



Relationship between isoseismal area and magnitude of historical earthquakes in Greece by a hybrid fuzzy neural network method

G-A. Tselentis and E. Sokos

University of Patras, Seismological Lab, Greece

Correspondence to: G-A. Tselentis (tselenti@upatras.gr)

Received: 15 September 2011 – Accepted: 30 September 2011 – Published: 4 January 2012

Abstract. In this paper we suggest the use of diffusion-neural-networks, (neural networks with intrinsic fuzzy logic abilities) to assess the relationship between isoseismal area and earthquake magnitude for the region of Greece. It is of particular importance to study historical earthquakes for which we often have macroseismic information in the form of isoseisms but it is statistically incomplete to assess magnitudes from an isoseismal area or to train conventional artificial neural networks for magnitude estimation. Fuzzy relationships are developed and used to train a feed forward neural network with a back propagation algorithm to obtain the final relationships. Seismic intensity data from 24 earthquakes in Greece have been used. Special attention is being paid to the incompleteness and contradictory patterns in scanty historical earthquake records. The results show that the proposed processing model is very effective, better than applying classical artificial neural networks since the magnitude macroseismic intensity target function has a strong nonlinearity and in most cases the macroseismic datasets are very small.

1 Introduction

A significant stage of modern seismic hazard techniques is the assessment of the magnitudes (M) of the most severe historic earthquakes. The difficulty in identifying large seismogenic faults corresponding to historical events and assessing their magnitudes increases the need to derive methodologies to estimate the magnitudes of historic earthquakes from the macroseismic information depicted in the historic records (i.e. isoseisms).

Obviously, there is no established consensus on how to get an objective estimate of the magnitude of an earthquake from macroseismic data alone. In the absence of a simple physical

link between M and macroseismic intensity, the only alternative is to investigate the validity of an empirical relationship using statistical methodologies.

During the past years, the relationship between macroseismic information and earthquake magnitude has been investigated by many researchers either by providing some functional form, (e.g. Gupta and Nuttli, 1976; Vassileva, 2001; Tselentis and Danciu, 2008), between seismic intensity, (here when we refer to seismic intensity we mean the Modified Mercalli intensity MMI), and M throughout regressive analysis. Destructive earthquakes are infrequent with very small probability of occurrence. Thus, observations used in estimating seismic magnitudes from intensity data are incomplete and form a small sample.

Recently, many researchers have established regression analysis models based on powerful statistical tools to develop various nonlinear relationships between MMI, M , and various seismic engineering parameters (e.g. Tselentis and Danciu, 2008; Tselentis and Vladutu, 2010; Tselentis, 2011). Furthermore, the empirical relationships derived from this types of analyses are normally characterized by large scattering of the data due to the inherent uncertainty of the intensity parameter and by the uncertainty on the hypocentral depths. It is well known that for the same magnitude earthquake, the deeper the hypocenter the smaller the effects on the Earth's surface, thus the magnitude of shallow earthquakes is overestimated while the magnitude of deeper events is underestimated.

To overcome this depth uncertainty, Galanopoulos (1961) proposed the use of the area or average radius of individual isoseisms, or the felt area (the area where the earthquake was actually felt by people). This methodology gives a better estimate of M than a single MMI value. Even with this methodology, however, the hypocentral depth plays a significant role because it controls the decay of intensity with distance.

Another approach which has also been used to estimate seismic magnitudes from macroseismic information (isoseisms) was to synthesize isoseisms numerically corresponding to various seismogenic faults (e.g. Suhadolc et al., 1988; Zahradnic, 1989; Sirovich, 1996) and compare it with observed data throughout a trial and error process. Recently, for non-linear modelling, neural networks and neuro-fuzzy modelling approaches have received a great deal of attention (Huang and Leung, 1999; Tselentis and Vladutu, 2010; Tselentis, 2011). Neuro-fuzzy modelling is concerned with the extraction of models from numerical data representing the behaviour of a system. The models in this case are rule-based and use the formalism of fuzzy logic, i.e. they consists of sets of fuzzy “if-then” rules with possibly several premises (Morga, 2000). These models, known as hybrid neural fuzzy, combine the stochastic and artificial intelligence approaches and they are particularly suited for data sets with very weak statistical properties.

In the present paper, we investigate the efficiency of hybrid fuzzy-neural-network models in correlating seismic magnitudes with the isoseismal area using a data set of 24 earthquakes from Greece (Table 1), for which we have well known macroseismic information in the form of isoseisms.

2 Data

Most of the intensity information was available through the European Strong Motion Database (Ambraseys et al., 2004) and completed by the macroseismic database developed by Kalogeras et al. (2004). The macroseismic database covers most of the strong earthquakes occurring in Greece and for each event, MMI values are assigned to every recording station.

The isoseisms corresponding to each earthquake were constructed using the kriging methodology. Kriging is a statistical technique that estimates unknown values at specific points in space using data values from known locations. The main assumption when using kriging is that the data analysed are samples of a regionalized variable, as is assumed to be the case with intensity data. A regionalized variable varies continuously in such a manner that points near each other have a certain degree of spatial correlation, but points that are widely separated are statistically independent.

