International Journal of Neural Systems, Vol. 17, No. 1 (2007) 13–33

World Scientific
www.worldscientific.com

NEURAL NETWORK MODELS FOR EARTHQUAKE MAGNITUDE PREDICTION USING MULTIPLE SEISMICITY INDICATORS

ASHIF PANAKKAT and HOJJAT ADELI*
Department of Civil and Environmental Engineering and Geodetic Science,
The Ohio State University, Columbus, Ohio 43210 USA
**adeli.1@osu.edu*

Neural networks are investigated for predicting the magnitude of the largest seismic event in the following month based on the analysis of eight mathematically computed parameters known as seismicity indicators. The indicators are selected based on the Gutenberg-Richter and characteristic earthquake magnitude distribution and also on the conclusions drawn by recent earthquake prediction studies. Since there is no known established mathematical or even empirical relationship between these indicators and the location and magnitude of a succeeding earthquake in a particular time window, the problem is modeled using three different neural networks: a feed-forward Levenberg-Marquardt backpropagation (LMBP) neural network, a recurrent neural network, and a radial basis function (RBF) neural network. Prediction accuracies of the models are evaluated using four different statistical measures: the probability of detection, the false alarm ratio, the frequency bias, and the true skill score or R score. The models are trained and tested using data for two seismically different regions: Southern California and the San Francisco bay region. Overall the recurrent neural network model yields the best prediction accuracies compared with LMBP and RBF networks. While at the present earthquake prediction cannot be made with a high degree of certainty this research provides a scientific approach for evaluating the short-term seismic hazard potential of a region.

Keywords: Earthquake prediction; neural networks; seismicity; seismology.

1. Introduction

Prediction of the time of occurrence, magnitude, and epicentral location of future large earthquakes has been the subject of a number of scientific efforts with distinctly different conclusions in recent years. While some scientists have concluded that the prediction of time of occurrence, magnitude, or location of a single future earthquake cannot be done,[29] others have suggested several procedures for predicting these parameters.[18,41,52,25,15,16,47,48,28,54,36,72,31,51,60] Such procedures are based on either the study of precursory phenomena before earthquakes such as seismic quiescence, changes in magnetic and electric signals recorded at seismic sites, and abnormal animal behavior, or the analysis of historical earthquake data recorded in seismic catalogs.

Earthquake magnitude prediction studies based on the analysis of historical earthquake data assumes a temporal distribution model for earthquake magnitude for the seismic region considered in the study. Such models describe the frequencies of occurrence of seismic events as functions of their magnitudes. The most widely used magnitude-frequency model for hazard estimation is that based on the Gutenberg-Richter inverse power law.[33] For example, it is the basis for the new U.S. National Earthquake Hazard Maps.[54] The law is based on an original empirical power law put forth by Ishimito and Iida.[37]

The Gutenberg-Richter inverse power-law establishes an inverse linear relationship between the magnitude of seismic events and the logarithm of frequency of occurrence of events of magnitude equal or lower than that magnitude. Probabilistic earthquake magnitude-frequency distributions such as those based on the Gamma and Weibul distributions have also been employed to relate earthquake

*Corresponding author.

magnitudes and frequencies for certain seismic zones.[6,32]

Another temporal earthquake distribution model that is being given increasing scientific attention in recent years is the characteristic earthquake distribution model originally proposed by Kagan and Jackson.[42] Several active seismic zones exhibit a "recurring" or "characteristic" trend when it comes to major seismic events (also known as characteristic events). In such regions, characteristic events are punctuated by almost uniform time intervals. Several recent attempts have been made to quantify the characteristic magnitude and intervening time period for major seismic zones.[31,58,62]

Earthquake parameter prediction based on the analysis of observed precursor signals prior to major earthquakes is also a popular field of research. For example, earthquake precursors such as foreshocks before large earthquakes have been used to predict earthquake parameters.[14] Foreshocks can be defined as a series of seismic events of magnitudes higher than the magnitude of normal or background seismic activity for a particular region that precede a large earthquake.

Another important focus of earthquake prediction studies has been the phenomenon of seismic quiescence observed in certain regions.[16] Seismic quiescence is defined as the apparent lull in the normal seismicity of a region for some time before a major earthquake. Normal seismicity for a seismic zone is usually measured in terms of the earthquake energy released or the rate of its release. The extent of quiescence has been related to the magnitude of the succeeding earthquake based on the elastic rebound theory.[64] The rate of release of seismic energy or its square root may provide a means of predicting future earthquakes.

However, due to the extreme non-linear and complex geophysical processes that lead to earthquake occurrence, there is no accurate mathematical or empirical relationship between any physically recordable parameter and the time of occurrence, magnitude or location of a future earthquake. There are several factors that contribute to the non-linearity of earthquake occurrence time, location, and magnitude such as stress state of the fault after the last earthquake, healing of gouge and other chemical processes, changes in pore pressure due to compaction and fluid migration.[54] Therefore, such relationships, if existing, are expected to be highly non-linear and complex.

In general, neural network modeling has been found to be an effective solution approach when (a) many variables of diverse types need to be included in problem-formulation, (b) the form of relationship between the dependent and independent variables is unknown, and (c) future data needs to be included in the model.[63] Neural networks have been used for solution of a variety of complex problems from image recognition,[2] to design automation,[7] construction scheduling,[4] automatic detection of traffic incidents in intelligent freeway systems,[5] and forecasting nonlinear natural phenomena such as earthquakes[49,50,52,10,46,51] and avalanches.[63]

In this work, three neural network models are presented for predicting the magnitude of the largest seismic event in the following month based on the analysis of eight mathematically defined seismicity indicators. For a given seismic region, eight seismicity indicators calculated from a pre-defined number of significant seismic events (here-on called simply events) before the beginning of each month are used as inputs to predict the occurrence or non-occurrence of an earthquake of a pre-defined threshold magnitude or more during that month using a multi-layer neural network.

2. Seismicity Indicators

In this section, eight mathematically-defined seismic parameters known as seismicity indicators are presented. These indicators can be used to evaluate the seismic potential of a region. Three of the eight parameters are independent of the temporal distribution of the earthquake magnitude assumed. They are the time elapsed over a predefined number (n) of events (T), the mean magnitude of the last n events (M_{mean}), and the rate of release of square root of energy ($dE^{1/2}$).

Three indicators are based on the Gutenberg–Richter inverse power law temporal magnitude distribution. They are the slope of the Gutenberg-Richter inverse power-law curve, known as the b value, the summation of the mean square deviation about the regression line based on the Gutenberg-Richter inverse power law, known as the η value, and the magnitude deficit or the difference between observed and expected magnitudes based on the Gutenberg-Richter inverse power law or the ΔM value.[33]

The remaining two indicators are based on the characteristic temporal earthquake magnitude distribution. They are the mean time between

Table 1. Seismicity indicator symbols and their description and magnitude distribution (GR represents Gutenberg-Richter power-law distribution and C represents characteristic earthquake distribution).

Seismicity indicator	Description	Assumed distribution
T	Elapsed time	—
M_{mean}	Mean magnitude	—
$dE^{1/2}$	Rate of square root of seismic energy	—
β	Slope of magnitude-log (frequency) Plot	GR
η	Mean square deviation	GR
ΔM	Magnitude deficit	GR
μ	Mean time	C
c	Coefficient of variation	C

characteristic or typical events, known as the μ value and the coefficient of variation of the mean time or the aperiodicity of the mean, known as the c value.[42] The eight indicators, their descriptions, and the magnitude distribution assumed as their basis are summarized in Table 1.

(1) The T value

The time elapsed over the last n events of magnitude greater than a predefined threshold value is defined as

$$T = t_n - t_1 \tag{1}$$

where t_n is the time of occurrence of the n^{th} event and t_1 is the time of occurrence of the 1^{st} event. Most earthquakes are preceded by a significant precursor activity such as a series of smaller magnitude quakes called foreshocks.[14] In fact, some of the most popular earthquake prediction models such as the colliding cascades model[72] and other earthquake dynamic studies are based on observing the frequency and intensity of foreshocks. The T value can be a measure of the frequency of foreshocks depending on the threshold value chosen for the magnitude. In this case a large T value indicates a relative lack of foreshocks which in many seismic regions may indicate a lower probability of occurrence of a forthcoming large seismic event. Inversely a small T value indicates a relatively high foreshock frequency and a higher probability of occurrence of a forthcoming large seismic event.

(2) The Mean Magnitude

The mean of the Richter magnitudes of the last n events is defined as

$$M_{mean} = \Sigma M_i/n \tag{2}$$

Together with the T value (which is a measure of the frequency of foreshocks), the mean of the magnitudes of foreshocks is also a crucial indicator of an impending earthquake in some regions. According to the accelerated release hypothesis[18] and its modifications,[40,67] the energy released from a fractured fault increases exponentially as the time to earthquake occurrence gets shorter. In other words, the observed magnitudes of foreshocks increase immediately before the occurrence of a major earthquake.

(3) The rate of square root of seismic energy released ($dE^{1/2}$)

The rate of square root of seismic energy released over time T is defined as

$$dE^{1/2} = \Sigma E^{1/2}/T \tag{3}$$

where $E^{1/2}$ is the square root of seismic energy (E) calculated from the corresponding Richter magnitude using the following empirical relationship[45]

$$E = 10^{(11.8+1.5M)}\text{ergs} \tag{4}$$

Most seismic regions can be approximated as open physical systems with gradual build-up of energy through the movement of lithospheri plates. Such systems remain in relative equilibrium if this gradual built-up is released through regular low-magnitude (or background) seismic activity.[58] If the background seismic activity is disrupted for significantly long periods of time (seismic quiescence) due to frictional or mechanical reasons, the physical system accumulates energy that will be released abruptly in the form of a major seismic event when the stored energy reaches a threshold.[65] Figure 1 shows the plot

Table 2. Earthquake Magnitude Recordings in Southern California for the 10 years between Jan. 1st 1987 and Dec. 31st 1996, showing the distribution similar to Magnitude-Frequency inverse-power law distribution of Gutenberg and Richter.

