function. The level of activity generated at the output neurons is the network's solution to the problem presented at the inputs. The following set of equations provides a generalized description of the mode of operation of the above network, independently on the number of neurons in each layer

$$\begin{aligned} h_j &= f[\textstyle\sum_i (w_{ij} x_i) + b_j], \\ o_k &= f[\textstyle\sum_j (w_{jk} h_j) + b_k], \end{aligned} \tag{2}$$

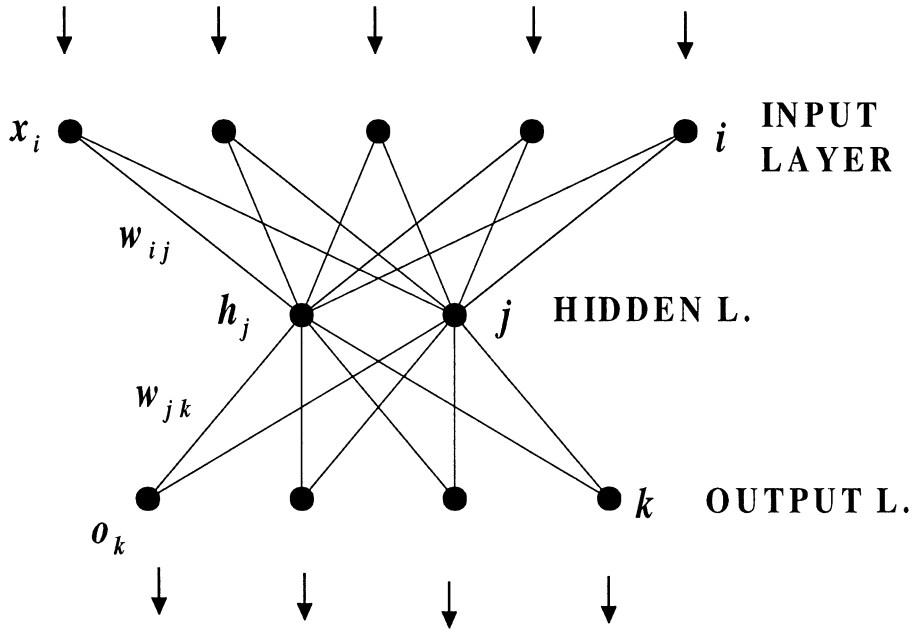

Fig. 8. Configuration of a three-layered feed-forward neural network.

where h_j is activity level generated at the jth hidden neuron, o_k the activity level generated by the kth output neuron, w_{ij} and w_{jk} the weights on the connections of the input-hidden and hidden-output layers of neurons, respectively, b_j and b_k are weighted biases. Eqs. (2) show that the total input to a unit j in the hidden layer, for example, is a linear function of the outputs from the units that are connected to j and of the weights w_{ij} on these connections (the bias b_j can be considered as the weight to an additional input with the value of 1). The output, however, depends on the activation function f. In principle, any monotonically increasing and continuously differentiable function can be used as an activation function for back-propagation type networks. However, the most commonly used ones are the nonlinear sigmoidal functions: hyperbolic tangent, and, like in present case, the 'logistic' function

$$f(z) = \frac{1}{1 + e^{-z}}, \tag{3}$$

where z stands for the weighted sum of outputs, the terms in the square brackets in Eqs. (2). The derivatives of such functions have a Gaussian shape that helps stabilize the network, and they are convenient also because of some finer features of the correction procedure of the connecting weights of the network (Anderson, 1995). Since the logistic function operates in the range from 0 to 1 for all values of z, the real input and output values used in training and testing must be scaled in this range.

The major concern in the development of a neural network is the determination of an appropriate set of weights that make it produce a desired output function. Error back propagation involves two phases: a feed-forward phase in which the external input information (in our case seismicity rates in selected time intervals, and occasionally also in some magnitude intervals) at the input neurons is propagated forward to compute the output information signal at the output units (occurrence times, or more precisely the actual time intervals between subsequent $M \geqslant 6.0$ earthquakes). Then follows a backward phase in which modifications to the connection strengths (weights) are made based on the differences between the computed and observed signals at the output units. It is assumed that the network does not have any a priori knowledge about the solution. The connection weights of the network are learned through a process called training. Initially, the network weights are assigned random values. Training actually means an iteration process, in which the feed-forward operation calculates an output pattern for each input pattern and then compares it with the correct (known from observations) output. The total error E based on the squared difference between predicted and actual outputs is computed for the whole training set as

