Large Earthquake Occurrence Estimation Based on Radial Basis Function Neural Networks

Alex Alexandridis, *Member, IEEE*, Eva Chondrodima, *Student Member, IEEE*, Evangelos Efthimiou, Giorgos Papadakis, Filippos Vallianatos, and Dimos Triantis

***Abstract*—This paper presents a novel scheme for the estimation of large earthquake event occurrence based on radial basis function (RBF) neural network (NN) models. The input vector to the network is composed of different seismicity rates between main events, which are easy to calculate in a reliable manner. Training of the NNs is performed using the powerful fuzzy means training algorithm, which, in this case, is modified to incorporate a leave-one-out training procedure. This helps the algorithm to account for the limited number of training data, which is a common problem when trying to model earthquakes with data-driven techniques. Additionally, the proposed training algorithm is combined with the Reasenberg clustering technique, which is used to remove aftershock events from the catalog prior to processing the data with the NN. In order to evaluate the performance of the resulting framework, the method is applied on the California earthquake catalog. The results show that the produced RBF model can successfully estimate interevent times between significant seismic events, thus resulting to a predictive tool for earthquake occurrence. A comparison with a different NN architecture, namely, multilayer perceptron networks, highlights the superiority of the proposed approach.**

***Index Terms*—Clustering methods, earthquakes, interevent times, neural networks (NNs), radial basis function (RBF).**

I. Introduction

THE LONG-TERM prediction of natural disaster occurrence is one of the most sought-after goals in geoscience. Succeeding in such a goal involves obviating a multitude of difficulties; not only should the proper variables that will act as precursors be recognized and measured, but more importantly, the correlations between those variables and disaster occurrence should be identified. In spite of the significant progress over the last 20 years [1], the determination of such correlations remains a difficult endeavor as the governing relationships are usually rather complex and nonlinear [2]–[4], and the mechanisms creating the respective correlations are only recently coming to be understood [5], [6]. Although the use of methods based on first principle equations is a promising direction toward predicting natural disaster occurrence, the aforementioned facts, together with the large number of unknown involved parameters, impede the practical implementation of such methods in civil protection [1], [7].

On the other hand, neural networks (NNs) are powerful mathematical tools [8] that simulate the way that the human brain deals with information and the procedure of learning. NNs have the ability to identify and learn highly complex and nonlinear relationships from input–output data only, without the use of first principle equations describing the system. Due to these properties, NNs have been thoroughly exploited by many researchers for developing predictive models in geoscience: In [9], an NN-based approach is applied to predict rainfall, using radar reflectivity and tipping-bucket data as inputs. A method for aerosol optical depth retrieval in the Earth's atmosphere based on NNs is presented in [10], together with a method for estimating retrieval uncertainty. Three different NN modeling techniques are applied in [11] for the prediction of species-specific forest attributes, yielding high prediction accuracy levels. In [12], an adaptive neuro-fuzzy inference system is utilized for improving the accuracy of atmospheric profiles of temperature and humidity, which are retrieved from infrared sounder observations. In [13], a multiple NN model approach is applied to construct a predictive model for sea clutter, based on radial basis function (RBF) networks. The development of NN-based models for estimating the error variances of individual infrared and microwave atmospheric temperature and humidity profile retrievals and, thus, potentially improving their assimilation into numerical weather prediction models is presented in [14]. In [15], leaf area index is estimated from time-series remote sensing data using general regression NNs. NNs are used in [16] for estimating passive microwave brightness temperatures over snow-covered land in North America. In [17], a partial least squares NN model is applied to the remote estimation of chlorophyll-a concentration for turbid inland waters. NN algorithms have been also successfully used in conjunction with synthetic aperture radar images for predicting soil moisture content [18], [19], classifying sea ice types [20], and automatically detecting changes in suburban areas [21].

In a similar context, NN techniques have been also used for predicting and assessing the risk of natural disaster occurrence.

Manuscript received June 3, 2013; revised September 25, 2013; accepted October 23, 2013. Date of publication December 6, 2013; date of current version March 5, 2014. This work was supported by the THALES Program of the Ministry of Education of Greece and the European Union in the framework of the project entitled "Integrated understanding of seismicity, using innovative methodologies of fracture mechanics along with earthquake and nonextensive statistical physics—Application to the geodynamic system of the Hellenic Arc" (SEISMO FEAR HELLARC).

A. Alexandridis, E. Chondrodima, E. Efthimiou, and D. Triantis are with the Department of Electronics, Technological Educational Institute of Athens, 12210 Athens, Greece (e-mail: alexx@teiath.gr).

G. Papadakis is with the Institute for Risk and Disaster Reduction, University College London, London WC1E 6BT, U.K.

F. Vallianatos is with the Laboratory of Geophysics and Seismology, Technological Educational Institute of Crete, 73133 Chania, Greece, and also with the Institute for Risk and Disaster Reduction, University College London, London WC1E 6BT, U.K.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TGRS.2013.2288979

An NN model is developed in [22], in order to derive the tropical cyclone heat potential in the Indian Ocean using satellite images, sea surface temperature, and climatological depth of 26 °C isotherm. In [23], recurrent NNs are used in the prediction of severe meteorological events and, more specifically, to the rainfall field nowcasting.

Earthquake prediction is a subject of primary interest for the geophysical community, and many efforts have been made in this direction [1], [24]. Most of these approaches are based upon earthquake recurrence times [25], as changes in seismic activity reflect stress changes and the preparation stages of large earthquakes [26]–[29]. However, the uncertainty of the physics describing the earthquake generation and the lack of causal relationships between seismicity patterns and related crustal environments are limiting factors for the development of similar methods. Not surprisingly, the NN data-driven approach has been used as an alternative to classical approaches for the prediction of earthquake occurrence, with the number of relative papers increasing during the last few years: In [30], feedforward NNs are used to predict the magnitude of earthquakes, using as inputs the concentration of the soil-sourced radon gas emergence, together with other variables, including the earthquake's location and depth. In [31], seismic electric signals are used as inputs in a feedforward NN in order to predict the magnitude of impeding seismic events. In [32], feedforward NN models are constructed to analyze earthquake occurrence and predict the origin times of large earthquakes. Two recent publications have presented a method based on NNs for estimating the probabilities that an earthquake of magnitude larger than a threshold value happens and the probability that an earthquake of a limited magnitude interval might occur; the method is evaluated in Chile [33] and in seismogenic areas of the Iberian Peninsula [34]. Various NN architectures, including feedforward, recurrent, and probabilistic networks, have been used for predicting the magnitude and/or location of the largest earthquake in a predefined future time period using several seismicity indicators as inputs [35]–[37].

A common problem in the attempts for building NN-based earthquake predictive models involves the proper validation of the NN predictions; due to the very limited number of available training data, usually the same earthquake events that have been used during the training stage for calculating the NN model parameters are afterward employed for evaluating the model's performance. In some cases, a single earthquake event may be omitted from the training data set in an attempt to validate the prediction; however, this is not enough as an indicator of the model's success. As pointed out in [38], one of the main problems when using NN techniques on earthquake data is the small size of the training set compared with the number of parameters needed for determining the NN model.

In this paper, we present a novel NN-based methodology for estimating the occurrence of significant earthquake events, with the ability to cope with the small size of the available data. The method employs the RBF NN architecture, which presents several advantages, including higher estimation accuracy levels. Training of the networks is performed using the powerful fuzzy means (FM) algorithm, which, in this case, has been modified to incorporate a leave-one-out cross-validation procedure. This modification enables us to effectively use a small data set, allowing for an adequate number of data points for model training, while still evaluating the model on training-independent data. Furthermore, the Reasenberg clustering method [39] is utilized in order to decluster the available data and remove the aftershocks, as the inclusion of these events would significantly deteriorate the network performance.

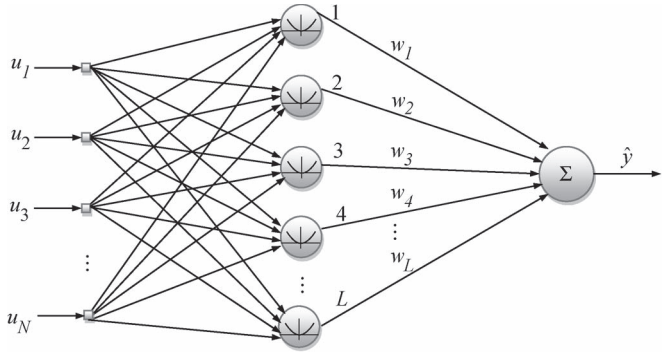

Fig. 1. Schematic overview of an RBF network.