The kriging estimator applied in the macroseismic dataset considered in the present paper is given by

$$I_j = \sum_{i=1}^n w_{ij} MMI_i \tag{1}$$

where I_j is the predicted intensity value at any grid node, n is the number of points used to interpolate at each node, MMI_i is the intensity value at the i -th point and w_{ij} is the weight associated with the i -th data value when estimating I_j . The weights are solutions of a system of linear equations which

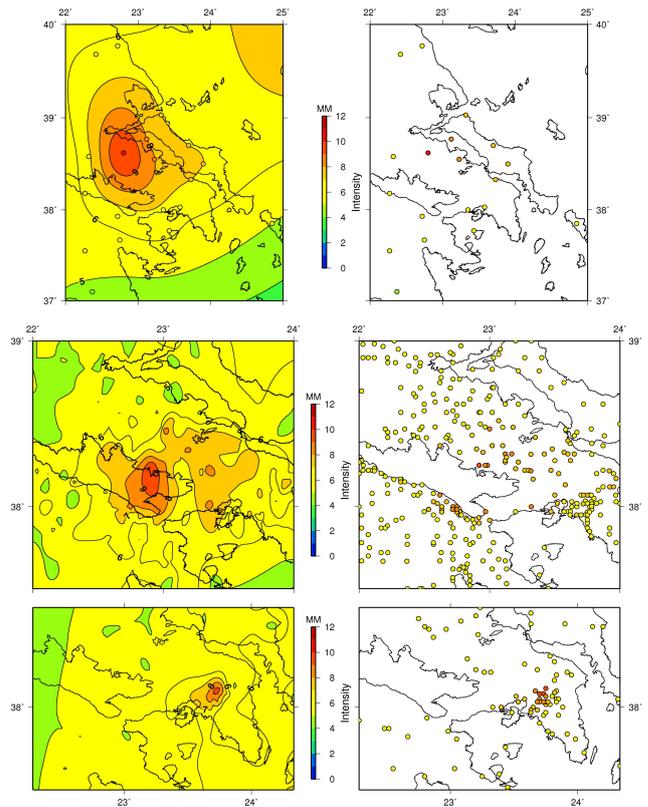


Fig. 1. Example of Isoseisms constructed by the kriging methodology and corresponding macroseismic data for 3 earthquakes in Attica.

are obtained by assuming that I is a sample-path of a random process and that the error of prediction is minimal.

The kriging algorithm assigns weights to each point based on the distance between the point to be interpolated and the data location (h), as well as the inter-data spacing. Other parameters, such as length scale, repeatability, and direction dependence of data are also considered for assigning weights. These parameters are entered into the algorithm via the variogram $\gamma(h)$, which is an analytical tool that quantifies the degree of spatial autocorrelation of data.

In the present investigation, the isoseismals that more accurately represented the observed intensity data field were chosen by modelling a simple linear variogram based on the kriging options of Surfer Package from Golden Software. A detailed explanation of the kriging algorithm and the variogram parameters can be found in De Rubeis et al. (2005). Figure 1 depicts some examples of the estimated isoseisms for the case of 3 earthquakes in Attica (Central Greece) and the corresponding macroseismic data.

Table 1. Earthquake data and calculated isoseismal areas for MMI = 6.

Date	Location	Magnitude	Area	Log (Area)
07/08/1915	Kefalonia	6.7	817	2.912222057
01/03/1941	Larisa	6.3	216	2.334453751
18/03/1953	nw Turkey	7.4	4567	3.659631012
09/07/1956	Amorgos	7.5	2060	3.31386722
19/02/1968	Ag. Efsratos	7.1	1007	3.003029471
20/06/1978	Thessaloniki	6.5	243	2.385606274
24/02/1981	E. Corinthos	6.7	919	2.963315511
07/09/1999	Attica	6	104	2.017033339
5/8/1766	Chora	7.6	20928	4.320727727
21/2/1858	Corinthos	6.5	178	2.250420002
1/8/1870	Arachova	6.8	411	2.613841822
3/4/1881	Nenita	6.5	566	2.752816431
27/8/1886	Philiatra	7.3	3145	3.49762065
20/4/1894	Atalanti	6.6	702	2.846337112
27/4/1894	St. Constantinos	7	5155	3.71222867
22/1/1899	Kuparissia	6.5	531	2.725094521
6/3/1737	Ezine	7.2	5935	3.773420723
21/8/1859	Imroz	6.9	1676	3.224274014
29/7/1880	Menemen	6.7	743	2.870988814
3/4/1881	Chios	6.5	745	2.872156273
15/10/1883	Tsesme	6.8	562	2.749736316
26/10/1889	Chidyra	6.8	1322	3.121231455
20/4/1894	Martino	6.6	3345	3.524396122
29/1/1898	Balikesir	?	14738	4.168438552
7/3/1867	Kloumidados	6.8	1087	3.036229544

3 Artificial neural network approximation

Since the relation between isoseismal area and seismic magnitude is strongly nonlinear, artificial neural networks (ANN) are particularly suited for treating macroseismic data (Tung et al., 1994; Davenport, 2004; Tselentis 2011).

