DOI 10.4430/bgta0144

Artificial Adaptive Systems to predict the magnitude of earthquakes

P.M. Buscema[1,2], G. Massini[1] and G. Maurelli[1]

[1] *Semeion Research Centre of Sciences of Communication, Rome, Italy*
[2] *Dept. Mathematical and Statistical Sciences, University of Colorado, Denver (CO), U.S.A.*

(Received: April 11, 2014; accepted: December 22, 2014))

ABSTRACT Currently, in the geological studies it is clear that the generation process and the dynamics of development of an earthquake belong to the highly nonlinear and non-stationary phenomena. For this reason, in recent years the authors, experts in the development of mathematical models based on Artificial Neural Networks (ANNs), decided to apply these mathematical models to forecast earthquakes. The aim of this experimental study was to test the capability of advanced ANNs and machine learning to estimate the magnitude of the events recorded daily. Features that describe each event are: origin time (UTC), latitude, longitude, depth, and magnitude. With seismic event means an event between 0.1 and 5.9 magnitude, in the database. We have tested the ANN technology on different data sets: a) USGS data from 1976 to 2002; b) USGS and ISIDe data together from 2005 to 2011; c) ISIDe data from 2005 to 2013. This paper aims at demonstrating as the ANNs are a promising technique for earthquake prediction and as an ANN training on the global data on earthquakes is also much more effective for a local earthquake prediction, than an ANN training on local data. In fact, the results show that the ANNs have very good performances both in functional approximation, than in pattern recognition when the training set represents a sample of worldwide earthquakes: 10% of absolute error of magnitude estimation and about 90% of correct classification (1 of 3 classes) in pattern recognition task. The results using only the Italian ISIDe data set are also promising, although the few information available, but less precise than the previous ones: about 99% of correct predictions for events with $M \leq 2.0$, around 75% for moderate events ($2.0 < M < 3.0$), and a rate of correct classification between 30% and 40% with events where $M \geq 3.0$. This last result is not surprising, due to the small number of events with this magnitude available in the Italian data set (ISIDe). These results can also be the starting point for the development of a system based on ANNs to provide the daily estimation of possible future seismic events.

Key words: earthquake prediction, Artificial Adaptive Systems.

1. Introduction

For several years, there are studies on the predictability of earthquakes (Kagan, 1997; Kagan and Jackson, 2000; Jordan *et al.*, 2011), but the scientific community is still far from being able to say that we have achieved significant results. However, as it is often the case with other events assumed at the time scientifically unpredictable, many geologists, physicists and mathematicians are working with the aim of eventually reach ing a result that can be operationally useful. As an example, we can mention the field of weather forecasting in the first half of the last century that was considered impossible, but it has now reached a level of operational predictability widely used in the field of environmental safety and economic viability.

It is now clear that the process of generation and the dynamics of development of an earthquake belong to highly nonlinear and non-stationary observable phenomena. For this reason, in the last years many scientists tried to apply Artificial Neural Networks (ANNs) to the issues concerning earthquakes, obtaining interesting and promising results (Sharma and Arora, 2005; Ashif *et al.*, 2007; Suratgar *et al.*, 2008).

In recent years, the Semeion Research Centre is working at the experimental level (Buscema and Benzi, 2011), to apply to earthquake prediction new and advanced mathematical models which come from the field of Artificial Intelligence, in particular the so-called Natural Computation and Artificial Adaptive Systems (ANNs, Evolutionary Algorithms, Artificial Organisms). Further, Pattern Informatics (PI) modelling has shown a way to provide intermediate forecasting about earthquakes (Crampin, 2012; Peresan *et al.*, 2012). We think that the PI approach is a serious way to code the time, space and magnitude of the big quakes, but it could be improved with a more complex technique of mathematical modelling using advanced ANNs for function approximation.

In this paper, we have only been inspired from PI. Our main target was to test the capability of a new ANN to make deep learning of a simple earthquake data set, in order to estimate the magnitude of quakes at short term (one day/week before).

The objective of this research is the application of advanced ANN models for the estimation of the magnitude of the registered daily events. In particular, we focused on the prediction of the events recorded worldwide, using the USGS data, and in the entire Italian territory, using data coming only from Italian data sets.

2. The databases

The first data set is formed by the seismic events recorded worldwide by USGS (http://www.usgs.gov/) from 1976 up today.

The second reference database is formed by the seismic events recorded on the Italian territory since 1981. In particular, the events coming from the archive of the Istituto Nazionale di Geofisica e Vulcanologia (INGV), the Italian seismic bulletin that is part of the Italian seismic instrumental and parametric database (ISIDe Working Group, 2010, http://ISIDe.rm.ingv.it). Temporal data available are structured as follows (table 1):

- catalogue of the Italian seismicity (CSI 1.1) for the period 1981-2002 (Castello *et al.*, 2006);
- seismic bulletin (BS), revising data from the Italian national seismic network, for the period 2003-2005;
- Italian seismic instrumental and parametric database (ISIDe, 2010; http://iside.rm.ingv.it/iside/standard/index.jsp), for the period 2005-2013.

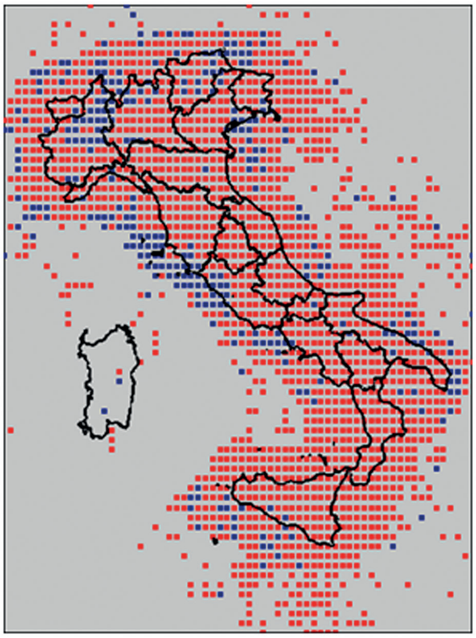

Fig. 1 - The 2054 quadrants of size 20×20 km considered for a prediction test on the Italian territory: red squares show places where at least one event with $H > 2.0$ was recorded.

Table 1 - Catalogues available for Italian events.

Database	Starting from	up to	Events
CSI	1981-01-01	2002-12-31	39534
BS	2003-01-01	2005-04-15	3551
ISIDe	2005-04-16	2013-06-30	105637
Total events			**148722**

In our experimentations we have considered only the ISIDe data, that is from 2005 to 2013.

For both the data sets, features that describe each event are: origin time (UTC), latitude, longitude, depth, and magnitude. Seismic event means an event between 0.1 and 9.0 of magnitude, in the USGS database, and between 0.1 and 5.9 of magnitude, in the ISIDe database.

3. The Italian database

For the Italian data, daily forecasts of all recorded events from July 1, 2012 have been carried out using ANNs.

In order to verify the daily prediction of neural networks on those areas of the country where no events have occurred (true negatives), another data set of artificial events was created considering 2054 quadrants of size 20×20 km, where in 1755 of which at least one event with magnitude larger than 2.0 has been recorded, during the reporting period (2005-2012) (Fig. 1).

In order to generate artificial events with magnitude equal to zero, we have considered the period starting from January 2005 until June 2012 included. Then, in the period considered (January 2005-June 2012), about 475,000 artificial “not-events” were included to train the

ANNs: every day the coordinates of the 20×20 km box where no event occurred were added as a "not-event" to the data set.

Subsequently, the entire data set for training and testing the ANNs includes globally 580,637 events (105,637 from ISIDe and 475,000 boxes of 20×20 km, for each day when no event happened). Further, from July 1, 2012, every day the global database has been increased, as well as the events actually recorded, also from about 2054 artificial "not-events".

4. The models

For the daily forecast, a Supervised Contractive Map [Sv-Cm: Buscema and Benzi, (2011)] and neural networks with supervised feed forward topology (Rumdhart *et. al.*, 1986; Buscema, 1998a, 2013) were used.

An SV-Cm is an advanced type ANN especially suitable for deep learning (Hinton et. al, 2006; Bengio, 2009). Here below we show the forward transfer equations of the signal from the input to the output vector and the consequent equations for the weight matrices updating (learning equations).

Legend:
$[l]$ = number or name of the ANN layer;
$u_i^{[l]}$ = values of the all i-th nodes of the l-th layer;
$w_{i,j}^{[l]}$ = weight matrix connecting the layer $[l-1]$ to the layer $[l]$;
$C^{[l]}$ = number of nodes of l-th layer;
t_i = value of i-th of the dependent variable;
$LCoef$ = ANN learning rate.

Signal transfer from input layer to output layer:

$$CNet_i^{[l]} = \sum_j^{C^{[l-1]}} u_j^{[l-1]} \cdot \left(1 - \frac{w_{ij}^{[l]}}{C^{[l-1]}}\right) \tag{1}$$

$$INet_i^{[l]} = \sum_j^{C^{[l-1]}} u_j^{[l-1]} \cdot w_{ij}^{[l]} \tag{2}$$

$$u_i^{[l]} = \sin\left\{ INet_i^{[l]} \cdot \left[1 - \frac{\sin\left(CNet_i^{[l]}\right)}{C^{[l-1]}}\right]\right\} \tag{3}$$

Weights update:

$$\delta_i^{[out]} = \left(t_i - u_i^{[out]}\right) \cdot \cos\left\{ INet_i^{[out]} \cdot \left[1 - \frac{\sin\left(CNet_i^{[out]}\right)}{C^{[out-1]}}\right]\right\} \tag{4}$$

$$\delta_i^{[hid]} = \sum_k^{Num^{[hid+1]}} \left(\delta_k^{[hid+1]} \cdot w_{ki}^{[hid+1]} \right) \cdot \cos\left\{ INet_i^{[hid]} \cdot \left[1 - \frac{\sin\left(CNet_i^{[hid]}\right)}{C^{[hid-1]}} \right] \right\} \tag{5}$$

$$\Delta w_{ij}^{[l]} = LCoef \cdot \delta_i^{[l]} \cdot u_j^{[l-1]} \cdot \left(1 - \frac{w_{ij}^l}{C^{[l-1]}} \right) \tag{6}$$

The SV-Cm calculates two net inputs for each node: a classic weighted input [Eq. (1)] and a contractive input [Eq. (2)].This second net input tends to decay or to increase when the positive or negative value of the weight (w) becomes close to a specific constant (C).

Eq. (3) activates each node according to a sine function of the two net inputs (the contractive input works as a harmonic modulation of the weighted input). The advantages and the disadvantages of the sine transfer function to work properly into the topology of Multilayer Perceptron were already analyzed in the scientific literature (Le Cun *et al.*, 1991, 1998).

Eq. (4) shows a typical error calculation using the distance between the desiderate output and the estimated output, times the first derivative of sine transfer function.

Eq. (5) works in the same way of Eq. (4), but using the chain rule to calculate the local error of each hidden unit.

Eq. (6) updates the weight matrices, using typical back error propagation, with a contractive factor useful to limit an extreme growing of each weight value.

This neural network has been trained every day in different ways depending on the encoding of information of events in the input vector:

- the first network with 7 inputs (with USGS data and when USGS data are missing with ISIDe data);
- the second one with 15 inputs (with ISIDe data).

In both cases, the networks were structured with 3 levels of 48 hidden units, in order to improve a deep learning of the training set (Bengio, 2009; Raiko *et al.*, 2012) and the learning coefficient (*LCoef*) was fixed at 0.01 for each layer.

The SV-Cm was trained:

a. each day before the prediction phase for 1000 epochs (about 2 hours of computer time - two cores CPU), with the inclusion of the new data occurred with the ISIDe data and for the real time Italian quakes prediction;
b. once for 500 epochs with the USGS data for a retrospective prediction task.

The prediction phase runs in a few seconds each day in both cases.

Fig. 2 shows an example of Supervised Neural Network with 3 hidden levels.