The rest of this paper is structured as follows. In the next section, the RBF NN architecture is presented, followed by a description of the FM algorithm and an elaboration on the modification of the algorithm for performing leave-one-out cross-validation (LOO-CV). Section III briefly describes the Reasenberg technique for declustering earthquake catalogs. Section IV presents the application of the proposed framework to the California earthquake catalog, including a short description of the catalog, the selection of input and output variables for the RBF model, the aftershock removal procedure and, finally, the results on the estimation of earthquake occurrence. The paper concludes by outlining the advantages of the proposed approach.

II. RBF Networks

RBF NNs form a class of NNs, which presents certain advantages, including better approximation capabilities, simple network structures, and faster learning algorithms. On the other hand, RBF networks suffer from the curse of dimensionality, which refers to the fact that there is an exponential increase in the number of parameters to be identified with the dimension of the input space.

An RBF network can be considered as a special three-layer NN, which is linear with respect to the output parameters, after fixing all the RBF centers and nonlinearities in the hidden layer. The typical structure of an RBF network is shown in Fig. 1. The input layer distributes the N input variables to the L nodes of the hidden layer. Each node in the hidden layer is associated with a center, which is equal in dimension with the number of input variables. Thus, the hidden layer performs a nonlinear transformation and maps the input space onto a new higher dimensional space. The activity $\mu_l(\mathbf{u}_k)$ of the lth node is the Euclidean norm of the difference between input vector $\mathbf{u}_k$ and the respective node center $\hat{\mathbf{u}}_l$. The output function of the node

is a radially symmetric function; in this paper, the thin-plate spline function is employed.

The final output $\hat{y}_k$ of the RBF network for the kth data point is produced by a linear combination of the hidden node responses, after adjusting the weights of the network appropriately.

A. FM Algorithm

The FM algorithm was introduced a decade ago, as an improvement over the standard k-means algorithm in the selection of the hidden layer nodes [40], [41]. The FM algorithm presents several advantages compared with the typical approach, including faster computational times and automatic determination of the size of the network, and has been successfully used in a number of applications [42], [43]. Recently, a nonsymmetric version of the algorithm with improved prediction abilities has been introduced [44]. A brief discussion about the FM algorithm is given below, whereas the interested reader is referred to the original publications.

Consider a system with N normalized input variables u_i, where $i = 1, \ldots, N$. The domain of each input variable is partitioned into an equal number of 1-D triangular fuzzy sets, i.e., c. Each fuzzy set can be written as

$$A_{i,j} = \{a_{i,j}, \delta\alpha\}, \quad i = 1, \ldots, N,\ j = 1, \ldots, c \tag{1}$$

where $a_{i,j}$ is the center element of fuzzy set $A_{i,j}$, and $\delta\alpha$ is half of the respective width (due to the symmetric partition, all the widths are equal). This partitioning technique creates a total of c^N multidimensional fuzzy subspaces $\mathbf{A}^l$, where $l = 1, \ldots, c^N$. The multidimensional fuzzy subspaces are generated by combining N 1-D fuzzy sets, one for each input direction.

Each one of the produced fuzzy subspaces is a candidate for becoming an RBF center, but only a subset of them will be finally selected, depending on the distribution of data within the input space. The selection is based on the idea of the multidimensional membership function $\mu_{\mathbf{A}^l}(\mathbf{u}_k)$ of an input vector $\mathbf{u}_k$ to a fuzzy subspace $\mathbf{A}^l$, which is given by Nie [45], i.e.,

$$\mu_{\mathbf{A}^l}(\mathbf{u}_k) = \begin{cases} 1 - r_l(\mathbf{u}_k), & \text{if } r_l(\mathbf{u}_k) \le 1 \\ 0, & \text{otherwise} \end{cases}, \quad k = 1, 2, \ldots, K \tag{2}$$

where K is the total number of data, and $r_l(\mathbf{u}_k)$ is the Euclidean relative distance between $\mathbf{A}^l$ and the input data vector $\mathbf{u}_k$, i.e.,

$$r_l(\mathbf{u}_k) = \sqrt{\sum_{i=1}^{N} \left(a^l_{i,j_i} - u_{i,k}\right)^2} \Bigg/ \delta\alpha\sqrt{N}. \tag{3}$$

Equation (3) defines a hypersphere on the input space with a radius equal to $\delta\alpha\sqrt{N}$. The objective of the training algorithm is to select a subset of fuzzy subspaces as RBF centers, so that all the training data are covered by at least one hypersphere. Expressing this requirement in terms of (2), the subset of fuzzy subspaces is selected so that there is at least one fuzzy subspace that assigns a nonzero multidimensional degree to each input training vector. The maximum possible number of selected RBF centers is equal to the number of training data, although, depending on the distribution of data in the input space, a smaller number of centers is usually produced.

Algorithm 1 presents an overview of the FM algorithm.

Algorithm 1—FM Algorithm

Input: $\{\mathbf{U}_{\text{train}}, \mathbf{Y}_{\text{train}}\}$: Training Data set,
s: Number of fuzzy sets for partitioning each input dimension

Output: L_s: Number of selected RBF centers,
$\hat{\mathbf{U}} = [\hat{\mathbf{u}}_1,\ \hat{\mathbf{u}}_2,\ \ldots,\ \hat{\mathbf{u}}_L]^T$: Selected RBF center locations

1: Take the first data point: $k \leftarrow 1$
2: Begin calculations for the first RBF center: $L \leftarrow 1$
3: **For** $i = 1 : N$ **Do:**
4: Calculate the fuzzy set with maximum membership in each dimension i: $A_i^1 = \{a_i^1, \delta a\} \leftarrow \max\limits_{1 \le j \le s_i} [\mu_{A_{i,j}}(u_i(1))]$
5: **End For**
6: Generate the first RBF center $\hat{\mathbf{u}}_1$: $\hat{\mathbf{u}}_1 = [a_1^1,\ a_2^1,\ \ldots,\ a_N^1]$
7: **For** $k = 2 : K$ **Do:**
8: **If** data point k lies outside the hyperspheres defined by the already selected centers: $\min\limits_{1 \le l \le L} [r_l(\mathbf{u}(k))] > 1$, where r_l is calculated by (3)
9: Add a new RBF center: $L \leftarrow L + 1$
10: **For** $i = 1 : N$ **Do:**
11: Calculate the fuzzy set with maximum membership in each dimension i: $A_i^L = \{a_i^L, \delta a\} \leftarrow \max\limits_{1 \le j \le s_i} [\mu_{A_{i,j}}(u_i(k))]$
12: **End For**
13: Generate the Lth RBF center $\hat{\mathbf{u}}_L$: $\hat{\mathbf{u}}_L = [a_1^L,\ a_2^L,\ \ldots,\ a_N^L]$
14: **End If**
15: **End for**
16: Finalize the number of selected RBF centers: $L_s \leftarrow L$

Following the determination of centers by the FM algorithm, the synaptic weights are calculated using linear regression of the hidden layer outputs to the real measured outputs (target values). The regression problem can be trivially solved using linear least squares in matrix form.

B. Applying the FM Algorithm for Small Data Sets

Normally, in order to apply the FM algorithm for training the network, the available data should be split into, at least, two data sets, as dictated by a family of methods known as cross-validation, the two most popular variants being holdout and k-fold (or multifold) cross-validations [8]. Cross-validation is necessary in order to avoid a phenomenon known as overfitting, where the model is excessively fitted to the training data. In the case of RBF networks, overfitting can occur by increasing the number of RBF centers; such an increase in the model degrees of freedom makes the surface produced by the RBF

network more flexible, which results in minimizing, or even eliminating, the modeling error as far as training data are concerned. Unfortunately, this comes at the expense of losing generality, as the network ends up in learning the noise that is present in the training data and its ability to model new different data points is severely impaired.