Magnitude (M) range	Count per M range	Cumulative total above lower M in range
2.5–2.9	9471	13590
3.0–3.4	2784	4119
3.5–3.9	912	1335
4.0–4.4	285	423
4.5–4.9	90	128
5.0–5.4	32	48
5.5–5.9	10	16
6.0–6.4	3	6
6.5–6.9	2	3
7.0–7.4	1	1

Table 3. Peak Magnitude and corresponding seismic energy for the 12 months of 1994.

Month	Peak recorded magnitude	Seismic energy released (ergs)
Jan 94	6.70	4.31×10^{20}
Feb 94	4.09	8.91×10^{17}
Mar 94	5.24	4.57×10^{19}
Apr 94	4.78	9.66×10^{18}
May 94	4.40	2.6×10^{18}
Jun 94	4.97	1.79×10^{19}
Jul 94	3.93	4.95×10^{17}
Aug 94	4.85	1.18×10^{19}
Sep 94	3.76	2.75×10^{17}
Oct 94	4.18	1.17×10^{18}
Nov 94	4.21	1.34×10^{18}
Dec 94	4.88	4.07×10^{18}

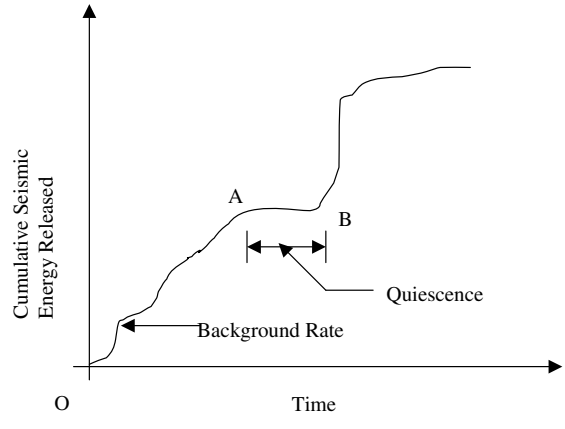

Fig. 1. Cumulative seismic energy released versus time illustrating the earthquake occurrence process for a hypothetical seismic region that exhibits seismic quiescence.

of the cumulative seismic energy released versus time for a hypothetical seismic region, representative of the earthquake occurrence process in several seismic regions.[68,73,20] In Fig. 1, the region of the graph between points O and A is approximately linear and depicts background seismic activity in the region. The graph between A and B is a plateau showing disruption in background seismic activity or seismic quiescence. Cumulative energy released increases abruptly at point B (end of quiescence where stored energy reaches a region-dependent threshold) indicating a major earthquake. Therefore, the seismic energy released during such earthquakes can be approximated from the rate of background seismic activity (slope of the approximately linear portion between points O and A of Fig. 1) and the period of quiescence (distance along the time axis between points A and B).[73] Rate of seismic energy released is therefore a crucial seismicity indicator in regions that show quiescence.

(4) Slope of the log of the earthquake frequency versus magnitude curve (b value)

This parameter is based on the so-called Gutenberg-Richter inverse power law for earthquake magnitude and frequency which is expressed as

$$\log_{10} N = a - bM \tag{5}$$

where N is the number of events of magnitude M or greater, and a and b are constants. The parameter b (known in earthquake prediction literature as the b-value) is the slope of the approximately linear plot between earthquake magnitude and logarithm of frequency of occurrence of events of equal or greater magnitude. As an example, Fig. 2 shows the plot of earthquake magnitude versus the logarithm of the frequency of occurrence of events of equal or greater magnitude for a sample set of data recorded in Southern California for the 10-year period between Jan 1st 1986 and Dec 31st 1995. The figure illustrates the Gutenberg-Richter inverse power law.

The values of a and b can be calculated using the linear least squares regression analysis as follows[23]:

$$b = \frac{(n\Sigma(M_i \log N_i) - \Sigma M_i \Sigma \log N_i)}{((\Sigma M_i)^2 - n\Sigma M_i^2)} \tag{6}$$

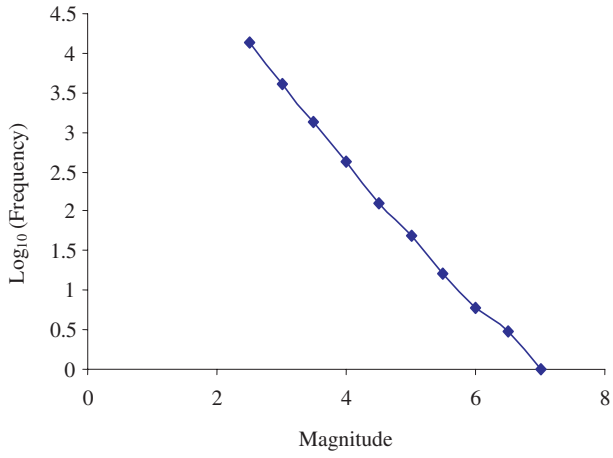

Fig. 2. Earthquake magnitude versus the logarithm of the frequency of occurrence of events of equal or greater magnitude for a sample set of data recorded in southern California for the 10-year period between Jan 1st 1986 and Dec 31st 1995.

$$a = \Sigma(\log_{10} N_i + bM_i)/n \tag{7}$$

where M_i is magnitude of the i^{th} event, N_i is the number of events with magnitude M_i or greater, and n is the total number of seismic events.

(5) Summation of the mean square deviation from the regression line based on the Gutenberg-Richter inverse power law (η value)

This parameter is defined based on the Gutenberg-Richter magnitude-frequency relationship as follows:

$$\eta = \frac{\Sigma(\log_{10} N_i - (a - bM_i))^2}{(n-1)} \tag{8}$$

This is a measure of the conformance of the observed seismic data to the Gutenberg-Richter inverse power-law relationship. The lower the η value, the more likely that the observed distribution can be estimated using the inverse power law whereas a high η value indicates higher randomness and the inappropriateness of the power-law for describing the magnitude-frequency distribution.

(6) Magnitude deficit or the difference between the largest observed magnitude and the largest expected magnitude based on the Gutenberg-Richter relationship (ΔM value)

This is defined as

$$\Delta M = M_{max,observed} - M_{max,expected} \tag{9}$$

where $M_{max,observed}$ is the maximum observed magnitude in the last n events and $M_{max,expected}$ is the maximum magnitude in the last n events based on the inverse power-law relationship. Since an event of the largest magnitude will likely occur only once among the n events, $N = 1$, $\log N = 0$ and Eq. (3) yields

$$M_{max,expected} = a/b \tag{10}$$

(7) Mean time between characteristic or typical events (μ value)

This is the average time or gap observed between characteristic or typical events among the last n events. Several seismic zones including the well-studied Parkfield, California, exhibit periodic trends in the gradual stress built-up and subsequent release through large earthquakes according to the elastic rebound hypothesis.[57] For the Parkfield region, Kagan and Jackson[42] found that the intervening times between large earthquakes are relatively constant. Such large earthquakes are known as characteristic events. In this context magnitudes are defined within a given range of approximation. For example, earthquakes of magnitude 7 to 7.5 are grouped together as one characteristic magnitude. Characteristic events should ideally be separated by approximately equal time periods. The mean time μ is then given by

$$\mu = \Sigma(t_{i\,characterisic})/n_{characteristic} \tag{11}$$

where $t_{i\,characteristic}$ is the observed elapsed time between characteristic events of magnitude M_i and $n_{characteristic}$ is the total number of characteristic events.

(8) Coefficient of variation of the mean time between characteristic events (μ), also known as the aperiodicity of the mean (c value)

This parameter is a measure of the closeness of the magnitude distribution of the seismic region to the characteristic distribution and is defined mathematically as

$$c = \text{standard deviation of the observed times}/\mu \tag{12}$$

A high c value indicates a large difference between the calculated mean time and the observed mean time between characteristic events and vice versa.

3. Neural Network Models for Categorical Earthquake Magnitude Prediction

There have been a number of efforts using neural networks to predict earthquake parameters during the past decade or so. Lakkos *et al.*,[49] used a feed-forward backpropagation (BP) neural network for predicting earthquake magnitude using variations of electrotelluric fluid (also called seismic electric signals), occurring between a few days to a few hours prior to major earthquakes as input. The succeeding earthquake magnitude and epicentral depth are the predicted output. The BP neural network was trained using seismic electric signals recorded in Western Greece. The authors report predictions were within 0.5 Richter in magnitude and 0.3^0 in epicentral location when tested using data not part of the training dataset.

Leach and Dowla[50] use velocity, amplitude, and spectral characteristics of primary or P waves as input in a BP network to predict the arrival time, intensity and duration of secondary S as well as Raleigh and Love waves. When tested using earthquake wave data recorded in Landers, CA, the network yielded correlation coefficients in the range 0.6–0.9. However, the network can only begin prediction after the arrival of P waves thus providing warning time in the order of seconds which is usually too short for life saving measures and initiating emergency management efforts.

Dai and Macbeth[22] use the BP neural network to differentiate between P and S waves arriving at a recording station. Network training is done by converting P and S "noise bursts" into special training data based on their degree of polarization. The authors report 60–75% accuracy in identifying P and S waves when the model was tested using data from two different recording stations. It should be noted that in this work neural network has been used as a classification tool to play the role of an expert seismologist and not for earthquake parameter prediction.

Ma *et al.*[52] use the BP neural network to predict magnitudes of future large earthquakes in Northern and Southwestern China using six seismicity indicators, namely, T, M_{mean}, $dE^{1/2}$, b, η and ΔM as network input. The model yielded an average R score, defined later in this paper, of 0.12 for Northern China and 0.10 for Southwest China. The seismic parameters used are limited to those based on the Gutenberg-Richter inverse power law which is not applicable in all seismic regions.

Negarestani *et al.*[55] use a BP neural network to differentiate environmentally induced changes in soil radon concentration from those caused by earthquakes. The authors propose the model as a tool for earthquake time prediction based on measuring soil radon concentration. The study reports improvement in results using neural networks compared with a linear time-series analysis when used on soil radon readings from a site in Thailand.

Kerh and Chu[46] use seismic data recorded in the Kaohsiung region of Taiwan in a BP neural network to predict peak ground acceleration values. The network inputs are epicentral distance from measuring stations, focal depths, and magnitudes of past earthquakes. The neural network model yielded correlation coefficients in the range 0.6–0.98 compared with 0.6–0.75 obtained from other non-linear regression analysis techniques.

