A Novel Method for Detection of Seismic Dual-Zones with Application to Earthquake Forecasting

Aref Bali-Lashak[1*], Mehdi Zare[2], Arash Andalib[3], Kazem Pourbadakhsh[4], and Yaser Radan[5]

1. PhD, International Institute of Earthquake Engineering and Seismology (IIEES), Iran,
* Corresponding Author; email: arefbali@ymail.com
2. PhD, Seismological Research Center, International Institute of Earthquake Engineering and Seismology (IIEES), Iran
3. MSc, Seismological Research Center, International Institute of Earthquake Engineering and Seismology (IIEES), Iran
4. PhD Student, Seismological Research Center, International Institute of Earthquake Engineering and Seismology (IIEES), Iran
5. PhD, Ferdowsi University of Mashhad, Mashhad, Iran

Received: 21/08/2010
Accepted: 19/06/2012

ABSTRACT

In this paper, we introduce a new approach to prepare the forecasting of earthquakes with magnitudes higher than a threshold level. This method can recognize the world's dual seismicity zones, where an earthquake in one zone acts as a precursor to other events in some other zone(s). To do so, we first, divide the entire global plane into well-defined sub-regions, and then create a matrix whose different cells correspond to different spatial-temporal seismic attitudes. In this matrix, each cell identifies the total number of events occurred in that sub-region within that specified period of time. The method, then proposes a procedure to measure the possibility or likelihood of an event in those regions by looking through the current situation of the reference region. On the other hand, the method can forecast future status of the reference region by searching the database of earthquakes, which have occurred already, and this would further result in prediction of other double-seismicity regions. Validity of the new forecasting approach is confirmed by the last year's events data recorded in NEIC catalogue.

Keywords:
Seismic probability; Dual zone; Precursor earthquake; Sparse matrix; Forecasting

1. Introduction

Geologists believe that the Earth is a complex system, and its physical parameters happen to show various nonlinear, chaotic and stochastic behaviors [1] many of which yet to be discovered and some of them not totally justified thus far. Therefore, forecasting the events introduced by such a complicated network is absolutely elusive. This area of challenge has engaged three different panels of scientists. The first group believes that the earthquake is quite an unpredictable phenomenon [2]. However, researchers involved in the second group believe that it has definitely some predictable attitudes, which must be searched for the proper statistical and precursory forecasting methods [3-4]. Finally, the third group has taken up a prudent point of view that is not based on the above-mentioned ideas [5]. Nevertheless, there is a wide variety of approaches for earthquake forecasting which are under investigation. Some of these methods consider anomalous signatures regarding specific physical quantities as precursory phenomena. Noticeable changes in electrical and magnetic fields of the Earth [6], significant changes in the emission of gases such as radon [7], changes observed in groundwater quality [8], electromagnetic radiation of the Earth [9], shaping of extraordinary earthquake clouds [10], foreshocks, or even the subtle changes seen in seismic activity/quiescence and unusual animal behaviors different than the normal ethological patterns [11] may be studied for this purpose [12-13]. Although some degrees of success have been reported for some of these precursory phenomena, they are not generally true for all cases but merely case-dependent. Furthermore, it is impossible to monitor most of these quantities continuously. As an alternative, however, some statistical methods have lately been introduced for earthquake prediction. In [14], Akasheh et al, present a method which monitors the events before a strong earthquake to raise an alarm for an event in the future. The Algorithm *M*8 that is a mid-term earthquake prediction method makes use of pattern recognition techniques for analysing the dynamics of the seismic behaviour preceding an earthquake event of magnitudes 8.0 or higher occurs around the world [15-16]. This method is then tested retrospectively in the vicinities of 143 points of which 132 are recorded as epicenters for these events with magnitudes $M = 8$ or greater [17]. In [18], Vorobieva, et al, employ a new scheme of spatially stabilized *M*8, named *M*8S, for earthquake prediction in Italy. The Algorithm MSc or "The Mendocino Scenario" is designed [19] by retroactive analysis of the regional seismic catalogue prior to the Eureka earthquake (1980, $M = 7.2$) near Cape Mendocino in California, hence its name. Given a TIP (Times of Increased Probability) diagnosed for a certain region U at the time T, the algorithm is designed to find a smaller area V which lies within U, where the predicted earthquake would be expected. An application of the algorithm requires a reasonably complete catalogue of earthquakes of magnitudes $M \geq (M0\text{-}4)$, which is lower than the minimal threshold usually used by *M*8. Here, Predictions are, firstly, made by *M*8 algorithm, and then, the areas of alarm are downsized by MSc at the cost that some earthquakes might well be missed in the second approximation of fore-casting process. The SSE algorithm [20] is another method introduced for prediction of relatively large earthquakes following a strong earthquake. A subsequent strong earthquake can be an aftershock or a main shock with a larger magnitude. In [21] a new algorithm, namely CN, is structured according to a pattern recognition scheme to allow a diagnosis of TIP's for the occurrence of strong earthquakes. This indicates the probability of an occurrence - inside a given region and time window- of events with magnitudes greater than a fixed threshold M_0, based on a quantitative analysis of the seismic flow. Hence, CN makes use of the information given by small and moderate earthquakes, having quite good statistics within the delimited region, to predict the stronger earthquakes which are rare events. While the results of some of these methods are encouraging, but further improvements are still desirable. Another approach based on pattern informatics is proposed for earthquake forecasting in [22-23]. However, this method only extracts local and regional seismicity patterns. Therefore, it may lose much valuable global information. Furthermore, it is not obvious to what region the method must be applied. Here, we describe a new method for finding dual zones and next, we will propose a new approach to forecast the mid-term and short-term earthquakes. Our method tries to find spatial-temporal patterns globally, including different regions and different time intervals. It also uses some heuristic methods with a view to reducing the complexity of the search algorithm.