In the holdout variant, data are split in two subsets, where the first one is used for the training phase of the network, i.e., for calculating the network parameters, whereas the second one is used for model selection and evaluation of the network performance after it has been trained. However, in cases where training data are scarce, it may be difficult to produce two data sets with adequate size in order to apply the holdout method. The k-fold cross-validation variant, on the other hand, divides the available data into k mutually exclusive subsets and fully trains the network using all subsets, except for one, whereas the resulting model is evaluated on the subset left out. This procedure is repeated k times, each time using a different subset for evaluation.

The most extreme form of k-fold cross-validation, which is known as LOO-CV, emerges when the number of subsets k is selected to be equal to the number K of training data. During this procedure, one data point $[\mathbf{x_i}\ y_i]$ at a time is excluded from the set, and the remaining data are employed for fully training an RBF network. The trained RBF network is then used to predict the output y_i, i.e., the ith output, which corresponds to the data point being left out. This procedure is recommended when dealing with very small data sets [8], as it employs the maximum amount of training data, which still allows for evaluating the model on training-independent examples, i.e., $K-1$ training data. When this procedure is completed for all data points, the leave-one-out cross-validated root-mean-square error, i.e., $\mathrm{RMSE_{CV}}$, is calculated by

$$\mathrm{RMSE_{CV}} = \sqrt{\frac{\sum_{i=1}^{K}(\hat{y}_i - y_i)^2}{K}}. \tag{4}$$

The network prediction for the ith data point while it has been removed from the training data is denoted by $\hat{y}_i$. In the same manner, one could calculate the cross-validated coefficient of determination R^2_{CV} for the entire data set using the following formulas:

$$\begin{aligned} R^2_{\mathrm{CV}} &= 1 - \frac{SS_{\mathrm{err}}}{SS_{\mathrm{tot}}} \\ SS_{\mathrm{err}} &= \sum_{i=1}^{K}(y_i - \hat{y}_i)^2 \\ SS_{\mathrm{tot}} &= \sum_{i=1}^{K}(y_i - \bar{y})^2. \end{aligned} \tag{5}$$

In the preceding equation, $\bar{y}$ stands for the mean value of the output variable. Both $\mathrm{RMSE_{CV}}$ and R^2_{CV} can be used as indicators for selecting the most appropriate model, which, in this case, means selecting the appropriate network size, and for evaluating the network performance. Algorithm 2 presents the incorporation of the LOO-CV procedure to the FM algorithm.

Algorithm 2—LOO Algorithm

Input: $\{\mathbf{U}, \mathbf{Y}\}$: Entire data set,
$s_{\min}$, $s_{\max}$: Minimum and maximum number of fuzzy sets
Output: $\mathrm{RMSE_{CV}}$: Cross-validated RMSE,
R^2_{CV}: Cross-validated R^2
For $s = s_{\min}: s_{\max}$ **Do:**
  For $k = 1 : K$ **Do:**
    Assign data point k to the validation data set: $\{\mathbf{U}_{\mathrm{val}}, \mathbf{Y}_{\mathrm{val}}\} \leftarrow \{\mathbf{U}(k), \mathbf{Y}(k)\}$
    Assign all data points, except from data point k, to the training data set: $\{\mathbf{U}_{\mathrm{train}}, \mathbf{Y}_{\mathrm{train}}\} \leftarrow \{\mathbf{U}(i), \mathbf{Y}(i)\}|_{i=1, i\neq k}^{K}$
    Pass the training data set $\{\mathbf{U}_{\mathrm{train}}, \mathbf{Y}_{\mathrm{train}}\}$ and s to algorithm 1, in order to calculate the total number of RBF centers $L_s(k)$ and their locations $\hat{\mathbf{U}}(k)$
    Calculate the synaptic weights $\mathbf{w}(k)$ using linear regression.
    Calculate the prediction $\hat{y}(k)$ of the produced network for data point k
  End for
  Calculate $\mathrm{RMSE}(s)$ and $R^2(s)$ using (4) and (5), respectively
End for
Calculate the best values for $\mathrm{RMSE_{CV}}$ and R^2_{CV}:
$\mathrm{RMSE_{CV}} \leftarrow \min_{s_{\min}\le s\le s_{\max}}[\mathrm{RMSE}(s)]$,
$R^2_{\mathrm{CV}} \leftarrow \max_{s_{\min}\le s\le s_{\max}}[R^2(s)]$

In the case of earthquake modeling, the very limited number of large events in the catalog dictates the application of the LOO-CV procedure, which allows for performing the training task with a small number of data and independently evaluating the network's performance at the same time.

III. Declustering

There is a general consideration on seismicity studies with regard to the earthquake events as independent or dependent on each other, such as aftershocks and foreshocks [46]. Independent earthquakes are also known as background earthquakes, main shocks, or parent earthquakes and are assumed to be mostly caused by the tectonic loading. Dependent earthquakes correspond to earthquakes triggered by mechanical processes that are controlled by previous earthquakes. It has been estimated that aftershocks account for about 30%–40% of the total number of earthquakes in word catalogs [47].

Seismicity declustering is the process of separating an earthquake catalog into foreshocks, mainshocks, and aftershocks. The identification of background earthquakes is important for many applications in seismology with regard to seismic hazard assessment, development of clustered seismicity models, earthquake prediction research, and seismicity rate change estimation [46].

The most applied declustering algorithms are those of Gardner and Knopoff [48] and of Reasenberg [39]. All

declustering methods identify the aftershocks based on their spatiotemporal proximity to previous earthquakes and on the fact that they occur at rates greater than the average seismicity rate.

In this paper, we use the cluster method introduced by Reasenberg [39]. This method identifies the aftershocks by linking earthquakes to clusters according to spatial and temporal interaction zones. Moreover, Reasenberg's procedure is free from assumptions with regard to the spatial aftershock distribution and describes their migration given that the background seismicity is low.

The spatial extent of the interaction zone is chosen according to stress distribution near the mainshock area. Reasenberg's local nearness of events is described by the spatial threshold d, depending on magnitude according to

$$\log(d) = 0.4M_0 - 1.943 + k \tag{6}$$

where M_0 is the magnitude, k is equal to 1 for the distance to the largest earthquake and equal to 0 for the distance to the last event, and d is expressed in kilometers. The model describing this relationship is a simple circular fault model of radius d. The Keilis–Borok formula defines the seismic moment for static cracks as follows: $16/7\ \Delta\sigma\ d^3$, where $\Delta\sigma$ is the stress drop. Its temporal extension is based on Omori's law. Each subsequent event is linked with the largest event or with the last one in each cluster, which has formed until current time. It should be also mentioned that overlapping clusters are joined [47]. The waiting interval τ in order to obtain a confidence of p_1 for observing the next event in the sequence is

$$\tau = -\ln(1-p_1)t/10^{2(\Delta M-1)/3} \tag{7}$$

where $\Delta M = M_{\text{mainshock}} - M_c$, and M_c is the magnitude of completeness.

It has become common practice in seismological studies to use standard parameter values provided by seismological software packages (e.g., ZMAP); however, it is recommended to analyze a declustered catalog based on the results of the applied method and the effect of varying the parameter values. Moreover, the advantages and disadvantages of any method and the parameter selection for them depend on the final goals of research.

IV. Case Study: California Earthquake Catalog

A. Catalog

The catalog used in this study is known as the Southern California Seismic Network catalog, and it is archived by the Southern California Earthquake Data Center. It contains historic seismological data dating from 1932 to present. The magnitude of completeness for this catalog is estimated from 3.25 in the early years to 1.8 or better at present [49]. The full catalog is available for download through the center's website (www.data.scec.org).

