Applied Soft Computing 13 (2013) 1314–1328

Contents lists available at SciVerse ScienceDirect

Applied Soft Computing

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

Neural networks to predict earthquakes in Chile

J. Reyes [a], A. Morales-Esteban [b], F. Martínez-Álvarez [c,*]

[a] TGT, Camino Agrícola 1710, Torre A, Of 701, Santiago, Chile
[b] Department of Continuum Mechanics, University of Seville, Spain
[c] Department of Computer Science, Pablo de Olavide University of Seville, Spain

ARTICLE INFO

Article history:
Received 23 December 2011
Received in revised form
26 September 2012
Accepted 31 October 2012
Available online 23 November 2012

Keywords:
Earthquake prediction
Artificial neural networks
Seismic risk
Time series

ABSTRACT

A new earthquake prediction system is presented in this work. This method, based on the application of artificial neural networks, has been used to predict earthquakes in Chile, one of the countries with larger seismic activity. The input values are related to the *b*-value, the Bath's law, and the Omori–Utsu's law, parameters that are strongly correlated with seismicity, as shown in solid previous works. Two kind of prediction are provided in this study: The probability that an earthquake of magnitude larger than a threshold value happens, and the probability that an earthquake of a limited magnitude interval might occur, both during the next five days in the areas analyzed. For the four Chile's seismic regions examined, with epicenters placed on meshes with dimensions varying from $0.5^\circ \times 0.5^\circ$ to $1^\circ \times 1^\circ$, a prototype of neuronal network is presented. The prototypes predict an earthquake every time the probability of an earthquake of magnitude larger than a threshold is sufficiently high. The threshold values have been adjusted with the aim of obtaining as few false positives as possible. The accuracy of the method has been assessed in retrospective experiments by means of statistical tests and compared with well-known machine learning classifiers. The high success rate achieved supports the suitability of applying soft computing in this field and poses new challenges to be addressed.

1. Introduction

Chile emerges as, perhaps, the highest seismic region in the world, with the possible exception of Japan. The earthquake of magnitude $9.5M_w$ occurred in Chile in 1960 is, probably, the largest earthquake in world history [42]. The mega-earthquake caused more than 2000 deaths, more than 2 million people lost their homes, and the tsunami generated caused damage in distant locations, such as Japan and Hawaii [8]. Fifty years later, on 27th February 2010, the south of Chile was affected by another mega-earthquake of moment magnitude 8.8. The earthquake and the ensuing tsunami caused claimed together more than 500 lives and severe damage along the adjacent coasts [43].

Despite forecast and prediction are used as synonymous in many fields, there are subtle differences, as discussed in [48]. In this sense, it has to be pointed that this work is about earthquake prediction. Given a set of inputs, a prediction consists in the interaction of such inputs through laws or well defined rules such as thermodynamic, rigid body mechanics, etc. As a result, the future has to be calculated with a high degree of accuracy as kinematics describes the trajectory of a projectile. In seismology the input values correspond to the stress point to point and the asperities or plates sub-topography, which are almost impossible to obtain.

Although several works claim to provide earthquake prediction, an earthquake prediction must provide, according to [3], the following information:

1. A specific location or area.
2. A specific span of time.
3. A specific magnitude range.
4. A specific probability of occurrence.

That is, an earthquake prediction should state when, where, how big, and how probable the predicted event is and why the prediction is made. Unfortunately, no general useful method to predict earthquakes has been found yet. And it may never be possible to predict the exact time when a damaging earthquake will occur, because when enough strain has built up, a fault may become inherently unstable, and any small background earthquake may or may not continue rupturing and turn into a large earthquake.

This is, precisely, the challenge faced in this work and its main motivation. In particular, the main novelty of this contribution is the development of a system capable of predicting earthquake occurrence for periods of time statistically significant, in order to allow the authorities to deploy precautionary policies.

Indeed, a system able to generate confident earthquake predictions allows the shutdown of systems susceptible to damage such

* Corresponding author.
E-mail addresses: daneel@geofisica.cl (J. Reyes), ame@us.es (A. Morales-Esteban), fmaralv@upo.es (F. Martínez-Álvarez).

1568-4946/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.asoc.2012.10.014

as power stations, transportation and telecommunications systems [16], helping the authorities to focalize their efforts to reduce damage and human looses.

Since false alarms are cause for deep concern in society, the proposed approach has been especially adjusted to generate as few false alarms as possible. In other words, when the output of the system is there is an upcoming earthquake, it is highly probable an earthquake occurs, as observed in the four Chilean areas examined.

It is worth noting that the system proposed is almost real-time, therefore, the authority receiving this information does not need to distinguish between mainshock and aftershock. That is, both mainshocks and aftershocks are considered because they are retrospective designations that can only be identified after an earthquake sequence has been completed [62].

To achieve the goal of predicting earthquakes, artificial neuronal networks (ANN) are used to predict earthquakes when the probability of exceeding a threshold value is sufficiently high. The ANN's presented in this paper fulfill the requirements above discussed. In particular, the specific location or area varies from $0.5^\circ \times 0.5^\circ$ to $1^\circ \times 1^\circ$. The specific span of time is within the next five days. The specific magnitude range is determined from the magnitude that triggers the alarm (the average magnitude of the area plus 0.6 times the typical deviation, as discussed in Section 4). Finally, the rate of earthquakes successfully predicted has been empirically calculated and varies from 18% to 92% (Section 6) depending on the area under analysis.

The main contribution of this work can be then summarized as follows. A new earthquake prediction system, based on ANN's, has been proposed. This system makes use of a combination seismicity indicators never used before, based on information correlated with earthquake occurrence, as input parameters. Especially remarkable is the small spatial and temporal uncertainty of this system, thus contrasting with other existing methods that mainly predict earthquakes in huge areas with high temporal uncertainty (see Section 2 for further information).

The rest of the work is divided as follows. Section 2 introduces the methods used to predict the occurrence of earthquakes. The geophysical fundamentals are described in Section 3. A description of the ANN's used in this work is provided in Section 4. In Section 5, the training of the ANN's is explained as well as the method to evaluate the networks. In Section 6, the results obtained for every ANN in the testing process are shown, as well as a comparative analysis with other classifiers. Also, a fractal hypothesis that explains the solid performance observed is presented. Section 7 is included to assess the robustness of the method considering the possible influence of earthquake clustering and different time periods of application. Finally, Section 8 summarizes the conclusions drawn.

2. Related work

Despite the great effort made and the multiple models developed by different authors [83], no successful method has been found yet. Due to the random behavior of earthquakes generation, it may never be possible to ascertain the exact time, magnitude and location of the next damaging earthquake.

Neural networks have been successfully used for solving complicated pattern recognition and classification problems in many domains such as financial forecasting, signal processing, neuroscience, optimization, etc. (in [1] neural applications are detailed). But only very few and very recent studies have been published on the application of neural networks for earthquake prediction [4,45,63,64,82].

Actually, there is no consensus among researchers in how to build a time-dependent earthquake forecasting model [17] nowadays, therefore, it is impossible to define a unique model. Hence, the Regional Earthquake Likelihood Model (RELM) has developed 18 different models.

The U.S. Geological Survey (USGS) and the California Geological Survey (CGS) have developed a time-independent model based on the assumption that the probability of the occurrence of an earthquake, in a given period of time, follows a Poisson distribution [66]. The forecast for California is composed of four types of earthquake sources: two types of faults, zones, and distributed background seismicity. The authors also presented a time-dependent model based on the time-independent 2002 national seismic hazard model and additional recurrence information.

Kagan et al. [37] presented a smoothed-seismicity model for southern California. This is a five-year forecast of earthquakes with magnitudes 5.0 and greater, based on spatially smoothed historical earthquake catalog [36]. The forecast is based on observed regularity of earthquake occurrence rather than on any physical model. The model presented in [29] is similar to that of Kagan et al. [37] and it is based on a smoothed seismicity model. This model is based on the assumption that small earthquakes are used to map the distribution of large damaging earthquakes and removing aftershocks.

Shen et al. [80] present a probabilistic earthquake forecast model of intermediate to long times during the interseismic time period, constrained by geodetically derived crustal strain rates. Moreover, the authors in [7] propose simple methods to estimate long-term average seismicity of any region. For California, they compute a long-term forecast based on a kinematic model of neotectonics derived from a weighted least-squares fit to available data. On the other hand, Console et al. [11] present two 24-h forecasts based on two Epidemic-Type Aftershock Sequence model (ETAS). These models are based on the assumption that every earthquake can be regarded as both triggered by previous events and as a potential triggering event for subsequent earthquakes [10].

The model presented by Rhoades [70] is a method of long-range forecasting that uses the previous minor earthquakes in a catalog to forecast the major ones. This model is known as EEPAS (Every Earthquake a Precursor According to Scale). Alternatively, Ebel et al. [15] have presented a set of long term (five-year) forecast of $M \geq 5.0$ earthquakes and two different methods for short-term (one-day) forecasts of $M \geq 4.0$ earthquakes for California and some adjacent areas. Five different methods for the RELM were presented by Ward [92]. The first one is based on smoothed seismicity for earthquakes of magnitude larger or equal to 5.0. The second is based on GPS derived strain and Kostrov's formula. The third one is based on geological fault slip-data. The fourth model is simply an average of the three aforementioned models and, finally, the last one is based on an earthquake simulator. The last model of the RELM was presented by Gerstenberguer et al. [19] and is a short term earthquake probability model that is a 24-h forecast based on foreshock/aftershock statistics.

The work in [95] presented and Asperity-based likelihood model for California. This method is based on three assumptions: the b-value has been shown to be inversely dependent of applied shear stress so the b-value can be used as a stress–meter within the Earth's crust where no direct measurements exist [78]. Secondly, asperities are found to be characterized by low b-values [77]. Finally, data from various tectonic regimes suggest that the b-value of micro-earthquakes is quite stationary with time [75]. Equally remarkable is the pattern informatics (PI) method [30] that does not predict earthquakes but forecasts the regions where earthquakes are most likely to occur in the relatively near future (5–10 years).

Many other methods have been proposed by different authors to predict earthquakes: From unusual animal behavior [38] to electromagnetic precursors [89] to any imaginable precursor. Despite some correct predictions and many efforts done, errors are predominant (i.e. Parkfield [22] or Loma Prieta [91]). Nevertheless, there are

some exceptions such as the Xiuyen prediction [99] or the prediction by Matsuzawa et al. [49], who predicted that an earthquake of magnitude 4.8 ± 0.1 would hit Sanriku area (NE Japan) with 99% probability on November 2001. An earthquake of magnitude 4.7 actually occurred on November 13th 2001, as expected. The method used by the authors was the comparison of recurrent seismograms every 5.3 ± 0.53 years altogether with the deduction of the existence of a small asperity with a dimension of $\approx$1 km surrounded by stable sliding area on the plate boundary.

Despite the vast variety of methods attempting to predict earthquakes, the analysis of the *b*-value as seismic precursor has stood out during the last decade. Indeed, the authors in [50] applied artificial intelligence to obtain patterns which model the behavior of seismic temporal data and claimed that this information could help to predict medium-large earthquakes. The authors found some patterns in the temporal series that correlated with the future occurrence of earthquakes. This method was based on the work by [54], where the authors had found spatial and temporal variation of *b*-values preceding the NW Sumatra earthquake of December 2004. Also, the use of artificial intelligence techniques can be found in [47]. In particular, the authors discovered the relation existing between the *b*-value and the occurrence of large earthquake by extracting numeric association rules and the well-known *M*5′ algorithm, obtaining also linear models.