$$E = \sum_{p=1}^{P} E_p = \sum_{p=1}^{P} \sum_{k=1}^{N} (T_{pk} - O_{pk})^2, \tag{4}$$

where P is the number of input–output pairs, E_p is the error for the pth input pattern, N is the number of outputs, and T_{pk}, O_{pk} are actual and predicted outputs, respectively. In the second phase of training, the adjustment of the connections strengths has been carried out using the standard error back-propagation algorithm (Rumelhart et al., 1986: Flood and Kartam, 1994), which minimizes the total error E with the steepest descent. The pattern errors, E_p, can be

assumed to be a function of the multidimensional weight space, visualized as a surface of peaks and valleys. The valleys are the minima in pattern errors that are sought by the gradient descent procedure. At the beginning of training, the location on the error surface will be near peaks, with movement to a minima being achieved by progressively correcting the interconnection weights by small amounts $\Delta_p w_{ij}$ calculated from the gradient descent rule:

$$\Delta_p w_{ij} = -\eta \frac{\partial E_p}{\partial w_{ij}}, \tag{5}$$

so that when a new input–output pattern pair is presented to the network, the error is reduced and the output is closer to the required one. The latter expression shows that the change in an interconnection weight between arbitrary layers i and j on pattern p is proportional to the slope of error surface ($\partial E_p/\partial w_{ij}$), and the proportionality constant η (usually called learning rate) plays an important role. If it has small values, oscillations in weight corrections during training can be avoided, but the number of iterations may become very large. Convergence can be accelerated with the introduction of a so called momentum term with a characteristic constant α, which makes the current weight change in iteration $n+1$ proportional to the previous weight change at iteration n. In the present work, the learning rate and the momentum constant were set to 0.5 and 0.9, respectively (values consistent with experience in the literature; e.g. Smith and Eli, 1995). To find the optimum number of neurons in the hidden layer, an algorithm for dynamic creation of hidden-layer neurons during training has been used (Hirose et al., 1991).

5. Results

Even from the very superficial and short information on the seismotectonic features and earthquake data in Section 2 one might have the impression that the selected study areas seem markedly different from this point of view. We believe that the considered Aegean earthquake data set is much more homogeneous and may be more representative of the real situation in the area than its Hungarian counterpart. This turns out particularly from the fact that we could retain for analysis only a little fraction of the available data set ($M \geqslant 4.0$ events, with $n = 375$) from the Carpathian region, compared to the much larger set of the $M \geqslant 3.5$ earthquakes (including 7951 events) in the Aegean area. Besides, a general shortcoming of the whole treatment of earthquake phenomenology in the present study should also be emphasized. We are concerned with seismicity rates and earthquake occurrence times, i.e. treat only an essentially one-dimensional aspect of a multidimensional process, and this always invites bias (Davis et al., 1989). It seems reasonable to assume that the larger the spatial extent of a region, the larger this effect may appear. Therefore we consider the Aegean region case more informative, and prefer to start the discussion of our results with it.

5.1. Aegean area

In the present work, we are concerned with ANNs which can predict earthquakes of magnitudes $M \geqslant 6.0$, the largest events in our data sets. Nine earthquakes in the magnitude range 6.0–

6.9 have occurred in the Peloponnesos-Aegean area for the considered time interval, with the mean recurrence time of 1.7 yr. (see also Fig. 5). The ANN used is a three-layer feed-forward network. We had the intention to learn/simulate the occurrence times for large earthquakes. Thus, the output values are the time intervals between subsequent large earthquakes. As already mentioned above, the input information used for training the network are mean seismicity rates in selected time intervals (acually, within the time intervals between the $M \geqslant 6.0$ earthquakes), and occasionally (when such data are representative) also in selected magnitude ranges. Such a choice needs some explanation.