Let's assume the Earth as a system that follows the global behavioral patterns [24], which any variation observed in one zone is simply transmissible to some other specific areas of this network. In a causal system like this, a relative pattern can be achieved by analyzing the current system status and then comparing the results to its situation in the past. This procedure would finally pave the way to predict the conditions of the system in the future. The planet Earth is complex and well prone to diverse changes taking place through its core as well as its lithosphere, i.e. the mantle and the crust. On the other hand, regarding a broader global view, we shall consider the different atmospheric layers and the ionized belt (Ionosphere sub-layer) surrounding the planet that may cause temporal changes to electromagnetic density, which in turn might eventually and gradually result in triggering some later earthquake events [25-27]. In this research, we are looking forward neither to discover any global basis nor to identify any global connection among the structural parts of this complex network; instead, the main goal of this paper is to locate the quakes likely to happen in the future, based upon a systematic search in the world earthquake catalogue for possible relations in seismic activity of different regions. We call these locations as dual zones (duals) to each other and name our suggested method as Bali-Zare Earthquakes Forecasting Method (BZEFM). This method, which is introduced in [28] in detail, can be used to evaluate the occurrence pattern of the events. The objective of the current research is neither to represent a new mechanism of the process, nor to express the relationship between two subsequent earthquakes - happened in different areas of the World, in terms of plate tectonic theory. However, this subject can also be studied further in another survey.

Here, by means of statistical analysis, we will try to answer the question that if it is possible for a certain region on earth to be disturbed and synchronized by another region's seismicity within a period of time delay. In fact, the problem is all about to answer whether a large earthquake in some typical areas would be able to act as a mid-term or as a short-term precursor to any other quake occurrence(s) in any other dual zone(s). If yes, then we may consider spatial-temporal clusters of the earthquakes that have been studied in the following sections of this paper. In a proper condition with the catalogue covering the time and when it is based on enough experiences, the great earthquake occurrences in the present time in some regions of the world can be supposed as the precursors to similar great earthquakes in other regions, and this will help in warning people, in advance, of an imminent event. In other words, BZEFM prepares to forecast the seismicity attributes for the reference region based on the current knowledge that, for example, a great earthquake has taken place in at least one of the two regions, which was well-known to make dual zones to the reference zone. Therefore, we will also be able to introduce possible spots for the big upcoming earthquakes. In this research, the data and statistical figures are obtained from NEIC catalogue (http://earthquake.usgs.gov/regional/neic/).