Due to the multiple evidences discovered, the variations of the *b*-value have been decided to be part of the input data used in the ANN's designed.

3. Geophysical fundamentals

This section exposes the fundamentals that support the methodology used to build the multilayer ANN's. That is, the information underlying the input/output parameters of the designed ANN's. Firstly, the Gutenberg–Richter law is explained. Then, the *b*-value of the Gutenberg–Richter law is examined as seismic precursor. Later, the Omori/Utsu and Bath's laws, that are related to the frequency and magnitude of the aftershocks, are presented. Finally, the six seismic regions of Chile are described [69], and their year of completeness calculated.

3.1. Gutenberg–Richter's law

The size distribution of earthquakes has been investigated since the beginning of the 20th century. Omori [61] illustrated a table of the frequency distribution of maximum amplitudes recorded by a seismometer in Tokyo. Moreover, Ishimoto and Iida [32] observed that the number of earthquakes, *N*, of amplitude greater or equal to *A* follows a power law distribution defined by:

$$N(A) = aA^{-B} \tag{1}$$

where *a* and *B* are the adjustment parameters.

Richter [71] noted that the number of shocks falls off very rapidly for the higher magnitudes. Later, Gutenberg and Richter [24] suggested an exponential distribution for the number of earthquakes versus the magnitude. Again, Gutenberg and Richter [25] transformed this power law into a linear law, expressing this relation for the magnitude frequency distribution of earthquakes as:

$$\log_{10} N(M) = a - bM \tag{2}$$

This law relates the cumulative (or absolute) number of events $N(M)$ with magnitude larger or equal to *M* with the seismic activity, *a*, and the size distribution factor, *b*. The *a*-value depends of the volume and the temporal window chosen and is the logarithm of the number of earthquakes with magnitude larger or equal to zero. The *b*-value is a parameter that reflects the tectonic of the area under analysis [40] and has been related with the physical characteristics of the area. A high value of the parameter implies that the number of earthquakes of small magnitude is predominant and, therefore, the region has a low resistance. Contrarily, a low value shows a smaller difference between the relative number of small and large events, implying a higher resistance of the material. The coefficient *b* usually takes a value around 1.0 and has been widely used by investigators [87].

It is assumed that the magnitudes of the earthquakes that occur in a region and in a certain period of time are independent and identically distributed variables that follow the Gutenberg–Richter law [67]. If earthquakes with magnitude M_0 and larger are used, this hypothesis is equivalent to suppose that the probability density of the magnitude *M* is exponential:

$$f(M, \beta) = \beta e^{-\beta(M - M_0)} \tag{3}$$

Gutenberg and Richter [26] used the least squares method to estimate coefficients in the frequency–magnitude relation. Shi and Bolt [81] pointed out that the *b*-value can be obtained by this method but the presence of even a few large earthquakes influences the result significantly. The maximum likelihood method, hence, appears as an alternative to the least squares method, which produces estimates that are more robust when the number of infrequent large earthquakes changes.

If the magnitudes are distributed according to the Gutenberg–Richter law and we have a complete data for earthquakes with magnitude M_0 and larger, the maximum likelihood estimate of *b* is Aki [2] and Utsu [86]:

$$b = \frac{\log e}{\bar{M} - M_0} \tag{4}$$

where $\bar{M}$ is the mean magnitude and M_0 is the cutoff magnitude [67].

The authors in [81] also demonstrated that for large samples and low temporal variations of *b*, the standard deviation of the estimated *b* is:

$$\sigma^2 = 2.30 b^2 \sigma(M) \tag{5}$$

where

$$\sigma^2(M) = \frac{\sum_{i=1}^{n} (M_i - \bar{M})^2}{n} \tag{6}$$

where *n* is the number of events and M_i is the magnitude of every event. Note that the *b*-value uncertainty is not required in this work because it is based on references [50,54] where only variations of *b*-value are used as seismic precursor, not considering its uncertainty.

It should be pointed out that different databases provide different locations, magnitudes and focal mechanisms in the catalogs [90]. Nevertheless, the data used in this study is that of provided by Chile's National Seismological Service, as it is considered the most confident regarding Chile seismicity.

Since Gutenberg and Richter estimated the *b*-value in various regions of the world, there is controversy between the investigators about the spatial and temporal variations of the *b*-value. Some authors [34,60] suggest that the value is universal and constant whereas some other authors [5,19,84,93] assert that there are spatial and temporal variations of the *b*-value. It is true that for large regions and for long periods of time $b \approx 1$. However, significant variations have been observed in limited regions and for short periods of time [40]. Local earthquakes have been associated with variations of the *b*-value of days or even hours [23].

From all the aforementioned possibilities, the maximum likelihood method has been selected for the estimation of the *b*-value in this work. The cutoff magnitude used is $M_0 = 3.0$.

It is expected that the ANN's will be able to incorporate the Gutenberg–Richter law from the *b*-value variation series with the

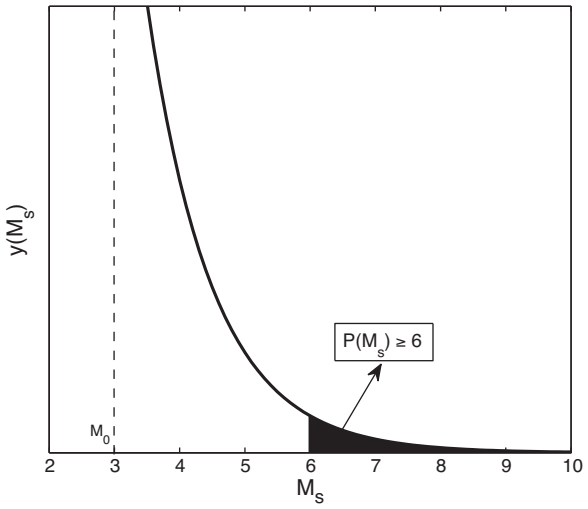

Fig. 1. Probability of observing an earthquake with magnitude larger or equal to 6 (from the inclusion of Gutenberg–Richter's law to ANN's).

theoretical probability to predict an earthquake of magnitude equal or larger to 6.0M_s, in concordance with the current configuration of the probability density function, considering the last 50 earthquakes in the calculation procedure (see Fig. 1).

3.2. The b-value as seismic precursor

The *b*-value of the Gutenberg–Richter law is an important parameter, because it reflects the tectonics and geophysical properties of the rocks and fluid pressure variations in the region concerned [40,100]. Thus, the analysis of its variation has been often used in earthquake prediction [54]. It is important to know how the sequence of *b*-values has been obtained, before presenting conclusions about its variation. The studies of both Gibowitz [20] and Wiemer et al. [94] on the variation of the *b*-value over time refer to aftershocks. The found an increase in *b*-value after large earthquakes in New Zealand and a decrease before the next important aftershock. In general, they showed that the *b*-value tends to decrease when many earthquakes occur in a local area during a short period of time.

Schorlemmer et al. [78] and Nuannin et al. [54] infer that the *b*-value is a stress–meter that depends inversely on differential stress. Hence, Nuannin et al. [54] presented a very detailed analysis on *b*-value variations. They studied the earthquakes in the Andaman–Sumatra region. To consider variations in *b*-value, a sliding time–window method was used. From the earthquake catalog, the *b*-value was calculated for a group of 50 events. Then the window was shifted by a time corresponding to five events. They concluded that earthquakes are usually preceded by a large decrease in *b*, although in some cases a small increase in this value precedes the shock. Previously, the authors in [74] clarified the stress changes in the fault and the variations on the stress changes in the fault and the variations of the *b*-value surrounding an important earthquake. In fact, they claimed: "A systematic study of temporal changes in seismic *b*-values has shown that large earthquakes are often preceded by an intermediate-term increase in *b*, followed by a decrease in the months to weeks before the earthquake. The onset of the *b*-value can precede earthquake occurrence by as much as seven years".

The neural networks used in this work to predict earthquakes do not use the temporal sequence of magnitude of earthquakes but the temporal sequence of variations of the *b*-value (calculated from the sequence of magnitude of earthquakes), as the correlation between the variations of the *b*-value and the later occurrence of earthquakes is proved to be high.

3.3. Aftershocks

In the theoretical Poisson law dependent data (aftershocks and foreshocks) must be removed [51,68]. However, it is known that almost all larger earthquakes trigger aftershocks with a temporal decaying probability [27]. In this study, the complete database of earthquakes has been used because the variation of the *b*-value (which is the input data for the ANN's) is strongly correlated with the aftershocks and foreshocks. It is intended that the ANN's presented in this paper incorporate the Omori–Utsu [88] and Bath's [6] laws from the sequence of the *b*-value variations altogether with the value of the largest earthquake registered in the last seven days.

3.3.1. Omori/Utsu's law

The rate of aftershocks with time follows Omori's law [88] in which the aftershock frequency decreases by roughly the reciprocal of time after the main shock according to:

$$N(t) = \frac{k}{c+t} \tag{7}$$

where N is the occurrence rate of aftershocks, t indicates the elapsed time since the mainshock [88], k is the amplitude, and the c-value is the *time offset* parameter which is a constant typically much less than one day.

The modified version of Omori's law was proposed by Utsu [85], which is known as Omori–Utsu law:

$$N(t) = \frac{k}{(c+t)^p} \tag{8}$$

The p-value is the *fitness parameter*, that modifies the decay rate, and typically falls in the range 0.8–1.2 [88].

These patterns describe only the statistical behavior of aftershocks. The actual times, numbers and locations of the aftershocks are stochastic, while tending to follow these patterns.

Alternative models for describing the aftershock decaying rate have been proposed [21,52], however the Omori–Utsu law generally provides a very good adjustment to the data [27].

3.3.2. Bath's law

The empirical Bath's law [6] states that the average difference in magnitude between a mainshock and its largest aftershock is approximately constant and equals to 1.1–1.2M_w, regardless of the mainshock magnitude.

It is expected that the ANN's, presented in this paper, will be able to deuce Bath's law from the incorporation of the magnitude of the largest earthquake registered in the last seven days.

3.4. Seismic areas in Chile

A global analysis of mainland Chile would involve a huge spatial uncertainty since its surface is 756,000 km^2. Thus, the predictions would lack of interest because it is obvious that earthquakes are likely to occur daily in a so vast area with an associated high rate occurrence.

Also, Chile's geography presents different particularities; therefore, a global analysis of all regions does not seem adequate in this situation. Some of these peculiarities are now listed:

1. Every seismogenic area described in [69] is characterized by different histograms based on Gutenberg–Richter law.
2. In Northern Chile the speed of subduction is 8 cm/year, whereas Southern Chile presents a speed of 1.3 cm/year.

Table 1
Seismic Chilean areas analyzed.

Parameters	Seismic area	Vertices	Dimensions
Talca	#3	(35CS, 72° W) and (36° S, 71° W)	$1^\circ \times 1^\circ$
Pichilemu	#4	(34° S, 72.5° W) and (34.5° S, 72° W)	$0.5^\circ \times 0.5^\circ$
Santiago	#5	(33° S, 71° W) and (34° S, 70° W)	$1^\circ \times 1^\circ$
Valparaíso	#6	(32.5° S, 72° W) and (33.5° S, 71° W)	$1^\circ \times 1^\circ$

3. Chile is longitudinally divided into coastal plains, coast mountains, intermediate depression, Andes mountains, which is related with different seismogenic sources: interplate thrust, medium depth intraplate, cortical and outer-rise.
4. Complex fault systems can be found across Continental Chile (Mejillones, San Ramón, Liquiñe-Ofqui faults, etc.), which means that every area has a specific fractal complexity, possibly characterized by Hurt's exponent (Section 6.6).