The times when large earthquakes occur are typically associated with changes in seismic activity. However, assumptions underlying the choice of possible precursors for large earthquakes may be quite different. An attempt of quantifying the predictive utility of various seismicity parameters has been made by Eneva and Ben-Zion (1997a) for synthetic earthquake catalogs. The parameters most commonly employed in reports of precursory changes of seismicity are the *b*-value of the frequency-size statistics of earthquakes and seismicity rates. In the present analysis the mean seismicity rates within the intervals between large earthquakes, calculated from the cumulative activity curves by least-squares fitting, have been used as input values to the network. The use of mean seismicity rates for input parameters makes the present approach different from more usual ones. Most existing approaches focus on changes in the vicinity of large earthquakes and/or threshold values. In such a case, however, seismicity changes only in the neighbourhood of a large event are taken into consideration. Efforts should be made to distinguish between anomalies which precede or follow large events. It is often difficult to make distinction between precursory effects that precede a given large event and aftereffects that follow the previous large event, especially when the frequency of events to be "predicted" increases. Generally, it is not clear how many observations one should make to identify any precursory trend for a subsequent large event. On the other hand, mean seismicity rates characterize the strain accumulation and release process on the average, can be calculated reliably enough, and permit to include into modelling time intervals far before the occurrence of large earthquakes.

For successful training the input-output set (the training sample) should be representative of the general characteristics of the modelled problem. Numerical experiments have been performed to find the best configuration of the input set. Use of two seismic rates, v_1 for events in the magnitude range of $3.5 \leqslant M < 4.0$ (with $n_1 = 5157$) and v_2 for quakes with $M \geqslant 4.0$ ($n_2 = 2794$), ensured a better performance of the network. The neural network training could not be carried out to sufficient training tolerance when the input consists of the single set of seismic rate values for the full magnitude range. This is probably a result of the existence of clear differences in the variations of seismic rates with the size of events (see Section 3, and also Figs. 4 and 5).

We have considered seismic rates $v_{1,p}$ and $v_{2,p}$, where p denotes the time interval between the pth and the next large earthquakes. For the Aegean data set $p = 1,...,8$. Experiments with different input layer sets have shown that the optimum configuration consisted of four neurons: $v_{1,p-1}$, $v_{2,p-1}$, $v_{1,p}$, v_{2p}. Scatter diagram 9, illustrating observed vs. predicted time intervals between neighbouring large earthquakes, shows the results of network training. Since the activity function falls in the range of 0–1 for all values of parameters, the input-output amounts have been normalized to this range. Seismicity rates v_1 and v_2 showed values in the ranges of 180–540 and 120–1300 events/yr, respectively. The time intervals between large earthquakes were scaled from the actually observed range of 0.004–4.997 yr into 0–1.

The more perfectly the neural network is trained, the closer a point falls on the straight line in Fig. 9. On that line the output of neural network would exactly coincide with the observed values. Fig. 9 shows that training becomes poorer in the case when only one input pair, $v_{1,p}$ and $v_{2,p}$, was used in the procedure. On the other hand, the use of three input pairs did not improve reasonably the performance of the network. The specific structure of the network input which is able to produce adequate occurrence times for large earthquakes generates the question of whether this particular configuration may have some identifiable relationship to the physics of the earthquake generation process. A possible explanation for the use of more than one intervals between subsequent earthquakes may be that, by itself, even a large earthquake cannot release most of the accumulated strain.

As already mentioned above, in present case we had eight sets of data available for training the neural network. Generally, the number of training patterns required is determined by the nature of the problem and the anticipated performance of a trained ANN. Although some rules give guidance in this respect (Baum and Haussler, 1989), experience plays a crucial role in selecting the suitable number of training sets. Too small number results in a poorly trained ANN with the total error exceeding ~10%, generally accepted in ANN-modelling (Koons and Gorney, 1990).