As reported by Gardner and Knopoff [48], the residual California catalog following aftershock removal is mainly influenced by an apparent Poissonian character, as far as the smallest

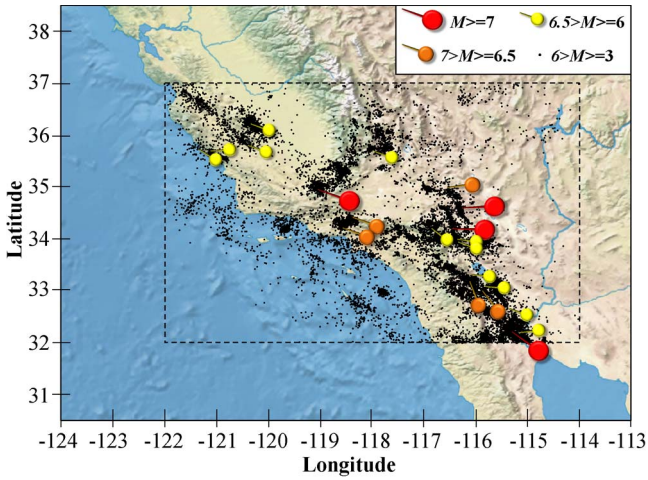

Fig. 2. Geographical locations for events with magnitude greater than or equal to 3 within the study area.

shocks are concerned. The larger shocks, however, were found to have only a small influence on the statistical character of the ensemble, whereas there are, as yet, insufficient events with $M > 5.8$ to draw any conclusions. In fact, it is generally hard to prove or disprove the time independence (Poisson) hypothesis because of the scarce number of large earthquakes for seismotectonic areas. Recently, alternative methodologies have been proposed, suggesting that the stationary epidemic-type aftershock sequence (ETAS) model is better than the Poisson model to describe the time distribution of worldwide large earthquakes, highlighting that the background of the ETAS model is not always a stationary Poisson model, but it can present time variations [50]. The latter is supported by a recent independent approach based on the application of nonextensive statistical physics for the time evolution of global seismicity in relation to mega events [2].

The space window used in this study contains the area between geographic coordinates 114°–122° W longitude and 32°–37° N latitude, whereas the time window contains all events occurring between January 1935 and June 2012. This corresponds to a total of 313 068 seismic events ranging in magnitude from 1.5 to 7.5. This window includes four events with magnitude greater than or equal to 7 and 21 events with magnitude greater than or equal to 6. Geographical locations for all events with magnitude greater than or equal to 3 are visually depicted in Fig. 2.

B. Input–Output Variable Selection

As the objective in this paper is to estimate the occurrence of large earthquakes, the output (target) variable of the RBF network was selected to be the interevent time between two consecutive main events. Thus, a threshold M was defined, and every event exceeding M in magnitude was considered to be a main event. We define the interevent time T_M^k as follows:

$$T_M^k = t^k - t^{k-1} \tag{8}$$

where t^k denotes the time of occurrence of the kth event in the catalog, which is greater in magnitude than the value M.

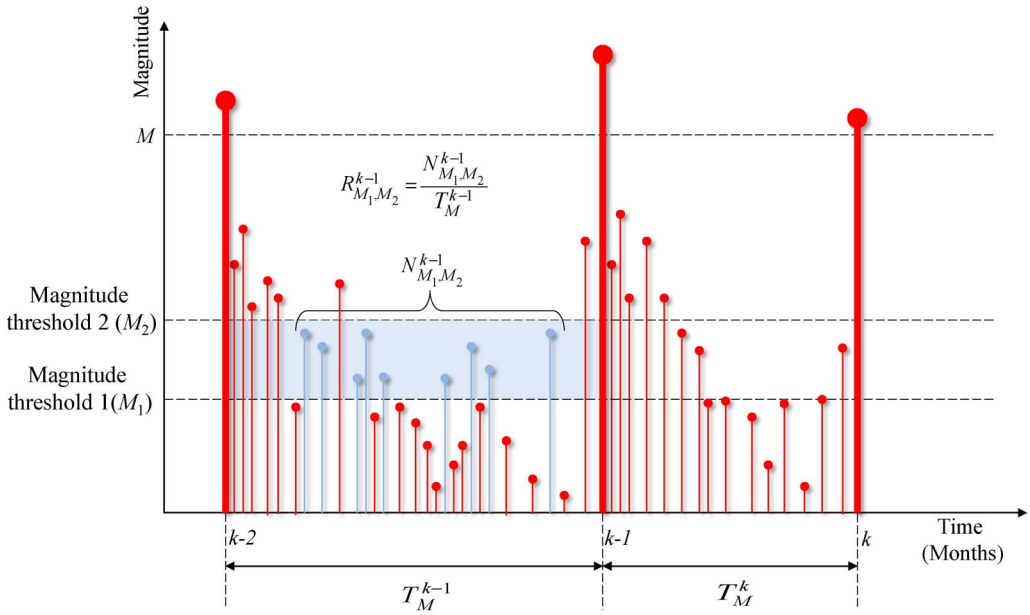

Fig. 3. Visual example for the definition of the seismicity rate $R^{k-1}_{M_1,M_2}$, in relation with main events numbered $k-1$and $k-2$.

Selecting the proper input variables is of paramount importance for the prediction abilities of the resulting NN model. In this work, the inputs to the model were selected as the seismicity rates between main events. There are many reasons for justifying this selection: 1) Seismicity rates can characterize the strain accumulation and release process on the average, 2) they are very easy to calculate, and 3) their calculation is reliable [32]. The latter is rather important in data-driven techniques such as NNs since the only source of information lies in the training examples.

We define the seismicity rate $R^k_{M_1,M_2}$ as the rate of earthquake occurrence in the time interval between the main events numbered k and $k-1$, i.e.,

$$R^k_{M_1,M_2} = \frac{N^k_{M_1,M_2}}{T^k_M} \tag{9}$$

where $N^k_{M_1,M_2}$ is the number of earthquake events that are larger in magnitude than the value M_1 and smaller than the value M_2 and have occurred between main events numbered k and $k-1$.

Previous studies [32] showed that estimation accuracy increases when more than one seismicity rate is given as input, attributing the fact to the existence of clear differences in the variations of seismic rates with the size of events. As the results of our experiments were consistent with this hypothesis, we allowed the input vector to the network to contain several values of different seismicity rates, which were calculated for different values of M_1, M_2, and k. The exact number of seismicity rates to be used as input variables was optimized based on numerical experiments, as will be discussed later.

It would be intuitive to include as input seismicity rates of the type $R^k_{M_1,M_2}$ for the prediction of the corresponding interevent time T^k, as they contain the most recent information preceding the main event numbered k. However, such an inclusion would significantly diminish the practical value of the resulting NN as a predictive tool. The reason is that, in order to make a prediction for the interevent time T^k between main events $k-1$ and k, one would have to wait until main event k has already occurred, so that the rate $R^k_{M_1,M_2}$ is available. In this case, prediction of the interevent time T^k would be of no practical use. Thus, the input vector for prediction of the interevent time T^k should contain past seismicity rates from the period between main events $k-1$ and $k-2$, i.e., $R^{k-1}_{M_1,M_2}$. More than one period between older main events could be also utilized to produce additional input variables. Fig. 3 depicts a visual example for the definition of the seismicity rate $R^{k-1}_{M_1,M_2}$, in relation with main events numbered $k-1$ and $k-2$.

The bounding parameters M_1 and M_2 define the range of magnitudes used for calculating the corresponding seismicity rate $R^{k-1}_{M_1,M_2}$. The selection of the lower bounding parameter M_1, which acts as a cutoff magnitude, is important in order to avoid producing a nonhomogenous set of training data. A sufficiently high value for M_1 should be applied, guaranteeing that variations in the calculated seismicity rate are not due to differences in the way earthquakes are recorded (e.g., changes in the sensitivity of earthquake recorders), but in fact reflect a change in seismicity dynamics. The importance of securing the homogeneity of the data set by properly selecting the cutoff magnitude is discussed in [32]. It is also possible to introduce additional bounding parameters $M_3, M_4, \ldots, M_z$, thus segmenting the magnitude range in more than one zone, i.e., $M_3 - M_4, \ldots, M_{z-1} - M_z$. In this case, the total number of seismicity rates to be used as input variables for each period between older main events is equal to $z-1$. Assuming that p periods between older main events are taken into account, a total of $p \cdot (z-1)$ possible input variables to the NN is produced.