Therefore, the zonification proposed by Reyes and Cárdenas [69] is used in this work. Based on Kohonen-Self Organized Map (SOM), the authors demonstrated that considering the similarity of histograms, frequency versus magnitude, Chile can be divided into six seismic regions.

The whole continental territory of Chile was thus divided into 156 squares of $1^\circ \times 1^\circ$ and the SOM defined six seismic regions. Later, it will be demonstrated that it is possible to predict earthquakes in Chile using multilayer ANN's without creating 156 ANN (one for each square of the grid). Instead of the 156 ANN's, a prototype for each seismic region is created. However, the Southern regions 1 and 2 are characterized by a low seismic activity (the sum of earthquakes of magnitude equal or larger to $4M_s$, for both regions, in the period 2001–2010, is only 32). The joint surface of both zones, corresponding to the Chilean Patagonia, is approximately 245,000 km^2, producing $1.3{\cdot}10^{-5}$ earthquakes per year and km^2. Furthermore, the subduction speed is 1.5 cm/year approximately or, in other words, there occur mainly low magnitude earthquakes (central Chile's speed of subduction is 7.5 cm/year). Moreover, Chilean Patagonia does not generate the minimum number of events to create the training sets. In short, the Southern regions lack of interest due its huge spatial uncertainty, its low earthquake occurrence, and the low magnitude characterizing such shocks. In this study, only the four ANN's, corresponding to regions #3–6, are presented.

In Section 4, it will be demonstrated that a minimum of 70 + 122 earthquakes is necessary to create an ANN. Furthermore, another 50 extra earthquakes are required to test every ANN. The lapse of time, from the database, used in this study, is limited from 1st January 2001 to March 31st 2011 although the precise limits depend on the square under study.

To accumulate the largest quantity of earthquakes, for the statistical analysis of an epicentral region, the observed area must be extended, which increases the spatial uncertainty of the prediction. In order to achieve the maximum spatial accuracy a plot that shows the quantity of earthquakes observed for 2010 is shown in Fig. 2.

The quantity of earthquakes is maximized for the 34.5°S latitude. This location guarantees the minimum quantity of data necessary (approximately 70 + 122 + 50) from the information contained in relatively small areas.

Finally, the four seismic regions of Chile that will be modeled by the ANN's are summarized in Table 1.

3.5. Chile's earthquake database

The database of earthquakes used in this paper has been obtained from the Chile's National Seismological Service. In order to calculate the *b*-value from [4] the database must be complete. To guarantee the completeness of the data the analysis should only include earthquakes recorded after the year of completeness of the database. So, once the seismic database has been obtained, the first step is to calculate, for every area, the year of completeness of the seismic catalog, defined as the year from which all the earthquakes of magnitude larger or equal to the minimum magnitude have been recorded. The Chilean earthquake database only contains earthquakes of magnitude equal or larger to 3.0. The year of completeness of the seismic catalog can be graphically calculated. With that purpose the time is inversely plot in the *x*-axis and the accumulated number of earthquakes in the *y*-axis, as illustrated in Fig. 3.

It can be observed that the slope of the curve, that represents the annual rate, is approximately constant until 2001, been reduced for the previous years. If the arrival of earthquakes is admitted to follow a Poisson stationary process, the annual rate should be time independent, so a systematic decrease of this rate, when it approaches the initial years of the database must be concluded as a lack of completeness, ought to the small number of seismic stations available at that moment. In this case, 2001 is the year of completeness of the catalog for the four areas for a magnitude larger or equal to 3.0.

4. ANN applied to predict earthquakes in Chile

This section discusses the ANN's configuration for predicting earthquakes in Chile. In particular, the number of layers and neurons per layer, the activation function, the topology and the learning paradigm are introduced.

4.1. ANN's configuration

One neural network has been used for each seismic area. Note that these areas are tagged accordingly to the main city existing in their area of interest or cells: Talca (seismic area #3), Pichilemu (seismic area #4), Santiago (seismic area #5) and Valparaíso (seismic area #6).

Although one different ANN has been applied to each area, they all share the same architecture. The common features are listed in Table 2.

Note that all the four ANN's have been constructed following the scheme discussed and successfully applied in [65], where the

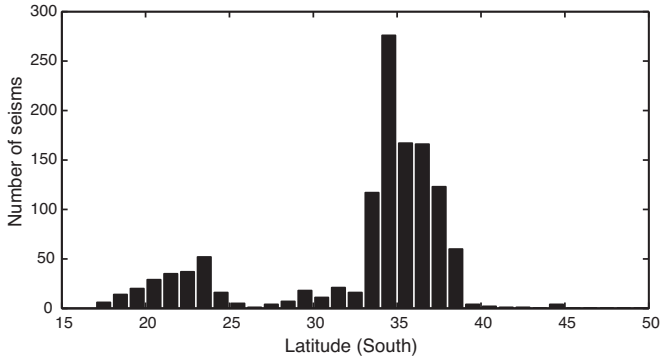

Fig. 2. Seismicity in Chile during 2010.

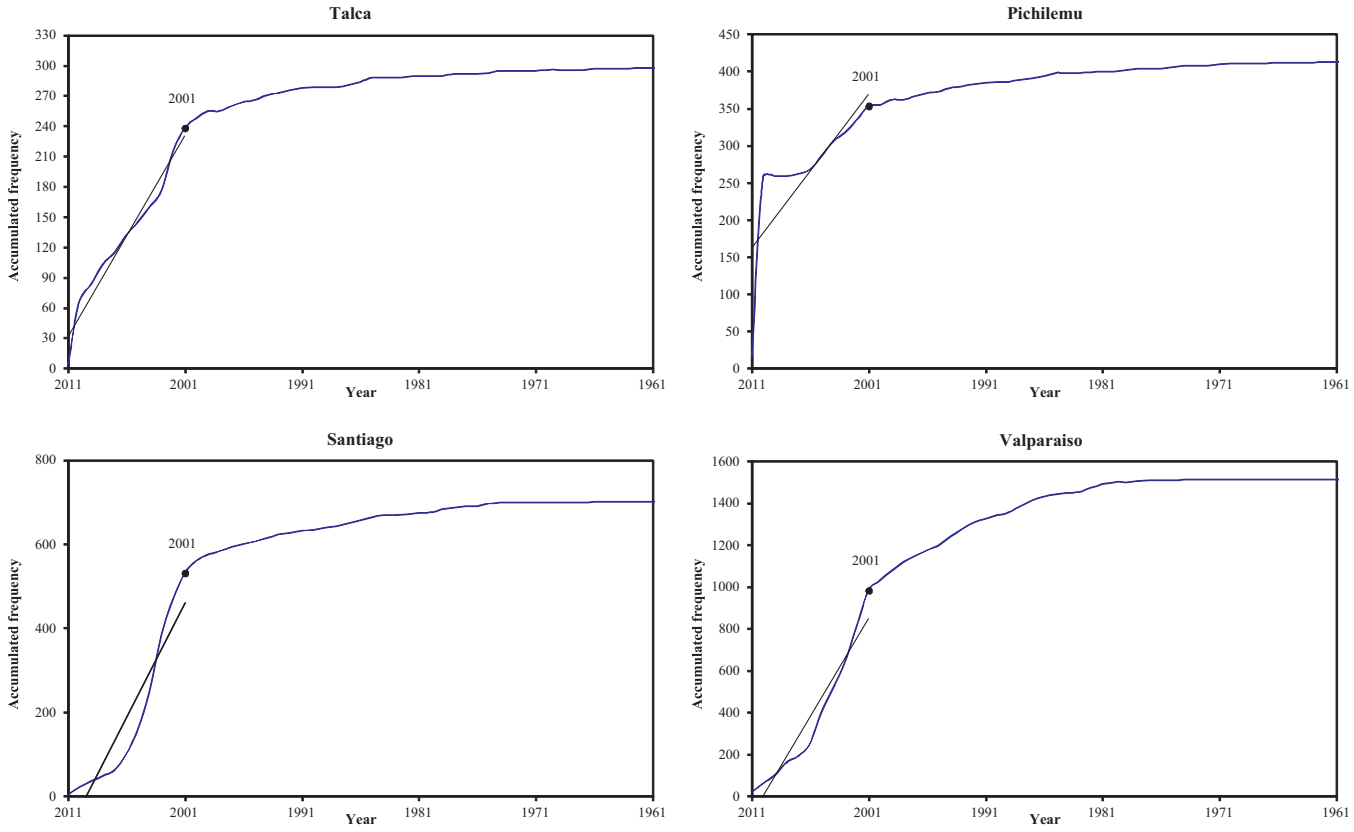

Fig. 3. Year of completeness for all the studied areas.

authors used such an ANN configuration to predict atmospheric pollution.

4.2. Input neurons

Every time an earthquake occurs at the cell subjected to analysis a new training vector, composed of seven inputs and one output, is created. First, the Gutenberg–Richter law's b-value is calculated using the last 50 quakes recorded [53]:

$$b_i = \frac{\log(e)}{(1/50)\sum_{j=0}^{49} M_{i-j} - 3} \tag{9}$$

where M_i is the magnitude for the ith earthquake and three is the reference magnitude, M_0.

Then, increments of b are calculated:

$$\Delta b_{1i} = b_i - b_{i-4} \equiv x_{1i} \tag{10}$$

$$\Delta b_{2i} = b_{i-4} - b_{i-8} \equiv x_{2i} \tag{11}$$

$$\Delta b_{3i} = b_{i-8} - b_{i-12} \equiv x_{3i} \tag{12}$$

$$\Delta b_{4i} = b_{i-12} - b_{i-16} \equiv x_{4i} \tag{13}$$

$$\Delta b_{5i} = b_{i-16} - b_{i-20} \equiv x_{5i} \tag{14}$$

Table 2
Common features in ANN's.

Parameters	Values
Input neurons	7
Neurons in hidden layer	15
Output neurons	1
Activation function	Sigmoid shape
Topology of the network	Feedforward
Learning paradigm	Backpropagation

From these equations it can be concluded that 70 earthquakes are required to calculate the x_i values, therefore, to obtain the five of the ANN inputs.

The sixth input variable, x_{6i}, is the maximum magnitude M_s from the quakes recorded during the last week in the area analyzed. The use of this information as input is to indirectly provide the ANN's with the required information to model Omori/Utsu and Bath's laws.

$$x_{6i} = \max\{M_s\}, \quad \text{when} \quad t \in [-7, 0) \tag{15}$$

where the time t is measured in days.

Note that Omori/Utsu law has been used to design the ANN architecture, but the equality in Eq. (8) is not applied as such, since the correlation between x_{6i} and the rate of aftershocks is quite high. Indeed, if the Gutenberg–Richter law was considered in any particular area, it would be reported that, in a period equals to the completeness time, large earthquakes would generate many more aftershocks than small ones. Therefore, it is expected that the information contained in Omori/Utsu law (number aftershocks per unit of time) will be codified in variable x_{6i}.

The last input variable, x_{7i}, identifies the probability of recording an earthquake with magnitude larger or equal to 6.0. The addition of this information as input is to include Gutenberg–Richter's law in a dynamic way. It is calculated from the probability density function (PDF):

$$x_{7i} = P(M_s \geq 6.0) = e^{-3b_i/\log(e)} = 10^{-3b_i} \tag{16}$$

Finally, there is one output variable, y_i, which is the maximum magnitude M_s observed in the cell under analysis, in the next five days. Note that y_i has been set to 0 for such situations where

no earthquake with magnitude equal or greater to 3, $M_s \geq 3$ was recorded. Formally:

$$y_i = \max\{M_s\}, \quad \text{when} \quad t \in (0, 5] \tag{17}$$

where the time t is measured in days.