An ANN is an information processing paradigm that is inspired by the way biological nervous systems, such as the brain, process information. ANN can be understood as a mapping $f : R_n \rightarrow R_m$, defined by $y = f(x) = g(W \cdot x)$, where x is the input vector, y is the output vector, W is the weight matrix and g is a nonlinear activation function. The mapping f can be decomposed into a series of mappings resulting in a multi-layer network (Fig. 2):

$$R_n \rightarrow R_p \rightarrow R_q \rightarrow \dots \rightarrow R_m$$

The algorithm for computing W is the training algorithm. The most popular ANN (this kind of ANN will be used in the present investigation) are the multi layer back propagation networks (Rumelhart and McClelland, 1986), whose training algorithm is the well-known gradient descent method. In the learning phase of training such a network, we present the pattern $x_p = \{i_{pi}\}$ as an input and ask that the network adjust the set of weights in all the connecting links and also all the thresholds in the nodes such that the desired outputs $y_p = \{t_{pk}\}$ are obtained at the output nodes. When this adjustment

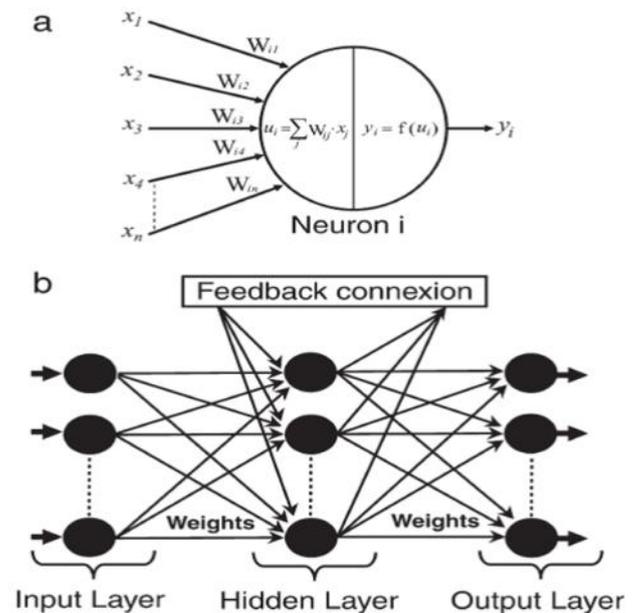


Fig. 2. General topology of a feed-forward ANN with one hidden layer (b).

is completed, we present to the network another pair x_p, y_p and ask that the network also learns this association. In other words, we ask that the network find a single set of weights \mathbf{W} and biases that will satisfy all the input-output pairs which are presented to it.

In general, the output $\{o_{pk}\}$ of the network will not be the same as the target values $\{t_{pk}\}$. For each pattern, the square of the error is

$$E_p = \sum_k (t_{pk} - o_{pk})^2 \quad (2)$$

And the average system error is given by

$$E = \frac{1}{P} \sum_p \sum_k (t_{pk} - o_{pk})^2 \quad (3)$$

where P is the sample size. A true gradient search for minimum system error should be based on the minimization of Eq. (3).

Thus, an ANN is a learning machine whose function depends on the training examples, it does not recognize any functional relation between the input data but it determines a numerical relation among the state parameters. According to the principle of information diffusion, (Huang, 2002), we can increase the certainty of the determined relation if we increase the number of the training examples with the help of an appropriate information scattering function. ANN trained in this manner are called diffusion neural networks (e.g. Huang and Moraga, 2004).

However, neural information processing models generally assume that the patterns used for training an ANN are compatible. If the patterns are contradictory, the neural network does not converge because the adjustments of the weights and thresholds do not know where to turn.

In theory, neural networks, and fuzzy systems are equivalent in that they are convertible, yet in practice each has its own advantages and disadvantages. For neural networks, the knowledge is automatically acquired by the back propagation algorithm, but the learning process is relatively slow and analysis of the trained network is difficult (black box). Neither is it possible to extract structural knowledge (rules) from the trained neural network, nor can we integrate special information about the problem into the neural network in order to simplify the learning procedure.

Fuzzy systems are more favourable in that their behaviour can be explained based on fuzzy rules and thus their performance can be adjusted by tuning the rules. But since, in general, knowledge acquisition is difficult and also the universe of discourse of each input variable needs to be divided into several intervals, applications of fuzzy systems are restricted to the fields where the number of input variables is small.

Various researchers (e.g. Monostori and Egresits, 1994; Hernandez et al., 1995; Radeva and Radev, 2002; Radeva, 2002) have developed fuzzy neural networks with stronger nonlinear mapping abilities than the conventional ANN. This

kind of ANN have promising application prospects in nonlinear modeling, fuzzy identification and self-organizing fuzzy control for complex systems such macroseismic data sets (Huang and Liu, 1985).

Hybrid-fuzzy-neural-networks are expected to be very effective in estimating the relationship between isoseismal area and earthquake magnitude although data are usually scanty, incomplete and contradictory. The basics of information diffusion and fuzzy theory will be presented in the next section.

4 Information diffusion

If there are only few data available in the examination of a phenomenon, we can assign these to some already existing statistical distribution (e.g. the Bayes method), and the structured sample will have an informational value. The question arises: what to do in the case when we do not know a priori statistical distribution? From a small data sample, any classical ANN cannot recognize a nonlinear function. This is the case we mostly encounter when we deal with macroseismic datasets.

In such a case, the theory of *fuzzy sets* can be applied with a very good efficiency. This theory enables the processing of uncertain information, to be more precise, it writes down the fuzzy logical assertions in an exact mathematical form (Zadeh, 1974).