All neural network approaches for earthquake parameter prediction reported in the literature use the simple BP neural network with the exception of one recent conference proceedings paper by Liu *et al.*[51] who use an ensemble of radial basis function (RBF) neural networks for earthquake magnitude prediction using past earthquake magnitude data as network input. A bootstrap aggregating algorithm is used to combine and average predictions from the component neural networks. When tested on 30 past earthquakes recorded in China, the RBF ensemble yielded lower prediction errors than single networks for all 30 earthquakes.

Sharma and Arora[60] use a multi layer feed-forward BP network to model seismicity cycles in the Himalayan region. The authors divide the Himalayas into six seismogenic sub-regions. The recorded earthquake magnitude data for these six sub-regions between 1970 and 1998 are used as a time series in order to recognize patterns using the BP network. The authors report moderate success in predicting the 1999 earthquake of Sinchen, Tibet, with magnitude $M = 5.4$, where the time of occurrence was predicted within 100 days of the actual occurrence of the earthquake.

In the present work, three neural network models are presented and tested for predicting the largest earthquake magnitude during the following month

in a seismic region using eight seismicity indicators described earlier as input simultaneously. They are a feed-forward Levenberg-Marquardt backpropagation (LMBP) neural network, a recurrent neural network, and a radial basis function (RBF) neural network. For all three networks, output is either 1 representing the occurrence of an earthquake of predefined threshold magnitude or greater during the following month, or 0 otherwise. Network operation is repeated by gradually increasing the threshold magnitude in increments of 0.5. The largest earthquake magnitude during the following month is the particular threshold magnitude above which network output changes to 0.

3.1. *Levenberg-Marquardt backpropagation network*

Backpropagation was developed by generalizing the Widrow-Hoff learning rule to multi-layer networks using nonlinear differentiable transfer functions.[13,59] Architecture of the BP neural network model for predicting the occurrence of an earthquake of threshold magnitude or greater during the following month is shown in Fig. 3.

Each training dataset corresponds to one month in the historical record and contains an eight by one input vector and a single output. Each input node represents one of the eight seismicity indicators calculated for each month for a predefined number of seismic events prior to that month using the definitions described in the previous section. Network output is obtained as

$$O_{pi} = \sum_{j=1}^{n} f(\mathbf{s}_i.\mathbf{w}_j) \tag{13}$$

where O_{pi} is the predicted occurrence (1) or non-occurrence (0) of an earthquake of threshold magnitude or greater during the i^{th} month, $\mathbf{s}_i$ is the 8×1 input vector of seismicity indicators for the i^{th} month, $\mathbf{w}_j$ is the weight vector associated with the j^{th} hidden layer, f is the transfer function, and n is the total number of hidden layers. In this research the tan-sigmoid function was found to be the most suitable transfer function. The number of hidden layers and number of nodes in each hidden layer are determined by trial-and-error to obtain the best results.

Backpropagation is essentially a gradient-descent algorithm in which the network weights are moved along the negative of the gradient of a performance or error function. The problem is formulated as an unconstrained optimization problem, that is, to minimize a performance function (z) defined as the mean square difference between the predicted and observed (desired) output.

$$z = \frac{1}{N}\sum_{i=1}^{N}(O_{oi} - O_{pi})^2 \tag{14}$$

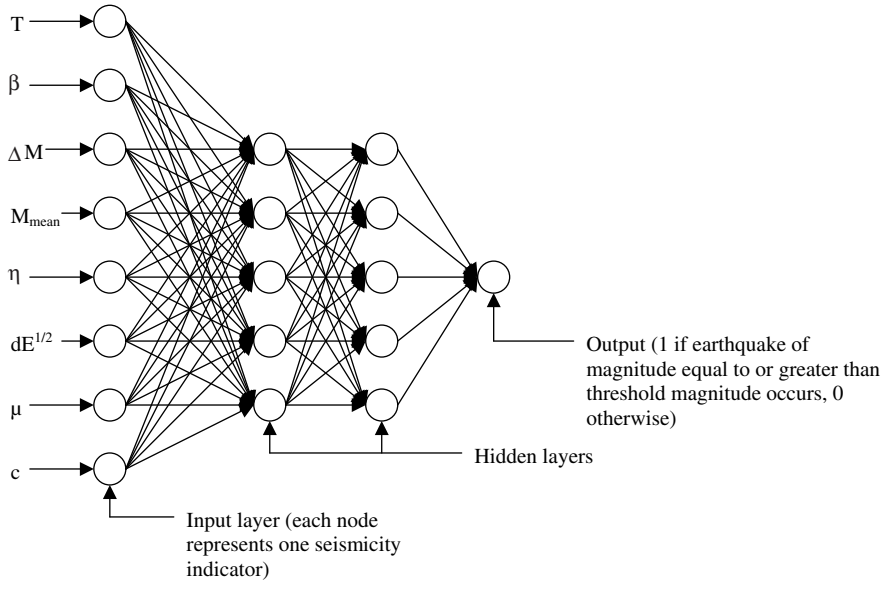

Fig. 3. Architecture of the backpropagation neural network for predicting the occurrence of an earthquake of threshold magnitude or greater during the following month.

where O_{oi} is the observed occurrence (1) or non-occurrence (0) of an event of threshold magnitude or greater during the i^{th} month, and N is the total number of training datasets.

Batch training using gradient descent algorithms is often too slow for large or complicated problems.[2] Faster training can be accomplished using other numerical optimization-based algorithms such as the conjugate gradient,[2] quasi-Newton,[17] and the Levenberg-Marquardt[34] algorithms which attain convergence anywhere between 10 to a 100 times faster than the standard BP algorithm. The Levenberg-Marquardt backpropagation algorithm is used as the learning rule in this work.

There are three basic parameters that determine the accuracy and training time for the Levenberg-Marquardt BP algorithm. They are (a) the number of iterations to be performed before training is stopped or the number of *epochs*, (b) the value of the mean square error function at which training stops, and (c) the minimum magnitude of the gradient below which training stops.

3.2. *Recurrent network*

BP networks are also called static networks because the order in which training datasets are used in the network is immaterial. For instance, seismicity indicators corresponding to each month in the historical record (input) and the observed occurrence or non-occurrence of an event of threshold magnitude or greater during that month (output) can be used in a simple BP network in any random order. In contrast to BP networks, recurrent networks have the capacity to retain past results by incorporating a time-delay.[24] Because of this ability to operate not only on the input space, but also on an internal *state* space, recurrent networks have been used in a number of problems involving time series, e.g., acoustic phonetic decoding of continuous speech[27] and prediction of volatile stock trading trends.[30] Temporal earthquake magnitude recordings also form time-series data and as such, recurrent neural networks are suitable for predicting earthquake trends. To the best of the authors' knowledge, no research has been published on the prediction of earthquake parameters using recurrent neural networks.

Architecture of the recurrent network for predicting the occurrence of an earthquake of threshold magnitude or greater during the following month is shown in Fig. 4. Similar to the BP neural network, the input layer has eight nodes representing the eight seismicity indicators corresponding to a given month in the historical record. The number of hidden layers

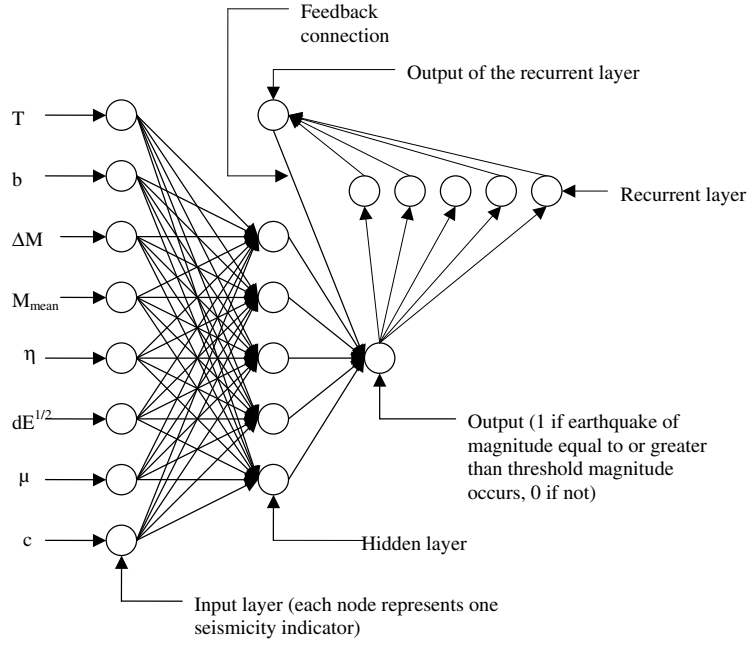

Fig. 4. Architecture of the recurrent neural network developed for predicting the occurrence of an earthquake of threshold magnitude or greater during the following month.

and the number of nodes in each hidden layer are determined by trial-and-error for best results. In a recurrent neural network, in every iteration, the network output is passed through a recurrent layer and the output of the recurrent layer is added to the output of the hidden layer and the sum is used as the argument of the transfer function to obtain the network output in the succeeding iteration as follows:

$$O_{pi} = \sum_{j=1}^{n} f[\mathbf{s}_i.\mathbf{w}_j + O_{p(i-1)}.\mathbf{w}_r] \tag{15}$$

where $\mathbf{w}_r$ is the vector of weights of the links connecting the nodes in the recurrent layer to the output node. Therefore, network output is obtained as a function of not only the input vector, but also the predicted occurrence (1) or non-occurrence (0) of an earthquake of threshold magnitude or greater during the preceding time period (month). The Levenberg-Marquardt BP algorithm is used for training the network and finding two sets of weights, $\mathbf{w}_j$ and $\mathbf{w}_r$, which minimize the mean square error function.