C. Aftershock Removal

The removal of aftershock events is an important preprocessing procedure, which should precede the presentation of the

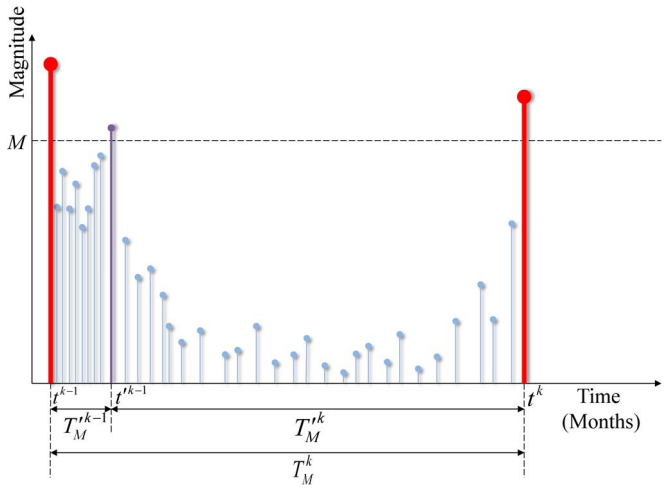

Fig. 4. Seismic activity between two main events. Selecting an aftershock as a main event results in estimating two erroneous interevent times $T_M'^{k-1}$ and $T_M'^{k}$.

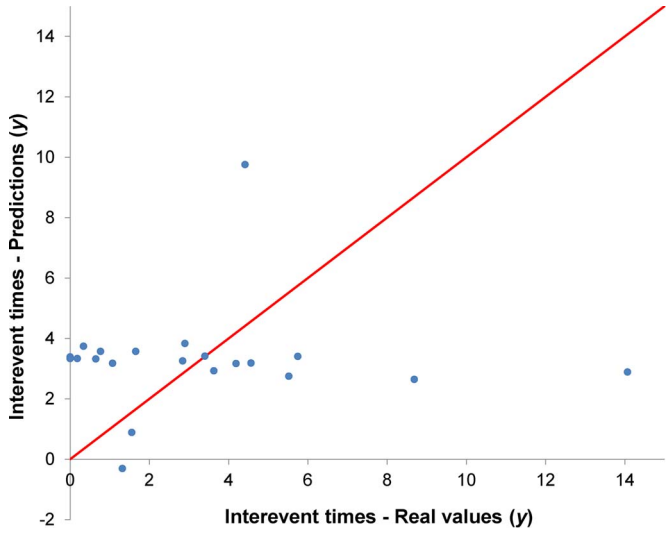

Fig. 5. Raw catalog. Real versus predicted interevent times.

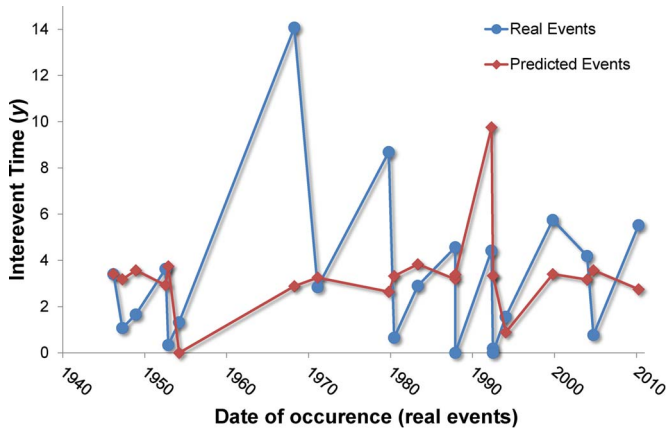

Fig. 6. Raw catalog. Real and predicted interevent times versus time of real event occurrence.

data to the NN. Application of the aftershock removal stage is crucial, particularly in the context of the output variable selection adopted in this study: In many cases, the aftershock events are of important size that may exceed the threshold M indicating a main event; thus, the particular aftershock events would be erroneously considered as main events, and the corresponding interevent times between the main event and the aftershock event would be presented to the NN as training examples. This not only produces new erroneous input examples but also alters the values of interevent times corresponding to the real main events, thus totally skewing the data set. This phenomenon is visualized in Fig. 4, where the seismic activity between two main events is presented.

In order to illustrate the importance of removing the aftershocks before presenting the training examples to the NN, we applied the training procedure to the original California catalog, without using any declustering technique. The threshold parameter M for selecting the main events was set equal to 6, and then, the choice of input variables was optimized for this particular case. The parameters to be optimized were the number of periods between older main events p and the number and values for magnitude bounding parameters $M_1, M_2, \ldots, M_z$. Optimization was performed following an exhaustive search procedure, testing all possible combinations for values of p that ranged between 1 and 3 and values of z that ranged between 2 and the maximum recorded magnitude in the catalog. For each possible configuration of p and z, the full magnitude range between M_1 and M_z was split in increments of 0.1, and all possible combinations for the values of $M_1, M_2, \ldots, M_z$ were also tested. Each combination of input variables was evaluated using the LOO-CV procedure, and the best input vector $\mathbf{x}_{\mathrm{br}}$ was found to contain the following inputs:

$$\mathbf{x}_{\mathrm{br}} = \left[R^{k-1}_{3.2,3.6} \quad R^{k-1}_{3.6,4} \quad R^{k-1}_{4,7.2}\right]. \tag{10}$$

The RBF network model was trained and tested using the LOO-CV procedure described by Algorithm 2. The best network using this procedure was the one using an input partition of 20 fuzzy sets, which resulted in 13 hidden node centers. The results for this configuration are depicted in Figs. 5 and 6, showing the real versus predicted interevent times and the real and predicted interevent times versus the year of the real event occurrence, respectively. The correlation coefficient R^2_{CV} and $\mathrm{RMSE}_{\mathrm{CV}}$ were found to be equal to the values -5.2 and $2.13y$, respectively. Additionally, a multilayer perceptron (MLP) network was also trained using the same optimized input vector. In this case, a standard two-hidden-layer architecture was postulated, and then, an exhaustive search was performed, testing all possible combinations when the number of nodes in each hidden layer ranged from 5 to 60. Table I contains the results for both networks. The results show that despite optimizing the input vector, the produced NNs cannot detect any significant correlation between the input values and the interevent times.

In order to remove the aftershocks and improve the network predictions, the Reasenberg declustering technique was applied, resulting to a declustered catalog of 152 641 events, 16 out of which were main events with $M \geq 6$. The decluster analysis was performed using the ZMAP software [51]. The

TABLE I
COMPARATIVE RESULTS FOR ALL CASES

Case	Input vector	# of data	NN Architecture	# of fuzzy sets	# of hidden layer nodes[a]	R^2_{CV}	$RMSE_{CV}$ (y)
Raw catalogue, M>6	$\left[R^{k-1}_{3.2,3.6}\ R^{k-1}_{3.6,4}\ R^{k-1}_{4,7.2}\right]^T$	21	RBF	20	13	-5.2	2.13
			MLP	-	[42 18]	-5.7	2.21
Declustered catalogue, M>6	$\left[R^{k-1}_{3.2,3.5}\ R^{k-1}_{3.5,4}\ R^{k-1}_{4,7.2}\right]^T$	16	RBF	10	13	0.875	1
			MLP	-	[36 16]	0.627	1.26
Declustered catalogue, M>6.2	$\left[R^{k-1}_{3.2,3.5}\ R^{k-1}_{3.5,3.7}\ R^{k-1}_{3.7,7.2}\right]^T$	12	RBF	7	10	0.705	1.28
			MLP	-	[58 41]	0.432	1.62

[a]The number of nodes for the MLP networks is given in the form: [1st layer nodes 2nd layer nodes]

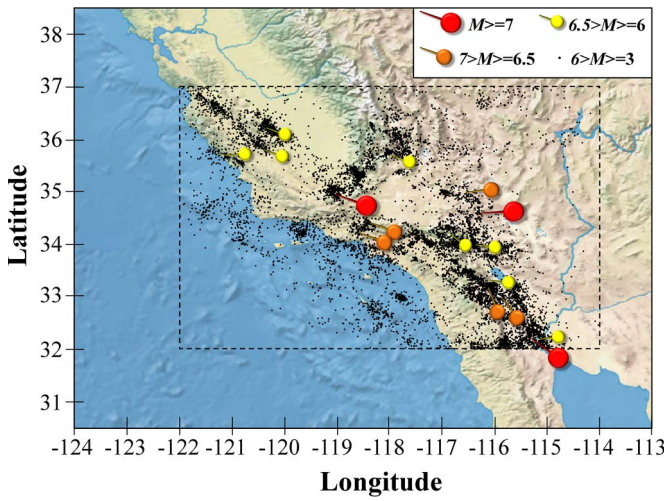

Fig. 7. Geographical locations for events with magnitude greater than or equal to 3 within the study area, following the implementation of the declustering procedure.

parameters used in the declustering method were the default ones suggested by the software, except from some slight changes concerning the number of crack radii surrounding each earthquake r_{fact}, which has been set equal to 5, and the depth and epicenter errors, which have been set equal to 0. Fig. 7 presents the geographical locations of the remaining events with magnitude greater than or equal to 3, following the implementation of the declustering procedure.