Before deciding to choose an SV-Cm for the prediction task shown in this paper, we have compared its performances with other algorithms and using different earthquake catalogues (USGS and ISIDe).

4.1. Test 1: function approximation with USGS data only

The first comparison considers the USGS catalogue from 1976 to 2002: after a short pre-processing we have represented each event by 7 independent variables (time and space: year, month, day, hour, latitude, longitude, depth) and one dependent variable (magnitude). In this case, we have tried a functional approximation of magnitude of each event considering only its space and time of occurrence.

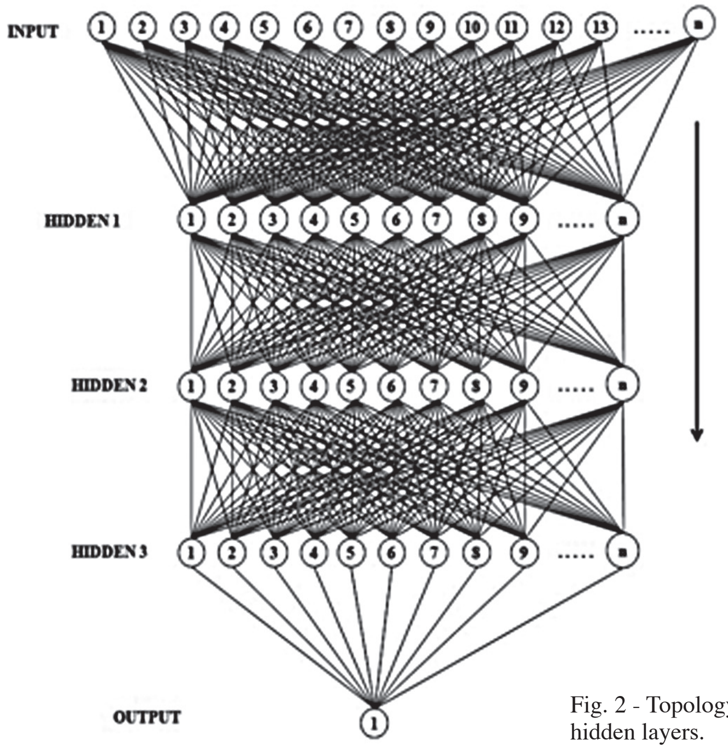

Fig. 2 - Topology of an SV-Cm with 3 hidden layers.

The entire data set was composed of 324,542 events whose distribution is shown in Fig. 3.

The entire data set was split into two halves randomly: a subset was used to train the algorithms and the second subset was used as blind validation test. Each algorithm was evaluated only in test phase using the following much known cost measures:

1. The Root Mean Square Error (*RMSE*), a traditional measure for ANNs:

$$RMSE = \sqrt[2]{\frac{\frac{1}{2} \cdot \sum_{k}^{M} \left(t_k - y_k\right)^2}{M}} \tag{7}$$

with: M = record number;
t_k = the k-th real magnitude; $t \in [0,1]$;
y_k = the k-th predicted magnitude; $t \in [0,1]$.
2. The Linear Correlation Index (*LC*):

$$LC = \frac{\sum_{k=1}^{M} \left(t_k - \bar{t}\right) \cdot \left(y_k - \bar{y}\right)}{\sqrt{\sum_{k=1}^{M} \left(t_k - \bar{t}\right)^2 \cdot \sum_{k=1}^{M} \left(y_k - \bar{y}\right)^2}} \tag{8}$$

with $-1 \leq LC \leq= +1$.

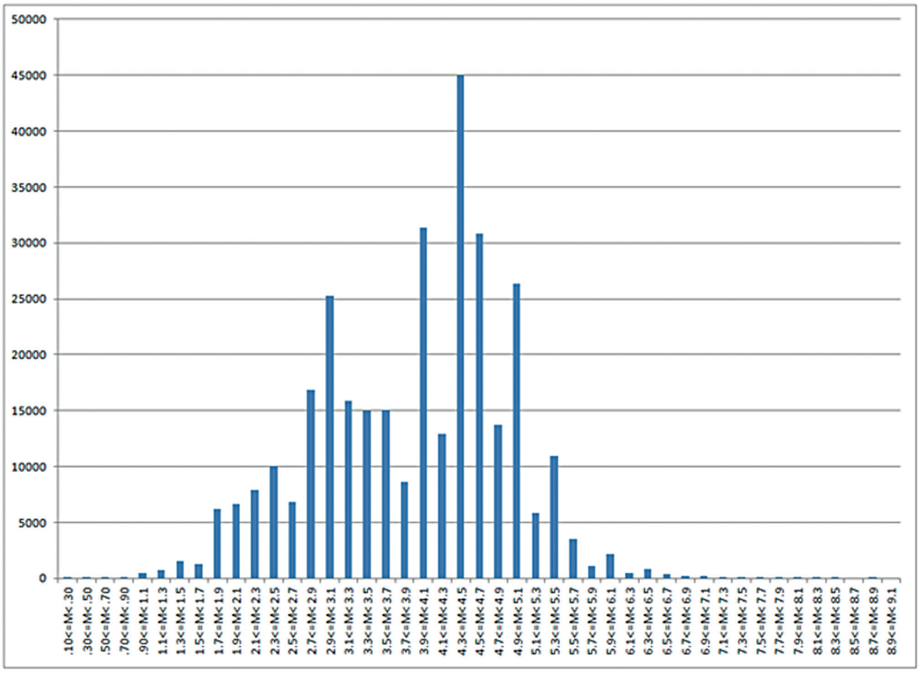

Fig. 3 - Distribution of magnitude of earthquakes from 1996 to 2002 (source: USGS).

3. The Absolute Mean Error (*AbsErr*):

$$AbsErr = F\left(\frac{\sum_{k}^{M}\left|t_k - y_k\right|}{M}\right); \tag{9}$$

with $F()$ = linear function to re-scale the error into the original interval of magnitude.

4. The Weighted Error (*Tau*):

$$Tau = -\sum_{k}^{M}\frac{\left(t_k - y_k\right)^2}{2\sigma^2}; \tag{10}$$

with: $-\infty \leq Tau \leq 0$;

σ^2 = variance.

The following algorithms were chosen for the comparison:

a) an SV-Cm (Buscema and Benzi, 2011);

b) a Back Propagation Multilayer Perceptron (Buscema, 1998a; Le Cun *et al.*, 1998);

c) a Linear Regression (Seber, 2003);

d. a Cart Decision Tree (Breiman *et al.*, 1984; Quinlan, 1986).

Table 2 shows the results: SV-Cm overperforms the other algorithms from all the cost function point of view.

Table 2 - Results of the blind validation of the compared algorithms (in brackets the number of hidden units of the ANNs).

Learning Machine	RMSE	Square Corr.	Linear Corr.	Magnitude ERROR	TAU	%ABS_Err
SV-Cm(32x32x32)	0.043075	0.730514	0.854701	0.387299	-22368.72461	10.62%
MLP_Bp(48)	0.047574	0.671461	0.819427	0.435622	-22368.72461	12.24%
CART	0.050017	0.655396	0.809565	0.453704	-31781.97266	12.29%
Linear Regression	0.068000	0.312867	0.559345	0.659792	-179580.8125	20.28%

4.2. Test 2: function approximation with USGS and ISIDe data together

In this second test, we consider a hybrid data set mixing the data of two catalogues, ISIDe and USGS, from 2005 to 2011. The global data set includes 203,108 events, represented in the same way of Test 1:7 independent variables (space and time) and magnitude as dependent variable, to be estimated. But in this test we split the data set according to a temporal criterion: events from 2005 to 2010 to be used as training set (200,825 events) and the 2011 events to be used for blind validation test (2283 events). In this test, we have compared the performances of the best two algorithms of Test 1: SV-Cm and MLP-Bp.

Table 3 shows the results of this new comparison: SV-Cm has again the best performance also predicting the magnitude of events occurred many months after its training data.

Table 3 - Results of the blind validation of the ANNs on 2001 events occurred one year after the training data.

ANN	RMSE	Absolute Error	TAU	Square Corr.	Linear Corr.
SV-Cm(32x32x32)	0.04672195	0.43691791	-154.0453796	0.88547373	0.94099611
MLP_Bp(48)	0.04697858	0.44407987	-155.7422485	0.87074155	0.93313533

This test has many limits, but it may put in evidence the capability of ANNs to model highly nonlinear processes represented by uncertain, mixed, and imprecise values.

One evident limit of Test 2 is the contribution of many small events to increase the accuracy of the ANN estimation. To reduce this bias, we have repeated the same experiment removing all the events whose magnitude is less than 2.

According to this criterion, the training set is represented by 152,931 events (from 2005 to 2010, ISIDe and USGS catalogues) and the testing set is represented by 1323 events (year 2011, both the catalogues).

Table 4 shows that: there is a decrement of the ANN performances, but the SV-Cm estimation remains quite good.

Table 4 - Results of the blind validation of the ANNs on 2001 events occurred one year after the training data, after the removal from the entire data set of the events with M<2.

ANN	RMSE	Absolute Error	TAU	Square Corr.	Linear Corr.
SV-Cm(32x32x32)	0.05822717	0.40955503	-171.8819122	0.75952047	0.87150472
MLP_Bp(48)	0.06839291	0.50974736	-237.1380615	0.70042348	0.83691305

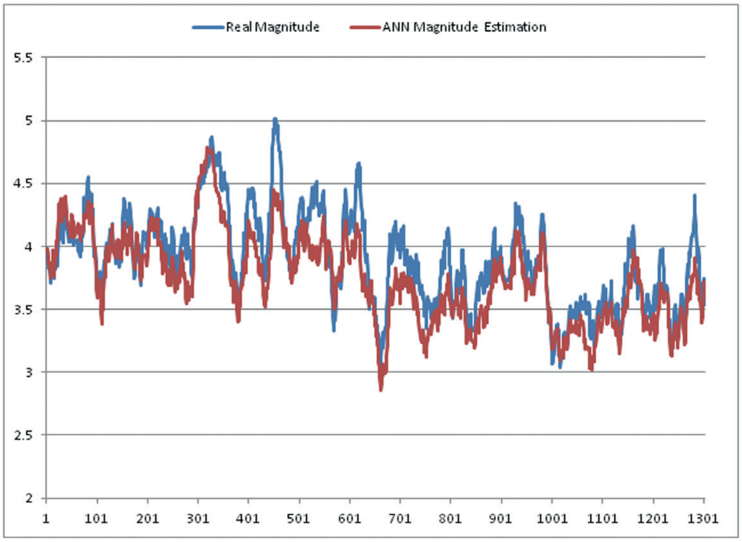

Fig. 4 - Mobile average of the real magnitude and of the ANN estimation of the events in 2011 recorded by USGS and ISIDe catalogues.

If we smooth the real magnitude and the ANN estimations independently with an average mobile window, *W*, where *W*=10, it is possible to see also visually how good are the ANN estimations (see Fig. 4).

$$\overline{m_i} = \frac{1}{W}\sum_t^{W=10} m_t; \qquad (11)$$

$$\overline{e_i} = \frac{1}{W}\sum_t^{W=10} e_t \qquad (12)$$

where:
m_t = real magnitude of the t-th event;
m_i = mobile average of real magnitude of first W events;
e_t = estimated magnitude of the t-th event by ANN;
e_i = mobile average of estimated magnitude of first W events.

These two tests have shown the capability of an advanced ANN to interpolate and to extrapolate the magnitude of many events, defined only by time and space features, from imprecise, mixed and uncertain data coming from different catalogues.

4.3. Test 3: pattern recognition and classification with USGS and ISIDe data together

This new test has been implemented to evaluate the capability of SV-Cm to execute also good pattern recognition (Bishop, 1995; Duda *et. al.*, 2001): we have used the same catalogues of the previous Test 2, but we have split each event into one of three classes, according to the magnitude (see: Tables 5 and 6).

Table 5 - Frequency of distribution of training and testing events in three classes.