3.3. *Radial basis function (RBF) network*

In an RBF neural network, the output is computed as a function of the Euclidian distance between the input vector and a prototype vector.[3,5,13,43] The prototype vector is conveniently defined as the vector of weights of links connecting the input layer to the hidden layer (or one hidden layer to the next hidden layer). Usually, RBF neural networks require more nodes than BP networks, but provide significantly improved results especially when a large number of training datasets is available. Also, an RBF network can usually be trained in a fraction of the time required to train a BP network.[3]

Architecture of the RBF neural network for predicting the occurrence of an earthquake of threshold magnitude or greater during the following month is shown in Fig. 5. The network has an eight-node input layer and a single output. In most applications of RBF neural networks, one hidden layer is found to be sufficient. In this research, however, we also investigate the improvement in prediction accuracy using more than one hidden layer. The Gaussian function, which is the most commonly used radial basis transfer function, is used to compute network output as

$$O_{pi} = \Phi(\mathbf{s}_i - \mathbf{w}_j) = \sum_{j=1}^{n} e^{\left\| \frac{s_i - w_j}{\sigma_j} \right\|^2} \tag{16}$$

where σ_j is known as a width factor associated with the j^{th} hidden layer. Theoretically, w_j and σ_j determine the center and spread of the Gaussian bell curve, respectively. In this research, it was observed through numerical experimentation that the width factor σ does not contribute to the network's ability for function approximation and therefore its value is set equal to one. The Levenberg-Marquardt training algorithm is used to minimize the mean-square error function.

4. Prediction Verification and Hypothesis Testing

Earthquake parameter prediction is a complicated problem involving a large number of unknown variables. No prediction model is suitable for all seismic regions. Emergency management organizations such as the U.S. Federal Emergency Management Agency in the U.S. base their public warning and evacuation procedures on the probability associated with a prediction. In this research, prediction accuracies obtained from the three different neural network models are compared using four different statistical measures: the probability of detection (POD), false alarm ratio (FAR), frequency bias (FB), and R scores.[66] Prediction accuracies are also compared with the probability of detection for each magnitude obtained using the Poisson's null hypothesis. Testing is limited to the prediction of earthquakes of Richter magnitudes 4.5 or greater since seismic events of lower magnitudes do not have much structural engineering or emergency management significance.

The POD (or *hit* rate), FAR and FB are calculated for categorical predictions (such as prediction of the occurrence of earthquake of a threshold magnitude or greater) using the following equations

$$\text{POD} = \frac{N_{pc}}{N_{pc} + N_{ni}} \tag{17}$$

$$\text{FAR} = \frac{N_{pi}}{N_{pc} + N_{pi}} \tag{18}$$

$$\text{FB} = \frac{N_{pc} + N_{pi}}{N_{pc} + N_{ni}} \tag{19}$$

where N_{pc} (predicted-correct) is the number of months during which an earthquake of threshold

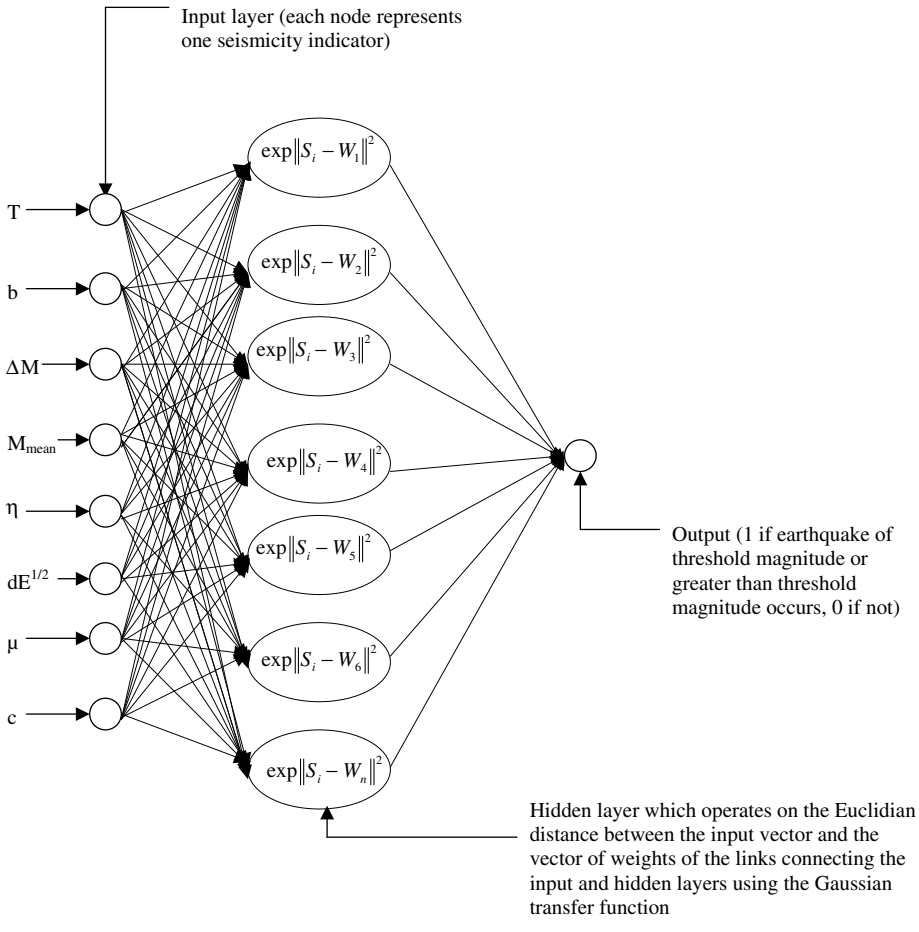

Fig. 5. Architecture of the radial basis function neural network for predicting the occurrence of an earthquake of threshold magnitude or greater during the following month.

magnitude or greater occurred and was predicted; N_{pi} (predicted-incorrect) is the number of months during which an earthquake of threshold magnitude or greater did not occur but was predicted, N_{nc} (not predicted-correct) is the number of months during which an earthquake of threshold magnitude or greater did not occur and was not predicted, and N_{ni} (not predicted-incorrect) is the number of months during which an earthquake of threshold magnitude or greater occurred but was not predicted. These variables are summarized in Table 4.

Another commonly used forecast verification method is computing the so-called skill scores. Skill scores are a measure of the skill (or competence) of the prediction tool (neural network modeling in this case) in predicting a particular parameter (earthquake magnitude in this case).[56] In this research, the Hanssen-Kuiper skill score, also known as the real skill or R score, is used. It is defined as the difference between the probability of detection and false alarm ratio for each predicted magnitude:

$$R = POD - FAR = \frac{N_{pc}}{N_{pc} + N_{pi}} - \frac{N_{pi}}{N_{pi} + N_{nc}} \qquad (20)$$

The R score is -1 if no correct predictions are made and $+1$ if all predictions are correct. This score is considered advantageous over POD and FAR

Table 4. Parameters used for the computation of POD, FAR, FB and R scores. "Yes" denotes predicted or observed occurrence of an earthquake of threshold magnitude or greater and "No" denotes the predicted or observed non-occurrence of an earthquake of threshold magnitude or greater.

Observed \ Predicted	Yes	No
Yes	N_{pc}	N_{pi}
No	N_{ni}	N_{nc}

because it includes an equal representation of both correct and incorrect predictions.

Further, the models are evaluated using a hypothesis verification technique: the probability of detection calculated for different predicted magnitudes using the three neural networks are compared with the probability of occurrence for those magnitudes based on the Poisson's null hypothesis. Assuming that earthquakes occur at completely random intervals, the probability that an earthquake of a certain magnitude occurring during a given month based on the Poisson's null-hypothesis (p_0) is given by[38]

$$p_0 = 1 - e^{-rt} \tag{21}$$

where r is the frequency of occurrence of earthquakes of the particular magnitude in the historical record for a time period of t (for example $t = 1$ month). A prediction model for an earthquake of a given magnitude is superior to the Poisson's null hypothesis only if the POD for that magnitude is greater than p_0.

5. Parametric Analysis of Earthquake Prediction

For each neural network model, the effects of any input parameter on the prediction accuracies are investigated by removing that parameter from the input vector. In that case the input layer will have seven nodes. A comparison of results obtained from the neural network with a 7-node input layer with those of the network with an 8-node input layer will provide the significance of the removed parameter. Significant parameters provide a larger POD and R score and lower FB and FAR for predicted magnitudes. Parametric analysis in this research is divided into three main categories as follows.

5.1. *Parametric analysis for determining the most suitable frequency-magnitude relationship*

Parametric analysis of seismicity indicators will help to determine what magnitude-frequency relationship is the most suitable for the seismic region under study. If T, η or ΔM are significant parameters, the Gutenberg-Richter inverse power law is probably the most suitable magnitude-frequency relationship for the region. On the other hand, if μ and c are significant parameters, characteristic earthquake distribution is probably the most suitable for the region.

5.2. *Parametric analysis for determining seismic quiescence patterns*

If higher POD and R scores and lower FAR and FB are calculated when $dE^{1/2}$ is part of the input vector compared to when it is not, it can be concluded that the seismic region exhibits seismic quiescence characteristics.

5.3. *Parametric analysis for determining the optimum threshold magnitude and the size of dataset for calculating seismicity indicators*

Threshold magnitude and the number of seismic events of threshold magnitude or greater used for computing seismicity indicators corresponding to each month in the historical record are varied to determine their statistical significance. For example, what is the relative statistical significance of 100 events of magnitude 4.0 or more compared with 200 events of magnitude 3.5 or more in computing seismicity indicators corresponding to each month?

6. Example Applications

The Southern California Earthquake-Data Center (SCEC), a joint venture between the United States Geological Survey (USGS) and the California Institute of Technology, maintains a list of earthquake recordings (time of occurrence, geographic location of the epicenter, depth and magnitude) for the region. Their catalog is available over the internet at www.scec.org, and is searchable for different magnitude, time, location and depth ranges. Table 5 shows sample readings from the SCEC catalog with search parameters, parameter values, and events that matched the search criteria.

Data recorded for two seismic regions obtained from the SCEC website is used to train and test the neural network models developed. The first region is southern California described by the geographic coordinates 32 and 36 N latitude and 114 and 120 W longitude.[69] The second region is the San-Francisco bay area which has been observed historically to have different seismicity patterns from southern California.[39] In this research, the San Francisco bay region is defined by the geographic coordinates 37.5 and 40 N latitude and 116 and 123.5 W

Table 5. Sample readings from the SCEC catalog showing search parameters, parameter values and events that matched the search criteria.