D. Interevent Time Estimation Using RBF Models

The first step toward obtaining the RBF model was to formulate the input vector, which was optimized using the same procedure described in the previous section. In this case, the best input vector $\mathbf{x}_{\text{bd6}}$ contained the following inputs:

$$\mathbf{x}_{\text{bd6}} = \left[R^{k-1}_{3.2,3.5} \quad R^{k-1}_{3.5,4} \quad R^{k-1}_{4,7.2}\right]. \tag{11}$$

Following the optimization of the input vector, Algorithm 2 was applied for training and evaluating the model. The best network was found to have 13 hidden layer nodes, which correspond to a partition of ten fuzzy sets per input dimension. During the execution of Algorithm 2, models with a higher number of RBF centers were also tested; however, they were rejected as they produced a higher cross-validated prediction error, as a result of overfitting during the LOO-CV procedure. The R^2_{CV} and RMSE_{CV} pointers for the optimal configuration were equal to 0.875 and $1y$, respectively. The results for this case are depicted in Figs. 8 and 9, showing the real versus predicted interevent times and the real and predicted interevent times versus the year of the real event occurrence, respectively. For comparison purposes, an MLP network was also trained based on the same input vector configuration. The results for both NN models are summarized in Table I. It should be noted that, although the MLP network can also provide successful estimations, the superiority of the RBF model is obvious, as it provides higher accuracy, combined with a simpler network structure.

The benefits of incorporating a declustering stage before formulating the input vector and presenting it to the NN model become obvious, as, in contrast to the previous case where the raw catalog was used, the results now reveal a strong correlation between the seismicity rates used as input variables and the interevent times between significant events. This is supported not only by the high values of the R^2_{CV} coefficient but also by the RMSE_{CV} value of $1y$, taking into account that the uncertainty for interevent time prediction when building a precursory model

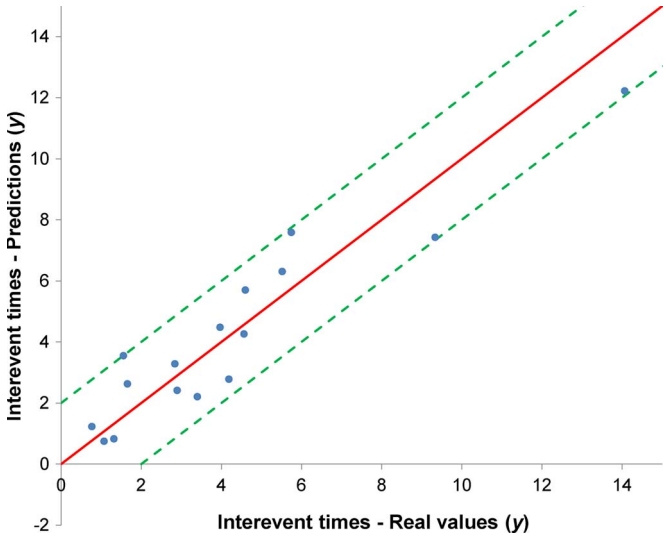

Fig. 8. Declustered catalog. Real versus predicted interevent times, with aftershock removal. The dashed lines represent $\pm 2y$ error bounds, and the straight line represents the ideal model with zero error.

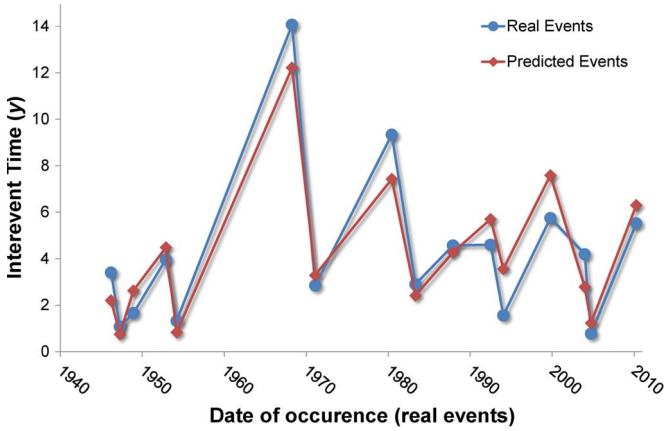

Fig. 9. Declustered catalog. Real and predicted interevent times versus time of real event occurrence, with aftershock removal.

for earthquakes with $M >= 6$ was estimated to be $\sim 2.5y$ [32]. It should be emphasized that, in contrast to similar studies, this result is not produced by evaluating the network predictions on the same data used for training. Therefore, the prediction for each event is independent from the training data set—a fact which adds to the statistical significance of the results. It should be also noted that the applied data set contains a wide range of interevent times spanning from $0.77y$ to $14.07y$.

Apart from serving as an estimator for earthquake occurrence, the produced model can be also used to shed some light in the relationship between seismic rate and interevent time. An indication for this relationship can be obtained through univariate analysis, where the effect of a single input variable on the output is separately examined, whereas the rest of the input variables are kept at constant values. In this case, the produced RBF model was applied in order to obtain estimations for the interevent time between significant earthquakes, when the seismicity rate for events with magnitudes larger than 3.2 and smaller than 3.5, i.e., input $R^{k-1}_{3.2,3.5}$, spanned the domain [40 75]. The specific input domain was selected to be within the

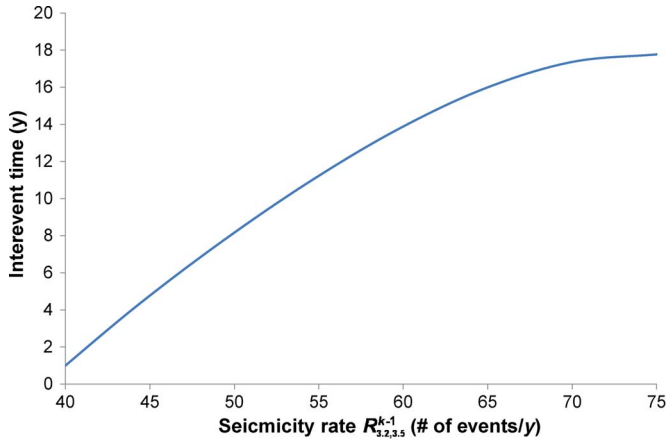

Fig. 10. Effect of the seismicity rate $R^{k-1}_{3.2,3.5}$ on the interevent time, as predicted by the RBF model using univariate analysis.

training limits for the particular input variable, thus avoiding extrapolation. The results are depicted in Fig. 10, showing a nonlinear relationship. It should be noted, however, that this technique does not take into account possible correlations between the input variables.

In order to demonstrate the robustness of the proposed approach, the value of the threshold M for selecting the main events was slightly changed from 6 to 6.2, and the input vector optimization and training procedures were repeated. Following the optimization procedure, the best input vector $\mathbf{x}_{\mathrm{bd6.2}}$ was found to contain the following inputs:

$$\mathbf{x}_{\mathrm{bd6.2}} = \left[R^{k-1}_{3.2,3.5} \quad R^{k-1}_{3.5,3.7} \quad R^{k-1}_{3.7,7.2} \right]. \tag{12}$$

After repeating the training procedure for both RBF and MLP networks, the best network was an RBF network with ten hidden layer nodes, with a partition of seven fuzzy sets per input dimension. The R^2_{CV} and $\mathrm{RMSE}_{\mathrm{CV}}$ pointers were now equal to 0.705 and $1.28y$, respectively. The results for this case are also summarized in Table I, where the superiority of the RBF network predictor is confirmed.

As a result of shifting the threshold value M, the quality of the models' estimations has obviously deteriorated, a fact that can be attributed to the decrease in the total number of training data from 16 to 12; the RBF model, however, is still acceptable as a predictor tool. The MLP network, on the other hand, essentially fails to capture the relationship between input and output variables as it exhibits a significantly lower R^2_{CV} coefficient. It should be noted that, despite changing the threshold value M, the calculated optimal values for the number of input variables and the magnitude bounding parameters $M_1, M_2, \ldots, M_z$ were similar to the ones resulting from the rest of the cases examined in this work. This does not only attest to the robustness of the proposed method but could be also an indication of an underlying relationship to the physics of earthquake generation.