Suppose that we have n observations \mathbf{X} , of magnitudes m_i and isoseismal areas s_i obtained from n historical earthquakes. Our task is to find some sort of relation R between them. Let f be an operator (mathematical procedure) of the observations which is employed to estimate R . Let $R(f, \mathbf{X})$ denote the estimation of R by f .

If n is large, then f can be a probabilistic or statistical operator and $R(f, \mathbf{X})$ is a statistical relationship. In the case that n is small (inadequate data sample), we generally cannot obtain a statistical or physical relation from \mathbf{X} . Under this situation we need to employ fuzzy relationship based on \mathbf{X} . Let us suppose that we are given a sample \mathbf{X} of n real valued observations, x_i ($i = 1, n$), which have two components, earthquake magnitude m_i and isoseismal area A_i , whose underlying relationship is to be estimated

$$\mathbf{X} = \{x_1, x_2, \dots, x_n\} = \{(A_1, m_1), (A_2, m_2), \dots, (A_n, m_n)\} \quad (4)$$

To reduce scattering it is preferable to use the logarithm of the isoseismal area

$$s = \log_{10} A \quad (5)$$

and in this case our sample becomes

$$\mathbf{X} = \{x_1, x_2, \dots, x_n\} = \{(s_1, m_1), (s_2, m_2), \dots, (s_n, m_n)\}$$

Let $\mathbf{S} = \{s_1, s_2, \dots, s_n\}$ be a sample taken from the universe of discourse \mathbf{U} , any mapping from $S \times \mathbf{U}$ to $[0, 1]$

$$\begin{aligned} \mu : S \times \mathbf{U} &\rightarrow [0, 1] \\ (s, u) &\rightarrow \mu(s, u); s, u \in S \times \mathbf{U} \end{aligned} \quad (6)$$

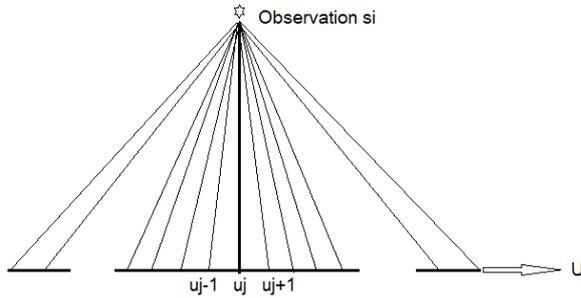


Fig. 3. The information diffusion function is used to diffuse information s_i to all the elements of the universe of discourse U .

is called an information diffusion of S on U if it satisfies the following condition for every s_i in S and for every u', u'' in U if $|u' - s| \leq |u'' - s|$

$$\text{then } \mu(s, u') \geq \mu(s, u'') \quad (7)$$

$\mu(s_i, u)$ is called an information diffusion function of S on U . If U is discrete, μ can be written as $\mu(s_i, u_j)$. As we can deduce from Eq. (7), μ is a convex function about U .

In other words μ defines a fuzzy subset on the universe of discourse $S \times U$. Obviously, an available fuzzy relation would approximately reveal the information structure implied by the observations X .

The trivial diffusion function is defined as
1, if $u = s$

$$\mu(s, u) = s \in S, u \in U \quad (8)$$

0, otherwise

The simplest diffusion function is the linear distribution function with respect to a discrete monitoring space U . When $S = \{s_i | i = 1, n\}$ and the monitoring space $U = \{u_j | j = 1, m\}$ has steps of equal length Δ we get the 1-dimensional linear information distribution which can be written as

$$1 - |s - u|/\Delta, \text{ if } |s - u| \leq \Delta$$

$$\mu(s, u) = s \in S, u \in U \quad (9)$$

0, otherwise

where $\Delta = u_{j+1} - u_j$.

The principle of information diffusion (Huang and Shi, 2002) alerts that, when we use an incomplete data set to estimate a relationship, there must exist reasonable diffusion means to change observations into fuzzy sets to partly fill the gap caused by incompleteness and improve the original estimate.

Obviously, if our macroseismic observations (data set) X is incomplete, this implies that the patterns are insufficient. In other words, we need more patterns to train the BP network for obtaining a more accurate estimate of input-output relation. The simplest model to derive the required patterns is based on the similarities of information and molecules.

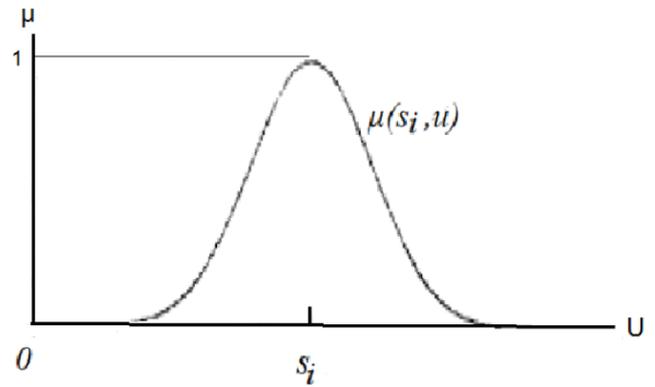


Fig. 4. Information diffusion function of s_i on a continuous universe of discourse U .