Mathematically, the training vector associated to the ith earthquake can be expressed as:

$$T_i = \{x_{1i}, x_{2i}, x_{3i}, x_{4i}, x_{5i}, x_{6i}, x_{7i}, y_i\} \tag{18}$$

The minimum number of input vectors linearly independent forming the training set depends on the number of synaptic weights, discussed in Section 4.3.

4.3. Hidden layer neurons

Although the four ANN's have one input layer, the number of hidden layers, and in particular the number of neurons forming those layers, may vary.

The number of input neurons has already been discussed and set to 7. Also, following with the scheme introduced in [65], the number of hidden layers is 1. To determine the number of neurons that compose the hidden layer, the Kolmogorov's theorem [39] has been applied. In particular, this theorem was adapted to ANN's by Hecht-Nielsen [28]. The theorem states that an arbitrary continuous function $f(x_1, x_2, \ldots, x_n)$ on an n-dimensional cube (of arbitrary dimension n) can be represented as a composition of addition of $2n+1$ functions of one variable. Therefore, the number of neurons in the hidden layer is $2 \cdot 7 + 1 = 15$, as there are $n = 7$ input neurons.

In general, the number of synaptic connections of an ANN is calculated as: [input neurons]·[hidden layer neurons] + [hidden layer neurons]·[number of hidden layers] + 2. Consequently, the number of connections is $7 \cdot 15 + 15 \cdot 1 + 2 = 122$. This number means that a minimum of 122 training vectors, linearly independent, are required to find all these 122 synaptic connections. However, the training set may include more vectors particularly selected to improve the learning under certain situations (i.e. to increase or decrease the mean value of y_i during the selftest).

4.4. Output neuron

Regarding the outputs, the ANN's only have one: the maximum value observed in the quakes occurred during the next five days, in their corresponding cell. The procedure followed to trigger the ANN is now detailed. The mean magnitude $\bar{M}_s$ is calculated from the 122 earthquakes observed during the training set and, then, 0.6 times the standard deviation is added to this value. Therefore this value is the threshold that activates the ANN's output.

4.5. Activation function

The activation function selected is the sigmoid. Mathematically, this function is formulated as:

$$\phi(X_i) = \frac{1}{1 + e^{-X_i}} \tag{19}$$

where

$$X_i = \phi\left(\sum_i w_{ij} x_i\right) \tag{20}$$

and w_{ij} are the connection weights between unit i and unit j, and u_i are the signals arriving from unit i. The signal generated by unit i is sent to every node in the following layer or is registered as an output if the output layer is reached.

4.6. Network topology

A feedforward neural network is an ANN where connections between the units do not form a directed cycle (unlike Recurrent Neural Networks, RNN [31]).

The feedforward neural network was the first and simplest type of artificial neural network devised. In this network, the information moves in only one direction, forward, from the input nodes, through the hidden nodes and to the output nodes. There are no cycles or loops in the network.

4.7. Learning paradigm

The learning method selected is backpropagation [73], a common supervised method to train ANN's, especially useful for feedforward networks, the topology chosen for these ANN's. Backpropagation usually consists of two main phases: propagation and weight update.

5. Training the ANN's

This section introduces the quality parameters used to train the ANN's. Also, the results obtained after the training phase are presented and discussed for every area.

As stated in [33], any prediction should be tested at two stages: A training period, when suitability of the method is ascertained, and values of adjustable parameters established, and a evaluation stage, where no parameter fitting is allowed. This is exactly what it is done in this work, presenting in this section the first stage (the training of the method) and in Section 6 the second stage (evaluation of the model).

5.1. Performance evaluation

To assess the performance of the ANN's designed, several parameters have been used. In particular:

1. *True positives (TP)*. The number of times that the ANN predicted an earthquake and an earthquake did occur during the next five days.
2. *True negatives (TN)*. The number of times that the ANN did not predict an earthquake and no earthquake occurred.
3. *False positives (FP)*. The number of times that the ANN predicted an earthquake but no earthquake occurred during the next five days.
4. *False negatives (FN)*. The number of times that the ANN did not predict an earthquake but an earthquake did occur during the next five days.

If the output does not exceed the threshold and the maximum observed magnitude M_s during the next five days is less than the threshold ($M_s < T_k$ in next five days), this situation is called *zero-level hit*, and denoted by P_0. On the other hand, if the output $\geq$ threshold and the maximum observed magnitude M_s during the next five days is larger than the threshold ($M_s \geq T_k$ in next five days), the situation is called *one-level hit*, and denoted by P_1. Both probabilities, widely used by seismologists to evaluate the performance of approaches, are calculated as follows:

$$P_0 = \frac{\#(\text{zero-level hits})}{N_0} = \frac{TN}{TN + FN} \tag{21}$$

$$P_1 = \frac{\#(\text{one-level hits})}{N_1} = \frac{TP}{TP + FP} \tag{22}$$

where $N_0 = TN + FN$ denotes the times that the ANN predicted the zero-level and $N_1 = TP + FP$ the times that the ANN predicted the

Table 3
Training values for Talca.

Parameters	Value
TP	17
TN	77
FP	10
FN	18
P_1	63.0%
P_0	81.1%
S_n	48.6%
S_p	88.5%
Average	70.3%

one-level. Obviously, $N = N_0 + N_1$ denotes the total number of possible predictions.

Additionally, two more parameters have been used to evaluate the performance of the ANN's, as they correspond to common statistical measures of supervised classifiers performance. These two parameters, sensitivity or rate of actual positives correctly identified as such (denoted by S_n) and specificity or rate of actual negatives correctly identified (denoted by S_p), are defined as:

$$S_n = \frac{TP}{TP + FN} \tag{23}$$

$$S_p = \frac{TN}{TN + FP} \tag{24}$$

It is worth noting that the use of the first two parameters (P_0, P_1) is particularly suitable for earthquake prediction as they satisfy the prediction criterion provided by several authorities [3,97].

Also note that several empirical rules, derived from a 17-year work in the University of Santiago's National Center of Environment [57,65], are used to accurately train the networks. These rules are:

1. P_0 and P_1 values greater than 95% in the training set involves overlearning, that is, the ANN may not be capable of generalizing results. This effect can be avoided by modifying this set repeating or adding new vectors to the original training set until reaching satisfactory probabilities, but never greater than 95%.
2. The lower P_0 value in the training set must be greater than 70%. If this value is not reached, the training set has to be modified until reaching acceptable values.
3. Sometimes, the ANN may not be able to predict particularly low or high values in the training set (probability = 0/0). This situation can be avoided whether adding vectors with very low or very high outputs.

5.2. ANN for Talca

The time interval used to train this ANN was from June 19th 2003 to March 21st 2010. However, when the selftest was carried out (to test using the training set), P_1 was equal to 0%, meaning that the ANN was unable to generate high values of y_i. This issue was solved by adding three times to the training set the 20 vectors with higher values (magnitude larger or equal to 4.2) and training for 500 epochs. This new situation allowed the ANN to obtain satisfactory P_1 values during the selftest, allowing thus the learning of situations when the ANN's output was activated.

The values of the quality parameters obtained during the training of this ANN are summarized in Table 3. Note that when the output reached values less than 3.0, these values were arbitrarily substituted by 2.9, since no values less than 3.0 are present in the dataset.

Table 4
Training values for Pichilemu.

Parameters	Value
TP	56
TN	39
FP	11
FN	16
P_1	70.9%
P_0	83.6%
S_n	77.8%
S_p	78.0%
Average	77.6%

Table 5
Training values for Santiago.

Parameters	Value
TP	10
TN	94
FP	2
FN	26
P_1	83.3%
P_0	78.3%
S_n	27.8%
S_p	97.9%
Average	71.8%

5.3. ANN for Pichilemu

The time interval used to train this ANN was from August 10th 2005 to March 31st 2010. In this area, the vectors presented a high logic coherence, as evidenced the results obtained after 500 epochs. Such results are shown in Table 4.

5.4. ANN for Santiago

The time interval used to train this ANN was from May 13th 2003 to June 2nd 2004. After the first 500 training epochs, P_1 was equal to 0%. This problem was overcome by adding 10 linear independent vector, registered before May 13th 2003, and 10 repeated vectors from the original set with a particular feature: they represented the maximum possible values registered of y_i. The results obtained after the modification of the original training set are shown in Table 5.

5.5. ANN for Valparaíso

The time interval used to train this ANN was from 31st January 2006 to December 19th 2008. The training results are shown in Table 6.

6. Results

This section shows the results obtained when the proposed ANN's were applied to the test sets representing the four

Table 6
Training values for Valparaíso.

Parameters	Value
TP	15
TN	77
FP	5
FN	25
P_1	75.0%
P_0	75.5%
S_n	37.5%
S_p	93.9%
Average	70.5%

seismogenic Chilean zones analyzed. These sets can be downloaded from the University of Chile's National Service of Seismology [55].

First, the type of predictions performed by the ANN are introduced. Then, the results for every area are summarized in terms of the quality parameters described in Section 5.1. Finally, the last subsection is devoted to discuss the results achieved.

Note that, to the authors' knowledge, there is no other approach capable of predicting earthquake using only the information analyzed in this work. However, for the sake of comparison, the authors have turned the prediction problem into a classification one, so that well-known classifiers can be applied and compared. Note that this conversion has already been used in [1]. To achieve this goal, a new attribute has been added: A label has been assigned to every sample (composed of the attributes described in Section 4.2) indicating whether an earthquake with magnitude larger than T occurred within the following seven days or not. That is, the values represented by the label are 1 if such an earthquake occurred, and 0 if not.

The methods selected for the comparative analysis are K-nearest neighbors (KNN) [13], support-vector machines (SVM) [12], and classification via K-means clustering algorithm [44]. All these methods are implemented on the Weka platform [59] and a guide to optimally set these classifiers can be found in [96]. These methods have been chosen because they cover the different classification strategies based on machine learning.

6.1. Type of predictions

The ANN's designed are capable of providing two kind of predictions:

1. There is a probability $P_{1'}$ that an earthquake occurs with M_s magnitude larger or equal to T, during the next five days, in the area subjected to analysis.
2. There is a probability P_1'' that an earthquake occurs with magnitude in interval $M_s \in [T, E]$, during the next five days, in the area subjected to analysis.

T denotes the *threshold* that activates the ANN's output, calculated as the addition of the mean value of the 122 earthquakes observed during the training period and 0.6 times its standard deviation:

$$T = \text{mean}(M_s) + 0.6\sigma_{M_s} \tag{25}$$

where M_s is the set of magnitudes of the earthquakes in the training set.