Search parameter	Parameter range	Event 1	Event 2	Event 3
Magnitude	2.75–3.50	3.20	2.82	2.97
Date, Time	05/15/2005, 00:00:00–23:59:59	13:29:56	16:14:08	18:34:46
Latitude	32^0–36.5^0	33.45	34.36	32.35
Longitude	-114.75^0–121^0	−116.63	−116.91	−115.32
Epicentral Depth	0 m–900 m	196 m	215 m	36 m

longitude (this region includes most of north-central California). The three neural network models are trained and tested separately using historical seismic data recorded at two different regions: southern California and the San Francisco Bay region as defined in a previous section. For training the networks eight seismicity indicators corresponding to each month between January 1950 and December 1990 (four hundred ninety two 8×1 input vectors, corresponding to 492 months) for each region are calculated and stored in respective two-dimensional arrays where each column represents one set of seismicity indicators. The indicators are calculated from a predetermined number, say 100, events of certain threshold magnitude, say 4.5, or greater prior to each month. The desired network output which is the observed occurrence (1) or non-occurrence (0) of an earthquake of threshold magnitude or greater during each of the 492 months for each one of the two regions is stored in respective one-dimensional arrays. For each region, these two arrays form 492 training datasets. Sample training datasets for the 12 months between January and December of 1994 for southern California showing eight-element input vectors and the corresponding desired outputs (M = 4.5 or greater) are presented in Table 6.

Convergence criteria for training each network include limiting the value of the mean square error function to 0.001 and the maximum number of iterations to 1000. Network operation is repeated by increasing the threshold magnitude by increments of 0.5 to obtain the largest predicted magnitude.

6.1. *Best network architecture*

For each seismic region, all three neural networks are initially modeled with one hidden layer and one recurrent layer in the case of the recurrent neural network with four nodes in each layer using the tan-sigmoid transfer function. Once trained, the networks with this base architecture are used to

Table 6. Sample training datasets for the 12 months between January and December of 1994 for southern California showing eight-element input vectors and the corresponding desired outputs (M = 4.5 or greater).

Month	Input vector								Desired output
	T (Days)	b	η	ΔM	M_{mean}	$dE^{1/2}$ ($\times 10^{19}$ergs)	μ (Days)	c	
Jan 94	1875	0.89	0.34	1.48	4.41	0.023	28	0.50	1
Feb 94	1688	0.91	0.28	2.50	4.99	0.015	35	0.37	0
Mar 94	1753	0.80	0.15	0.44	5.10	0.009	17	0.55	1
Apr 94	1225	0.72	0.56	0.58	5.15	0.067	38	0.78	1
May 94	1311	0.77	0.48	0.63	4.37	0.004	55	0.25	0
Jun 94	988	0.92	0.18	2.48	4.10	0.101	64	0.12	1
Jul 94	1028	0.83	0.67	1.14	4.01	0.001	58	0.87	0
Aug 94	1191	0.98	0.59	0.35	4.85	0.055	29	0.66	1
Sep 94	1226	0.99	0.46	0.88	4.27	0.126	37	0.54	0
Oct 94	1365	0.92	0.80	1.60	4.09	0.005	44	0.68	1
Nov 94	1028	0.88	0.15	1.71	4.66	0.098	19	0.81	0
Dec 94	988	0.78	0.29	1.01	4.58	0.085	34	0.78	1

predict the occurrence of an earthquake of magnitude 4.5 or greater for the nine months between January and September of 2005 using the 8×1 input vectors of seismicity indicators for these months as input. The R scores for the nine predictions are computed for each of the three neural network models. The best network architectures are obtained by gradually increasing the number of hidden layers and the number of nodes in each hidden layer until the increase in R scores by doing so is less than 0.01. The best architectures for the BP and recurrent neural network models were found to be the same for both seismic regions whereas different best architectures were found in the case of the RBF neural network model for the two regions.

Since the chronological order of input has no significance in BP and RBF networks, network weights are updated after all 492 training datasets are applied. The best architecture for the BP neural network has two hidden layers, each with eight nodes. The best RBF network architecture has one hidden layer with eight nodes for the Southern California region and two hidden layers, each with eight nodes, for the San Francisco Bay region. Chronological order of input data is maintained in the recurrent neural network, and network weights are adjusted after each training dataset is applied. The best recurrent neural network architecture has one hidden layer with eight nodes and one recurrent layer with four nodes.

6.2. *Prediction results*

The neural network models were tested by comparing the network output with observed earthquake magnitudes for the 177 months between January 1991 and September 2005 (testing period). The accuracy of each network in predicting large, moderate, and small earthquakes is studied using the statistical measures described in an earlier section.

Southern California

The computed values of POD, FAR, FB, and R score for different magnitudes using each of the three neural network models are summarized in Table 7. This table also includes the values of the probabilities of occurrence of earthquakes of different magnitudes in a given month based on the Poisson's null hypothesis (p_0).

(a) Prediction of large earthquakes (Magnitude 6.5 or greater)

There were two months in the testing period during which an earthquake of magnitude greater than 7.0 occurred (June 1992 and October 1999). The BP neural network did not predict either of these earthquakes and did not make any false prediction of an earthquake of magnitude 7.0 or greater yielding a POD and FAR of 0.0. The recurrent neural network correctly predicted both of these earthquakes yielding a POD of 1.0 and did not falsely predict any earthquake of magnitude 7.0 or greater yielding an FAR of 0.00. The RBF neural network correctly predicted the event of June 1992, but did not predict the event of October 1999 yielding a POD of 0.5. The network also falsely predicted an earthquake of magnitude 7.0 or greater during one month (March 2000), yielding an FAR of 0.5.

There was a single month in the testing period during which an earthquake of magnitude between 6.5 and 7.0 occurred (January 1994). The BP neural network failed to predict this event yielding a POD of 0.0. Both the recurrent neural network and the RBF neural network correctly predicted this event yielding a POD of 1.0. None of the networks falsely

Table 7. Computed values of p_0,POD, FAR, FB, and R scores for different magnitudes using the three neural network models for southern California.

Magnitude	p_0	Backpropagation network				Recurrent network				Radial-basis network			
		POD	FAR	FB	R Score	POD	FAR	FB	R Score	POD	FAR	FB	R Score
4.5	0.301	0.52	0.44	0.86	0.08	0.67	0.31	0.90	0.36	0.44	0.45	0.90	−0.01
5.0	0.131	0.46	0.33	0.68	0.13	0.80	0.29	1.00	0.51	0.64	0.27	0.79	0.37
5.5	0.052	0.50	0.50	1.00	0.00	0.75	0.25	1.00	0.50	0.75	0.40	0.86	0.35
6.0	0.021	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
6.5	0.011	0.00	0.00	0.00	0.00	1.00	0.00	1.00	1.00	1.00	0.00	1.00	1.00
7.0	0.003	0.00	0.00	0.00	0.00	1.00	0.00	1.00	1.00	0.50	0.50	1.00	0.00

predicted the occurrence of an earthquake of magnitude between 6.5 and 7.0 yielding an FAR of 0.0.

(b) Prediction of moderate earthquakes (Magnitude 5.5 or greater but less than 6.5)
There was a single month in the testing period during which an earthquake of magnitude between 6.0 and 6.5 occurred (April and June 1992). The BP neural network did not predict this earthquake and did not make any false prediction of an earthquake of magnitude between 6.0 and 6.5 yielding a POD and FAR of 0.00. The recurrent network did not predict this event yielding a POD 0.0 and did not falsely predict any earthquake of magnitude between 6.0 and 6.5 yielding an FAR of 0.00. The RBF neural network did not predict the event yielding a POD of 0.00 and did not falsely predict any event of magnitude between 5.5 and 6.5 yielding an FAR of 0.00.

There were four months in the testing period during which an earthquake of magnitude between 5.5 and 6.0 occurred. The BP network correctly predicted two of these events yielding a POD of 0.50 and falsely predicted earthquakes of magnitude between 5.5 and 6.0 during two months yielding an FAR of 0.50. The recurrent network correctly predicted three of these earthquakes yielding a POD of 0.75 and falsely predicted earthquakes of magnitude between 5.5 and 6.0 during one month only yielding an FAR of 0.25. The RBF neural network correctly predicted three of the four events yielding a POD of 0.75 and incorrectly predicted earthquakes of magnitude between 5.5 and 6.0 during two months yielding an FAR of 0.40.

(c) Prediction of small earthquakes (Magnitude 4.5 or greater but less than 5.5)
There were 25 months in the testing period during which an earthquake of magnitude between 5.0 and 5.5 occurred. The BP neural network correctly predicted 12 of these events yielding a POD of 0.46 and falsely predicted earthquakes of magnitude between 5.0 and 5.5 during six months yielding an FAR of 0.33. The recurrent network correctly predicted 18 of the 25 events yielding a POD of 0.72 and falsely predicted earthquakes of magnitude between 5.0 and 5.5 during eight months yielding an FAR of 0.30. The RBF neural network correctly predicted 16 of the 25 events yielding a POD of 0.64 and falsely predicted earthquakes of magnitude between 5.0 and 5.5 during six months yielding an FAR of 0.27.

There were 27 months in the testing period during which an earthquake of magnitude between 4.5 and 5.0 occurred. The BP neural network correctly predicted 14 of these events yielding a POD of 0.52 and falsely predicted earthquakes of magnitude between 4.5 and 5.0 during 11 months yielding an FAR of 0.44. The recurrent neural network correctly predicted 18 of the 27 events yielding a POD of 0.67 and falsely predicted earthquakes of magnitude between 4.5 and 5.0 for eight months yielding an FAR of 0.31. The RBF neural network correctly predicted 12 of the 27 events yielding POD of 0.44 and falsely predicted earthquakes of magnitude between 4.5 and 5.0 for nine months yielding an FAR of 0.45.

(d) Parametric Analysis
Parametric studies are performed by removing the seismicity indicators one at a time. The computed values of R score for prediction of earthquakes of magnitude 5.0, 6.0 and 7.0 using the three neural network models are presented in Table 8. It is observed that for the southern California region, seismicity indicators obtained from the Gutenberg-Richter earthquake magnitude-frequency relationship (b, η, and ΔM) have very little impact on the computed R score and therefore the prediction accuracy whereas indicators based on the characteristic earthquake distribution (μ and c) have a significant impact on the prediction accuracy. It is therefore concluded that the characteristic earthquake distribution model better suits seismic patterns in southern California. The indicator corresponding to seismic energy ($dE^{1/2}$) has significantly less effect on the R score of the predicted magnitudes for the southern California region than the San Francisco bay region (to be discussed in the next section) therefore suggesting that the San Francisco region shows significantly more seismic quiescence patterns.