Taking into consideration the molecular diffusion theory, we consider the following normal diffusion function

$$\mu(s_j, u) = e^{-\frac{(u-s_j)^2}{2h^2}} \quad s \in S, u \in U \quad (10)$$

h is the normal diffusion coefficient which can simply be calculated (Huang, 1997) by

$$h = \begin{cases} 1.6987(b-a)/(n-1) & \text{for } 1 < n \leq 5 \\ 1.4456(b-a)/(n-1) & \text{for } 6 \leq n \leq 7 \\ 1.4230(b-a)/(n-1) & \text{for } 8 \leq n \leq 9 \\ 1.4208(b-a)/(n-1) & \text{for } 10 \leq n \end{cases} \quad (11)$$

Where $b = \max\{s_i\}, 1 \leq i \leq n; a = \min\{s_i\}, 1 \leq i \leq n;$ and n the number of pairs (s_i, m_i) .

Thus, it is obvious that with the help of the normal diffusion function, we can transform any one input-output observation (s_i, m_i) into two fuzzy subsets

$$A_i = \int_U \frac{\mu(s_i, u)}{u}$$

$$B_i = \int_V \frac{\mu(m_i, v)}{v} \quad (12)$$

Obviously, for a particular pair of observations (s_i, m_i) we have $A_i \rightarrow B_i$.

Information diffusion can be represented schematically in Fig. 3, where observation s_i is diffused to every point of U with different values. In the case of a continuous universe of discourse, information diffusion of s_i can be represented by a fuzzy membership function $\mu(s_i, u)$ as presented in Fig. 4.

In order to preserve more information we use the correlation product encoding (Kosko, 1992), to produce the required fuzzy relationships, therefore we can write

$$\mu_{R_i}(u, v) = \mu_{A_i}(u) \mu_{B_i}(v), u \in U, v \in V \quad (13)$$

with R_i depicting the corresponding rule (relation between u, v). Thus, we get n fuzzy relationships from n earthquake observations.

For the case that we have n observations $(s_1, m_1), (s_2, m_2), \dots, (s_n, m_n)$ then, using the information distribution approach, we can obtain n fuzzy IF-THEN rules $(R_i) : A_1 \rightarrow B_1, A_2 \rightarrow B_2, \dots, A_n \rightarrow B_n$ (as it is shown in Fig. 5).

Let us suppose that s_o is a known crisp input value and that we want to find a way to estimate the corresponding magnitude m_o from s_o and R_i . Considering that the universe of discourse \mathbf{U} where all s_i belong (the monitoring space) is discrete. Its elements u_j are the controlling points. Thus, s_o is not just equal to some value u_j in universe \mathbf{U} . We can apply the information distribution formula (9) to get the following fuzzy subset (information diffusion of s_o)

$$\mu_{s_o}(u_j) = \begin{cases} 1 - |s_o - u_j|/\Delta & \text{if } |s_o - u_j| \leq \Delta \\ 0 & \text{if } |s_o - u_j| \geq \Delta \end{cases} \quad (14)$$

where $\Delta = u_{j+1} - u_j$

Then, a fuzzy consequent $\sim m_o$ from $\sim s_o$ and R_i can be written as

$$\mu_{m_o}(v) = \sum_v \mu_{s_o}(v) \mu_{R_i}(u, v) \quad (15)$$

Changing the intensity component $\mu_{B_i}(v)$ in Eq. (13) into the following fuzzy subsets:

$$\mu_{B_i}(v) = \begin{cases} 1 & \text{if } v = m_i \\ 0 & \text{if } v \neq m_i \end{cases} \quad (16)$$

And

$$\mu_{R_i}(u, v) = \begin{cases} \mu_{A_i}(v) & \text{if } v = m_i \\ 0 & \text{if } v \neq m_i \end{cases} \quad (17)$$

Equation (15) can be written as

$$\mu_{m_o}(v) = \begin{cases} \mu_{s_o}(v) & \text{if } v = m_i \\ 0 & \text{if } v \neq m_i \end{cases} \quad (18)$$

Let

$$W_i = \sum_v \mu_{s_o}(v) \mu_{A_i}(v) \quad (19)$$

W_i , can be considered as the possibility that component m_o may be m_i .

Then integrating all results coming from R_1, R_2, \dots, R_n the required value m_o can be derived from the following (gravity center)

$$m_o = \frac{\sum_{i=1}^n w_i m_i}{\sum_{i=1}^n W_i} \quad (20)$$

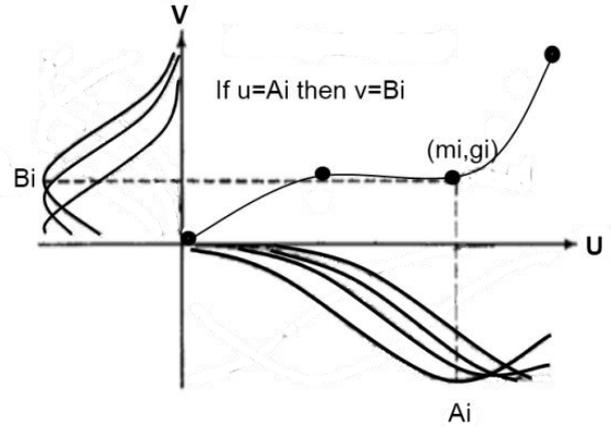


Fig. 5. Fuzzy rule “IF $u = A$ THEN $v = B$ ” through observation s_i, m_i using information diffusion.