E could be arbitrarily selected and represents the *end* PDF's cut-off value. However, considering the social impact that a false alarm may cause, the value for E chosen is the maximum y_i value observed during the training period Eq. (26) shows this value:

$$E = \max(y_i) \quad \text{such that} \quad y_i \in \text{Training period} \tag{26}$$

P_1' can be calculated according to two different strategies. First, P_1' can be the observed P_1 value during the training set. Alternatively, if the historical data the area under study was long enough (if many earthquakes were available to construct the system), it would be desirable to construct several ANN's, each one considering at least 122 training vectors chronologically ordered. That is, if the historical data is such that the number of training vectors is much greater than 122, consecutive ANN's are constructed using only 122 training vectors each (i.e. first ANN would use the first 122 training vectors, second ANN vectors 123–244, and so on), leading to robust estimations [58,65]. For this second situation:

$$P_1' = QP_1^{\text{CTS}} \tag{27}$$

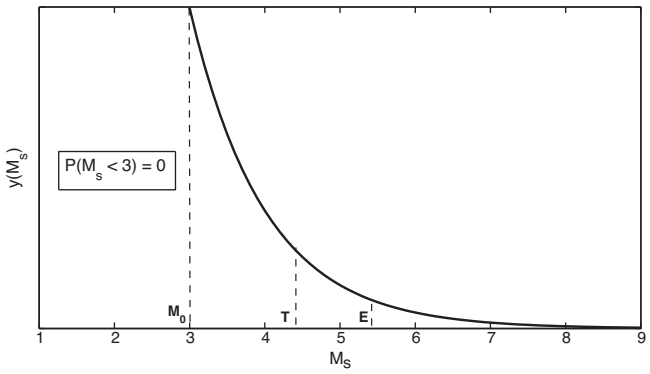

Fig. 4. Gutenberg–Richter's law. Obtention of the probability that an earthquake with $M_s \in [T, E]$ occurs from the PDF.

where P_1^{CTS} is the P_1 associated with the current training set, and Q is the average quotient of the observed values of P_1 in the n consecutive ANN's constructed before the current one is analyzed (containing the $122n$ training vectors occurred before the 122 ones selected to train the current ANN). Formally:

$$Q = \frac{1}{n}\sum_{i=1}^{n} \frac{P_{\text{test}}^{(i)}}{P_{\text{training}}^{(i)}} \tag{28}$$

Unfortunately, the datasets analyzed (periods 2001–2011) do not meet the required condition in any area, therefore $P_1' = P_1$ for all the four areas.

Fig. 4 illustrates how to calculate P_1'' from the Gutenberg–Richter's law. Mathematically, P_1'' can be formulated as:

$$P''_1 = P_{1'}P \tag{29}$$

where P is:

$$P = \frac{\int_T^E dM_s y(M_s)}{\int_T^\infty dM_s y(M_s)} \tag{30}$$

6.2. Earthquake prediction in Talca

For this cell, the training set contained the 122 linearly independent vectors occurred from June 19th 2003 to March 21st 2010. Analogously, the test set included the 45 vectors generated from March 24th 2010 to 4th January 2011.

Table 7 shows the results in terms of the quality parameters described in Section 5.1 as well as the results obtained by the methods the ANN is compared to. Note that both KNN and SVM were unable to distinguish between earthquake and no earthquake, in other words, they classified all the samples as absence of earthquake (obviously, the specificity is then 100%). By contrast, the K-means classified all the samples in the test set as earthquake (obviously, the sensitivity is then 100%). The dash means that the value could not be calculated as a division by zero was encountered.

Table 7
ANN's performance for Talca.

Parameters	ANN	KNN	SVM	K-means
TP	3	0	0	8
TN	23	37	37	0
FP	14	0	0	37
FN	5	8	8	0
P_1	17.6%	–	–	17.8%
P_0	82.1%	82.2%	82.2%	–
S_n	37.5%	0.0%	0.0%	100%
S_p	62.2%	100%	100%	0.0%
Average	49.86%	–	–	–

Table 8
ANN performance for Pichilemu.

Parameters	ANN	KNN	SVM	*K*-means
TP	13	20	0	16
TN	91	38	93	84
FP	2	55	0	9
FN	16	9	29	13
P_1	86.7%	26.7%	–	64.0%
P_0	85.0%	80.9%	76.2%	86.6%
S_n	52.0%	69.0%	0.0%	55.2%
S_p	97.8%	40.9%	100%	90.3%
Average	78.6%	54.3%	–	74.0%

P_1 = 17.6% might be understood as an unsatisfactory result. However, it is important to highlight that in the seismology field logarithmic scales are usually applied in earthquake occurrence probabilities [56].

6.3. Earthquake prediction in Pichilemu

For this cell, the training set contained the linearly independent vectors occurred from August 10th 2005 to March 31st 2010. Analogously, the test set included the vectors generated from April 1st 2010 to October 8th 2011.

Both sets contain 122 vectors. However, every new earthquake deeply affects the soil geometry in this area and, therefore, to predict the earthquake 123 the ANN had to be retrained. In particular, the 122 first vectors of the initial test set were used as a new training set for earthquake 123 and successive earthquakes. In practical terms the prediction of a maximum of 50 earthquakes is advised for further research work. Anyway, this paper is focused on studying the extreme situation, where the test set was composed of 122 vectors.

Table 8 shows the results in terms of the quality parameters described in Section 5.1, as well as the results obtained by the other machine learning classifiers. As occurred in Talca, the SVM classified all samples as no earthquake. KNN obtained poorer results than those of the ANN. As for the results of *K*-means, they were slightly worse than those of the ANN on average, however the number of FP was 9, more than four times greater than in then ANN.

6.4. Earthquake prediction in Santiago

For this cell, the training set contained the 122 linearly independent vectors occurred from May 13th 2003 to June 2nd 2004. Analogously, the test set included the 122 vectors generated from June 23rd 2004 to 16th January 2006.

Table 9 shows the results in terms of the quality parameters described in Section 5.1. The results of the application of the other classifiers is also summarized in this table. Again, the SVM was unable to distinguish between earthquake and no earthquake, and in particular, this time classified all the samples as earthquake. On the other hand, the ANN, KNN and *K*-means obtained similar results on average (KNN slightly worse and *K*-means slightly better).

Table 9
ANN performance for Santiago.

Parameters	ANN	KNN	SVM	*K*-means
TP	5	6	0	7
FN	101	93	108	95
FP	7	15	14	13
TN	9	8	0	7
P_1	41.7%	28.6%	0.0%	35.0%
P_0	91.8%	92.1%	100%	93.1%
S_n	35.7%	42.9%	–	50.0%
S_p	93.5%	86.1%	88.5%	88.0%
Average	65.7%	62.4%	–	66.5%

Table 10
ANN performance for Valparaíso.

Parameters	ANN	KNN	SVM	*K*-means
TP	20	30	35	43
TN	59	52	46	0
FP	3	10	16	62
FN	24	14	9	1
P_1	87.0%	75.0%	68.6%	41.0%
P_0	71.1%	78.8%	83.6%	0.0%
S_n	45.5%	68.2%	79.5%	97.7%
S_p	95.2%	83.9%	74.2%	0.0%
Average	74.7%	76.5%	76.5%	34.7%

However, to reach these values both methods reported twice more FP's than the ANN did, which was one of the main objectives set at the beginning: To have the smallest number of FP's possible.

6.5. Earthquake prediction in Valparaíso

For this cell, the training set contained the 122 linearly independent vectors occurred from 31st January 2006 to December 19th 2008. Analogously, the test set included the 106 vectors generated from December 19th 2008 to February 10th 2011.

Table 10 shows the results in terms of the quality parameters described in Section 5.1. The results derived from the application of other classifiers is also reported. This time was the *K*-means who showed its inability to distinguish between both situations. In particular it also classified one sample as no upcoming earthquake and it did it erroneously. On the contrary, KNN and SVM exhibited slightly better behavior on average. But again, when examining the FP's they obtained values much larger than those of the ANN.

6.6. Discussion of the results

The high values of P_0 and P_1 obtained for all the zones indicate that the input variables were, indeed, strongly correlated with the observed magnitude in a near future. The ANN's were capable of indirectly learning Omori/Utsu and Gutenberg–Richter's laws, confirming thus the great ability these techniques have in the seismology field [69]. This fact confirms that the choice of such input vectors was adequate.

With reference to the specificity, all the zones obtained values especially high (87.2% on average). This fact is of the utmost significance, as it is extremely important not to activate false alarms in seismology due to the social impact they may cause. The sensitivity achieved is, in general, not so high (40.9% on average). However, the scientific community accepts the fact of missing one-level hits [9] and such a sensitivity can be considered successful considering the application domain and the rest of the results.

From all the networks constructed, the one applied to Pichilemu was the easiest to construct and, also, the one with better performance for both training and test sets. This fact can be understood as that for Pichilemu the causality between one earthquake and the next one is greater than in the remaining areas. Following with a geological interpretation, it is probable that the subsoil configuration is such that the occurrence of seisms cracks the soil causing new seisms in the near surroundings. This could be the reason for the area of Pichilemu to be more predictable than the other three ones.

Regarding the comparison made to other methods, several conclusions can be drawn. First of all, none of them obtained better results in more than one area and when this fact happened, the results were just slightly better on average and triggered much more FP's. Secondly, there were many situations in which the algorithms classified all the samples into a same class. Finally, the results obtained could be considered satisfactory in other contexts

as some of them provides consistent results. This means that the input parameters have been properly chosen, providing thus one satisfactory solution for one of the most critical problems in earthquake prediction.

On the other hand, systems with memory whose previous states influence next ones (such as twisters, stock market or atmospheric pollution) present a fractal structure due to the inherent feedback in the variables causing these phenomena. In this sense, Omori/Utsu and Bath's laws reveal that aftershocks cause more aftershocks. That is, the land is characterized by feedback loops between seisms, geometrical alterations and cracks that affect its state. It can be concluded, then, that time series of earthquakes must present a fractal structure [46].

Therefore the authors wondered if, from a single chronologically ordered series of quakes, it could be mathematically induced that the earthquakes in Pichilemu were more predictable than in other areas. It is well-known that the fractal dimension of a time series expresses the degree of causal interdependence between one datum and another [41]. Moreover, the inverse fractal dimension of a time series, the so-called Hurst exponent (H), relates autocorrelations and lag pairs in time series. The application of H to the four areas revealed that Pichilemu has a value notably higher (Talca: $H = 0.0037$, Pichilemu: $H = 0.1580$, Santiago: $H = 0.0729$, Valparaíso: $H = 0.0538$), which supports the initial assumption and encourages further research in this way.

7. Additional remarks

This section is devoted to discuss the robustness of the results obtained, providing a discussion about possible earthquake clustering and analyzing different time intervals.

7.1. About the influence of earthquake clustering

One may think that the high performance of the designed ANN's is due to earthquake clustering influence. Indeed, Kagan [33] stated: "Any prospective earthquake prediction technique needs to demonstrate that its success is not due to the influence of earthquake clustering (foreshock–mainshock–aftershock sequences)." On the other hand, earthquakes are inherently presented in clusters. Any model that ignores this fact could be, therefore, pointed as incomplete. To this regard, Zechar and Jordan [97] claimed that: "Earthquakes cluster in space and time, and therefore, any forecast that captures this clustering behavior should outperform a uniform reference model."

In this section, the hypotheses obtained by comparing the results described in Section 6 with those obtainable by means of the null hypothesis is discussed. To complete this task a two-step strategy has been used:

1. First, the Poisson probability for each hit supposing pure chance is calculated.
2. Second, the probability of obtaining a result better than that of summarized in Tables 7–10 is calculated, using the probability obtained in the first step.

Even if there is evidence that the Poisson distribution is not ideal for describing earthquake number variation [35,76], such reference model has been used as done by Zechar and Zhuang [98], due to "its uniform applicability to all alarms and because of its simplicity-namely, it includes few assumptions and it is easily interpreted".

Actually, the authors listed 29 RTP alarms in this paper and they accepted as one-level hits statements with a maximum Poisson probability of 29.12%. As shown in Table 11, the probabilities reported by the constructed ANN's are below this threshold. The following summarizes the Poisson probability calculations using only the information prior to the test set, and taking into consideration that the horizon of prediction was set to five days:

Table 11
Influence of earthquake clustering in the proposed models.