To determine the optimum value of the threshold magnitude and the number of events used to calculate seismicity indicators, the three neural networks are trained and tested and the results are compared for three cases: 200 events of magnitude 3.5 or more, 100 events of magnitude 4.5 or more, and 50 events of magnitude 5.0 or more prior to each month. No substantial improvement in prediction results was observed as a result of considering smaller magnitude

Table 8. Computed values of R scores for different predicted magnitudes, using different input vectors obtained by removing one parameter at a time for the southern California region.

Predicted magnitude	Parameter removed from input vector	BP	Recurrent	RBF
5.0	None	0.13	0.51	0.37
	b	0.13	0.44	0.33
	η	0.11	0.42	0.31
	ΔM	0.14	0.33	0.28
	μ	0.01	0.27	0.12
	c	0.04	0.20	0.22
	$dE^{1/2}$	0.11	0.39	0.32
6.0	None	0.00	0.00	0.00
	b	0.00	0.00	0.00
	η	0.00	0.00	0.00
	ΔM	0.00	0.00	0.00
	μ	−1.00	−1.00	−1.00
	c	0.00	0.00	−1.00
	$dE^{1/2}$	−1.00	0.00	0.00
7.0	None	0.00	1.00	0.00
	b	0.00	1.00	0.00
	η	0.00	1.00	0.00
	ΔM	0.00	1.00	0.00
	μ	−1.00	0.00	0.00
	c	−1.00	0.00	−1.00
	$dE^{1/2}$	0.00	0.00	0.00

Table 9. Computed values of R scores for different predicted magnitudes, using input vectors (seismicity indicators) calculated form different threshold magnitudes and number of events of threshold magnitude or greater for the southern California region.

Predicted magnitude	Threshold magnitude	No. of events	BP	Recurrent	RBF
5.0	3.5	200	0.16	0.50	0.35
	4.5	100	0.13	0.51	0.37
	5.5	50	0.12	0.48	0.29
6.0	3.5	200	0.00	0.00	0.00
	4.5	100	0.00	0.00	0.00
	5.5	50	0.00	0.00	0.00
7.0	3.5	200	0.00	1.00	0.00
	4.5	100	0.00	1.00	0.00
	5.5	50	0.00	1.00	0.00

events in computing seismicity indicators as summarized in Table 9.

San Francisco bay region

(a) Prediction of large earthquakes (Magnitude 6.5 or greater)

There was no month in the testing period during which an earthquake of magnitude greater than 6.5 occurred. Therefore POD for all three networks for M = 6.5 and M = 7.0 is 0.0. None of the networks falsely predicted earthquakes of magnitude greater than 6.5 yielding an FAR of 0.0 for M = 6.5 and M = 7.0.

(b) Prediction of moderate earthquakes (Magnitude 5.5 or greater but less than 6.5)

There was no month in the testing period during which an earthquake of magnitude between 6.0 and 6.5 occurred. Therefore the POD for all three networks is 0.0 for M = 6.0. The BP neural network falsely predicted seven events of magnitude between 6.0 and 6.5 yielding an FAR of 1.00. The recurrent network did not falsely predict any event of magnitude between 6.0 and 6.5 yielding an FAR of 0.00. The RBF neural network falsely predicted 3 events of magnitude between 6.0 and 6.5 yielding an FAR of 1.00 for M = 6.0.

There was no month in the testing period during which an earthquake of magnitude between 5.5 and 6.0 occurred. Therefore the POD for all three networks is 0.00 for M = 5.5. The BP neural network falsely predicted 18 events of magnitude between 5.5 and 6.0 yielding an FAR of 1.00. The recurrent network did not falsely predict any earthquake of magnitude between 5.5 and 6.0 yielding an FAR of 0.00. The RBF network falsely predicted 10 earthquakes of magnitude between 5.5 and 6.0 yielding an FAR of 1.00 for M = 5.5.

(c) Prediction of small earthquakes (Magnitude 4.5 or greater but less than 5.5)

There were four months in the testing period during which an earthquake of magnitude between 5.0 and 5.5 occurred. The BP neural network did not predict any of these events yielding a POD of 0.00 and falsely predicted an earthquake of magnitude between 5.0 and 5.5 for three months yielding an FAR of 1.00 for M = 5.0. The recurrent neural network correctly predicted two of the four events yielding a POD of 0.50 and falsely predicted an earthquake of magnitude between 5.0 and 5.5 during one month yielding an FAR of 0.33 for M = 5.0. The RBF neural network predicted one of the four events yielding a POD of 0.25 and falsely predicted an earthquake of magnitude between 5.0 and 5.5 for one month yielding an FAR of 0.50 for M = 5.0.

There were eight months during which an earthquake of magnitude between 4.5 and 5.0 were recorded. The BP neural network predicted two of these events yielding a POD of 0.25 and falsely predicted earthquakes of magnitude between 4.5 and 5.0 during two months yielding an FAR of 0.50 for M = 4.5. The recurrent neural network predicted five of the eight events yielding a POD of 0.63 and falsely predicted earthquakes of magnitude between 4.5 and 5.0 for four months yielding an FAR of 0.44 for M = 4.5. The RBF neural network correctly predicted four of the eight events yielding a POD of 0.50 and falsely predicted two earthquakes of magnitude between 4.5 and 5.0 yielding an FAR of 0.33 for M = 4.5.

(d) Parametric Analysis

Similar to the Southern California region, parametric studies are performed by removing the seismicity indicators one at a time and in each case R scores are computed for different magnitudes. It is observed that prediction accuracies are influenced by seismicity indicators based on the Gutenberg-Richter earthquake magnitude-frequency distribution (b, η, and ΔM) and not significantly altered by seismicity indicators based on the characteristic earthquake distribution model (μ and c) leading to the conclusion that the Gutenberg-Richter earthquake magnitude-frequency relationship is better suited to the San Francisco bay region compared with the characteristic earthquake distribution. The R scores are significantly reduced if $dE^{1/2}$ is removed from the input vector leading to the conclusion that San Francisco bay region shows strong seismic quiescence trends.

As in the case of the southern California region, it is observed that no major improvement in prediction accuracies is achieved as a result of considering large numbers of low magnitude events for computing seismicity indicators.

7. Final Comments

Earthquake parameter prediction is a highly complex problem because it involves a large number of variables whose effects are not completely understood. Neural networks have not been used extensively for earthquake parameter prediction even though they have been used as an effective prediction tool in a variety of fields from economics to signal processing. In this paper, three neural network models are presented for earthquake magnitude prediction using eight mathematically computed seismic parameters as input. The goal is to predict the magnitude of the largest earthquake in a following period, for example, one month.

Overall the recurrent neural network model yields the best prediction accuracies compared with the

BP and RBF networks. This result is consistent with the fact that recurrent neural networks have the inherent capacity to model time-series data better compared with other networks. The study that is most closely related to the present work was undertaken by Ma *et al.*,[52] where six seismicity indicators were used as input to a BP neural network to predict the magnitude of the largest earthquake in the following month. However, the indicators were limited to those based on the Gutenberg-Richter magnitude-frequency relationship alone. Therefore, their model is not applicable in regions where seismicity data do not follow this relationship. When tested on the two seismic regions, all three neural network models developed in this research yielded larger and therefore superior R scores (Table 7) than those reported by Ma *et al.* (0.10 for Northern China and 0.12 for Southwestern China).

The following results from the parametric analysis are in agreement with general consensus in the field of seismology:

1. The San-Francisco bay region shows seismic quiescence characteristics.
2. The characteristic earthquake distribution is suited to model seismic data in Southern California
3. The occurrence of large earthquakes has a significant influence on future seismic activity in a region whereas the occurrence (even in large numbers) of small earthquakes usually does influence future seismic activity.

It appears that the networks make false predictions for the months for which there is significant high seismic activity immediately before or after the month. Also, since only one event is predicted per month, the networks tend to miss significant aftershocks and pre-shocks as these usually occur within 30 days of the main-shock. It is expected that the results can be improved by computing seismicity indicators for time-periods shorter than one month and for smaller regions.

The observed and predicted largest events using the recurrent neural network are compared for each month in the testing dataset for Southern California in Table 10. This table also presents the observed latitude and longitude of the largest earthquake during each month.

Table 10. Observed largest earthquake, and corresponding location compared with predicted magnitude for each month in the testing dataset for Southern California using the recurrent neural network. Blank cells indicate lack of observed or predicted events of magnitude 4.5 or greater during the corresponding month.