This result means that we have estimated the required value m_o , by fuzzy influence based on the diffusion method of the known value s_o . The above procedure is summarized in the flow chart depicted in Fig. 6.

Thus, throughout the information diffusion approximation reasoning technique, we can transform the original $\mathbf{X} = \{x_1, x_2, \dots, x_n\} = \{(s_1, m_1), (s_2, m_2), \dots, (s_n, m_n)\}$ sample of observations to $\sim \mathbf{X} = \{(s_1, \sim m_1), (s_2, \sim m_2), \dots, (s_n, \sim m_n)\}$, which is finally used to train the BP ANN (Fig. 7).

5 Results

After trying a various number of hidden layers of the BP ANN, we found that the optimum network which resulted in the least errors is that consisting of 7 hidden layers (Table 2, Fig. 8).

For supervised training of the ANN, a subset of two thirds of the total data was used. The individual sites assigned to the training set were selected at random from the complete set of records. The other third of the data was used for testing the ANN after it had been trained.

Next, we used Eq. (20) to calculate the new magnitudes by applying the information diffusion technique and used them again to train the ANN. Figure 9 depicts the three results of our model corresponding to conventional ANN, linear regression and the hybrid-neural-fuzzy and the neural network.

In order to compare the mean square errors E of the three estimators, the linear regression (LR), the hybrid fuzzy neural network (HFNN), and the conventional neural network estimator (NN) are computed as follows:

$$E_{LR} = \frac{1}{24} \sum_{i=1}^{24} (y_i - y_{LRi})^2 = 0.39$$

Table 2. Parameters of tested neural networks. Bold numbers correspond to the selected network.

ANN Type	# Weights	Fitness	Train <i>r</i> Error	Test Error	AIC	Correlation
[1-1-1]	4	3.2142	0.1476	0.3111	-78.460	0.8447
[1-7-1]	22	3.8909	0.1070	0.2570	-48.254	0.9196
[1-4-1]	13	1.4644	0.1110	0.6828	-65.593	0.9160
[1-5-1]	16	1.7609	0.1123	0.5678	-59.379	0.9151
[1-6-1]	19	1.3948	0.1116	0.7169	-53.480	0.9162

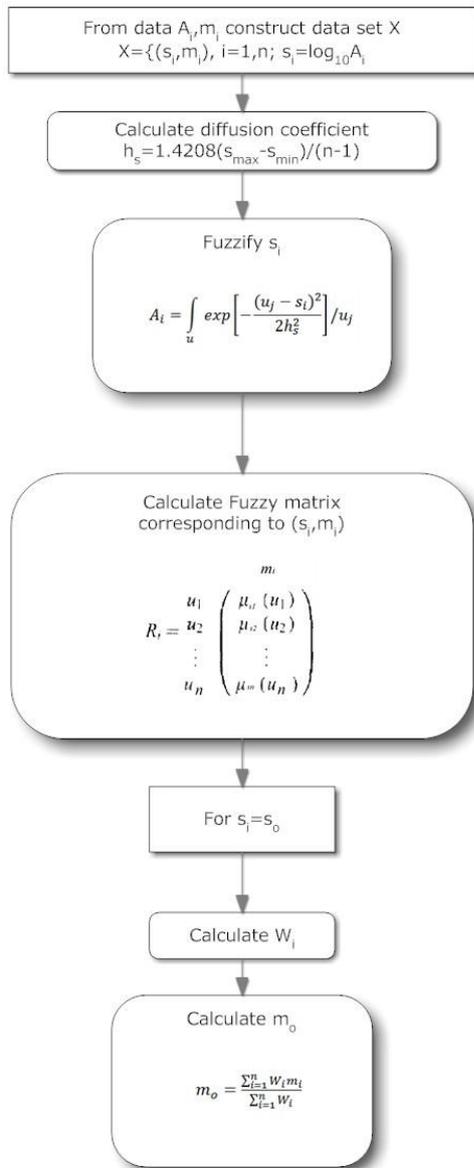


Fig. 6. Flowchart depicting the information diffusion method.

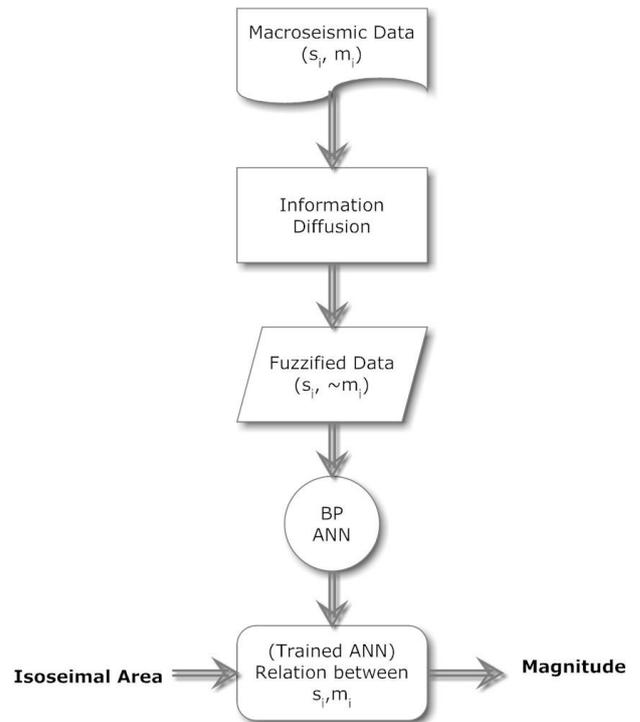


Fig. 7. System architecture of the hybrid neural-fuzzy network adopted in the present investigation.