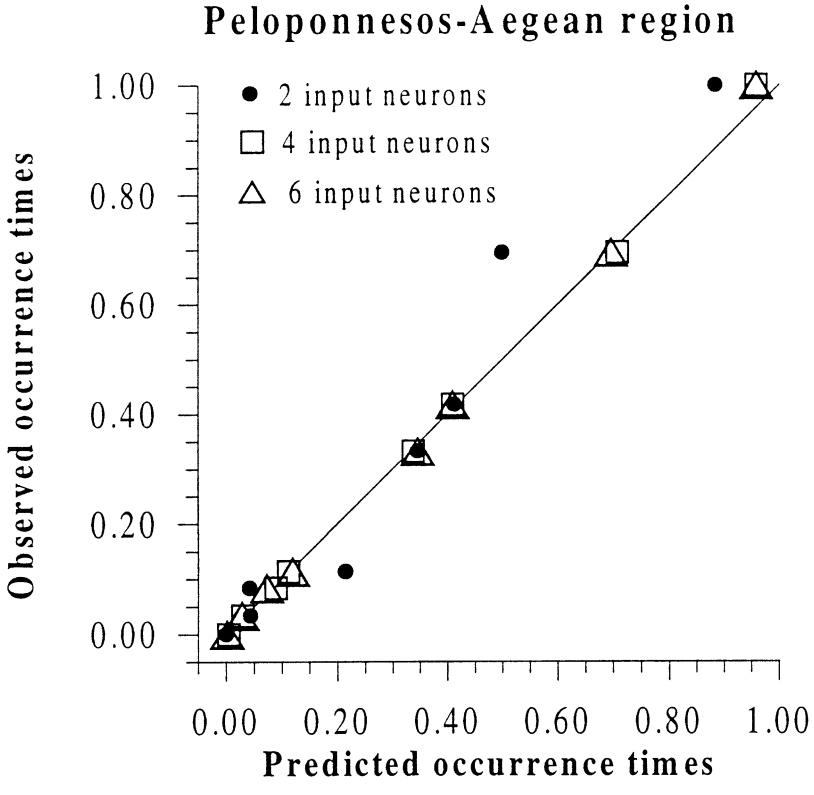

Fig. 9. Scatter diagram showing the predicted origin times versus the observed occurrence times for large earthquakes in the Peloponnesos–Aegean area. Black dots show solution when the network operates with two input neurons. Open squares: four input neurons. Triangles: six input neurons. Data are scaled in the 0–1 range.

On the other hand, too large a number leads to overtraining the network, which reduces its ability to generalize.

Fig. 10 gives a view of the accuracy of training achieved with the training set of four input seismicity rates designed for nine large earthquakes. To obtain convergence, the training criterion was set to 10% of the normalized range, which corresponds to the accuracy of about ± 0.5 yr for the earthquake occurrence times. The performance of network was tested by removing one of the earthquakes from the training set, carrying out retraining, and predicting the time of occurrence of the missing event. The results of testing when the particular $M = 6.1$ earthquake of 14 June 1995 (large event with the longest recurrence time; an extreme event in this respect) was removed are shown in Fig. 10 by crosses. Note that the largest recurrence time is predicted to be only 6% worse than that obtained previously with the whole training set.

After training is accomplished we obtain the final set of weights and thresholds, by means of which the network "remembers" each presented input and output patterns, and, taking into account training information, becomes able to respond upon unknown input. In this way the constructed network can be used to recognize and/or generate patterns given by new inputs.

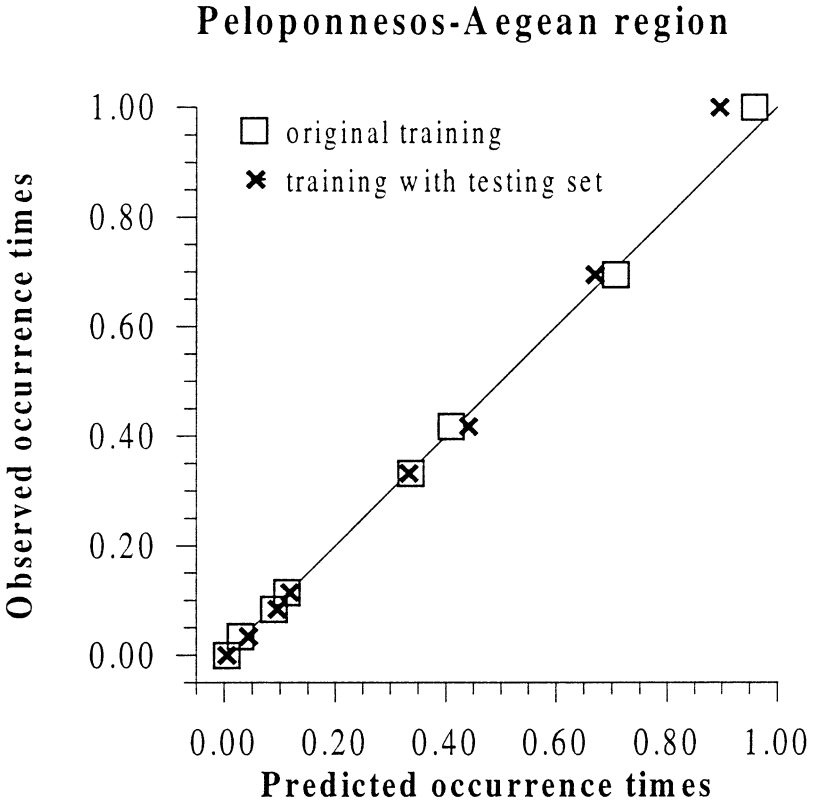

Fig. 10. Scatter diagram illustrating the performance of the constructed network with four input neurons for the Aegean data set. Open squares: the entire training set is used to predict the earthquake origin times. Crosses: prediction for earthquake occurrences when the last large earthquake is removed from the training set.