USGS + ISIDe catalogues	Class1	Class2	Class3	Tot
	M<3.5	3.5≤*M*<4.5	*M*≥4.5	
Training set (2005-2010)	9201	75095	68635	152931
Test set (2011)	114	492	717	1323

Table 6 - Frequency of distribution of training and testing events in three classes.

USGS + ISIDe catalogues	%	Class1	Class2	Class3	Tot
		M<3.5	3.5≤M<4.5	M≥4.5	
Training set (2005-2010)		6.02%	49.10%	44.88%	100.00%
Test set (2011)		8.62%	37.19%	54.20%	100.00%

We have tested 12 different and known algorithms, coming from five families of machine learning (Hastie *et al*., 2009):

a. ANNs, advanced and classic: SV-Cm (Buscema and Benzi, 2011), Sine Net (Sn) (Buscema *et al*., 2006) and the classic Multi Layer Perceptron with the Back Propagation Learning Law (MLP-Bp) (Buscema, 1998a; Le Cun *et al*., 1998);
b. Decision Trees (Breiman *et al*., 1984): Bagging (Breiman, 1996, 1998; Freund and Schapire, 1997), Random Forest (Breiman, 2001; Livingston, 2005), Logit Boost (Breiman *et al*., 1984), and J48 [also known as C4.5: Quinlan (1986, 1993, 1996)];
c. Instance Learning: kNN with N=3 and Euclidean distance (Kowalski and Bender, 1972; Aha *et al*., 1991);
d. Functions: Logistic Regression (Cessie *et al*., 1992) and a Linear Regression (McLachlan, 1992; Seber, 2003);
e. Probabilistic Nets: Bayes Net (Friedman *et al.,* 1997) and Naïve Bayes (Zhang, 2004).

We have used the following academic softwares to implement all the algorithms: Weka Data Mining Software (Hall *et al*., 2009) and Semeion Software (Buscema, 2013). Table 7 shows the results of this comparison.

Table 7 - Blind testing results of pattern recognition of 3 classes: a comparison among different learning machines (in brackets the number of hidden units of the ANNs).

Type of learning machine	M<3.5	3.5≤M<4.5	M≥4.5	A.Mean	W.Mean	# Errors
SV-Cm (48x48x48)	93.86%	90.04%	92.75%	92.22%	91.84%	108
Sn (48x48)	94.74%	89.63%	92.19%	92.19%	91.46%	113
MLP-Bp (48)	94.74%	88.01%	90.93%	91.23%	90.17%	130
Logit Boost	93.86%	89.63%	89.82%	91.10%	90.10%	131
J48 (C4.5)	88.60%	86.79%	94.14%	89.84%	90.93%	120
Bagging	87.72%	87.20%	94.28%	89.73%	91.08%	118
Random Forest	89.47%	86.79%	92.33%	89.53%	90.02%	132
Naive Bayes	92.98%	87.40%	80.89%	87.09%	84.35%	207
kNN_N=3_D=2.00	77.19%	84.15%	88.84%	83.39%	86.09%	184
Bayes Net	84.21%	79.27%	82.15%	81.88%	81.25%	248
Linear Regression	89.47%	83.54%	62.20%	78.40%	72.49%	364
Logistic	36.84%	89.84%	86.89%	71.19%	83.67%	216

These results show again that advanced ANNs (Sv-Cm and Sn) outperform the other algorithms. But they show also that a global and a worldwide training data set (also with the fusion of two different catalogues) is much useful for an individual and local prediction of the magnitude of a single event.

It is useful at this point to analyze in details the Confusion Matrix of the results generated by the SV-Cm, which has realized the best prediction performance (see Table 8).

Table 8 - Confusion Matrix of the test set for SV-Cm.

Type of algorithm: SV-Cm(48x48x48)	**ANN magnitude estimation**						
Real Magnitude	Confusion Matrix	M<3.5	3.5≤M<4.5	M≥4.5	Row Total	Class Errors	Class Accuracy
	M<3.5	107	7	0	114	7	93.86%
	3.5≤M<4.5	31	443	18	492	49	90.04%
	M≥4.5	5	47	665	717	52	92.75%
	Column Total	143	497	683	1323	Total errors=108	
Arithmetic Mean Accuracy	92.22%						
Weighted Mean Accuracy	91.84%						

The Confusion Matrix shown in Table 6 allows interesting observations:

a. all the events with M<3.5 are correctly predicted, but seven (6.14%), and these errors have all occurred in the close class;
b. the moderate events (3.5≤M<4.5) are sometimes confused with smaller events (6.3%), and only in 18 cases (3.6%) are confused with the big ones;
c. only a very small number of the big events (M≥4.5) are confused with the small ones (0.7%), and a reasonable number of big events are confused with the moderate ones (6.5%).

The behaviour of the SV-Cm and of the main part of other algorithms make evident that a pattern recognition of the earthquake magnitude at short term is at least possible, even not useful.

We understand that an isolated and a retrospective application cannot be a milestone. In any case, it shows a promising use of advanced ANNs in this field. We also understand that, to reach up a stable point for earthquake prediction, our analysis has to be integrated with a deep and a smart data collection with an expertise that we lack. We also think that ANN technology has to be embedded with other methodologies already known and validated in earthquake analysis (Keilis-Borok, 1996; Romanchkova *et al.*, 1998; Kanamori, 2003; Crampin, 2012; Peresan *et al.*, 2012; Radan *et al.*, 2013).

4.4. Test 4: pattern recognition with ISIDe data only

The experiment with the ISIDe database was implemented with a different protocol, because in this case we have the possibility to activate a non-retrospective prediction task: using the ISIDe database we were sure to have, day after day, the real magnitude of every earthquake in Italy.

4.4.1. Research protocol for the Italian database

The data are entered daily into a single database from which two training and testing subsets are extracted for the tuning phase of each neural network. The tuning phase of a network includes training and testing sub phases.

The training subset temporally consists of events until about one month before the day to predict; and the testing subset starts from the day following the last in training and, therefore, will consist of by the events of the last month. Thus, a training of each neural network is done every day. This operation has the task of learning the value of the magnitude of any actual event starting from the input vector, which is made up of the space-time information differently coded for each neural network.

During the training phase, the system verifies the predictive power of the network by tests on the testing subset always saving the network with the best *RMSE*.

After the tuning phase, in the recall phase, each neural network elaborates an artificial data set (prediction). This prediction data set is composed of *n* records equal to a number of quadrants of the affected area to the prediction of the next day. Clearly, each prediction data set will have an input vector congruous with the one used by the neural network in the tuning phase.

The results, therefore, are then calculated by comparing the value of the magnitude of the event really happened with those provided by the corresponding neural networks in each quadrant. The protocol used allows evaluating the daily prediction error obtained by each single neural network.

4.4.2. Input vector codifies

The first input vector consists of 7 variables: year, month, day, hour, latitude, longitude, and depth. The second input vector consists of 15 variables: the previous seven and 8 more statistical variables calculated on the events recorded in each quadrant analysed (Table 9).

Table 9 - Variables.

TRAIN statistical variables:	**TEST statistical variables:**	**PRED statistical variables:**
- TOT events (from T_0 to T_2)	- TOT events (from T_0 to T_{end})	- TOT events (from T_2 to T_{end})
- Max magnitude (from T_0 to T_2)	- Max magnitude (from T_0 to T_{end})	- Max magnitude (from T_2 to T_{end})
- Min magnitude (from T_0 to T_2)	- Min magnitude (from T_0 to T_{end})	- Min magnitude (from T_2 to T_{end})
- Mean magnitude (from T_0 to T_2)	- Mean magnitude (from T_0 to T_{end})	- Mean magnitude (from T_2 to T_{end})
- TOT events (from T_1 to T_2)	- TOT events (from T_2 to T_{end})	- TOT events (last 15 days)
- Max magnitude (from T_1 to T_2)	- Max magnitude (from T_2 to T_{end})	- Max magnitude (last 15 days)
- Min magnitude (from T_1 to T_2)	- Min magnitude (from T_2 to T_{end})	- Min magnitude (last 15 days)
- Mean magnitude (from T_1 to T_2)	- Mean magnitude (from T_2 to T_{end})	- Mean magnitude (last 15 days)

As already said, the file of prediction refers to the next day. Therefore, the only available information is the date that constitutes 3 variables (year, month and day). The other 4 variables (time, latitude, longitude, and depth) are artificially obtained by the information present in the database of real events. It is worth remembering that the number of records in prediction, every day, will be equal to the total number of quadrants where it happened at least a real seismic event.

Having divided the geographic area of interest into quadrants of 400 km^2 (20×20), we have considered the latitude and longitude of the central point of each quadrant. A random value between 0 and 23 was calculated for the time.

The other variable, depth, was calculated as the average value of all events recorded in each quadrant. In summary, the 7 variables considered in the prediction data set are:

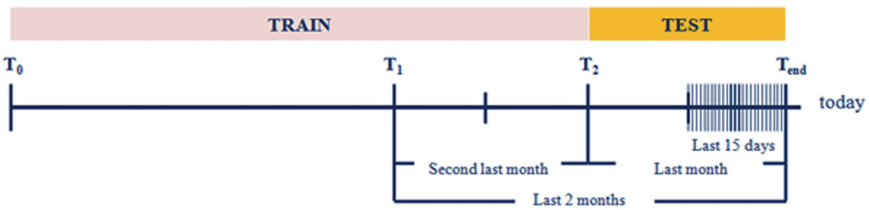

Fig. 5 - Explanation of the pre-processing coding system adopted to increase the 7 input variables of the data set into 15, in order to make daily prediction for each of the 2054 boxes of the assumed grid.

- year, month and day of a day to predict;
- time: random value between 0 and 23;
- latitude and longitude: coordinates of the central point of all quadrants considered;
- depth: average value calculated for the events recorded in each quadrant.

4.4.3. The results of Test 4

In this paragraph we report the prediction results obtained from the neural networks in the period from July 2012 to June 2013. For each day, we have calculated the network results by comparing them with the values of the real events recorded considering 3 various distinctions in Confusion Matrix (Figs. 4 and 5):

a. with respect to 2 classes:
   - low magnitude: $M \leq 2.0$;
   - high magnitude: $M > 2.0$;

b. with respect to 3 classes:
   - low magnitude: $M \leq 1.5$;
   - moderate magnitude: $1.5 < M < 3.0$;
   - high magnitude: $M \geq 3.0$;

c. with respect to 4 classes:
   - null: $M < 0.5$;
   - low magnitude: $0.5 \leq M \leq 1.5$;
   - moderate magnitude: $1.5 < M < 3.0$;
   - high magnitude: $M \geq 3.0$.

From the first Confusion Matrix (with respect to 2 classes, Fig. 7) we can calculate:

- global accuracy, sensitivity and specificity;
- probability of false alarm and probability of missed alarm.

$$Global\_accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)}; \tag{13}$$

$$Sensitivity = \frac{TP}{(TP + FN)}; \tag{14}$$

$$Specificity = \frac{TN}{(TN + FP)}; \tag{15}$$

$$Probability\_false\_alarm = \frac{FP}{(TP + FP)}; \tag{16}$$

$$Probability\_missed\_alarm = \frac{FN}{(TN + FN)}. \tag{17}$$

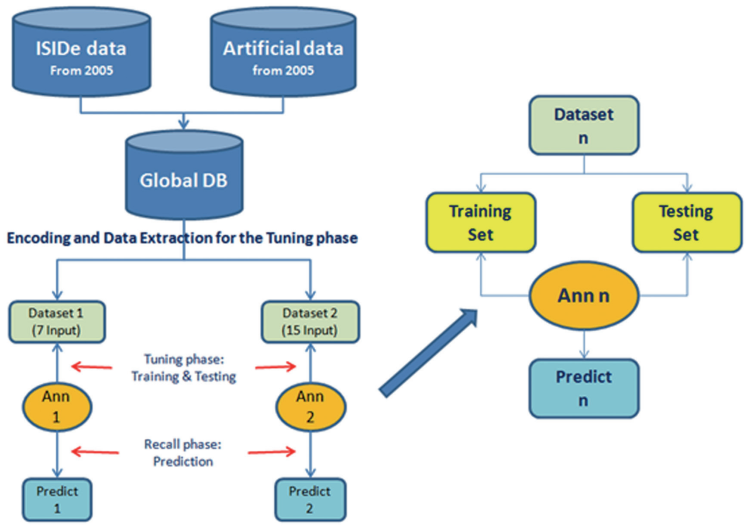

Fig. 6 - On research protocol for ANN evaluation and prediction: on the right side, the structure of the global work for tuning the ANNs and the daily prediction one day before the new event; on left the side, a detail of the tuning phase, where the ANNs are calibrated.