Parameter	Talca	Pichilemu	Santiago	Valparaíso
h	3	13	5	20
N	8	29	14	44
P_A	0.049	0.080	0.233	0.100
P	$5.47 \cdot 10^{-3}$	$1.09 \cdot 10^{-7}$	0.210	$1.61 \cdot 10^{-9}$

1. *Talca.* 25 one-level events (or target earthquakes), in a period of 2467 days. Therefore, the Poisson probability within five days is 4.9%.
2. *Pichilemu.* 28 one-level events (or target earthquakes), in a period of 1694 days. Therefore, the Poisson probability within five days is 8.0%.
3. *Santiago.* 19 one-level events (or target earthquakes), in a period of 386 days. Therefore, the Poisson probability within five days is 23.3%.
4. *Valparaíso.* 22 one-level events (or target earthquakes), in a period of 1053 days. Therefore, the Poisson probability within five days is 10.0%.

Now, the probability of obtaining results better than the hit rates presented in Tables 7–10 is calculated, by means of the technique introduced in [97]. This probability is calculated as follows:

$$P = \sum_{n=h}^{N} B(n|P_A) \tag{31}$$

where h is the number of hits, N is the number of target earthquakes, P_A is the true probability of occurrence of an earthquake exceeding the threshold in the area of study, and $B(n|P_A)$ is the binomial distribution, calculated as:

$$B(n|P_A) = P_A^U (1 - P_A)^{N-u} \tag{32}$$

Given this equation, Table 11 summarizes the parameters obtained for every area.

Although the probability in Santiago is 21.0%, the probability associated with other areas is very low ($p \ll 1$). Therefore, in all cases the hypothesis most likely to be correct is not the null hypothesis, but the one that states that the rates achieved by the ANN's are superior to pure chance. Note that even if the null hypothesis is not rejected with a small probability (typically 5% in sociological research [18] or 1% in biomedical research [72]), this does not mean it is true, as thoroughly discussed by Dimer de Oliveira [14].

7.2. About the use of different time intervals

Again, one may think that the results shown are the best ones found from a large dataset. This section is therefore to prove the robustness of the ANN's built over time, discussing the results obtained for different time intervals.

As discussed in Section 6.1, the generalization ability for any ANN decreases over time, fact reflected in the existence of the Q factor – see Eq. (28). Consequently, the synaptic weights have to be periodically recalculated so that the ANN is always updated, since ANN's can only be applied to predict the same number of items forming the training set, considering this number as its upper bound. In other words, the test set can contain maximum the same number of elements existing in the training set as shown, for instance, in [79].

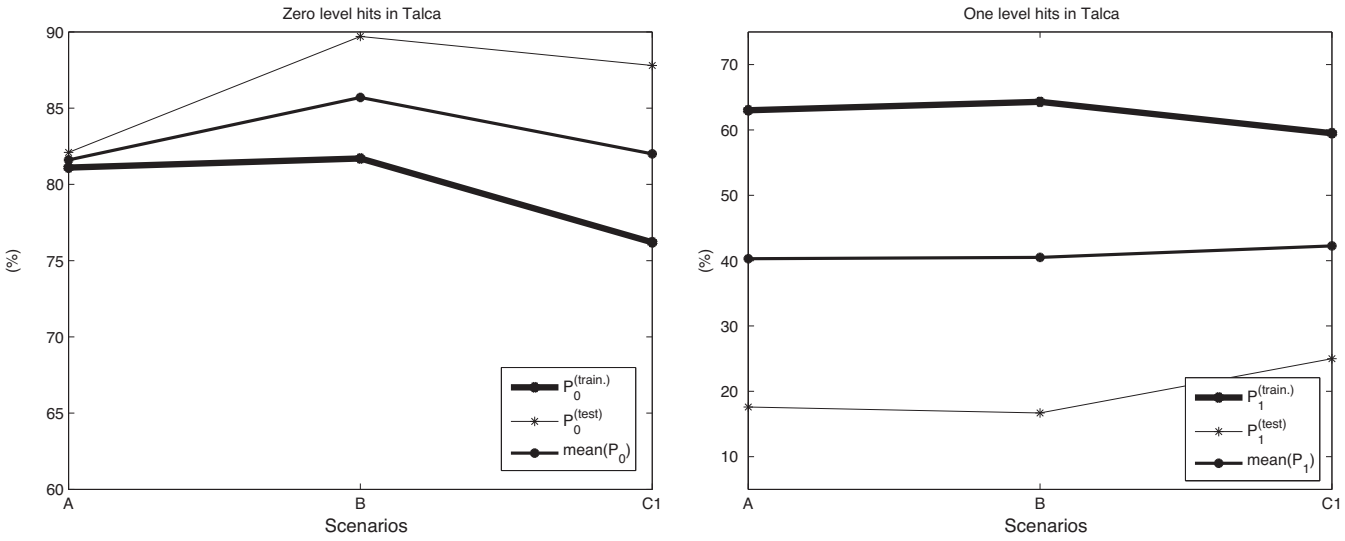

Fig. 5. Variations of P_0 and P_1 in Talca, for different scenarios.

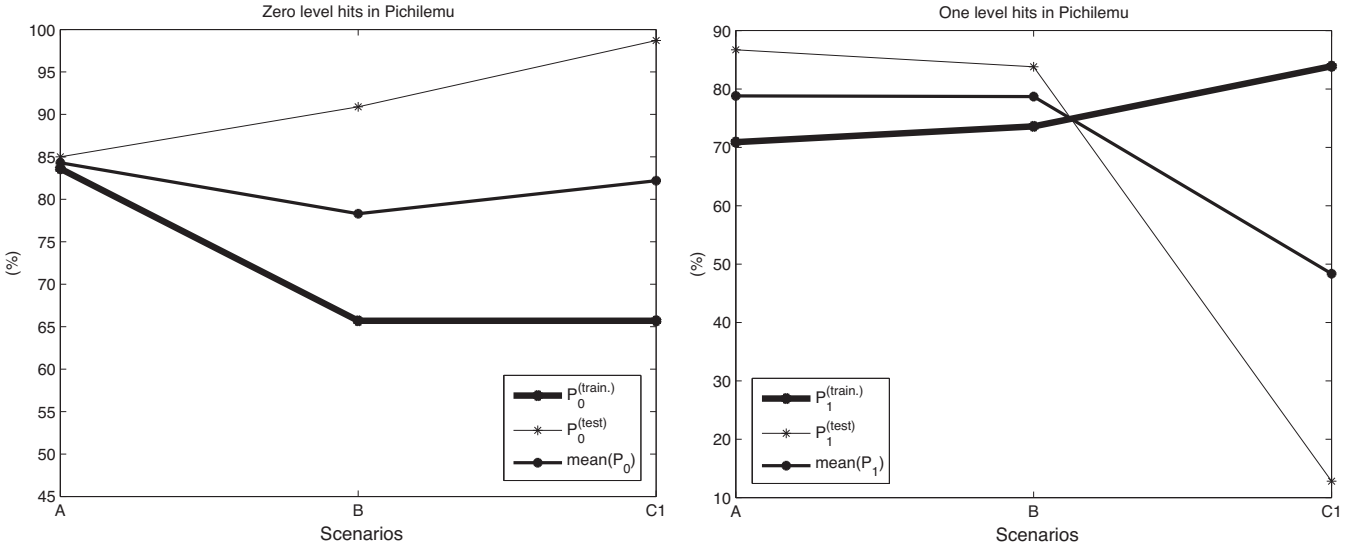

Fig. 6. Variations of P_0 and P_1 in Pichilemu, for different scenarios.

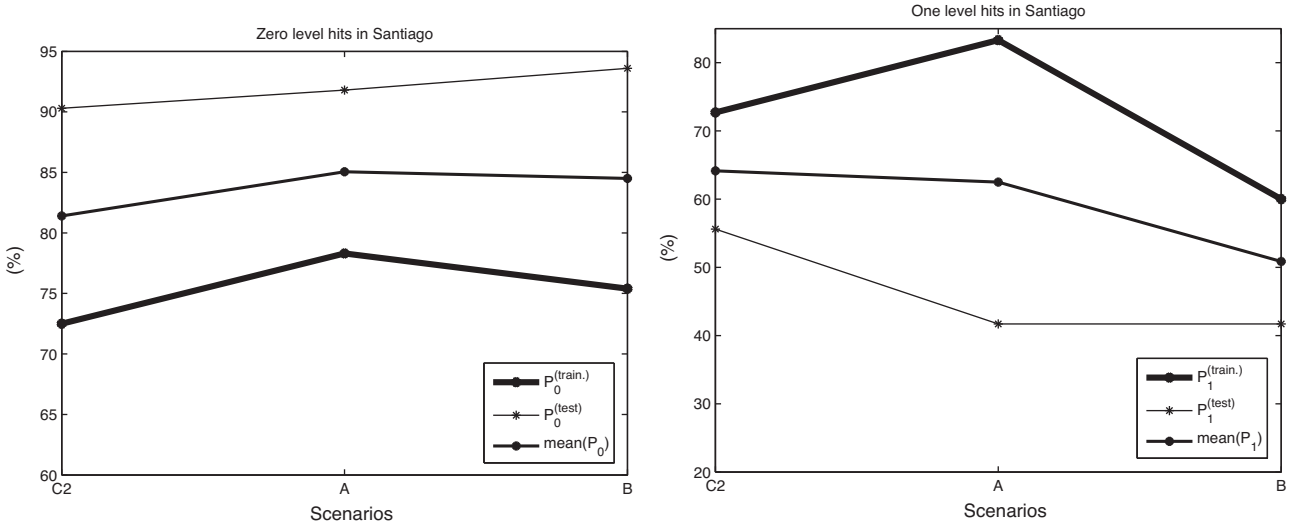

Fig. 7. Variations of P_0 and P_1 in Santiago, for different scenarios.

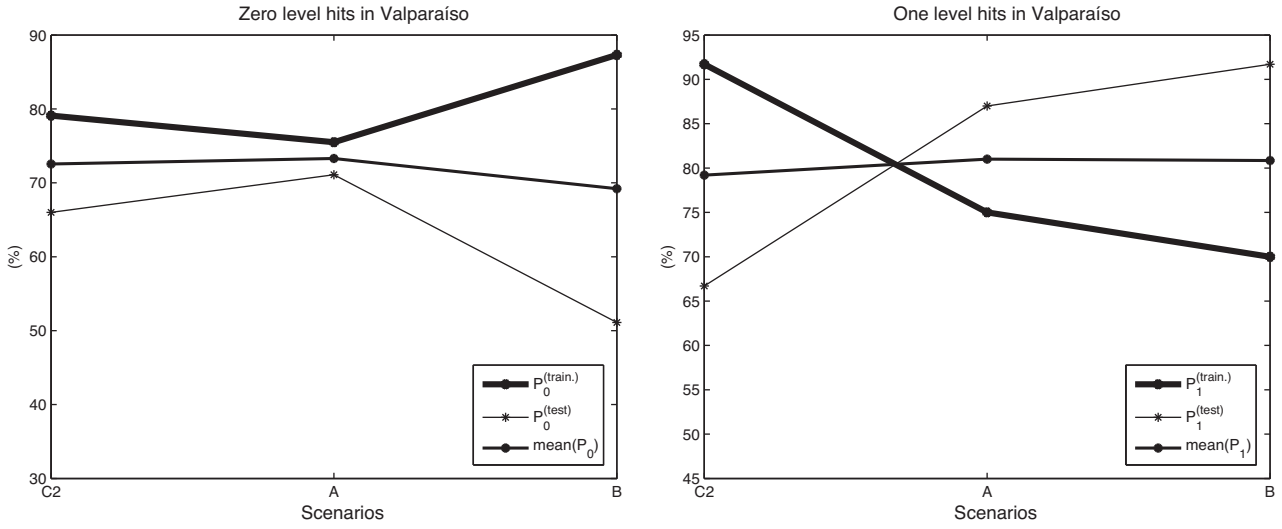

Fig. 8. Variations of P_0 and P_1 in Valparaíso, for different scenarios.