Month	Observed M_{max}	Observed location		Predicted M_{max}
		Latitude	Longitude	
Jan-91	—	—	—	—
Feb-91	—	—	—	—
Mar-91	—	—	—	—
Apr-91	—	—	—	—
May-91	—	—	—	—
Jun-91	5.8	34.27	−117.99	5.5
Jul-91	—	—	—	—
Aug-91	—	—	—	—
Sep-91	5.01	35.81	−119.42	—
Oct-91	—	—	—	—
Nov-91	—	—	—	—
Dec-91	4.68	35.20	−116.68	4.5
Jan-92	—	—	—	—
Feb-92	—	—	—	—
Mar-92	—	—	—	—
Apr-92	6.1	33.96	−116.32	5.5
May-92	4.98	33.94	−116.3	4.5
Jun-92	7.3	34.2	−116.44	7
Jul-92	5.67	35.21	−118.07	5.5
Aug-92	5.23	34.2	−116.86	5.0
Sep-92	5.26	34.06	−116.36	5.0
Oct-92	4.59	34.6	−116.64	—
Nov-92	5.29	34.34	−116.9	5.0
Dec-92	5.26	34.37	−116.9	5.0
Jan-93	—			—
Feb-93	4.5	35.03	−116.97	—
Mar-93	—	—	—	—
Apr-93	0			—
May-93	5.19	35.15	−119.1	5.0
Jun-93	—	—	—	—
Jul-93	—	—	—	—
Aug-93	5	34.03	−116.32	—
Sep-93	—	—	—	—
Oct-93	—	—	—	—
Nov-93	4.64	35.97	−120.00	4.5
Dec-93	—	—	—	—
Jan-94	6.7	34.21	−118.54	6.5
Feb-94	—	—	—	—
Mar-94	5.24	34.23	−118.48	5.0
Apr-94	4.78	34.19	−117.1	—
May-94	—	—	—	—
Jun-94	4.97	34.27	−116.4	4.5
Jul-94	—	—	—	—
Aug-94	4.85	34.64	−116.52	4.5
Sep-94	—	—	—	—
Oct-94	4.50	34.28	−116.42	—
Nov-94	—	—	—	—

Table 10. (*Continued*)

Month	Observed M_{max}	Observed location		Predicted M_{max}
		Latitude	Longitude	
Dec-94	4.88	35.91	−120.00	4.5
Jan-95	—	—	—	—
Feb-95	—	—	—	—
Mar-95	—	—	—	—
Apr-95	—	—	—	—
May-95	4.77	33.91	−116.29	4.5
Jun-95	5.02	34.39	−118.67	—
Jul-95	—	—	—	—
Aug-95	5.36	35.78	−117.66	5.0
Sep-95	5.75	35.76	−117.64	5.5
Oct-95	—	—	—	—
Nov-95	—	—	—	—
Dec-95	—	—	—	—
Jan-96	5.17	35.76	−117.65	5.0
Feb-96	—	—	—	—
Mar-96	—	—	—	—
Apr-96	—	—	—	—
May-96	—	—	—	—
Jun-96	—	—	—	—
Jul-96	—	—	—	—
Aug-96	—	—	—	—
Sep-96	—	—	—	—
Oct-96	—	—	—	—
Nov-96	5.3	36	−117.65	5.0
Dec-96	—	—	—	—
Jan-97	—	—	—	—
Feb-97	—	—	—	—
Mar-97	5.26	34.97	−116.82	5.0
Apr-97	5.07	34.37	−118.67	—
May-97	—	—	—	—
Jun-97	4.76	32.63	−118.11	4.5
Jul-97	4.89	33.4	−116.35	4.5
Aug-97	—	—	—	—
Sep-97	—	—	—	5.0
Oct-97	—	—	—	5.0
Nov-97	—	—	—	—
Dec-97	—	—	—	—
Jan-98	—	—	—	—
Feb-98	—	—	—	—
Mar-98	5.23	36	−117.64	5.0
Apr-98	—	—	—	—
May-98	—	—	—	—
Jun-98	—	—	—	—
Jul-98	4.75	35.95	−117.53	4.5
Aug-98	4.78	34.12	−116.93	—
Sep-98	—	—	—	—
Oct-98	4.82	34.32	−116.84	4.5
Nov-98	—	—	—	—
Dec-98	—	—	—	—
Jan-99	—	—	—	5.0
Feb-99	—	—	—	—
Mar-99	—	—	—	—

Table 10. (*Continued*)

Month	Observed M_{max}	Observed location		Predicted M_{max}
		Latitude	Longitude	
Apr-99	—	—	—	—
May-99	4.93	34.06	−116.37	4.5
Jun-99	4.92	32.38	−115.24	4.5
Jul-99	—	—	—	—
Aug-99	—	—	—	—
Sep-99	4.8	32.27	−115.23	4.5
Oct-99	7.1	34.59	−116.27	7
Nov-99	—	—	—	—
Dec-99	—	—	—	—
Jan-00	—	—	—	4.5
Feb-00	—	—	—	—
Mar-00	—	—	—	—
Apr-00	—	—	—	5.0
May-00	—	—	—	—
Jun-00	4.51	34.78	−116.3	—
Jul-00	—	—	—	—
Aug-00	—	—	—	—
Sep-00	—	—	—	4.5
Oct-00	—	—	—	—
Nov-00	—	—	—	—
Dec-00	—	—	—	—
Jan-01	—	—	—	—
Feb-01	5.13	34.29	−116.95	5.0
Mar-01	—	—	—	—
Apr-01	—	—	—	5.0
May-01	—	—	—	—
Jun-01	—	—	—	—
Jul-01	5.1	36.02	−117.87	5.0
Aug-01	—	—	—	—
Sep-01	—	—	—	—
Oct-01	5.09	33.51	−116.51	4.5
Nov-01	—	—	—	—
Dec-01	4.89	32	−115.01	—
Jan-02	—	—	—	—
Feb-02	5.7	32.32	−115.32	—
Mar-02	4.6	33.67	−119.33	4.5
Apr-02	—	—	—	—
May-02	—	—	—	5.0
Jun-02	4.87	36.69	−116.34	—
Jul-02	—	—	—	—
Aug-02	—	—	—	—
Sep-02	4.75	33.92	−117.78	4.5
Oct-02	4.77	34.8	−116.27	—
Nov-02	—	—	—	—
Dec-02	4.84	32.23	−115.8	4.5
Jan-03	4.54	35.32	−118.65	—
Feb-03	5.37	34.31	−116.85	5.0
Mar-03	4.64	34.36	−116.13	—
Apr-03	—	—	—	—
May-03	—	—	—	—
Jun-03	—	—	—	5.0
Jul-03	—	—	—	—

Table 10. (*Continued*)

Month	Observed M_{max}	Observed location		Predicted M_{max}
		Latitude	Longitude	
Aug-03	—	—	—	—
Sep-03	—	—	—	—
Oct-03	—	—	—	4.5
Nov-03	—	—	—	4.5
Dec-03	—	—	—	—
Jan-04	—	—	—	—
Feb-04	—	—	—	—
Mar-04	—	—	—	5.0
Apr-04	—	—	—	—
May-04	—	—	—	—
Jun-04	5.27	32.33	−117.92	5.5
Jul-04	—	—	—	—
Aug-04	—	—	—	—
Sep-04	5.04	35.39	−118.62	—
Oct-04	—	—	—	—
Nov-04	—	—	—	—
Dec-04	—	—	—	4.5
Jan-05	—	—	—	4.5
Feb-05	—	—	—	—
Mar-05	—	—	—	—
Apr-05	5.15	35.03	−119.18	5.0
May-05	—	—	—	—
Jun-05	5.2	33.53	−116.57	5.0
Jul-05	—	—	—	—
Aug-05	4.59	33.17	−115.64	4.5
Sep-05	5.11	33.16	−115.64	5.0

References

1. H. Adeli, Neural networks in civil engineering: 1989–2000, *Computer Aided Civil and Infrastructure Engineering* **16**(2) (2001) 126–142.
2. H. Adeli and S. L. Hung, *Machine Learning-Neural Networks, Genetic Algorithms and Fuzzy System* (Wiley, New York, NY, 1995).
3. H. Adeli and A. Karim, Fuzzy-wavelet RBFNN model for freeway incident detection, *Journal of Transportation Engineering*, ASCE **126**(6) (2000) 464–471.
4. H. Adeli and A. Karim, *Construction Scheduling, Cost Optimization, and Management — A New Model Based on Neurocomputing and Object Technologies* (Spon Press, London, UK, 2001).
5. H. Adeli and Karim, A. *Wavelets in Intelligent Transportation Systems* (John Wiley and Sons, New York, NY, 2005).
6. H. Adeli and J. Mohammadi, Seismic risk analysis based on Weibull distribution, in *Proceedings of the Eight World Conference on Earthquake Engineering*, Vol. 1, San Francisco, July 21–28 (1984), pp. 191–198.
7. H. Adeli and H. S. Park, *Neurocomputing for Design Automation* (CRC Press, Boca Raton, FL, 1998).
8. H. Adeli and G. F. Sirca, Neural network model for uplift capacity of metal roof panels, *Journal of Structural Engineering* **127**(11) (2001) 1276–1285.
9. K. Aki, Introduction to seismology for earthquake prediction, in *Proceedings of the Seventh Workshop on Non-Linear Dynamics and Earthquake Prediction*, September 29–October 11, Trieste, Italy (2003).
10. M. Anghel and Y. Ben-Zion, Non-linear system identification and forecasting of earthquake fault dynamics using artificial neural networks, in *Proceedings of the Fall Meeting of the American Geophysical Union*, December 10–14, San-Francisco, CA (2001).
11. W. Bakun and A. Lindh, The Parkfield, CA earthquake prediction experiment, *Science* **229**(4714) (1985) 619–624.
12. J. Barnes, An algorithm for solving non-linear equations based on the secant method, *Computer Journal* **8**(1) (1965) 66–67.
13. C. Bishop, *Neural Network for Pattern Recognition* (Oxford University Press, Oxford, UK, 1995).
14. D. Boore, Comparisons of ground motions from the 1999 Chi-Chi earthquake with empirical predictions largely based on data from California, *Bulletin of Seismological Society of America* **91**(5) (2001) 1212–1217.
15. D. Brehm and L. Braile, Intermediate-term earthquake prediction using the modified time-to-failure method in the New Madrid seismic zone, Bulletin of Seismological Society of America **88**(2) (1998) 564–580.
16. D. Brehm and L. Braile, Intermediate-term earthquake prediction using the modified time-to-failure method in Southern California, *Bulletin of Seismological Society of America* **89**(1) (1999) 275–293.
17. C. G. Broyden, A class of methods for solving nonlinear simultaneous equations, *Mathematics of Computation* **19**(1) (1965) 577–593.
18. C. Bufe and D. Varnes, Predictive modeling of seismic cycle in the greater San-Fransisco bay region, *Journal of Geophysical Research* **98** (1993) 9871–9983.
19. Y. Chen, J. Liu, B. Tsai and C. Chen, Statistical tests for pre-earthquake ionospheric anomaly, *Terrestrial Atmospheric and Oceanic Sciences* **15**(3) (2004) 385–396.
20. G. Chouliaras and G. Stavrakakis, Evaluating seismic quiescence in Greece, in *Proceeding of the European Geophysical Society Joint Assembly*, April 7–11, Nice, France (2003).
21. P. Cortez, M. Rocha and J. Neves, Evolving time-series forecasting neural network models, in *Proceedings of the International Symposium on Adaptive Systems: Evolutionary Computation and Probabilistic Graphical Models*, March 19–23, Havana Cuba (2001), pp. 84–91.