We applied the trained network to predict the time of occurrence of the next large earthquake that can be expected in the area. For this "prediction", in both of the mentioned magnitude bands the input data set contained seismicity rates for the interval between the last two observed large earthquakes at 1990.455 and 1995.452 ($v_{1,8}$, $v_{2,8}$) and corresponding seismic rates in the subsequent interval of 1995.452–1996.996 ($v_{1,9}$, $v_{2,9}$) lasting to the end of data series considered. Thus we use the intrinsic assumption that the considered mean seismicity rates for the interval between the last $M \geqslant 6.0$ event at 1995.452 and the forthcoming large earthquake remain the same as they were for the time section from the last big earthquake to the end of the available data set (i.e. between 1995.452 and 1996.996). We admit that, in principle, the best training pattern would be time dependent, since new events may change seismicity rates based on past experience. Allowing for this bias, the predicted date of the forthcoming large earthquake is 0.89768, which, when converted into absolute units, means that one can expect the next $M \geqslant 6.0$ earthquake in the study area at A.D. 1999.94, i.e. at around 10 December 1999. However, a supplement of the considered data set by the list of quakes for the recent time since 1997.0 can, in principle, change noticeably the seismicity rates used in the above prediction, and by means of it also the obtained date for the impending large earthquake. Since the time of performance of the above calculations, a disastrous, magnitude 5.9 earthquake did occur in the study area (with its epicenter near Athens) on 7 September 1999. The date of occurrence of this earthquake falls certainly well into the bounds of uncertainty of our estimation.

5.2. Carpathian–Pannonian region

Because of the scarcity of large earthquakes, with a mean recurrence time of the order of $\sim$12 years, the design of the neural network in this case is more problematic. Ten earthquakes with magnitudes of 6.0–7.2 have occurred in the area since 1880 A.D., all without exception at the periphery of the region, in Vrancea and Háromszék Mts., Transsylvania, etc. Due to heterogeneity of data, only the events with $M \geqslant 4.0$ (see above, Section 3) have been used for ANN training. This data set includes 375 events only, with two intervals between large earthquakes with only by three events in each. Thus, it was not possible to obtain reasonable values for seismicity rates in different magnitude bands. For neural network training we were able to use only a single seismic rate value in each time interval, obtained for all $M \geqslant 4.0$ events. The calculated seismicity rates range in the interval of 1–15 events per year, i.e. they are by about two orders of magnitude smaller than the values characteristic of the Aegean area.

As previously, we intended to model/predict the occurrence times of large events, thus the network's output values were the time intervals between neighbouring large earthquakes. According to numerical experiments, the best performance of the neural network was achieved when the system had the input: v_{p-1}, v_p, M_{p-1}, M_p, where v_p is the mean seismicity rate in the interval between the p-th and the next earthquakes, and M_p is magnitude of the pth event. As previously, the use of information about not one but two previous earthquakes proved indispensable for a successful network construction. It was not possible to achieve a sufficient training tolerance without considering the magnitude values. On the other hand, the inclusion of magnitudes in the input set for the Aegean data did not influence noticeably the performance of the network. This can be explained by the fact that most of the large Aegean earthquakes show magnitudes at

around 6.0–6.2, thus the higher magnitudes are not well represented in the training set. Magnitudes of the Carpathian earthquakes, however, are more uniformly distributed between 6.0 and 7.2.