On the other hand, a work based on the application of ANN's to predict atmospheric pollution [65] used different synaptic weights for a same architecture to predict different time intervals (see Tables 1–5 of this paper for further information).

The authors want to highlight that this standard weights updating are not related with the possibility of being tuning the model in retrospective experiments to fit the data, as training and test sets are not overlapped in any element (the intersection of both sets is $\varnothing$).

To show the robustness of the architecture designed, the ANN's are applied to three different time intervals. To make such a comparison, the original dataset was updated and extended so that the ANN's could be applied to new events: The new dataset time interval is from 1st January 2001 to March 14th 2012. The area subjected to analysis with less earthquakes is Talca (266), which led to the definition of three new scenarios:

1. *Scenario A*. The original scenario described in Section 6.
2. *Scenario B*. The new training set was slid 15 earthquakes from the original one forward.
3. *Scenario C_1*. The new training set was slid 30 earthquakes from the original one forward.
4. *Scenario C_2*. The new training set was slid 15 earthquakes from the original one backward.

The use of the scenarios C_1 or C_2 depends on the number of earthquakes found when the dataset was extended. As the dataset could not be slid 30 earthquakes forward in Santiago and Valparaíso, the scenario C_2 was chosen for these two areas and, contrarily, scenario C_1 was selected for both Talca and Pichilemu.

Obviously, the ANN's architectures did not change and the synaptic weights were calculated using only the information contained in the training sets (no information from the test sets were used to adjust the model), using the same restrictions described in Section 5.

Figs. 5–8 show the performance of the ANN's for the three scenarios described above. The parameters depicted on them mean:

1. $P_0^{(\text{train.})}$ is the P_0 value obtained in the training set.
2. $P_0^{(\text{test})}$ is the P_0 value obtained in the test set.
3. $\text{mean}(P_0) = (P_0^{(\text{train.})} + P_0^{(\text{test})})/2$
4. $P_1^{(\text{train.})}$ is the P_1 value obtained in the training set.

Table 12
P_0 values for different periods of time.

Scenario	Talca	Pichilemu	Santiago	Valparaíso	Mean (%)
A	82.1%	85.0%	91.8%	71.1%	82.5
B	89.7%	90.9%	93.6%	51.1%	81.3
C_1	87.8%	98.7%	–	–	93.3
C_2	–	–	90.3%	66.0%	78.2

Table 13
P_1 values for different periods of time.

Scenario	Talca	Pichilemu	Santiago	Valparaíso	Mean (%)
A	17.6%	86.7%	41.7%	87.0%	58.3
B	16.7%	83.8%	41.7%	91.7%	58.5
C_1	25.0%	12.8%	–	–	18.9
C_2	–	–	55.6%	66.7%	61.2

5. $P_1^{(\text{test})}$ is the P_1 value obtained in the test set.
6. $\text{mean}(P_1) = (P_1^{(\text{train.})} + P_1^{(\text{test})})/2$

From all these figures, it can be concluded that the ANN's presented robust and consistent results regardless of the time interval used to assess their accuracy. Nevertheless, Tables 12 and 13 are also provided to quantify these values for all the three studied scenarios.

As it can be observed, P_0 values remain almost constant in all scenarios, even exhibiting some degree of improvement on average when the training and test sets were slid forward. A similar situation occurs with P_1 values, except for the scenario C_1 in Pichilemu. However, this fact is partly due to the inclusion of the aftershocks occurred after the 8.8M_w earthquake in February 27th 2010. This involves that the threshold set during the training stage was much bigger than earthquakes reported in its associated test set – Omori/Utsu law, Eq. (8). It is also possible that the uncertainty in the training is high, which may negatively influence the results as stated by Wang [90]: "the uncertainties decrease with time, except that they increase temporarily after large earthquakes."

8. Conclusions

Three-layer feedforward artificial neural networks have been constructed to predict earthquakes in four seismogenic areas in Chile. These networks included as input parameters information

related to the *b*-value, the Bath's law and the Omori/Utsu's law. As output, the networks were able to predict the occurrence of earthquakes for a five-day horizon with great reliability, confirming thus the adequacy of using the combination of these parameters as seismicity indicators. The robustness of the method has been assessed by sliding the training sets several earthquakes both forward and backward as well as comparing the results with other well-known machine learning classifiers. Since the application of ANN's has not been exploited in this field, it is expected that the results obtained serve as a seed for further research. As for future work, the authors propose the application of Hurst's exponent to a priori determine the predictability of earthquake's series.

Acknowledgements

The authors want to thank TGT-www.geofisica.cl for the support through grants number 2210 and 2211. The financial support given by the Spanish Ministry of Science and Technology, projects BIA2004-01302 and TIN2011-28956-C02-01 are equally acknowledged. This work has also been partially funded by a JPI 2012 Banco Santander's grant. Finally, the authors want to thank the reviewers for their careful reviews and helpful suggestions.

References

[1] H. Adeli, A. Panakkat, A probabilistic neural network for earthquake magnitude prediction, Neural Networks 22 (2009) 1018–1024.
[2] K. Aki, Maximum likelihood estimate of b in the formula $\log N = a - bM$ and its confidence limits, Bulletin Earthquake Research Institute 43 (1965) 237–239.
[3] C.R. Allen, Responsibilities in earthquake prediction, Bulletin of the Seismological Society of America 66 (1982) 2069–2074.
[4] E.I. Alves, Earthquake forecasting using neural networks: results and future work, Nonlinear Dynamics 44 (1–4) (2006) 341–349.
[5] A. Ayele, O. Kulhanek, Spatial and temporal variations of seismicity in the horn of Africa from 1960 to 1993, Geophysical Journal International 130 (1997) 805–810.
[6] M. Bath, Lateral inhomogeneities in the upper mantle, Tectonophysics 2 (1965) 483–514.
[7] P. Bird, Z. Liu, Seismic hazard inferred from tectonics: California, Seismological Research Letters 78 (1) (2007) 37–48.
[8] H. Carroll-Talley, W.K. Cloud, United States Earthquakes 1960, U.S. Department of Commerce, Coast and Geodectic, Washington D.C., 1960.
[9] E. Cartlidge, Quake experts to be tried for manslaughter, Science 332 (6034) (2011) 1135–1136.
[10] R. Console, M. Murru, F. Catalli, Physical and stochastic models of earthquake clustering, Tectonophysics 417 (2006) 141–153.
[11] R. Console, M. Murru, F. Catalli, G. Falcone, Real time forecasts through an earthquake clustering model constrained by the rate-and-state constitutive law: comparison with a purely stochastic ETAS model, Seismological Research Letters 78 (1) (2007) 49–56.
[12] C. Cortes, V. Vapnik, Support-vector networks, Machine Learning 20 (3) (1995) 273–297.
[13] T.M. Cover, P.E. Hart, Nearest neighbor pattern classification, IEEE Transactions on Information Theory 13 (1) (1967) 21–27.
[14] F. Dimer de Oliveira, Can we trust earthquake cluster detection tests? Geophysical Research Letters 39 (2012) L17305.
[15] J.E. Ebel, D.W. Chambers, A.L. Kafka, J.A. Baglivo, Non-Poissonian earthquake clustering and the hidden Markov model as bases for earthquake forecasting in California, Seismollogical Research Letters 78 (1) (2007) 57–65.
[16] M. Erdik, Y. Fahjan, O. Ozel, H. Alcik, A. Mert, M. Gul, Istanbul earthquake rapid response and the early warning system, Bulletin of Earthquake Engineering 1 (2003) 157–163.
[17] E.H. Field, Overview of the working group for the Development of Regional Earthquake Likelihood Models (RELM), Seismological Research Letters 78 (1) (2007) 7–16.
[18] A. Gelman, D. Weakliem, Of beauty, sex, and power: statistical challenges in estimating small effects, American Scientist 97 (2008) 310–316.
[19] M.C. Gerstenberguer, L.M. Jones, S. Wiemer, Short-term aftershock probabilities: case studies in California, Seismological Research Letters 78 (1) (2007) 66–77.
[20] S.J. Gibowitz, Frequency–magnitude depth and time relations for earthquakes in Island Arc: North Island, New Zealand, Tectonophysics 23 (3) (1974) 283–297.
[21] S. Gross, C. Kisslinger, Estimating tectonic stress rate and state with Landers aftershock, Journal of Geophysical Research 102 (B4) (1994) 7603–7612.
[22] Council Working Group. Earthquake Research at Parkfield, California, for 1993 and Beyond – National Earthquake prediction Evaluation. Technical report, 1116, 14 pages. United States Government Printing Office, 1994.
[23] G. Grunthal, E. Hurtig, E. Ruge, Time dependence of statistical parameters: the aftershock sequence of the Friuli, northern Italy, 1976 earthquake and a section of the Montenegro, Yugoslavia, earthquake series 1979, Earthquake Prediction Research 2 (1982) 275–285.
[24] B. Gutenberg, C.F. Richter, Seismicity of the Earth, Geological Society of America, Special Paper 34 (1941) 1–131.
[25] B. Gutenberg, C.F. Richter, Frequency of earthquakes in California, Bulletin of the Seismological Society of America 34 (1944) 185–188.
[26] B. Gutenberg, C.F. Richter, Seismicity of the Earth, Princeton University, Princeton, 1954.
[27] S. Hainzl, D. Marsan, Dependence of the Omori–Utsu law parameters on mainshock magnitude: observations and modeling, Journal of Geophysical Research 113 (B10309) (2008) 12.
[28] R. Hecht-Nielsen, Kolmogorov's mapping neural network existence theorem, in: IEEE International Conference on Neural Networks, 1987, pp. 11–14.
[29] A. Helmstetter, Y.Y. Kagan, D.D. Jackson, High-resolution time-independent grid-based forecast for $M = 5$ Earthquakes in California, Seismological Research Letters 78 (1) (2007) 78–86.
[30] J.R. Holliday, C.C. Chen, K.F. Tiempo, J.B. Rundle, D.L. Turcotte, A. Donnellan, A RELM earthquake forecast based on pattern informatics, Seismological Research Letters 78 (1) (2007) 87–93.
[31] J.J. Hopfield, Neural networks and physical systems with emergent collective computational abilities, Proceedings of the National Academy of Sciences 79 (1982) 2554–2558.
[32] M. Ishimoto, K. Iida, Observations sur les seismes enregistrés par le microsismographe construit dernièrement, Bulletin Earthquake Research Institute 17 (1939) 443–478.
[33] Y.Y. Kagan, VAN earthquake predictions – an attempt at statistical evaluation, Geophysical Research Letters 23 (11) (1996) 1315–1318.
[34] Y.Y. Kagan, Seismic moment distribution revisited: statistical results, Geophysical Journal International 148 (2002) 520–541.
[35] Y.Y. Kagan, Statistical distributions of earthquake numbers: consequence of branching process, Geophysical Journal International 180 (2010) 1313–1328.
[36] Y.Y. Kagan, D.D. Jackson, Long-term probabilistic forecasting of earthquakes, Journal of Geophysical Research 99 (13) (1994) 685–700.
[37] Y.Y. Kagan, D.D. Jackson, Y. Rong, A testable five-year forecast of moderate and large Earthquakes in Southern California based on smoothed seismicity, Seismological Research Letters 78 (1) (2007) 94–98.
[38] J.L. Kirschving, Earthquake prediction by animals: evolution and sensory perception, Bulletin of the Seismological Society of America 90 (2000) 312–323.
[39] A.N. Kolmogorov, On the representation of continuous functions of several variables by superposition of continuous functions of one variable and addition, Doklady Akademii Nauk SSSR 114 (1957) 359–373.
[40] K. Lee, W.S. Yang, Historical seismicity of Korea, Bulletin of the Seismological Society of America 71 (3) (2006) 846–855.
[41] M. Li, Fractal time series – a tutorial review, Mathematical Problems in Engineering ID157264 (2010) 26.
[42] L.A. Lomnitz, Major Earthquakes of Chile: a historical survey, 1535–1960, Seismological Research Letters 75 (3) (2004) 368–378.
[43] S. Lorito, F. Romano, S. Atzori, X. Tong, A. Avallone, J. McCloskey, M. Cocco, E. Boschi, A. Piatanesi, Limited overlap between the seismic gap and coseismic slip of the great 2010 Chile Earthquake, Nature Geoscience 4 (2011) 173–177.
[44] J. MacQueen, Some methods for classification and analysis of multivariate observations, in: Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability, 1968, pp. 281–297.
[45] R. Madahizadeh, M. Allamehzadeh, Prediction of aftershocks distribution using artificial neural networks and its application on the May 12, 2008 Sichuan Earthquake, Journal of Seismology and Earthquake Engineering 11 (3) (2009) 111–120.
[46] B. Mandelbrot, Fractals: Form, Chance and Dimension, WH Freeman and Co., New York, 1977.
[47] F. Martínez-Álvarez, A. Troncoso, A. Morales-Esteban, J.C. Riquelme, Computational intelligence techniques for predicting earthquakes, Lecture Notes in Artificial Intelligence 6679 (2) (2011) 287–294.
[48] W. Marzocchi, J.D. Zechar, Earthquake forecasting and earthquake prediction: different approaches for obtaining the best model, Seismological Research Letters 82 (3) (2011) 442–448.
[49] T. Matsuzawa, T. Igarashi, A.A. Hasegawa, Characteristic small-earthquake sequence off Sanriku, northeastern Honshu, Japan, Geophysical Research Letters 29 (11) (2002) 1543–2147.
[50] A. Morales-Esteban, F. Martínez-Álvarez, A. Troncoso, J.L. de Justo, C. Rubio-Escudero, Pattern recognition to forecast seismic time series, Expert Systems with Applications 37 (12) (2010) 8333–8342.
[51] R.M. Musson, P.W. Winter, Seismic hazard maps for the UK, Natural Hazards 14 (1997) 141–154.
[52] C. Narteau, P. Shebalin, M. Holschneider, Temporal limits of the power law aftershock decay rate, Journal of Geophysical Research 107 (B12) (2002) 12.
[53] P. Nuannin. The potential of *b*-value variations as earthquake precursors for small and large events. Technical report, 183. Uppsala University, Sweden, 2006.
[54] P. Nuannin, O. Kulhanek, L. Persson, Spatial and temporal *b* value anomalies preceding the devastating off coast of NW Sumatra earthquake of December 26, 2004, Geophysical Research Letters 32 (2005).
[55] University of Chile. National Service of Seismology. http://ssn.dgf.uchile.cl/seismo.html