V. Conclusion

This paper introduces a new methodology to estimate the occurrence of significant earthquake events based on RBF NNs, given a database of historic seismicity data. In order to formulate the models, the input variables are selected among

different seismicity rates, whereas the output variable is the interevent time between significant seismic events. Training of the RBF models is performed using a novel variant of the FM algorithm, which applies LOO-CV in order to deal successfully with small data sets, while maintaining the advantages of the original FM algorithm, which include higher accuracy and parsimony of the produced networks compared with other approaches.

The resulting methodology is applied to the California earthquake catalog. The results stress the importance of removing the aftershock events since a reliable estimation cannot be obtained using the raw catalog. To this end, the Reasenberg technique is employed in order to decluster the raw catalog. Following the application of the Reasenberg technique, the input vector to the NN model is optimized, and the FM algorithm is used to train the model. The resulting predictions reveal a strong correlation of the input variables with the interevent times, thus confirming the applicability of the proposed approach for successfully estimating large earthquake occurrence. A different NN architecture, namely, the MLP architecture, is also tested on the same data. A comparison between the two techniques highlights the superiority of the RBF network models, in terms of higher estimation accuracy and simpler network structure.

REFERENCES

[1] A. Panakkat and H. Adeli, "Recent efforts in earthquake prediction (1990–2007)," *Natural Hazards Rev.*, vol. 9, no. 2, pp. 70–80, 2008.

[2] F. Vallianatos and P. Sammonds, "Evidence of non-extensive statistical physics of the lithospheric instability approaching the 2004 Sumatran–Andaman and 2011 Honshu mega-earthquakes," *Tectonophysics*, vol. 590, pp. 52–58, Apr. 2013.

[3] F. Vallianatos, P. Benson, P. Meredith, and P. Sammonds, "Experimental evidence of a non-extensive statistical physics behaviour of fracture in triaxially deformed Etna basalt using acoustic emissions," *Europhys. Lett. (EPL)*, vol. 97, no. 5, pp. 58002-1–58002-6, Mar. 2012.

[4] F. Vallianatos and P. Sammonds, "Is plate tectonics a case of non-extensive thermodynamics?" *Phys. A, Statist. Mech. Appl.*, vol. 389, no. 21, pp. 4989–4993, Nov. 2010.

[5] D. Pliakis, T. Papakostas, and F. Vallianatos, "A first principles approach to understand the physics of precursory accelerating seismicity," *Ann. Geophys.*, vol. 55, no. 1, pp. 165–170, 2012.

[6] P. A. Varotsos, N. V. Sarlis, and E. S. Skordas, *Natural Time Analysis the New View of Time: Precursory Seismic Electric Signals, Earthquakes and Other Complex Time Series*. Berlin, Germany: Springer-Verlag, 2011.

[7] G. Hloupis and F. Vallianatos, "Wavelet-based rapid estimation of earthquake magnitude oriented to early warning," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 1, pp. 43–47, Jan. 2013.

[8] S. Haykin, *Neural Networks: A Comprehensive Foundation*, 2nd ed. Upper Saddle River, NJ, USA: Prentice-Hall, 1999.

[9] A. Kusiak, X. Wei, A. P. Verma, and E. Roz, "Modeling and prediction of rainfall using radar reflectivity data: A data-mining approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 4, pp. 2337–2342, Apr. 2013.

[10] K. Ristovski, S. Vucetic, and Z. Obradovic, "Uncertainty analysis of neural-network-based aerosol retrieval," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 2, pp. 409–414, Feb. 2012.

[11] H. Niska, J. Skon, P. Packalen, T. Tokola, M. Maltamo, and M. Kolehmainen, "Neural networks for the prediction of species-specific plot volumes using airborne laser scanning and aerial photographs," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 3, pp. 1076–1085, Mar. 2010.

[12] K. S. Ajil, P. K. Thapliyal, M. V. Shukla, P. K. Pal, P. C. Joshi, and R. R. Navalgund, "A new technique for temperature and humidity profile retrieval from infrared-sounder observations using the adaptive neuro-fuzzy inference system," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 4, pp. 1650–1659, Apr. 2010.

[13] X. Nan, H. Leung, and H. Chan, "A multiple-model prediction approach for sea clutter modeling," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 6, pp. 1491–1502, Jun. 2003.

[14] Z. Tao, W. J. Blackwell, and D. H. Staelin, "Error variance estimation for individual geophysical parameter retrievals," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 3, pp. 1718–1727, Mar. 2013.

[15] Z. Xiao, S. Liang, J. Wang, P. Chen, X. Yin, L. Zhang, and J. Song, "Use of general regression neural networks for generating the GLASS leaf area index product from time-series MODIS surface reflectance," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 1, pp. 209–223, Jan. 2014.

[16] B. A. Forman, R. H. Reichle, and C. Derksen, "Estimating passive microwave brightness temperature over snow-covered land in North America using a land surface model and an artificial neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 1, pp. 235–248, Jan. 2014.

[17] K. Song, L. Li, S. Li, L. Tedesco, H. Duan, Z. Li, K. Shi, J. Du, Y. Zhao, and T. Shao, "Using partial least squares-artificial neural network for inversion of inland water chlorophyll-a," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 2, pp. 1502–1517, Feb. 2014.

[18] S. Paloscia, P. Pampaloni, S. Pettinato, and E. Santi, "A comparison of algorithms for retrieving soil moisture from ENVISAT/ASAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 10, pp. 3274–3284, Oct. 2008.

[19] M. Kseneman and D. Gleich, "Soil-moisture estimation from X-band data using Tikhonov regularization and neural net," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 7, pp. 3885–3898, 2013.

[20] N. Y. Zakhvatkina, V. Y. Alexandrov, O. M. Johannessen, S. Sandven, and I. Y. Frolov, "Classification of sea ice types in ENVISAT synthetic aperture radar images," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 5, pp. 2587–2600, May 2013.

[21] C. Pratola, F. Del Frate, G. Schiavon, and D. Solimini, "Toward fully automatic detection of changes in suburban areas from VHR SAR images by combining multiple neural-network models," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 5, pp. 2587–2600, May 2013.

[22] M. M. Ali, P. S. V. Jagadeesh, I. I. Lin, and H. Je-Yuan, "A neural network approach to estimate tropical cyclone heat potential in the Indian Ocean," *IEEE Geosci. Remote Sens. Lett.*, vol. 9, no. 6, pp. 1114–1117, Nov. 2012.

[23] F. S. Marzano, G. Rivolta, E. Coppola, B. Tomassetti, and M. Verdecchia, "Rainfall nowcasting from multisatellite passive-sensor images using a recurrent neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 11, pp. 3800–3812, Nov. 2007.

[24] R. E. Habermann, "Precursory seismic quiescence: Past, present, and future," *Pure Appl. Geophys.*, vol. 126, no. 2–4, pp. 279–318, 1988.

[25] J. M. Carlson, "Time intervals between characteristic earthquakes and correlations with smaller events: An analysis based on a mechanical model of a fault," *J. Geophys. Res., Solid Earth*, vol. 96, no. B3, pp. 4255–4267, Mar. 1991.

[26] M. Matthews and P. Reasenberg, "Statistical methods for investigating quiescence and other temporal seismicity patterns," *Pure Appl. Geophys.*, vol. 126, no. 2–4, pp. 357–372, 1988.

[27] E. Mariana and B. Z. Yehuda, "Techniques and parameters to analyze seismicity patterns associated with large earthquakes," *J. Geophys. Res., Solid Earth*, vol. 102, no. B8, pp. 17785–17795, Aug. 1997.

[28] C. G. Bufe and D. J. Varnes, "Predictive modeling of the seismic cycle of the Greater San Francisco Bay Region," *J. Geophys. Res., Solid Earth*, vol. 98, no. B6, pp. 9871–9883, Jun. 1993.

[29] M. Eneva and Y. Ben-Zion, "Application of pattern recognition techniques to earthquake catalogs generated by model of segmented fault systems in three-dimensional elastic solids," *J. Geophys. Res., Solid Earth*, vol. 102, no. B11, pp. 24513–24528, Nov. 1997.