It can be seen in the scatter diagram in Fig. 11 that the predicted earthquake occurrence times fall quite close to actual data. As previously, the training criterion was set to 10% of the range, which corresponds to ± 2.8 years in terms of earthquake occurrence time. This interval is by 5–6 times larger than the corresponding value in the ANN network for the Aegean data, but is not worse than the uncertainties of other methods. Eneva and Ben-Zion (1997b), for example, in a way of examining synthetic earthquake catalogs for earthquake precursors, have shown that earthquakes with $M \geqslant 6.0$ can generally be expected within 2.5 yr of reaching local extrema in the study parameters. Estimating the occurrence time of the next large earthquake in the Carpathian area in the way as described above, we obtained the date of 2012.05 (i.e. 18 January 2012) for this event. To calculate the mean seismicity rates necessary for this prediction, the time interval between 1986.663 (last large event) and 1994.922 (end of data series) and that between the preceding two large events have been considered. Again, newer data from more recent time in principle can modify the predicted date.

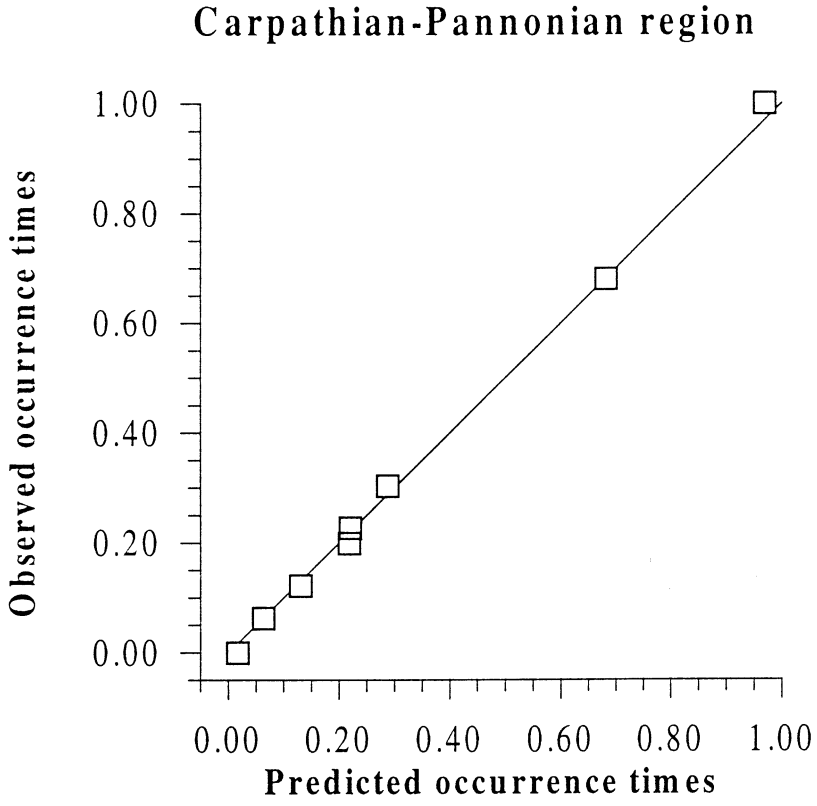

Fig. 11. Scatter diagram showing the predicted origin times versus the actual origin times for large earthquakes in the Carpathian–Pannonian region. Neural network operates with four input neurons.

6. Concluding remarks

Earthquake prediction is the greatest unsolved problem of seismology. The complexity of seismic phenomena justifies the attempts to use different approaches in which various models are constructed, each reproducing in one way or other certain observational results. The investigation of the long-term predictability of earthquakes is generally based on seismicity catalogs representing the most reliable basis for quantitative treatments in the search for precursory patterns. The success of ideas about statistical predictability of large earthquakes (concept of self-organized criticality; Bak and Tang, 1989) provides a rationale to expect predictability also in real sense. An attempt at such prediction may be the use of ANN models which are fairly schematic, but have proved an effective tool for reproducing many features of seismicity patterns. The need in such "black-box" models is motivated by the nonlinear multivariate nature of the earthquake generation process and our insufficient knowledge of its physics.

In the present work we have sampled two seismicity data sets and constructed ANN models suitable for the prediction of the origin times of large earthquakes. Regardless of certain inadequacy of the network training sets due to scarcity of available large events in the data sets, we found the performance of the constructed ANNs quite satisfactory, and this supports the usefulness of the application of this tool to similar problems.

Beside its inherent "black-box" nature, the present technique differs from other existing approaches by the use of mean seismicity rates over intervals between large earthquakes, instead of considering sequences of changes or empirically determined threshold (alarm) values to identify significant anomalies. This approach makes it possible to take in consideration also the information about previous, older large earthquakes. Of different parameters characterizing the spatial, temporal and size distribution of earthquake populations, in the present case a single seismic rate is used for analysis. Incorporation of other seismicity parameters into input sets for ANN training, may be a good objective for future research.