22. H. Dai and C. Macbeth, Application of backpropagation neural networks to identification of seismic arrival types, *Journal of Geophysical Research* **102**(B7) (1997) 15105–15113.
23. N. Draper and H. Smith, *Applied Regression Analysis* (Wiley, New York, NY, 1966).
24. J. L. Elman, Finding structure in time, *Cognitive Science* **14** (1999) 179–211.
25. W. Ellsworth, M. Mathews, R. Nadeau, S. Nishenko, P. Reasenberg and R. Simpson, A physically-based earthquake recurrence model for estimation of long-term earthquake probabilities, *Workshop on Earthquake Recurrence: State-of-the-art and Directions for the Future*, February 22–25, Rome, Italy (1999).
26. M. W. Firebaugh, *Artificial Intelligence, A Knowledge Based Approach*, Boyd and Fraser, Boston, MA (1988).
27. F. Freitag and E. Monte, Acoustic-phonetic decoding based on Elman predictive neural networks, in *Proceedings of the International Conference on Spoken Language Processing*, October 3–6, Philadelphia, USA (1996), pp. 522–533.
28. A. Gabrielov, I. Zaliapin, W. Newman and V. Kelis-Borok, Colliding cascades model for earthquake prediction, *Geophysical Journal International* **143**(2) (2000) 427–437.
29. R. J. Geller, D. D. Jackson, Y. Y. Kagan and F. Mulgaria, Earthquakes cannot be predicted, *Science* **275**(5306) (1997) 1616–1617.
30. C. Giles, S. Lawrence and A. Tsoi, Noisy time-series prediction using a recurrent neural network and grammatical inference, *Machine Learning* **44**(2) (2001) 161–183.
31. J. Gomez and A. Pacheco, The minimalist model of characteristic earthquakes as a useful tool for description of the recurrence of large earthquakes, *Bulletin of Seismological Society of America* **95**(5) (2004) 1960–1967.
32. A. Gonzalez, M. Vasquez-Prada, J. Gomez and A. Pacheko, Using synchronization to improve earthquake forecasting in cellular automation models, *Condensed Matter* **04**(035) (2004) 93.
33. B. Gutenberg and C. F. Richter, Earthquake magnitude, intensity, energy and acceleration, *Bulletin of the Seismological Society of America* **46**(1) (1956) 105–146.
34. M. T. Hagan, H. B. Demuth and M. Beale, *Neural Network Design* (PWS Publishing Company, Boston, MA, 1996).
35. R. A. Harris, Forecasts of the 1989 Loma Prieta, California earthquake, *Bulletin of the Seismological Society of America* **88**(4) (1998) 898–916.
36. Y. Honkura, M. Matsushima, N. Oshiman, M. Tuncer, S. Baris, A. Ito, Y. Iio and A. Isikara, Small electric and magnetic signals observed before the arrival of seismic wave, *Earth Planets Space* **54**(E) (2002) 9–12.
37. M. Ishimoto and K. Iida, Observations of earthquakes registered with the micro seismograph constructed recently, *Bulletin of Earthquake Research*, University of Tokyo, **17** (1939) 443–478 [Translated from Japanese].
38. D. Jackson, Hypothesis testing and earthquake prediction, *Earthquake Prediction: The Scientific Challenge*, February 10–11, Irvine, CA (1995).
39. D. Jackson and Y. Y. Kagan, Parkfield earthquake: Not likely this year, *Seismological Research Letters* **69**(2) (1998) 151.
40. S. Jaume, D. Weatherley and P. Mora, Accelerating moment release and the evolution of event time and size statistics, results from two cellular automation models, *Pure and Applied Geophysics* **157**(11) (2000) 2209–2226.
41. Y. Y. Kagan, VAN earthquake predictions, a statistical evaluation, *Geophysical Research Letters* 23: **11** (1996) 1315–1318.
42. Y. Y. Kagan and Jackson, D. Long-term earthquake clustering, *Geophysical Journal International* **104** (1991) 117–133.
43. A. Karim and H. Adeli, Radial basis function neural network for work zone capacity and queue estimation, *Journal of Transportaiton Engineering*, ASCE **129**(5) (2003) 494–502.
44. K. Kasahara, *Earthquake Mechanics* (Cambridge University Press, Cambridge, 1981).
45. V. I. Kelis-Borok, and V. G. Kossobokov, Premonitory activation of earthquake flow: Algorithm M8, *Physics of the Earth and Planetary Interiors* **61** (1990) 73–83.
46. T. Kerh, and D. Chu, Neural networks approach and micro-tremor measurements in estimating peak ground acceleration due to strong motion, *Advances in Engineering Software* **33**(11) (2002) 733–742.
47. J. Kirschvink, Earthquake prediction by animals, evolution and sensory perception, *Bulletin of the Seism Society of America* **90**(2) (2000) 312–323.
48. L. Knopoff, The magnitude distribution of declustered earthquakes in Southern California, in *Proceedings of the National Academy of Sciences* **97**(22) (2000), pp. 11880–11884.
49. S. Lakkos, A. Hadjiprocopis, R. Comley and P. Smith, A neural network scheme for earthquake prediction based on the seismic electric signals, *Neural Networks for Signal Processing, Proceedings of the 1994 IEEE Signal Processing Society Workshop*, September 6–8, Ermioni, Greece (1994), pp. 681–689.
50. R. Leach and F. Dowla, Earthquake early warning system using real-time signal processing, *Neural Networks for Signal Processing, Proceedings of the 1996 IEEE Signal Processing Society Workshop*, September 4–6, Kyoto, Japan (1996), pp. 463–472.
51. Y. Liu, Y. Wang, Y. Li, B. Zhang and G. Wu, Earthquake prediction by RBF neural network ensemble, in *Proceedings of the International Symposium*

on Neural Networks, August 19–21, Dalian, China (2004), pp. 962–969.
52. L. Ma, L. Zhu and Y. Shi, Attempts at using seismicity indicators for the prediction of large earthquakes by Genetic Algorithm-Neural Network method, *Asia-Pacific Economic Cooperation for Earthquake Simulation*, January 31–February 5, Brisbane, Australia (1999).
53. D. Marquardt, An algorithm for the least-squares estimation of non-linear parameters, *Journal of Applied Mathematics* **11** (1963) 431–441.
54. M. Matthews, W. Ellsworth and P. Reasenberg, A Brownian model for recurrent earthquakes, *Bulletin of the Seismological Society of America* **92**(6) (2002) 2233–2250.
55. A. Negarestani, S. Sestayeshi, M. Ghannadi-Maragheh and B. Akashe, Layered neural networks based analysis of radon concentration and environmental parameters in earthquake prediction, *Journal of Environmental Radioactivity* **62**(3) (2002) 225–233.
56. T. Read and N. Cressie, *Goodness-of-Fit Statistics for Discrete Multivariate Data* (Springer Verlag, New York, NY, 1988).
57. H. Reid, The mechanism of the earthquake; The California earthquake of April, 18, 1906, *Report of the State Earthquake Investigation Commission, Carnegie Institute of Washington*, Washington D.C. 2 (1910) 16–28.
58. E. Roeloffs, The Parkfield, California earthquake experiment, An update in 2000, *Current Science* **79**(9) (2000) 1226–1236.
59. D. Rumelhart, G. Hinton and R. Williams, Learning representations by back-propagating error, *Nature* **323**(9) (1986) 533–536.
60. M. Sharma and M. Arora, Prediction of seismicity cycles in the Himalayas using artificial neural networks, *Acta Geophysica Polonica* **53**(3) (2005) 299–309.
61. Y. Shi, J. Liu and G. Zhang, An evaluation of Chinese annual earthquake predictions, 1990–1998, *Journal of Applied Probability* **38**(A) (2001) 222–231.
62. K. Sieh, The repetition of large earthquake ruptures, in *Proceedings of the National Academy of Science* **93**(9) (1996) 3764–3771.
63. J. Stephens, E. Adams, X. Huo, J. Dent and J. Hicks, Use of neural networks in avalanche forecasting, in *Proceedings of the International Snow Science Workshop*, October 30–November 3, Snowbird, USA, (1994), pp. 327–340.
64. C. Thanassoulas, Earthquake prediction based on electrical signals recorded on ground surface, in *Proceedings: Possible Correlation between Electromagnetic Earth-Fields and Future Earthquakes*, Bulgarian Academy of Sciences, Sofia, Bulgaria (2001), pp. 19–30.
65. K. Tiampo, J. Rundle, S. McGinnis, S. Gross and W. Klein, Mean-field threshold systems and phase dynamics: An application to earthquake fault systems, *Europhysics Letters* **60** (2002) 481–487.
66. P. Varshney, *Distributed Detection and Data Fusion* (Springer-Verlag, New York, NY, 1997).
67. D. Vere-Jones, R. Robinson and W. Wang, Remarks on the accelerated release moment model: Problems of model formulation, simulation and estimation, *Geophysics Journal International* **144**(3) (2001) 517–531.
68. M. Wyss, P. Bodin and R. E. Haberman, Seismic quiescence at Parkfield; an independent indication of an imminent earthquake, *Nature* **345**(290) (1990) 426–428.
69. Working Group on California Earthquake Probabilities, Earthquake probabilities in the San Fransisco bay region, *United States Geological Survey Open-File Report*, 03–214 (2003).
70. S. Xu, Ability evaluation for earthquake prediction, *Science Books and Periodicals Press — Seismology Volume* 586–590 [Translated from Chinese] (1989).
71. I. Zaliapin, V. Kelis-Borok and G. Axen, Premonitory spreading of seismicity of faults' network in southern California: Precursor accord, *Journal of Geophysical Research* **107**(B10) (2002) 2221.
72. I. Zaliapin, V. Kelis-Borok and M. Ghil, A Boolean delay equation model of colliding cascades; Part II: Prediction of critical transitions, *Journal of Statistical Physics* **111**(3) (2003) 839–861.
73. G. Zoller, S. Hainzl and J. Kurths, A systematic test on precursory seismic quiescence in Armenia, *Natural Hazards* **26**(3) (2002) 245–263.