From the second and third Confusion Matrices, for each class, it is possible to calculate:

- global accuracy = (number of values correctly classified) / (number of total predictions);
- accuracy of X = (number of values correctly classified as class X) / (number of values belonging to the class X);
- true ratio of X = (number of values correctly classified as class X) / (number of values classified as class X).

$$Global\_accuracy = \frac{(TN+TL+TM+TH)}{(TN+TL+TM+TH+FN+FL+FM+FH)}; \tag{18}$$

$$TrueRatio\_Low = \frac{TL}{(TL+FL)}; \tag{19}$$

$$TrueRatio\_Moderate = \frac{TM}{(TL+FM)}; \tag{20}$$

$$TrueRatio\_High = \frac{TH}{(TL+FH)}; \tag{21}$$

$$Accuracy\_Low = \frac{TL}{(TL+FN+FM+FH)}; \tag{22}$$

$$Accuracy\_Moderate = \frac{TM}{(TM+FN+FL+FH)}; \tag{23}$$

$$Accuracy\_High = \frac{TH}{(TH+FN+FL+FM)}. \tag{24}$$

We combined the daily results in monthly tables. Thus, for each month the table reports the 3 Confusion Matrices that summarize the results. In addition, for each month, we show two maps of Italy divided into 2054 quadrants: one with the real events and the other with the network output.

Since in one quadrant it is possible that multiple events occur on the same day, to evaluate

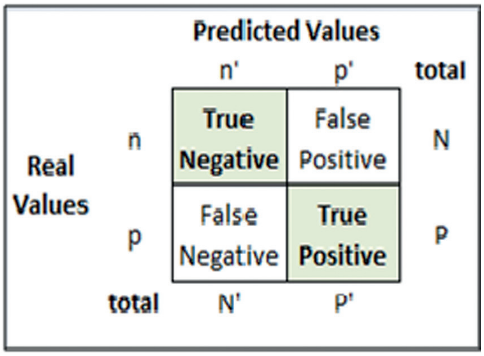

Fig. 7 - Framework of Confusion Matrix with 2 classes (null and moderate quakes): on the columns the ANN estimation and on the rows the real magnitude occurred.

the performance of the network, the maximum value of the network output for each quadrant is compared with the maximum value of magnitude recorded in that quadrant.

All the results obtained, summarized in monthly tables from July 2012 to June 2013, are reported in the Appendix.

5. Conclusions

The results of this research point out two types of considerations, one about the different experiments carried out in this paper, and the other about the next possible use of advanced ANN algorithms in earthquakes prediction.

The experiments show some limits and some interesting points:

a. when we use advanced ANNs with global and representative data (USGS and ISEIDe databases) to predict local events, we reach up interesting results, both when we need to predict the value of the magnitude of a single event, and when we need to classify the single event in a specific class (i.e., small, moderate, severe earthquake). These results do not mean that with these data and with ANN algorithms we are able and ready to make useful daily prediction. These results simply mean that research for earthquake prediction with ANNs, fused with other already developed methods (i.e., Pattern Informatics), and using large worldwide samples of data is a reasonable aim;
b. when we decide to implement the same task using only local data (i.e., the Italian database ISIDe), the performances of ANNs decrease (also if they are not statistically trivial). The reason why, for example, the severe events are systematically underestimated (30%-40% of class accuracy) is evident: in the ISIDe database there are few severe events and many moderate and small events;
c. reasonable conclusion: any local and specific earthquake prediction task has to be considered on a global scale, using, consequently, a worldwide sample of data. This is due to the fact that we do not know *a priori* "which area interacts directly or indirectly with which one". ANNs are suitable to approximate these side effects (Tastle, 2013);
d. earthquake catalogues represent a problem: problem of coding system, of precision and sensitivity, of completeness, etc. Nevertheless, each catalogue seems to present also a systematic error. Consequently, putting together different catalogues is not always a bad

		Predicted Values				
		Null'	Low'	Moderate'	High'	total
Real Values	Null	**True Null**	False Low	False Moderate	False High	N
	Low	False Null	**True Low**	False Moderate	False High	L
	Moderate	False Null	False Low	**True Moderate**	False High	M
	High	False Null	False Low	False Moderate	**True High**	H
	total	N'	L'	M'	H'	

Fig. 8 - Framework of Confusion Matrix with 3 classes (low, moderate and high quakes) and with 4 classes (null, low, moderate and high quakes).

practice. Especially if we mix data coming from different sources after a specific analysis and a consequent planning. ANNs are suitable also to process imprecise data collected with different criteria (Bengio, 2009). Through the statistics of the results of the blind testing phase, it is possible to establish how much the ANN has worked properly in these extreme cases;

e. the input information we consider in this paper is not the only information that needs to be coded for a prediction task and the code system that we adopted is not the only and the best way to code data for ANNs. In the next research, we should take into account local and global information about geology, atmospheric data, volcano dynamics, and electromagnetic fields. Advanced ANNs are able to select spontaneously the significant variables of a complex data set (Buscema *et al.*, 2013a) and they are also able to justify the reasons of their variable pruning [white box versus black box: Buscema *et al.* (2014)];

f. advanced ANNs, trained on similar or different data sets with the same target, may be assembled in an ensemble (Kuncheva, 2004), also with non-ANN algorithms, in order to compose a "parliament of judges" able to refine their decision according to the different competencies of their "members" (Buscema, 1998b; Buscema *et al.*, 2010, 2013b). This architecture could resolve the problem of the incompleteness of data, of their imprecision and also the scarcity of data in some geological field. But above all, this method would increase the precision of the predicted estimates.

In conclusion, this work represents a first step to implement and use advanced ANNs in the arena of earthquake prediction/forecasting. The next step is to expand our cooperative networks to professional geologists. We think that, from the intelligent fusion of ANN technology, other algorithms, and big variety of earthquake data, we will be able to contribute to the growing of the earthquake prediction research area. This area is not important only from a scientific viewpoint: a lot of people die every year because of big earthquakes. A culture that, in the last century, was able to collect the best of the science to build up a nuclear device to destroy human lives has the duty to make much more effort to collect the best of the science to save lives.

Acknowledgments. This study has benefited from funding provided by the Italian Presidenza del Consiglio dei Ministri - Dipartimento della Protezione Civile (DPC), project S3-2012. This paper does not necessarily represent DPC official opinion and policies.

REFERENCES

Aha D.W., Kibler D. and Albert M.K.; 1991: *Instance-based learning algorithms*. Mach. Learn., **6**, 37-66.

Ashif P. and Hojjat A.; 2007: *Neural network models for earthquake magnitude prediction using multiple seismicity indicators*. Int. J. Neural Syst., **17**, 13-33.

Bengio Y.; 2009: *Learning deep architectures for AI*. Mach. Learn., **2**, 1-127.

Bishop C.M.; 1995: *Neural networks for pattern recognition*. Oxford University Press, Oxford, UK, 504 pp.

Breiman L.; 1996: *Bagging predictors*. Mach. Learn., **24**, 123-140.

Breiman L.; 1998: *Arcing classifiers*. Ann. Stat., **26**, 801-849.

Breiman L.; 2001: *Random forest*. Mach. Learn., **45**, 5-32.

Breiman L., Friedman J.H., Olshen R.A. and Stone C.J.; 1984: *Classification and regression trees*. Wadsworth International Group, Belmont, CA, USA, 368 pp.

Buscema M.; 1998a: *Back propagation neural networks*. Subst. Use Misuse, **33**, 233-270.

Buscema M.; 1998b: *MetaNet: the theory of independent judges*. Subst. Use Misuse, **33**, 439-461.

Buscema M.; 2013: *Supervised ANNs & Organisms*. Semeion Software n. 12, ver. 21.5, Rome, Italy.

Buscema P.M. and Benzi R.; 2011: *Quakes prediction using highly non linear systems and a minimal dataset*. In: Buscema P.M. and Ruggieri M. (eds), Advanced networks, algorithms and modeling for earthquake prediction, River Publishers, Aalborg, Denmark, pp. 41-66.

Buscema M., Terzi S. and Breda M.; 2006: *Using sinusoidal modulated weights improve feed-forward neural network performances in classification and functional approximation problems*. WSEAS Trans. Inf. Sci. Appl., **3**, 885-893.

Buscema M., Terzi S. and Tastle W.J.; 2010: *A new meta-classifier*. In: Fuzzy Information Processing Society (NAFIPS), Annual Meeting of the North American, Toronto, Canada, pp. 1-7, doi:10.1109/NAFIPS.2010.5548298.

Buscema M., Breda M. and Lodwick W.; 2013a: *Training With Input Selection and Testing (TWIST) algorithm: a significant advance in pattern recognition performance of machine learning*. J. Intell. Learn. Syst. Appl., **5**, 29-38.

Buscema M., Tastle W.J. and Terzi S.; 2013b: *Meta Net: a new meta-classifier family*. In: Tastle W.J. (ed), Data mining applications using artificial adaptive systems, Springer Science+Business Media, New York, NY, USA, pp. 141-182, doi:10.1007/978-1-4614-4223-3_5.

Buscema M., Consonni V., Ballabio D., Mauri A., Massini G., Breda M. and Todeschini R.; 2014: *K-CM: a new artificial neural network. Application to supervised pattern recognition*. Chemom. Intell. Lab. Syst., **138**, 110-119.

Castello B., Selvaggi G., Chiarabba C. and Amato A.; 2006: *Catalogo della Sismicità Italiana (CSI) 1981-2002, versione 1.1*. INGV-CNT, Roma, Italy.

Cessie S. and Van Houwelingen J.C.; 1992: *Ridge estimators in logistic regression*. Appl. Stat., **41**, 191-201.

Crampin S.; 2012: *Comment on the report "Operational Earthquake Forecasting" by the International Commission on earthquake forecasting for Civil Protection*. Ann. Geophys., **55**, doi:10.4401/ag-5516.

Duda R.O., Hart P.E. and Stork D.G.; 2001: *Pattern classification*. Wiley and Sons, New York, NY, USA, 654 pp.

Freund Y. and Schapire R.E.; 1997: *A decision-theoretic generalization of on-line learning and an application to boosting*. J. Comput. Syst. Sci., **55**, 119-139.

Friedman N., Geiger D. and Goldszmidt M.; 1997: *Bayesian networks classifiers*. Mach. Learn., **29**, 131-163.

Hall M., Frank E., Holmes G., Pfahringer B., Reutmann P. and Witten I.H.; 2009: *The WEKA data mining software: an update*. In: ACM SIGKDD explorations newsletter, vol. 11, pp. 10-18.

Hastie T., Tibshirani R. and Friedman J.H.; 2009: *The elements of statistical learning: data mining, inference, and prediction*. SpringerVerlag, New York, NY, USA, 745 pp.

Hinton G.E., Osindero S. and Teh Y-W.; 2006: *A fast learning algorithm for deep belief nets*. Neural Comput., **18**, 1527-1554.

ISIDe Working Group; 2010: *Italian Seismological Instrumental and parametric DatabasE*. INGV, <http://ISIDe.rm.ingv.it>.