Acknowledgements

This research was financially supported by the National Science Research Foundation (OTKA) of Hungary through grant No. T 022954.

References

Anderson, J.A., 1995. An Introduction to Neural Networks (Neural Network Modelling and Connectionism). MIT Press, Cambridge, MA.

Bak, P., Tang, C., 1989. Earthquakes as self-organized critical phenomena. J. Geophys. Res. 94, 15635–15637.

Baum, E.B., Haussler, O., 1989. What size net gives valid generalization. Neural Comput 1, 151–160.

Bufe, C.G., Varnes, D.J., 1993. Predictive modeling of the seismic cycle of the great San Francisco Bay region. J. Geophys. Res. 98, 9871–9883.

Carlson, J.M., 1991. Time intervals between characteristic earthquakes and correlations with smaller events: an analysis based on mechanical model of a fault, to appear. J. Geophys. Res. 96, 4255–4267.

Chiaruttini, C., Roberto, V., Saitta, F., 1989. Artificial intelligence techniques in seismic signal interpretation. Geophys. J. Int. 98, 223–232.

Chiaruttini, C., Salemi, G., 1993. Artificial intelligence techniques in the analysis of digital seismograms. Comput.-Geosci. 19, 149–156.

Dai, H., MacBeth, C., 1997. The application of back-propagation neural network to automatic picking seismic arrivals from single-component recordings. J. Geophys. Res. 102, 15105–15113.

Davis, P.M., Jackson, D.D., Kagan, Y.Y., 1989. The longer it has been since the last earthquake, the longer the expected time till the next?. Bull. Seism. Soc. Am. 79, 1439–1456.

Eneva, M., Ben-Zion, Y., 1997a. Techniques and parameters to analyze seismicity patterns associated with large earthquakes. J. Geophys. Res. 102, 17785–17795.

Eneva, M., Ben-Zion, Y., 1997b. Application of pattern recognition techniques to earthquake catalogs generated by models of segmented fault systems in three-dimensional elastic solids. J. Geophys. Res. 102, 24513–24528.

Eneva, M., Habermann, R.E., Hamburger, M.W., 1994. Artificial and natural changes in the rates of seismic activity: a case study of the Garm region, Tadjikistan (CIS). Geophys. J. Int. 116, 157–172.

Flood, I., Kartam, N., 1994. Neural networks in civil engineering. I: principles and understanding. J. Comp. Civ. Eng. 8, 131–148.

Fraser-Smith, A.C., Bernardi, A., McGill, P.R., Ladd, M.E., Helliwell, R.A., Villard Jr, O.G., 1990. Low-frequency magnetic measurements near the epicenter of the M_s 7.1 Loma Prieta earthquake. Geophys. Res. Lett. 17, 1465–1468.

Geller, R., 1966a. Debate on VAN. Geophys. Res. Leu. 23, 1291–1452.

Geller, R., 1996b. Debate on evaluation of the VAN method: Editor's introduction. Geophys. Res. Lett. 23, 1291–1293.

Gutdeutsch, R., Aric, K., 1988. Seismicity and neotectonics of the East Alpine–Carpathian and Pannonian area. In: Royden, L.H., Horváth, F. (Eds.), The Pannonian Basin—A Study in Basin Evolution. AAPG Memoir 45. Am. Assoc. Petrol. Geol. Publ., Tulsa, OK, pp. 183–194.

Habermann, R.E., 1987. Man-made changes of seismicity rates. Bull. Seism. Soc. Am. 77, 141–159.

Habermann, R.E., 1988. Precursory seismic quiescence: past, present and future. Pure Appl. Geophys. 126, 279–318.

Habermann, R.E., 1991. Seismicity rate variations and systematic changes in magnitudes in teleseismic catalogs. Tectonophysics 193, 277–290.

Hirose, Y., Yamashita, K., Hijiya, S., 1991. Back-propagation algorithm which varies the number of hidden units. Neural Networks 4, 61–66.

Kagan, Y.Y., 1996. VAN earthquake predictions—an attempt at statistical evaluation. Geophys. Res. Lett. 23, 1315–1318.

Kagan, Y.Y., Jackson, D., 1991. Long-term earthquake clustering. Geophys. J. Int. 104, 117–133.