$$E_{HFN} = \frac{1}{24} \sum_{i=1}^{24} (y_i - y_{HFNi})^2 = 0.16$$

$$E_{LR} = \frac{1}{24} \sum_{i=1}^{24} (y_i - y_{NNi})^2 = 0.23$$

Obviously, the HFN estimator is better than the linear regression estimator and the conventional neural network estimator since it is more precise, nearer to real value, and more stable than the conventional neural estimator.

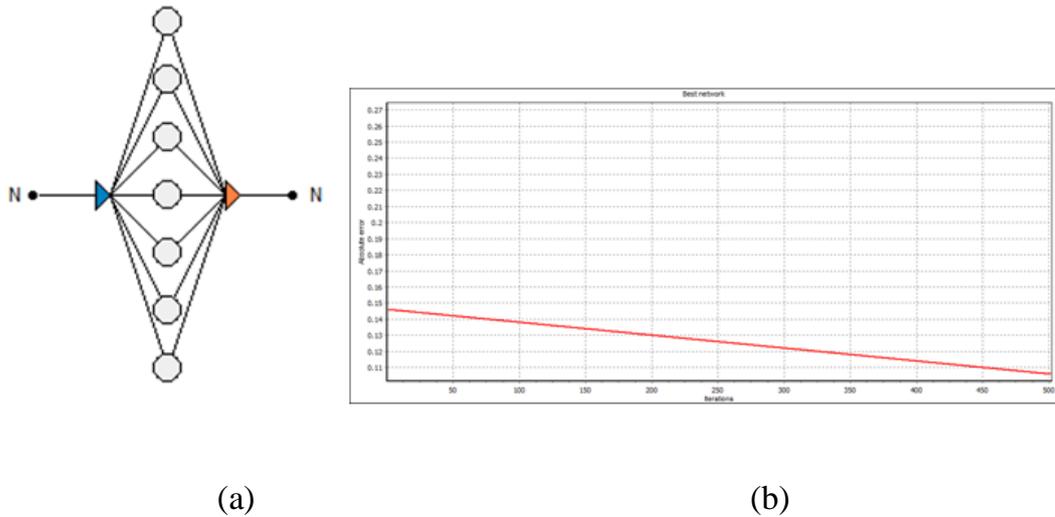


Fig. 8. The selected neural network consisting of 7 hidden layers (a) and the corresponding convergence curve (b).

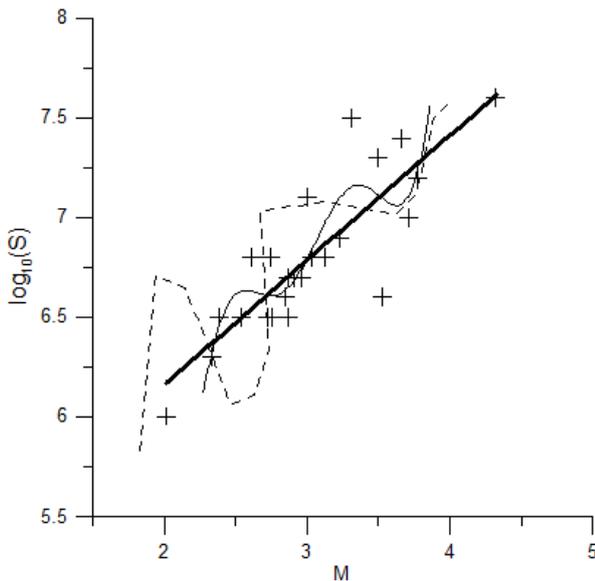


Fig. 9. Relationship between logarithmic isoseismal area and earthquake magnitude estimated by the back propagation NN (dashed curve) and the proposed hybrid fuzzy-neural methodology (solid curve). The lines corresponds to the list square fit of the data.

6 Conclusions

Neural information processing models largely assume that the samples for training a neural network are sufficient. Otherwise, there exists a non-negligible error between the real function and estimated function from a trained network. To reduce the error in this paper, we suggest a hybrid fuzzy neural network to learn from a small sample.

The obtained error is less than the error of the conventional ANN. The results show that the hybrid fuzzy neural model is very effective in the case of treating macroseismic M,MMI datasets of historic earthquakes where the target function has a strong nonlinearity and the given sample is very small.

Acknowledgements. This paper was motivated by the pioneering work of C. Huang on hybrid fuzzy neural theory. We also thank N. Chingsao for helping us to understand the intricacies of fuzzy systems. All ANN calculations were performed using Alyuda Neurointelligence algorithm.