Jordan T.H., Chen Y.T., Gasparini P., Madariaga R., Main I., Marzocchi W., Papadopoulos G., Sobelev G., Yamaoka K. and Zschau J.; 2011: *Operational earthquake forecasting: state of knowledge and guidelines for utilization*. Ann. Geophys., **54**, 315-391, doi:10.4401/ag-5350.

Kagan Y.Y.; 1997: *Are earthquakes predictable? Special section-assessment of schemes for earthquake prediction*. Geophys. J. Int., **131**, 505-525.

Kagan Y.Y. and Jackson D.D.; 2000: *Probabilistic forecasting of earthquakes*. Geophys. J. Int., **143**, 438-453.

Kanamori H.; 2003: *Earthquake prediction: an overview*. In: Lee W.H.K., Kanamori H., Jennings J.C. and Kisslinger C. (eds), Int. Handb. Earthquake Eng. Seismol., Academic Press, San Diego, CA, USA, Part B, 1453 pp.

Keilis-Borok V.I.; 1996: *Intermediate-term earthquake prediction*. Proc. Natl. Acad. Sci. U.S.A., **93**, 3748-3755.

Kowalski B.R. and Bender C.F.; 1972: *The K-Nearest Neighbor classification rule (Pattern Recognition) applied to nuclear magnetic resonance spectral interpretation*. Anal. Chem., **44**, 1405-1411.

Kuncheva L.I.; 2004: *Combining pattern classifiers: methods and algorithms*. Wiley and Sons, New York, NY, USA, 376 pp.

Le Cun Y., Kanter L. and Solla S.A.; 1991: *Second order properties of error surface learning time and generalization*. Adv. Neural Inf. Process. Syst., **3**, 918-924.

Le Cun Y., Bottou L., Orr G.B. and Muller K.R.; 1998: *Efficient backprop*. In: Orr G. and Muller K.R. (eds), Neural networks: tricks of the trade, Springer, New York, NY, USA, pp. 9-50.

Livingston F.; 2005: *Implementing Breiman's random forest algorithm into Weka*. In: Mach. Learn. Conf. Papers, ECE591Q.

McLachlan G.J.; 1992: *Discriminant analysis and statistical pattern recognition*. Wiley, New York, NY, USA, 526 pp.

Peresan A., Kossobokov V.G. and Panza G.F.; 2012: *Operational earthquake forecast/prediction*. Rendiconti Lincei, Scienze Fisiche e Naturali, **23**, 131-138, doi:10.1007/s12210-012-0171-7.

Quinlan J.R.; 1986: *Induction of decision trees*. Mach. Learn., **1**, 81-106.

Quinlan J.R.; 1993: *C4.5: programs for machine learning*. Morgan Kaufman Publ. Inc., San Mateo, CA, USA, 302 pp.

Quinlan J.R.; 1996: *Improve use of continuous attributes in C4.5*. J. Artif. Intell. Res., **4**, 77-90.

Radan M.Y., Hamzehloo H., Peresan A., Zare M. and Zafarani H.; 2013: *Assessing performances of pattern informatics method: a retrospective analysis for Iran and Italy*. Nat. Hazards, doi:10.1007/s11069-013-0660-8.

Raiko T., Valpola H. and LeCun Y.; 2012: *Deep learning made easier by linear transformations in perceptrons*. In: Proc. 15th Int. Conf. Artif. Intell. Stat., La Palma, Canary Islands, Spain, vol. 22, pp. 924-932.

Romachkova L., Kossobokov V.G., Panza G.F. and Costa G.; 1998: *Intermediate-term predictions of earthquakes in Italy: algorithm M8*. Pure Appl. Geophys., **152**, 37-55, doi:0033-4553/98/010037.

Rumelhart D.E., Hinton G.E. and Williams R.J.; 1986: *Learning internal representations by error propagation*. In: Rumelhart D.E. and McClelland J.L. (eds), Parallel Distributed Processing, MIT Press, Cambridge, MA, USA, vol. 1, pp. 318-362.

Seber G.A.F.; 2003: *Linear regression analysis, 2nd ed*. Wiley, Hoboken, NJ, USA, 557 pp.

Sharma M.L. and Arora M.K.; 2005: *Prediction of seismicity cycles in the Himalaya using artificial neural network*. Acta Geophys. Pol., **53**, 299-309.

Suratgar A.A., Setoudeh F., Salemi A.H. and Negarestani A.; 2008: *Magnitude of earthquake prediction using neural network*. In: Guo M.Z., Zhao L. and Wang L.P. (eds), 4th Int. Conf. Nat. Comput., Los Alamitos, CA, USA, vol. 2, pp. 448-452, doi:10.1109/ICNC.2008.781.

Tastle WJ. (ed); 2013: *Data mining applications using Artificial Adaptive Systems*. Springer Science + Business Media, New York., U.S.A., doi: 10.1007/978-1-4614-4223-3_1.

Zhang H.; 2004: *The optimality of naive Bayes*. In: Proc. 17th Int. Florida Artif. Intell. Res. Soc. Conf., Am. Ass. Artif. Intell., Menlo Park, CA, USA, pp. 562-567.

Corresponding author: Guido Maurelli
Semeion Research Centre of Sciences of Communication
Via Sersale 117, 00128 Roma, Italy
Phone: +39 06 50652350; fax: +39 06 5060064; e-mail: g.maurelli@semeion.it

Appendix: Results in tables and maps

Results of July 2012

July 2012		Predicted Values		Total
		M ≤ 2.0	*M > 2.0*	
Real values	*M ≤ 2.0*	**63127**	391	63518
	M > 2.0	18	**138**	156
Total		63145	529	**63674**

Accuracy	99.36%
Sensitivity	88.46%
Specificity	99.38%
Prob False Alarm	73.91%
Prob Missed Alarm	0.03%

Confusion Matrix with respect to 2 classes.

July 2012		Predicted Values			Total
		LOW	*Moderate*	*High*	
Real Values	*LOW (M ≤ 1.5)*	**63079**	198	16	63293
	Moderate (1.5 < M < 3.0)	61	**282**	20	363
	High(M ≥ 3.0)	3	11	**4**	18
Total		63143	491	40	**63674**

	Total	*LOW*	*Moderate*	*High*
Accuracy	99.51%	99.66%	77.69%	22.22%
True Ratio	-	99.90%	57.43%	10.00%

Confusion Matrix with respect to 3 classes.

July 2012		Predicted Values				Total
		Null	*LOW*	*Moderate*	*High*	
Real Values	*Null (M < 0.5)*	**62783**	56	0	0	62839
	LOW (0.5 ≤ M ≤ 1.5)	2	**238**	198	16	454
	Moderate (1.5 < M < 3.0)	5	56	**282**	20	363
	High(M ≥ 3.0)	2	1	11	**4**	18
Total		62792	351	491	40	**63674**

	Total	*Null*	*LOW*	*Moderate*	*High*
Accuracy	99.42%	99.91%	52.42%	77.69%	22.22%
True Ratio	-	99.99%	67.81%	57.43%	10.00%

Confusion Matrix with respect to 4 classes.

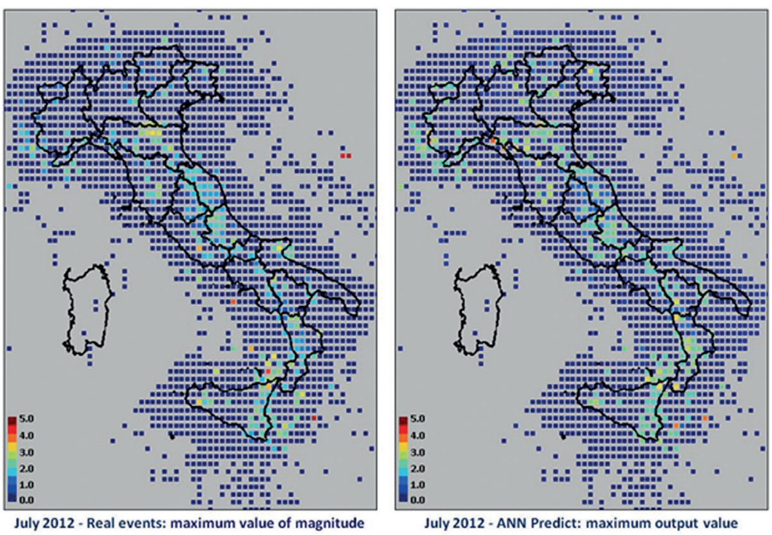

Fig. A1 - Results of July 2012: a) maximum magnitude of real events; b) maximum magnitude predicted by ANN.

Results of August 2012

August 2012		Predicted Values		Total
		$M \leq 2.0$	$M > 2.0$	
Real values	$M \leq 2.0$	**63127**	408	63535
	$M > 2.0$	26	**113**	139
Total		63153	521	**63674**

Accuracy	99.32%
Sensitivity	81.29%
Specificity	99.36%
Prob False Alarm	78.31%
Prob Missed Alarm	0.04%

Confusion Matrix with respect to 2 classes.

August 2012		Predicted Values			Total
		LOW	*Moderate*	*High*	
Real Values	*LOW* ($M \leq 1.5$)	**63075**	217	41	63333
	Moderate ($1.5 < M < 3.0$)	72	**201**	51	324
	High ($M \geq 3.0$)	4	7	**6**	17
Total		63151	425	98	**63674**

	Total	*LOW*	*Moderate*	*High*
Accuracy	99.38%	99.59%	62.04%	35.29%
True Ratio	-	99.88%	47.29%	6.12%

Confusion Matrix with respect to 3 classes.

August 2012		Predicted Values				Total
		Null	*LOW*	*Moderate*	*High*	
Real Values	*Null* ($M < 0.5$)	**62844**	18	4	0	62866
	LOW ($0.5 \leq M \leq 1.5$)	9	**204**	213	41	467
	Moderate ($1.5 < M < 3.0$)	4	68	**201**	51	324
	High ($M \geq 3.0$)	1	3	7	**6**	17
Total		62858	293	425	98	**63674**

	Total	*Null*	*LOW*	*Moderate*	*High*
Accuracy	99.34%	99.97%	43.68%	62.04%	35.29%
True Ratio	-	99.98%	69.62%	47.29%	6.12%

Confusion Matrix with respect to 4 classes.

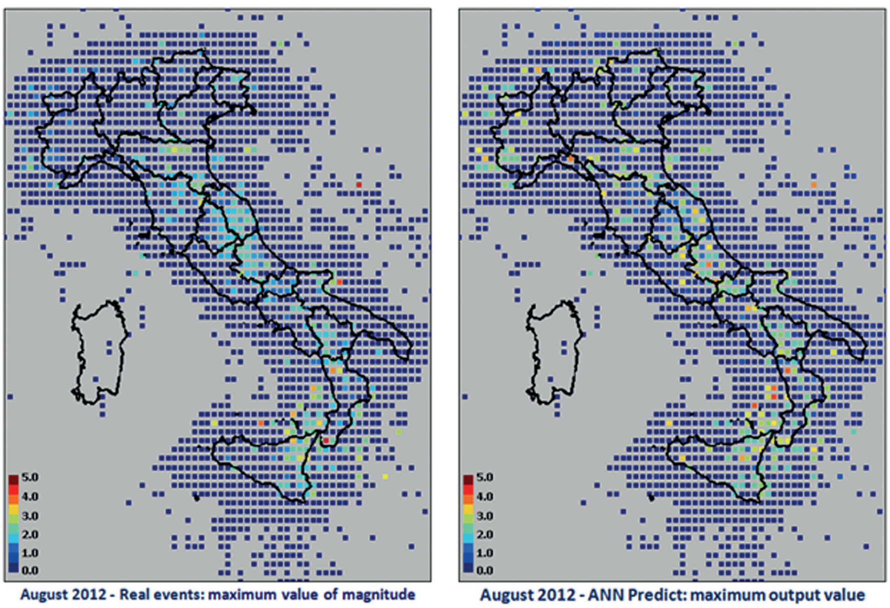

Fig. A2 - Results of August 2012: a) maximum magnitude of real events; b) maximum magnitude predicted by ANN.