Kiratzi, A., Papazachos, C.B., 1995. Active crustal deformation from the Azores triple junction to the Middle East. Tectonophysics 243, 1–24.

Koons, H.C., Gorney, D.J., 1990. A sunspot maximum prediction using a neural network. EOS, Trans. Am. Geophys. Union 71, 677–678.

Main, I.G., 1995. Earthquakes as critical phenomena: implications for probabilistic seismic hazard analysis. Bull.-Seism.Soc.Am. 85, 1299–1308.

Matthews, M.V., Reasenberg, P.A., 1988. Statistical methods for investigating quiescence and other temporal seismicity patterns. Pure Appl.Geophys. 126, 357–372.

Ogata, Y., 1988. Statistical models for earthquake occurrence and residual analysis for point processes. J.Am. Statist.Assoc. 83, 9–27.

Omori, F., 1894. On the aftershocks of earthquakes. J. Coll. Sci., Tokyo Imp. Univ. 7, 111–200.

Papazachos, B., Kiratzi, A., Papadimitriou, E., 1991. Regional focal mechanism for earthquakes in the Aegean sea. Pure Appl. Geophys. 136, 405–420.

Prozorov, A.G., Dziewonski, A.M., 1982. A method of studying variations in the clustering property of earthquakes: application to the analysis of global seismicity. J. Geophys. Res. 87, 2829–2839.

Raiche, A., 1991. A pattern recognition approach to geophysical inversion using neural networks. Geophys. J. Int. 105, 629–648.

Reasenberg, P., 1985. Second-order moment of Central California seismicity, 1969–1982. J. Geophys. Res. 90, 5479–5495.

Roeloffs, E.A., 1988. Hydrologic precursors to earthquakes: a review. Pure Appl. Geophys. 126, 177–209.
Royden, L.H., Horváth, F., 394h. The Pannonian Basin—A Study in Basin Evolution. AAPG Memoir 45. Am. Assoc. Petrol. Geol. Publ, Tulsa, OK.
Rumelhart, D.E., Hinton, G.E., Williams, R.S., 1986. Learning representations by back-propagating errors. Nature 323, 533–536.
Shlien, S., Toksöz, N.M., 1970. Clustering model for earthquake occurrences. Bull. Seism. Soc. Am. 60, 1765–1787.
Smith, J., Eli, R.N., 1995. Neural-network models of rainfall-runoff process. J. Water Res. Plan. Mngmnt. 121, 499–514.
Shaw, B.E., Carlson, J.M., Langer, J.S., 1992. Patterns of seismic activity preceding large earthquakes. J. Geophys. Res. 97, 479–488.
Smalley Jr., R.F., Chatelain, J.-L., Turcotte, D.L., Prévot, R., 1978. A fractal approach to the clustering of earthquakes: application to the seismicity of the New Hebrides. Bull. Seism. Soc. Am. 77, 1368–1381.
Sykes, L.R., 1983. Predicting great earthquakes. In: Kanamori, H., Bosch, E. (Eds.), Earthquakes: Observation, Theory and Interpretation. Proc. Enrico Fermi Int. School of Physics, Amsterdam, pp. 398–411.
Triep, E.G., Sykes, L.R., 1997. Frequency of occurrence of moderate to great earthquakes in intracontinental regions: implications for changes in stress, earthquake prediction, and hazard assessments. J. Geophys. Res. 102, 9923–9948.
Varotsos, P., Kulhanek, O., 1993. Measurement and theoretical models of the earth's electric field variations related to earthquakes. Tectonophysics (special issue) 224, 1–288.
Wallace, R.E., Davis, J.F., McNally, K.C., 1984. Terms for expressing earthquake potential, prediction, and probability. Bull. Seism. Soc. Am. 74, 1819–1825.
Wyss, M., Burford, R.O., 1987. Occurrence of a predicted earthquake on the San Andreas fault. Nature 329, 323–325.
Zsíros, T. 2000. Seismicity and Seismic Hazard in the Carpathian Basin: Hungarian Earthquake Catalog (456–1995). Seismol. Observatory, Geod. Geophys. Res. Ins, Hung. Acad. Sci. Publ., Budapest (in Hungarian).
Zsíros, T., Mónus, P., Tóth, L., 1988. Hungarian Earthquake Catalog (456-1986). Seismol. Observatory, Geod. Geophys. Res. Inst., Hung. Acad. Sci. Publ., Budapest.