[30] F. Külahcý, M. Ýnceöz, M. Doru, E. Aksoy, and O. Baykara, "Artificial neural network model for earthquake prediction with radon monitoring," *Appl. Radiation Isotopes*, vol. 67, no. 1, pp. 212–219, Jan. 2009.

[31] M. Moustra, M. Avraamides, and C. Christodoulou, "Artificial neural networks for earthquake prediction using time series magnitude data or seismic electric signals," *Expert Syst. Appl.*, vol. 38, no. 12, pp. 15032–15039, Nov./Dec. 2011.

[32] B. Bodri, "A neural-network model for earthquake occurrence," *J. Geod.*, vol. 32, no. 3, pp. 289–310, Oct. 2001.

[33] J. Reyes, A. Morales-Esteban, and F. Martínez-Álvarez, "Neural networks to predict earthquakes in Chile," *Appl. Soft Comput.*, vol. 13, no. 2, pp. 1314–1328, Feb. 2013.

[34] A. Morales-Esteban, F. Martínez-Álvarez, and J. Reyes, "Earthquake prediction in seismogenic areas of the Iberian Peninsula based on computational intelligence," *Tectonophysics*, vol. 593, pp. 121–134, May 2013.

[35] H. Adeli and A. Panakkat, "A probabilistic neural network for earthquake magnitude prediction," *Neural Netw.*, vol. 22, no. 7, pp. 1018–1024, Sep. 2009.

[36] A. Panakkat and H. Adeli, "Neural network models for earthquake magnitude prediction using multiple seismicity indicators," *Int. J. Neural Syst.*, vol. 17, no. 1, pp. 13–33, Feb. 2007.
[37] A. Panakkat and H. Adeli, "Recurrent neural network for approximate earthquake time and location prediction using multiple seismicity indicators," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 24, no. 4, pp. 280–292, May 2009.
[38] F. Aminzadeh, S. Katz, and K. Aki, "Adaptive neural nets for generation of artificial earthquake precursors," *IEEE Trans. Geosci. Remote Sens.*, vol. 32, no. 6, pp. 1139–1143, Nov. 1994.
[39] P. Reasenberg, "Second-order moment of central California seismicity, 1969–82," *J. Geophys. Res.*, vol. 90, no. B7, pp. 5479–5495, Jun. 1985.
[40] H. Sarimveis, A. Alexandridis, G. Tsekouras, and G. Bafas, "A fast and efficient algorithm for training radial basis function neural networks based on a fuzzy partition of the input space," *Ind. Eng. Chem. Res.*, vol. 41, no. 4, pp. 751–759, 2002.
[41] A. Alexandridis, H. Sarimveis, and G. Bafas, "A new algorithm for online structure and parameter adaptation of RBF networks," *Neural Netw.*, vol. 16, no. 7, pp. 1003–1017, Sep. 2003.
[42] K. Ninos, C. Giannakakis, I. Kompogiannis, I. Stavrakas, and A. Alexandridis, "Nonlinear control of a dc-motor based on radial basis function neural networks," in *Proc. IEEE INISTA*, Istanbul, Turkey, 2011, pp. 611–615.
[43] A. Alexandridis, "An evolutionary-based approach in RBF neural network training," in *Proc. IEEE Conf. EAIS*, Madrid, Spain, 2012, pp. 127–132.
[44] A. Alexandridis, E. Chondrodima, and H. Sarimveis, "Radial basis function network training using a nonsymmetric partition of the input space and particle swarm optimization," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 24, no. 2, pp. 219–230, Feb. 2013.
[45] J. Nie, "Fuzzy control of multivariable nonlinear servomechanisms with explicit decoupling scheme," *IEEE Trans. Fuzzy Syst.*, vol. 5, no. 2, pp. 304–311, May 1997.
[46] Ö. van Stiphout, J. Zhuang, and D. Marsan, "Seismicity declustering," in *Proc. Community Online Resource for Statistical Seismicity Analysis*, 2012, pp. 1–25.
[47] G. Molchan and O. Dmitrieva, "Aftershock identification: Methods and new approaches," *Geophys. J. Int.*, vol. 109, no. 3, pp. 501–516, Jun. 1992.
[48] J. K. Gardner and L. Knopoff, "Is the sequence of earthquakes in Southern California, with aftershocks removed, Poissonian?" *Bull. Seismol. Soc. Amer.*, vol. 64, no. 5, pp. 1363–1367, 1974.
[49] K. Hutton, J. Woessner, and E. Hauksson, "Earthquake monitoring in Southern California for seventy-seven years (1932–2008)," *Bull. Seismol. Soc. Amer.*, vol. 100, no. 2, pp. 423–446, Apr. 2010.
[50] A. M. Lombardi and W. Marzocchi, "Evidence of clustering and nonstationarity in the time distribution of large worldwide earthquakes," *J. Geophys. Res., Solid Earth*, vol. 112, no. B2, pp. B02303-1–B02303-15, Feb. 2007.
[51] S. Wiemer, "A software package to analyze seismicity: ZMAP," *Seismol. Res. Lett.*, vol. 72, no. 2, pp. 373–382, 2001.

Alex Alexandridis (M'12) received the Diploma degree in chemical engineering and the Ph.D. degree in computational intelligence and control from the National Technical University of Athens (NTUA), Athens, Greece, in 2000 and 2003, respectively.

He was a Postdoctoral Fellow with NTUA from 2004 to 2009. Since 2010, he has been an Assistant Professor with the Department of Electronics, Technological Educational Institute of Athens, Athens, where he heads the Computational Intelligence Unit. He has authored or coauthored more than 50 original research works. His research interests include computational intelligence, nonlinear system modeling and control with emphasis on model predictive control methods, intelligent control, and applications to process engineering, materials science, geoscience, and environmental science.

Eva Chondrodima (S'13) received the B.Sc. degree in electronic engineering from the Technological Educational Institute of Athens, Athens, Greece, in 2011 and the M.Sc. degree in advanced information systems from the University of Piraeus, Piraeus, Greece, in 2013. She is currently working toward the Ph.D. degree in computational intelligence at the National Technical University of Athens, Athens, Greece.

Her research interests include neural networks, evolutionary computation methods, and decision support systems.

Evangelos Efthimiou is working toward the B.Sc. degree in electronic engineering at the Technological Educational Institute of Athens, Athens, Greece, where he is a member of the Computational Intelligence Unit.

Upon graduation, he intends to work toward the M.Sc. degree in electronic engineering.

Giorgos Papadakis received the B.Sc. degree in geology, geosciences and geoenvironment and the M.Sc. degree in prevention, assessment and management of natural hazards from the National and Kapodistrian University, Athens, Greece, in 2005 and 2008, respectively. He is currently working toward the Ph.D. degree in geophysics in the Institute for Risk and Disaster Reduction, University College London, London, U.K.

His research interests include the application of the nonextensive statistical mechanics formalism in seismology, the strong motion seismology, and the assessment of natural hazards.

Filippos Vallianatos received the B.A. and Ph.D. degrees in physics from the University of Athens, Athens, Greece, in 1985 and 1989, respectively.

From 2009 to 2011, he was a Senior Research Fellow with the Department of Earth Sciences, University College London, London, U.K., where he is currently a Visiting Professor with the Institute for Risk and Disaster Reduction. He is also currently a Professor with the Laboratory of Geophysics and Seismology, Department of Natural Resources and Environment, Technological Educational Institute of Crete, Chania, Greece. His research interests include geophysics, physics of the Earth's interior, and natural hazards.

Dimos Triantis received the B.Sc. degree in physics, the M.Sc. degree in electronics, and the Ph.D. in solid-state physics from the University of Athens, Athens, Greece, in 1975, 1980, and 1983, respectively.

Since 1989, he has been a Professor with the Department of Electronics, Technological Educational Institute of Athens, Athens, where he heads the Laboratory of Electric Characterization of Materials and Electronic Devices and is currently the Dean of the Faculty of Technological Applications. He has authored or coauthored more than 180 papers in journals and conference proceedings. His research interests include electric properties of materials and nondestructive testing, dielectric spectroscopy of solid materials, detection and study of mechanically stimulated electric signals and acoustic emission, and new technologies in education.