[56] US Deparment of Interior. US Geological Survey Website. http://earthquake.usgs.gov, 2011.
[57] University of Santiago. National Centre of Environment. http://www.geofisica.cl/macam/cma_usach.htm
[58] University of Santiago de Chile. Web service: Online PM10 forecasting. http://fisica.usach.cl/pronostico/index.htm
[59] WEKA – The University of Waikatu. Data mining with open source machine learning software in Java. http://www.cs.waikato.ac.nz/ml/weka/
[60] E.A. Okal, B.A. Romanovicz, On the variation of *b*-value with earthquake size, Physics of the Earth and Planetary Interiors 87 (1994) 55–76.
[61] F. Omori, Macroseismic measurements in Tokyo, II and III, Earthquake Investigation Communications 11 (1902) 1–95.
[62] International Commission on Earthquake Forecasting for Civil Protection, Operational earthquake forecasting. State of knowledge and guidelines for utilization, Annals of Geophysics 54 (4) (2011) 21–27.
[63] A. Panakkat, H. Adeli, Neural network models for earthquake magnitude prediction using multiple seismicity indicators, International Journal of Neural Systems 17 (1) (2007) 13–33.
[64] A. Panakkat, H. Adeli, Recurrent neural network for approximate earthquake time and location prediction using multiple sesimicity indicators, Computer-Aided Civil and Infrastructure Engineering 24 (2009) 280–292.
[65] P. Pérez, J. Reyes, An integrated neural network model for PM10 forecasting, Atmospheric Environment 40 (2006) 2845–2851.
[66] M.D. Petersen, T. Cao, K.W. Campbell, A.D. Frankel, Time-independent and time-dependent seismic hazard assessment for the State of California: Uniform California Earthquake Rupture Forecast Model 1.0, Seismological Research Letters 78 (1) (2007) 99–109.
[67] G. Ranalli, A statistical study of aftershock sequences, Annali di Geofisica 22 (1969) 359–397.
[68] P.A. Reasenberg, L.M. Jones, Earthquake hazard after a mainshock in California, Science 243 (4895) (1989) 1173–1176.
[69] J. Reyes, V. Cárdenas, A Chilean seismic regionalization through a Kohonen neural network, Neural Computing and Applications 19 (2010) 1081–1087.
[70] D.A. Rhoades, Application of the EEPAS model to forecasting earthquakes of moderate magnitude in Southern California, Seismological Research Letters 78 (1) (2007) 110–115.
[71] C.F. Richter, An instrumental magnitude scale, Bulletin of the Seismological Society of America 25 (1935) 1–32.
[72] R. Romero-Záliz, C. Rubio-Escudero, J.P. Cobb, F. Herrera, O. Cordón, I. Zwir, A multiobjective evolutionary conceptual clustering methodology for gene annotation within structural databases: a case of study on the gene ontology database, IEEE Transactions on Evolutionary Computation 122 (6) (2008) 679–701.
[73] D.E. Rumelhart, G.E. Hinton, R.J. Willians, Learning Internal Representations by Error Propagation, MIT Press, Massachussets, 1986.
[74] P.R. Sammonds, P.G. Meredith, I.G. Main, Role of pore fluid in the generation of seismic precursors to shear fracture, Nature 359 (1992) 228–230.
[75] D. Schorlemmer, S. Wiemer, M. Wyss, Earthquake statistics at Parkfield: 1. Stationarity of b values, Journal of Geophysical Research 109 (2004) B12307.
[76] D. Schorlemmer, J.D. Zechar, M.J. Werner, E.H. Field, D.D. Jackson, T.H. Jordan, First results of the Regional Earthquake Likelihood Models experiment, Pure and Applied Geophysics 167 (2010) 8–9.
[77] D.G. Schorlemmer, S. Wiemer, Microseismicity data forecast rupture area, Nature 434 (7037) (2005) 1086.
[78] D.G. Schorlemmer, S. Wiemer, M. Wyss, Variations in earthquake-size distribution across different stress regimes, Nature 437 (7058) (2005) 539–542.
[79] F.M. Shah, M.K. Hasan, M.M. Hoque, S. Ahmmed, Architecture and weight optimization of ANN using sensitive analysis and adaptive particle swarm optimization, International Journal of Computer Science and Network Security 10 (8) (2010) 103–111.
[80] Z.Z. Shen, D.D. Jackson, Y.Y. Kagan, Implications of geodetic strain rate for future earthquakes, with a five-year forecast of M5 Earthquakes in Southern California, Seismological Research Letters 78 (1) (2007) 116–120.
[81] Y. Shi, B.A. Bolt, The standard error of the magnitude–frequency *b* value, Bulletin of the Seismological Society of America 72 (5) (1982) 1677–1687.
[82] S. Srilakshmi, R.K. Tiwari, Model dissection from earthquake time series: a comparative analysis using nonlinear forecasting and artificial neural network approach, Computers and Geosciences 35 (2009) 191–204.
[83] K.F. Tiampo, R. Shcherbakov, Seismicity-based earthquake forecasting techniques: ten years of progress, Tectonophysics 522–523 (2012) 89–121.
[84] C.I. Trifu, V.I. Shumila, A method for multidimensional analysis of earthquake frequency–magnitude distribution with and application to the Vrancea region of Romania, Tectonophysics 261 (1997) 9–22.
[85] T. Utsu, A statistical study on the occurrence of aftershocks, Geophysical Magazine 30 (1961) 521–605.
[86] T. Utsu, A method for determining the value of *b* in a formula $\log n = a - bM$ showing the magnitude–frequency relation for earthquakes, Geophysical bulletin of Hokkaido University 13 (1965) 99–103.
[87] T. Utsu, Representation and analysis of the earthquake size distribution: a historical review and some new approaches, Pure and Applied Geophysics 155 (1999) 509–535.
[88] T. Utsu, Y. Ogata, R.S. Matsu'ura, The centenary of the Omori formula for a decay law of aftershock activity, Journal of Physics of the Earth 43 (1995) 1–33.
[89] P. Varotsos, K. Alexopoulos, K. Nomicos, Seismic electric currents, Proceedings of the Academy of Athens 56 (1981) 277–286.
[90] Q. Wang, D.D. Jackson, Y.Y. Kagan, California Earthquakes, 1800–2007: a unified catalog with moment magnitudes, uncertainties, and focal mechanisms, Seismological Research Letters 80 (3) (2009) 446–457.
[91] P.L. Ward, R.A. Page, The Loma Prieta Earthquake of October 17, 1989. What happened – what is expected – what can be done. Technical report. United States Government Printing Office; 1990.
[92] S.N. Ward, Methods for evaluating earthquake potential and likelihood in and around California, Seismological Research Letters 78 (1) (2007) 121–133.
[93] S. Wiemer, J. Benoit, Mapping the *b* value anomaly at 100 km depth in the Alaska and New Zealand Subduction Zones, Geophysical Research Letters 23 (1996) 1557–1660.
[94] S. Wiemer, M. Gerstenberger, E. Hauksson, Properties of the aftershock sequence of the 1999M_w 7.1 Hector Mine Earthquake: implications for aftershock hazard, Bulletin of the Seismological Society of America 92 (4) (2002) 1227–1240.
[95] S. Wiemer, D. Schorlemmer, ALM: an asperity-based likelihood model for California, Seismological Research Letters 78 (1) (2007) 134–143.
[96] I.H. Witten, E. Frank, Data Mining. Practical Machine Learning Tools and Techniques, Elsevier, Amsterdam, 2005.
[97] J.D. Zechar, T.H. Jordan, Testing alarm-based earthquake predictions, Geophysical Journal International 172 (2008) 715–724.
[98] J.D. Zechar, J. Zhuang, Risk and return: evaluating reverse tracing of precursors earthquake predictions, Geophysical Journal International 182 (3) (2010) 1319–1326.
[99] S.D. Zhang, The 1999 Xiuyen–Haichung, Liaoning, M5.4 Earthquake, Beijing Seismological Press, Beijing, 2004.
[100] A. Zollo, W. Marzocchi, P. Capuano, A. Lomaz, G. Iannaccone, Space and time behaviour of seismic activity at Mt. Vesuvius volcano, southern Italy, Bulletin of the Seismological Society of America 92 (2) (2002) 625–640.