Results of September 2012

September 2012		Predicted Values		Total
		$M \le 2.0$	$M > 2.0$	
Real values	$M \le 2.0$	**60952**	542	61494
	$M > 2.0$	2	**124**	126
Total		60954	666	**61620**

Accuracy	99.12%
Sensitivity	98.41%
Specificity	99.12%
Prob False Alarm	81.38%
Prob Missed Alarm	0.00%

Confusion Matrix with respect to 2 classes.

September 2012		Predicted Values			Total
		LOW	*Moderate*	*High*	
Real Values	*LOW ($M \le 1.5$)*	**60911**	318	64	61293
	Moderate ($1.5 < M < 3.0$)	11	**250**	48	309
	High($M \ge 3.0$)	0	12	**6**	18
Total		60922	580	118	**61620**

	Total	*LOW*	*Moderate*	*High*
Accuracy	99.26%	99.38%	80.91%	33.33%
True Ratio	-	99.98%	43.10%	5.08%

Confusion Matrix with respect to 3 classes.

September 2012		Predicted Values				Total
		Null	*LOW*	*Moderate*	*High*	
Real Values	*Null ($M < 0.5$)*	**60855**	1	15	2	60873
	LOW ($0.5 \le M \le 1.5$)	4	**51**	303	62	420
	Moderate ($1.5 < M < 3.0$)	5	6	**250**	48	309
	High($M \ge 3.0$)	0	0	12	**6**	18
Total		60864	58	580	118	**61620**

	Total	*Null*	*LOW*	*Moderate*	*High*
Accuracy	99.26%	99.97%	12.14%	80.91%	33.33%
True Ratio	-	99.99%	87.93%	43.10%	5.08%

Confusion Matrix with respect to 4 classes

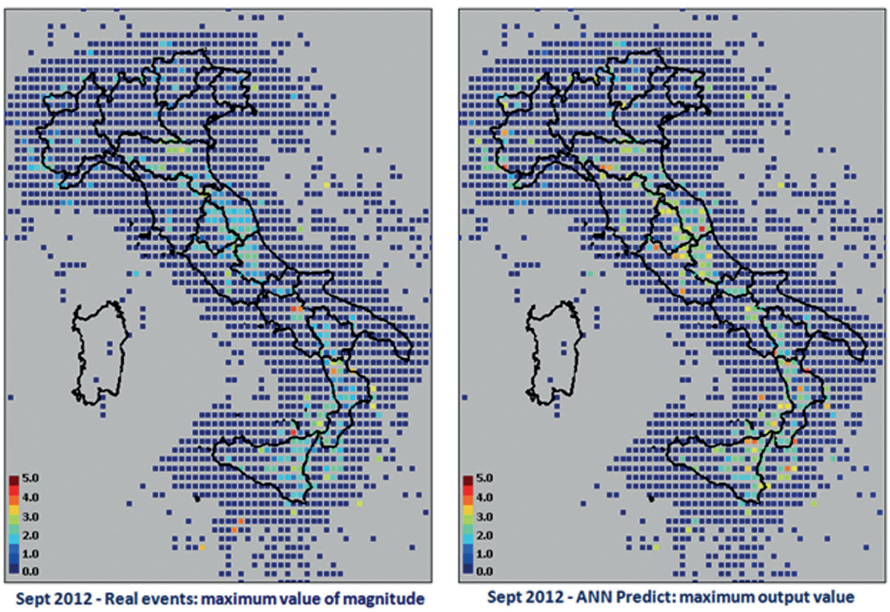

Fig. A3 - Results of September 2012: a) maximum magnitude of real events; b) maximum magnitude predicted by ANN.

Results of October 2012

October 2012		Predicted Values		Total
		M ≤ 2.0	*M > 2.0*	
Real values	*M ≤ 2.0*	**63007**	536	63543
	M > 2.0	10	**121**	131
Total		63017	657	**63674**

Accuracy	99.14%
Sensitivity	92.37%
Specificity	99.16%
Prob False Alarm	81.58%
Prob Missed Alarm	0.02%

Confusion Matrix with respect to 2 classes.

October 2012		Predicted Values			Total
		LOW	*Moderate*	*High*	
Real Values	*LOW (M ≤ 1.5)*	**62980**	286	63	63329
	Moderate (1.5 < M < 3.0)	35	**216**	74	325
	High(M ≥ 3.0)	1	11	**8**	20
Total		63016	513	145	**63674**

	Total	*LOW*	*Moderate*	*High*
Accuracy	99.26%	99.45%	66.46%	40.00%
True Ratio	-	99.94%	42.11%	5.52%

Confusion Matrix with respect to 3 classes.

October 2012		Predicted Values				Total
		Null	*LOW*	*Moderate*	*High*	
Real Values	*Null (M < 0.5)*	**62766**	137	2	3	62908
	LOW (0.5 ≤ M ≤ 1.5)	1	**76**	284	60	421
	Moderate (1.5 < M < 3.0)	5	30	**216**	74	325
	High(M ≥ 3.0)	0	1	11	**8**	20
Total		62772	244	513	145	**63674**

	Total	*Null*	*LOW*	*Moderate*	*High*
Accuracy	99.05%	99.77%	18.05%	66.46%	40.00%
True Ratio	-	99.99%	31.15%	42.11%	5.52%

Confusion Matrix with respect to 4 classes.

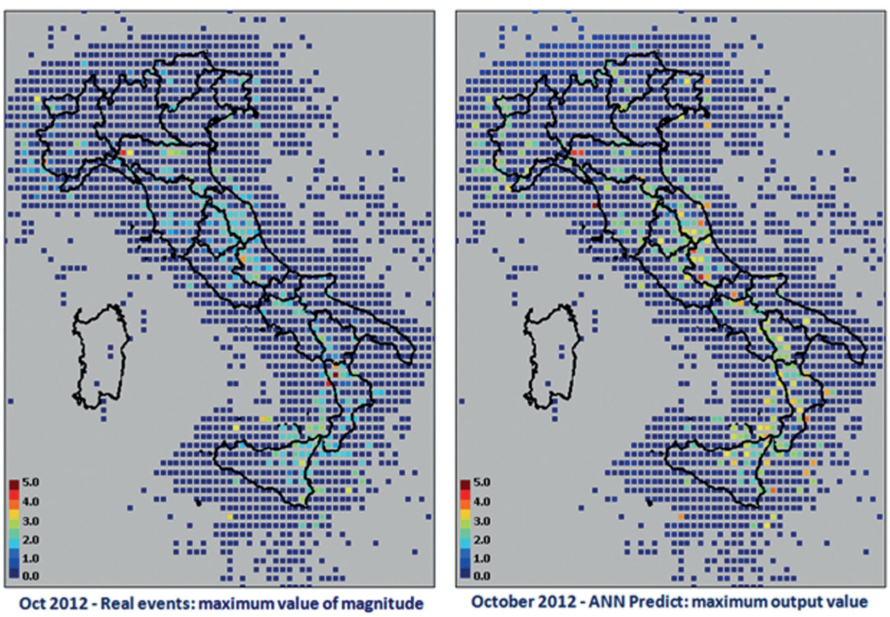

Fig. A4 - Results of October 2012: a) maximum magnitude of real events; b) maximum magnitude predicted by ANN.

Results of November 2012

November 2012		Predicted Values		Total
		$M \le 2.0$	$M > 2.0$	
Real values	$M \le 2.0$	**61092**	391	61483
	$M > 2.0$	8	**129**	137
Total		61100	520	**61620**

Accuracy	99.35%
Sensitivity	94.16%
Specificity	99.36%
Prob False Alarm	75.19%
Prob Missed Alarm	0.01%

Confusion Matrix with respect to 2 classes.

November 2012		Predicted Values			Total
		LOW	*Moderate*	*High*	
Real Values	*LOW* ($M \le 1.5$)	**61062**	209	35	61306
	Moderate ($1.5 < M < 3.0$)	37	**188**	64	289
	High($M \ge 3.0$)	1	15	**9**	25
Total		61100	412	108	**61620**

	Total	*LOW*	*Moderate*	*High*
Accuracy	99.41%	99.60%	65.05%	36.00%
True Ratio	-	99.94%	45.63%	8.33%

Confusion Matrix with respect to 3 classes.

November 2012		Predicted Values				Total
		Null	*LOW*	*Moderate*	*High*	
Real Values	*Null* ($M < 0.5$)	**60967**	3	3	1	60974
	LOW ($0.5 \le M \le 1.5$)	4	**88**	206	34	332
	Moderate ($1.5 < M < 3.0$)	0	37	**188**	64	289
	High($M \ge 3.0$)	1	0	15	**9**	25
Total		60972	128	412	108	**61620**

	Total	*Null*	*LOW*	*Moderate*	*High*
Accuracy	99.40%	99.99%	26.51%	65.05%	36.00%
True Ratio	-	99.99%	68.75%	45.63%	8.33%

Confusion Matrix with respect to 4 classes.

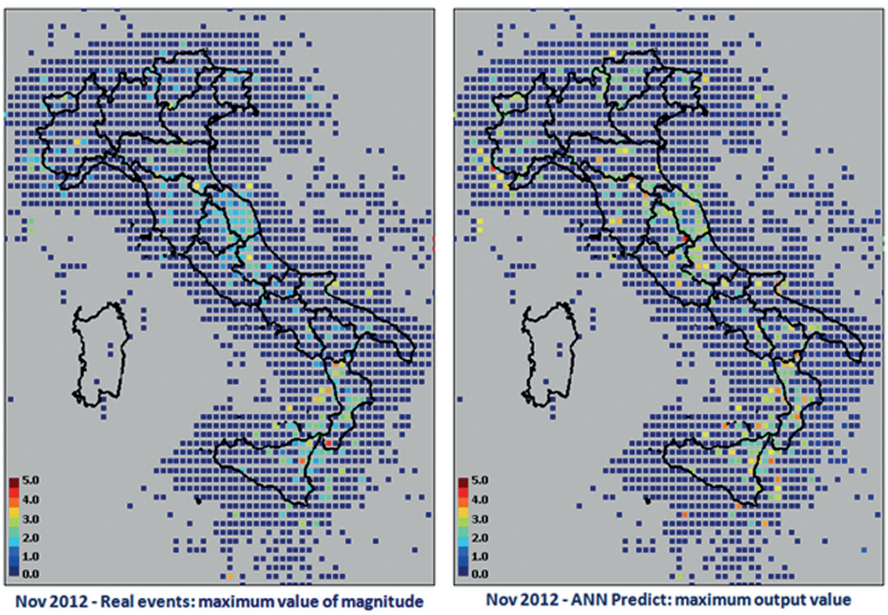

Fig. A5 - Results of November 2012: a) maximum magnitude of real events; b) maximum magnitude predicted by ANN.

Results of December 2012

December 2012		Predicted Values $M \le 2.0$	Predicted Values $M > 2.0$	Total
Real values	$M \le 2.0$	**63141**	401	63542
Real values	$M > 2.0$	18	**117**	135
Total		63159	518	**63677**

Accuracy	99.34%
Sensitivity	86.67%
Specificity	99.37%
Prob False Alarm	77.41%
Prob Missed Alarm	0.03%

Confusion Matrix with respect to 2 classes.

December 2012		Predicted Values *LOW*	Predicted Values *Moderate*	Predicted Values *High*	Total
Real Values	*LOW* ($M \le 1.5$)	**63119**	228	10	63357
Real Values	*Moderate* ($1.5 < M < 3.0$)	39	**261**	8	308
Real Values	*High*($M \ge 3.0$)	1	10	**1**	12
Total		63159	499	19	**63677**

	Total	*LOW*	*Moderate*	*High*
Accuracy	99.54%	99.62%	84.74%	8.33%
True Ratio	-	99.94%	52.30%	5.26%

Confusion Matrix with respect to 3 classes.