Edited by: M. E. Contadakis

Reviewed by: two anonymous referees

References

- Ambraseys, N. N., Smit, P., Douglas, J., Margaris, B., Sigbjörnsson, R., Ólafsson, S., Suhadolc, P., and Costa, G.: Internet site for European strong-motion data, *Boll. Geof. Teor. Appl.*, 45, 113–129, 2004.
- Davenport, P.: Neural network analysis of seismic intensity from instrumental records, 13th WCEE, Vancouver, B.C., Canada, paper 692, 2004.
- De Rubeis, V., Tosi, P., Gasparini, C., and Solipaca, A.: Application of the kriging technique to seismic intensity data, *Bull. Seism. Soc. Am.*, 66, 540–548, 2005.
- Galanopoulos, A. G.: On magnitude determination by using macroseismic data, *Ann. Geofis.*, 14, 225–253, 1961.
- Gupta, I. N. and Nuttli, O. W.: Spatial attenuation of intensities for central U.S. earthquake, *Bull. Seismol. Soc. Am.*, 66, 743–751, 1976.
- Hernandez, K., Moore, K., and Elphic, R.: Sensor fusion and non-linear prediction for anomalous event detection, *Proc. SPIE Internat. Soc. Opt. Eng.*, 2484, 102–112, 1995.

- Huang, C. F.: Information diffusion techniques and small sample problem, *Int. J. Inf. Tech. Deci.*, 1, 229–249, 2002.
- Huang, C. F. and Liu, Z. R.: Isoseismal area estimation of Yunnan Province by fuzzy mathematical method. *Fuzzy Mathematics in Earthquake Research*, edited by: Deyi, F. and Xihui, L., Seismological Press, Beijing, 185–195, 1985.
- Huang, C. F. and Moraga, C.: A diffusion neural network for learning from small samples, *Int. J. Approx. Reason*, 35, 137–161, 2004.
- Huang, C. F. and Shi, Y.: Towards efficient fuzzy information processing – using the principle of information diffusion, *Physica-Verlag, Heidelberg*, 2002.
- Huang, F. F.: Principle of information diffusion, *Fuzzy sets and Systems* 91, 1, 69–90, 1997.
- Kalogeras, I. S., Marketos, G., and Theodoridis, Y.: A tool for collecting, querying and mining macroseismic data, *Bull. Geol. Soc. Greece*, 36, 1406–1411, 2004.
- Monostori, L. and Egresits, C.: Modelling and monitoring of milling through neuro-fuzzy techniques, *Proc. Intelligent Manufacturing Systems, IMS'94*, 463–468, 1994.
- Moraga, C.: Neuro-fuzzy modelling of compensating systems, in: *Quo Vadis Computational Intelligence*, edited by: Sincak, P. and Vascak, J., Physica-Verlag, Heidelberg, 385–389, 2000.
- Radeva, S. T.: Forecasting the Behavior of Seismic Waves on the Base of Fuzzy Logic Models, *Proc. Automatics and Informatics*, Sofia, 1, 101–104, 2002.
- Radeva, S. and Radev, D.: Fuzzy Random Vibration of Hysteretic System Subjected to Earthquake, *Proc of the IEEE Symposium on Intelligent Systems Varna*, 1, 130–135, 2002.
- Rumelhart, D. E. and McClelland, J. L.: *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, MIT Press/Bradford Books, Cambridge, Mass., 1986.
- Sirovich, L.: A simple algorithm for tracing synthetic isoseismals, *Bull. Seism. Soc. Am.*, 86, 1019–1027, 1996.
- Suhadolc, P., Cernobori, L., Pazzi, G., and Panza, G. F.: Synthetic isoseismals: applications to Italian earthquakes, in: *Seismic Hazard in Mediterranean Regions*, edited by: Bonnin, J., Kluwer, Dordrecht, Holland, 105–128, 1988.
- Tung, A. T., Wong, F. S., and Dong, W.: Prediction of the spatial distribution of the modified Mercalli intensity using neural networks, *Earthq. Eng. Struct. Dynam.*, 23, 49–62, 1994.
- Tselentis, G. A.: Assessment of Arias intensity of historical earthquakes using modified Mercalli intensities and artificial neural networks, *Nat. Hazards Earth Syst. Sci.*, in press, 2011.
- Tselentis, G. A. and Danciu, L.: Empirical Relationships between Modified Mercalli Intensity and Engineering Ground-Motion Parameters in Greece, *Bull. Seismol. Soc. Am.*, 98, 1863–1875, 2008.
- Tselentis, G.-A. and Vladutu, L.: An attempt to model the relationship between MMI attenuation and engineering ground-motion parameters using artificial neural networks and genetic algorithms, *Nat. Hazards Earth Syst. Sci.*, 10, 2527–2537, doi:10.5194/nhess-10-2527-2010, 2010.
- Vassileva, S. T.: Predicting Earthquake Ground Motion Descriptions Through Artificial Neural Network for Testing the Constructions, *Struct. Eng., Mech. Comput.*, Elsevier, 2, 927–934, 2001.
- Zadeh, L. A.: On the analysis of large scale system, in: *Systems Approaches and Environmental Problems*, edited by: Gottinger, H., Vandenhoeck and Ruprecht, Göttingen, 23–37, 1974.
- Zahradnik, J.: Simple method for combined studies of macroseismic intensities and focal mechanism, *Pure Appl. Geophys.*, 130, 83–97, 1989.