December 2012		Predicted Values *Null*	Predicted Values *LOW*	Predicted Values *Moderate*	Predicted Values *High*	Total
Real Values	*Null* ($M < 0.5$)	**63082**	1	0	0	63083
Real Values	*LOW* ($0.5 \le M \le 1.5$)	4	**32**	228	10	274
Real Values	*Moderate* ($1.5 < M < 3.0$)	11	28	**261**	8	308
Real Values	*High*($M \ge 3.0$)	1	0	10	**1**	12
Total		63098	61	499	19	**63677**

	Total	*Null*	*LOW*	*Moderate*	*High*
Accuracy	99.53%	100.00%	11.68%	84.74%	8.33%
True Ratio	-	99.97%	52.46%	52.30%	5.26%

Confusion Matrix with respect to 4 classes.

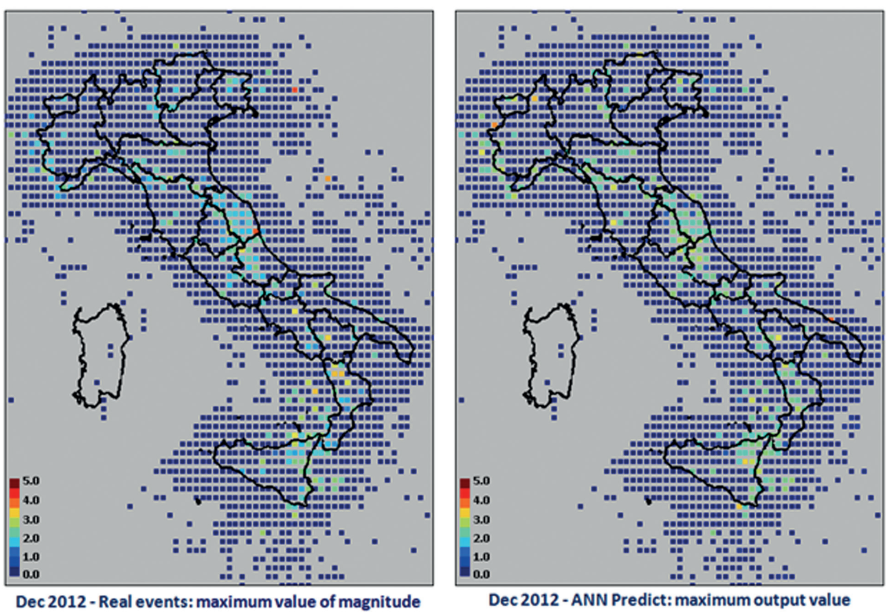

Fig. A6 - Results of December 2012: a) maximum magnitude of real events; b) maximum magnitude predicted by ANN.

Results of January 2013

January 2013		Predicted Values		Total
		$M \leq 2.0$	$M > 2.0$	
Real values	$M \leq 2.0$	**63239**	313	63552
	$M > 2.0$	16	**107**	123
Total		63255	420	**63675**

Accuracy	99.48%
Sensitivity	86.99%
Specificity	99.51%
Prob False Alarm	74.52%
Prob Missed Alarm	0.03%

Confusion Matrix with respect to 2 classes.

January 2013		Predicted Values			Total
		LOW	*Moderate*	*High*	
Real Values	*LOW* ($M \leq 1.5$)	**63209**	169	11	63389
	Moderate ($1.5 < M < 3.0$)	44	**184**	42	270
	High ($M \geq 3.0$)	1	13	**2**	16
Total		63254	366	55	**63675**

	Total	*LOW*	*Moderate*	*High*
Accuracy	99.56%	99.72%	68.15%	12.50%
True Ratio	-	99.93%	50.27%	3.64%

Confusion Matrix with respect to 3 classes.

January 2013		Predicted Values				Total
		Null	*LOW*	*Moderate*	*High*	
Real Values	*Null* ($M < 0.5$)	**63056**	1	1	0	63058
	LOW ($0.5 \leq M \leq 1.5$)	4	**148**	168	11	331
	Moderate ($1.5 < M < 3.0$)	4	40	**184**	42	270
	High ($M \geq 3.0$)	0	1	13	**2**	16
Total		63064	190	366	55	**63675**

	Total	*Null*	*LOW*	*Moderate*	*High*
Accuracy	99.55%	100.00%	44.71%	68.15%	12.50%
True Ratio	-	99.99%	77.89%	50.27%	3.64%

Confusion Matrix with respect to 4 classes.

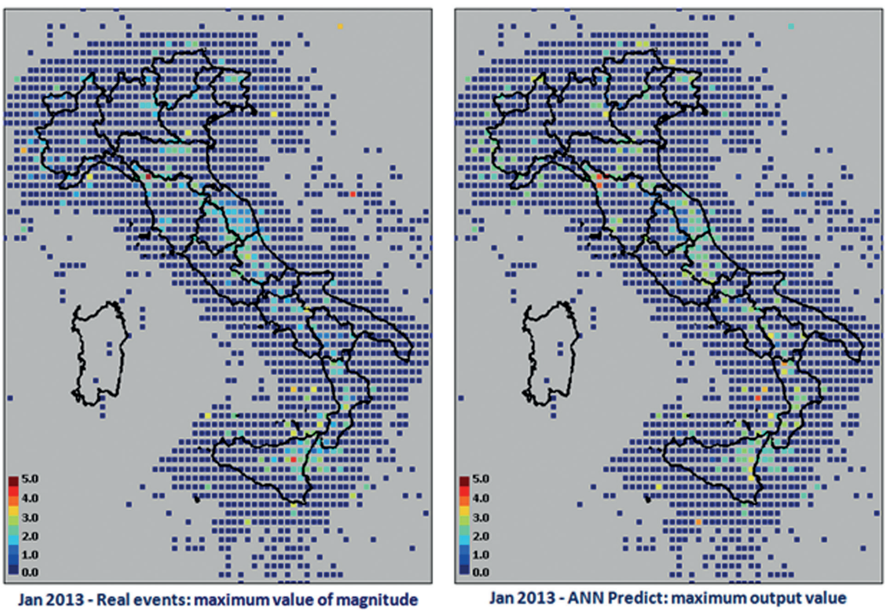

Fig. A7 - Results of January 2013: a) maximum magnitude of real events; b) maximum magnitude predicted by ANN.

Results of February 2013

February 2013		Predicted Values		Total
		$M \le 2.0$	$M > 2.0$	
Real values	$M \le 2.0$	**57091**	321	57412
	$M > 2.0$	9	**92**	101
Total		57100	413	**57513**

Accuracy	99.43%
Sensitivity	91.09%
Specificity	99.44%
Prob False Alarm	77.72%
Prob Missed Alarm	0.02%

Confusion Matrix with respect to 2 classes.

February 2013		Predicted Values			Total
		LOW	*Moderate*	*High*	
Real Values	*LOW* ($M \le 1.5$)	**57062**	159	9	57230
	Moderate ($1.5 < M < 3.0$)	35	**213**	21	269
	High ($M \ge 3.0$)	1	9	**4**	14
Total		57098	381	34	**57513**

	Total	*LOW*	*Moderate*	*High*
Accuracy	99.59%	99.71%	79.18%	28.57%
True Ratio	-	99.94%	55.91%	11.76%

Confusion Matrix with respect to 3 classes.

February 2013		Predicted Values				Total
		Null	*LOW*	*Moderate*	*High*	
Real Values	*Null* ($M < 0.5$)	**56949**	2	0	0	56951
	LOW ($0.5 \le M \le 1.5$)	4	**107**	159	9	279
	Moderate ($1.5 < M < 3.0$)	3	32	**213**	21	269
	High ($M \ge 3.0$)	0	1	9	**4**	14
Total		56956	142	381	34	**57513**

	Total	*Null*	*LOW*	*Moderate*	*High*
Accuracy	99.58%	100.00%	38.35%	79.18%	28.57%
True Ratio	-	99.99%	75.35%	55.91%	11.76%

Confusion Matrix with respect to 4 classes.

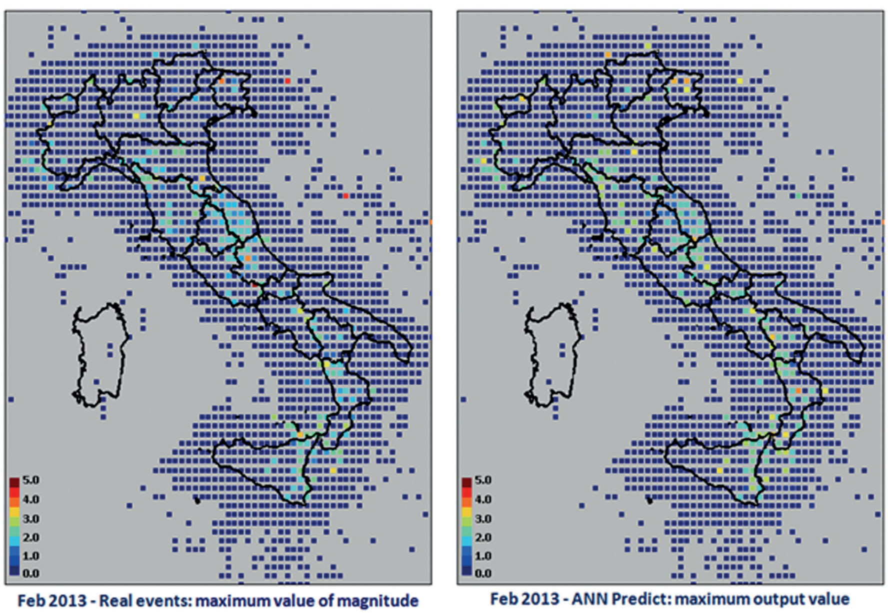

Fig. A8 - Results of February 2013: a) maximum magnitude of real events; b) maximum magnitude predicted by ANN.

Results of March 2013

March 2013		Predicted Values		Total
		$M \leq 2.0$	$M > 2.0$	
Real values	$M \leq 2.0$	**63236**	326	63562
	$M > 2.0$	10	**105**	115
Total		63246	431	**63677**

Accuracy	99.47%
Sensitivity	91.30%
Specificity	99.49%
Prob False Alarm	75.64%
Prob Missed Alarm	0.02%

Confusion Matrix with respect to 2 classes.

March 2013		Predicted Values			Total
		LOW	*Moderate*	*High*	
Real Values	*LOW* ($M \leq 1.5$)	**63202**	114	71	63387
	Moderate ($1.5 < M < 3.0$)	43	**139**	86	268
	High($M \geq 3.0$)	1	16	**5**	22
Total		63246	269	162	**63677**

	Total	*LOW*	*Moderate*	*High*
Accuracy	99.48%	99.71%	51.87%	22.73%
True Ratio	-	99.93%	51.67%	3.09%

Confusion Matrix with respect to 3 classes.

March 2013		Predicted Values				Total
		Null	*LOW*	*Moderate*	*High*	
Real Values	*Null* ($M < 0.5$)	**63058**	2	1	1	63062
	LOW ($0.5 \leq M \leq 1.5$)	5	**137**	113	70	325
	Moderate ($1.5 < M < 3.0$)	3	40	**139**	86	268
	High($M \geq 3.0$)	0	1	16	**5**	22
Total		63066	180	269	162	**63677**

	Total	*Null*	*LOW*	*Moderate*	*High*
Accuracy	99.47%	99.99%	42.15%	51.87%	22.73%
True Ratio	-	99.99%	76.11%	51.67%	3.09%

Confusion Matrix with respect to 4 classes.

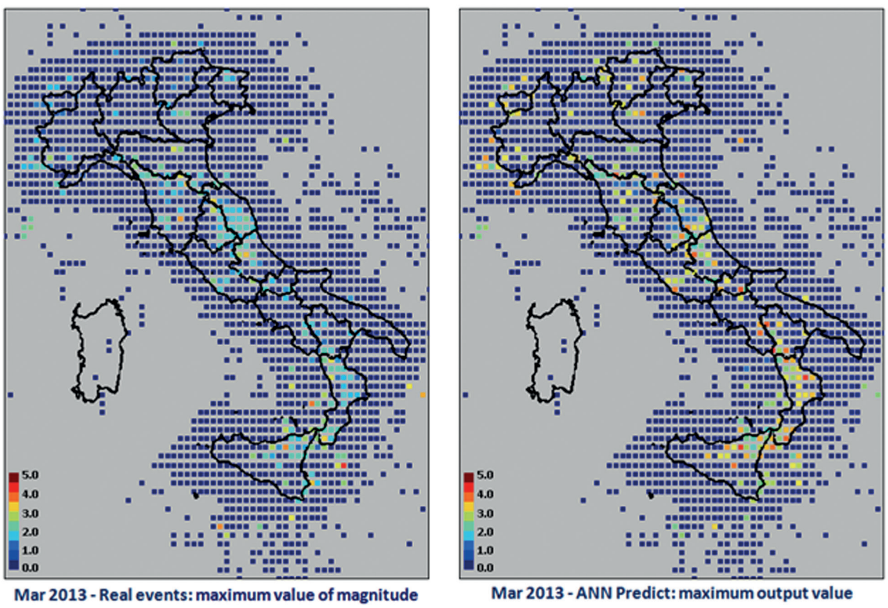

Fig. A9 - Results of March 2013: a) maximum magnitude of real events; b) maximum magnitude predicted by ANN.

Results of April 2013

April 2013		Predicted Values		Total
		$M \leq 2.0$	$M > 2.0$	
Real values	$M \leq 2.0$	**60995**	510	61505
	$M > 2.0$	14	**105**	119
Total		61009	615	**61624**

Accuracy	99.15%
Sensitivity	88.24%
Specificity	99.17%
Prob False Alarm	82.93%
Prob Missed Alarm	0.02%

Confusion Matrix with respect to 2 classes.

April 2013		Predicted Values			Total
		LOW	*Moderate*	*High*	
Real Values	*LOW ($M \leq 1.5$)*	**60958**	365	3	61326
	Moderate ($1.5 < M < 3.0$)	44	**221**	18	283
	High($M \geq 3.0$)	2	10	**3**	15
Total		61004	596	24	**61624**

	Total	*LOW*	*Moderate*	*High*
Accuracy	99.28%	99.40%	78.09%	20.00%
True Ratio	-	99.92%	37.08%	12.50%

Confusion Matrix with respect to 3 classes.

April 2013		Predicted Values				Total
		Null	*LOW*	*Moderate*	*High*	
Real Values	*Null ($M < 0.5$)*	**60890**	1	6	0	60897
	LOW ($0.5 \leq M \leq 1.5$)	7	**60**	359	3	429
	Moderate ($1.5 < M < 3.0$)	8	36	**221**	18	283
	High($M \geq 3.0$)	1	1	10	**3**	15
Total		60906	98	596	24	**61624**

	Total	*Null*	*LOW*	*Moderate*	*High*
Accuracy	99.27%	99.99%	13.99%	78.09%	20.00%
True Ratio	-	99.97%	61.22%	37.08%	12.50%

Confusion Matrix with respect to 4 classes.

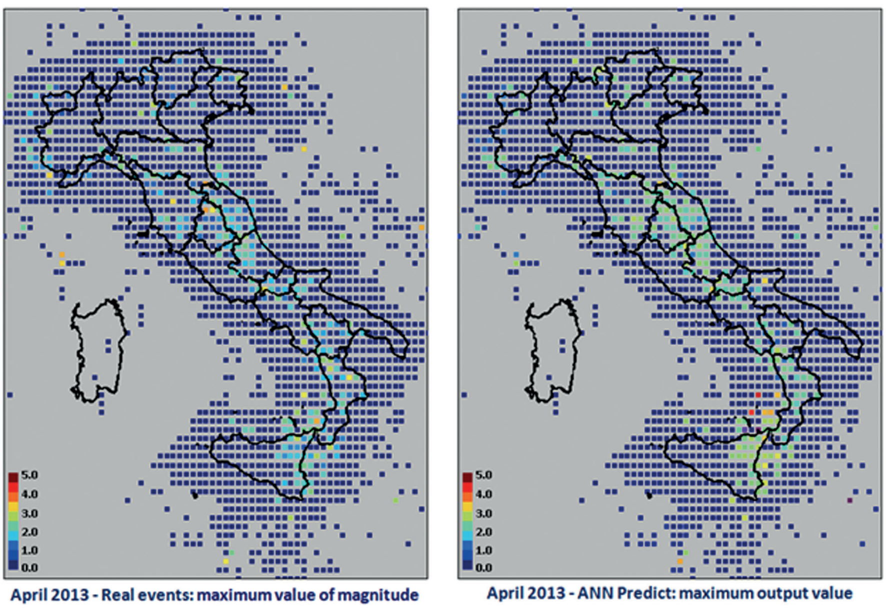

Fig. A10 - Results of April 2013: a) maximum magnitude of real events; b) maximum magnitude predicted by ANN.

Results of May 2013

May 2013		Predicted Values		Total
		$M \le 2.0$	$M > 2.0$	
Real values	$M \le 2.0$	62959	620	63579
	$M > 2.0$	7	92	99
Total		62966	712	63678

Accuracy	99.02%
Sensitivity	92.93%
Specificity	99.02%
Prob False Alarm	87.08%
Prob Missed Alarm	0.01%

Confusion Matrix with respect to 2 classes.

May 2013		Predicted Values			Total
		LOW	*Moderate*	*High*	
Real Values	*LOW ($M \le 1.5$)*	62944	414	18	63376
	Moderate ($1.5 < M < 3.0$)	18	242	25	285
	High($M \ge 3.0$)	2	9	6	17
Total		62964	665	49	63678

	Total	*LOW*	*Moderate*	*High*
Accuracy	99.24%	99.32%	84.91%	35.29%
True Ratio	-	99.97%	36.39%	12.24%

Confusion Matrix with respect to 3 classes.

May 2013		Predicted Values				Total
		Null	*LOW*	*Moderate*	*High*	
Real Values	*Null ($M < 0.5$)*	62887	5	5	0	62897
	LOW ($0.5 \le M \le 1.5$)	13	39	409	18	479
	Moderate ($1.5 < M < 3.0$)	6	12	242	25	285
	High($M \ge 3.0$)	1	1	9	6	17
Total		62907	57	665	49	63678

	Total	*Null*	*LOW*	*Moderate*	*High*
Accuracy	99.21%	99.98%	8.14%	84.91%	35.29%
True Ratio	-	99.97%	68.42%	36.39%	12.24%

Confusion Matrix with respect to 4 classes.

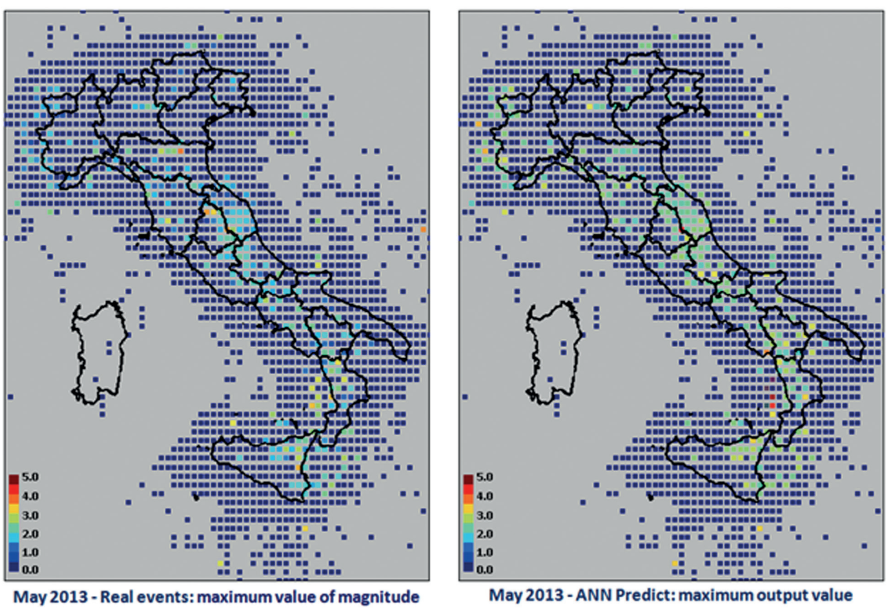

Fig. A11 - Results of May 2013: a) maximum magnitude of real events; b) maximum magnitude predicted by ANN.

Results of June 2013

June 2013		Predicted Values		Total
		$M \leq 2.0$	$M > 2.0$	
Real values	$M \leq 2.0$	61110	388	61498
	$M > 2.0$	19	107	126
Total		61129	495	61624

Accuracy	99.34%
Sensitivity	84.92%
Specificity	99.37%
Prob False Alarm	78.38%
Prob Missed Alarm	0.03%

Confusion Matrix with respect to 2 classes.

June 2013		Predicted Values			Total
		LOW	Moderate	High	
Real Values	LOW ($M \leq 1.5$)	60978	342	10	61330
	Moderate ($1.5 < M < 3.0$)	31	218	25	274
	High($M \geq 3.0$)	0	14	6	20
Total		61009	574	41	61624

	Total	LOW	Moderate	High
Accuracy	99.32%	99.43%	79.56%	30.00%
True Ratio	-	99.95%	37.98%	14.63%

Confusion Matrix with respect to 3 classes.

June 2013		Predicted Values				Total
		Null	LOW	Moderate	High	
Real Values	Null ($M < 0.5$)	60845	6	4	0	60855
	LOW ($0.5 \leq M \leq 1.5$)	9	118	338	10	475
	Moderate ($1.5 < M < 3.0$)	5	26	218	25	274
	High($M \geq 3.0$)	0	0	14	6	20
Total		60859	150	574	41	61624

	Total	Null	LOW	Moderate	High
Accuracy	99.29%	99.98%	24.84%	79.56%	30.00%
True Ratio	-	99.98%	78.67%	37.98%	14.63%

Confusion Matrix with respect to 4 classes.

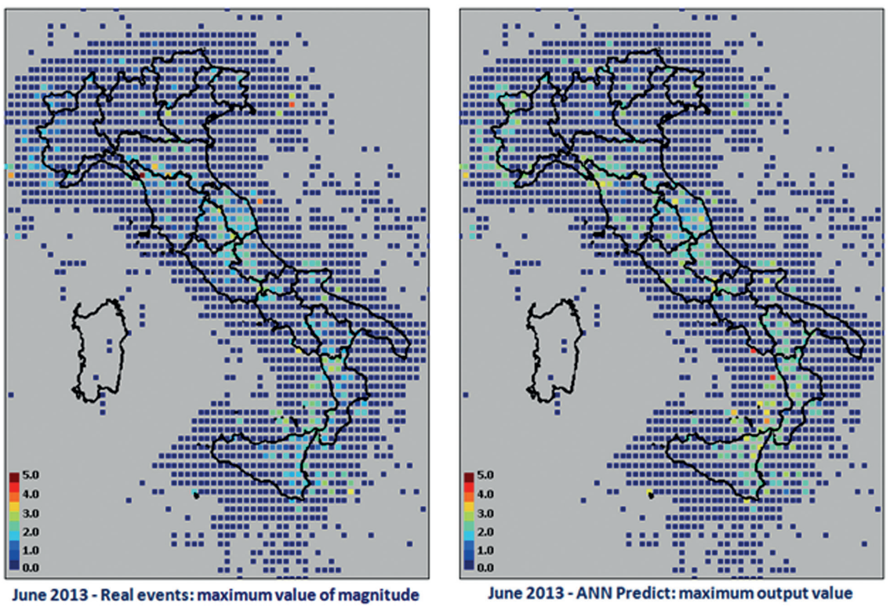

Fig. A12 - Results of June 2013: a) maximum magnitude of real events; b) maximum magnitude predicted by ANN.