2. BZEFM Approach

Let E indicate the set of vectors $E = [a_1, a_2, ..., a_i, ..., a_n]$, where a_i with i = 1, 2,…, n is the vector corresponded to the ith time interval of the catalogue, and $n = (T_F - T_s)/\tau$. Besides, T_s and T_F show the starting and ending date of the events recorded in the earthquake catalogue, respectively. τ is a desired constant parameter, which indicates the length of the time interval allocated to each vector a_i. According to BZEFM, the global map is divided into m cells with $\hat{l} \times \hat{l}$ degrees in size, and the event centers are at π_j, j = 1, 2, …, m. Then, only those events from the earthquake catalogue which have magnitudes greater than or equal to a threshold level $(M \geq M_T)$ and occurred within the desired time interval i, will be taken into consideration. The vector a_i includes $a_i = [n_1^i, n_2^i, ..., n_j^i, ..., n_m^i]^T$ where n_j^i is the number of earthquakes larger than M_T in the jth zone of π_j and the ith time interval, and $[\cdot]^T$ is the same vector, yet transposed. Therefore, the matrix E named here as “spatial-temporal matrix of earthquake events” or simply “event matrix”, has m rows and n columns, corresponding to the seismicity of m different cells at n different time intervals. This event matrix E might be referred to as E^τ, to emphasize the length of time intervals, which are considered to set up the event matrix. Figure (1) shows a typical event matrix. BZEFM uses such a matrix to extract the useful yet latent spatial-temporal patterns.

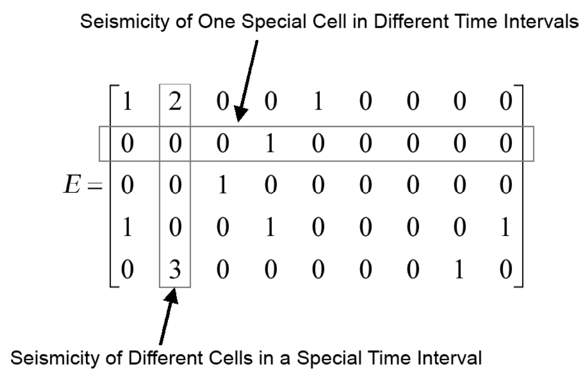

Figure 1. Example of an event matrix corresponding to the seismicity of m = 5 different cells at n = 9 different time intervals.

In this manner, one can study different attributes of the matrix from two different aspects: 1) row (cell)-based and 2) column-based approaches; both of which could be used efficiently to predict the probability of an imminent earthquake event in a specific geographical zone in the future.

1) Row-based approach: this method makes use of the recorded data related to the seismic behavior of one “specific zone”, e.g., Bandar-e-Abbas, Iran within “different time intervals”, e.g., as of the year 1973 through 2009.
2) Column-based approach: unlike the first proce-

dure, here the scientists use the available dataset in order to extract items of information recorded from a specific time event (e.g., time interval of Jan. 1973 through the end of March in the same year) around the world or within "several different zones".

By means of such a big two-dimensional spatial-temporal dataset, which has been embedded throughout the event matrix, there is a high chance for scientists to take a new analytical look at the recorded seismic behaviors, which is the main goal of the approach introduced in this paper. As such, a set of earthquake events (with spatial-temporal attributes) related to the past time, may be used as precursors to another event likely to occur in another place on the earth, and in this manner, a prediction becomes feasible. According to the same analysis, it is also possible for a series of earthquake events occurred previously in different spatial cells and temporal intervals to result in prediction of another event in a different cell and within some other future time intervals. For this to happen, each and every row of the matrix is compared one by one to the other rows conveying other items of information related to the past events, and this process continues until every desired entry of this matrix has been thoroughly compared to all the given entries from the other rows. So, at first, one row of the matrix that represent the seismic activity of one specific city, e.g., Bandar-e-Abbas, Iran, within different time intervals is picked up as the "target row", and then, the data embedded in the rest of the rows will be compared to the target row. This comparison takes effect via the seismic difference measure *A*:

$$A_{target,j} = \sum_{k=1}^{n} (E_{target,j} - E_{j,k})^2, \quad (j = 1, 2, ..., m) \qquad (1)$$

where *k* indicates time, $E_{target,j}$ and $E_{j,k}$ indicate respectively a specific entry *k* from the target row, and k-th entry from the j-th row of the spatial-temporal event matrix. This way, $A_{target,j}$ or shortly put $A_{t,j}$ gives a measure for comparing the seismic similarities between the target city and another city represented by the jth row; in such a way that the less the $A_{t,j}$, the higher the similarity of occurring the events between the two cities. By similarity, we mean simultaneous seismically-active time intervals with a similar number of events higher than the threshold. Meanwhile, there is also another parameter referred to as measure *B*, which indicates the seismic stillness similarity and is defined as follows:

$$B_{t,j} = \frac{\alpha_{t,j}}{2n}, \quad (j = 1, 2, ..., m) \qquad (2)$$

where $\alpha_{t,j}$ is the number of non-zero entries in the two rows *t* and *j*. Hence, $B_{t,j}$ shows the ratio of non-zero entries to the total number of rows in a matrix; e.g., if each row includes 30 entries, then comparing the two rows results in the assessment of data conveyed through the entire sum of 60 entries. Now, if only 15 entries out of the rest do have non-zero values, then B will be equal to 15/60 = 0.25. Concerning the explicit definitions for *A* and *B*, it is concluded that as far less and far more become the two measures *A* and *B*, then the seismicity behavior of the studied rows will be much more alike, and the amount of positive correlation between the two rows will also grow up. Therefore, the third measure *C* is obtained by combining the two previous definitions for *A* and *B*, as follows:

$$C_{t,j} = \frac{B_{t,j}}{A_{t,j}}, \quad (j = 1, 2, ..., m-1, \; j \neq t) \qquad (3)$$

Needless to mention that the bigger values for $C_{t,j}$ imply higher similarity between the two rows, which represent the events similarities between the target city and the other one under consideration. The descending vector $C_t = [C_{t,1}, C_{t,2}, ..., C_{t,m-1}]$ indicates the similarity among all different cells regarding the target cell.

Here, we use the vector C_t to define some of the main concepts discussed in this paper.

Definition#1 Given an event matrix *E* with the target cell *t*, the jth cell with a value $C_{t,j}$ greater than a specific threshold level is referred to as a "seismic dual zone to the target zone", or a "dual", in brief.

Definition#2 Given an event matrix E^{τ}, if the target row *t* is entirely shifted by an amount of $\tau = a\tau, (a = \pm 1, \pm 2, \pm 3, ...)$, and if afterward, it happens to the jth cell to make a dual to the newly shifted target cell, then the jth cell is referred to as a "seismic dual to the target zone with a time-shift of τ", or a "dual with a time-shift of τ", in brief.

Definition#3 Given an event matrix E^{τ}, the jth cell is a "precursor" of the target cell, if the jth cell is a dual to the target zone with a time-shift of $\tau = a\tau$, and $a > 0$.

Definition#4 Given an event matrix E^{τ}, the j[th] cell is a "postcursor" of the target cell, if the j[th] cell is a dual to the target zone with a time-shift of $\tau = a\tau$, and a<0.

Definition#5 If in the event matrix $E^{\tau_1}_{m\times n}$, the cell A is known as a dual to the cell B with a time-shift of $\tau_1 = a_1\tau_1$, and if in the event matrix $E^{\tau_2}_{m\times n}$, the cell B is known as a dual to the cell A with a time-shift of $\tau_2 = a_2\tau_2$, then the cells A and B are "resonant duals to each other, with respective time-shifts of τ_1 and τ_2", or as "resonant duals, with time-shifts of τ_1 and τ_2", in brief.

3. BZEFM for Earthquake Forecasting

BZEFM is based on searching for the seismic dual zones, and also the precursory and postcursory cells by means of the event matrix. However, after this stage, BZEFM approach will show us how to forecast or predict an earthquake event with the use of the precious data already compiled, i.e. how to predict the likelihood of an earthquake occurrence in a specific location within a specific time interval in the future. When two cells are determined dual, with the conditions required by BZEFM, then any earthquake event greater than the pre-selected threshold in one cell (reference region) in a time interval may then be considered as an alarm for an event in the dual cell (target region), in the same time interval. What comes hereafter is an explanation to two different BZEFM approaches for earthquake forecasting namely location-based and time-based event forecasting. In location-based forecasting, the aim is to find the likelihood of an earthquake event in a given region. However, in time-based forecasting, the goal is to declare earthquake alarms, based on the events occurred recently. These are two different kinds of predictions, which may be employed occasionally.

3.1. Location-Based Event Forecasting

In this section, it is assumed that we tend to perform an earthquake event prediction for a typical city *X*, which lies in a target cell of the event matrix under the same name *X*. To this aim and according to BZEFM approach, it is first needed to indicate those cells which act as precursors to the target cell *X* for all different amounts of time-shift $\tau = a\,\tau$; i.e. we need to get on with a multi-step procedure in order to indicate the 5-year precursory cells at first, then the 4-year precursory cells, and so forth until the entire precursory cells such as 3, 2, and 1-year long cells are properly indicated. This method will still run on for other monthly time-shifts as well; i.e., the 6-month and 3-month precursory cells shall be indicated too. For $\tau = 3$, the corresponding values of the coefficient a for these amounts of time-shift are $a = 1, 2, 3, 4, 8, 12, 16, 20$. The parameter a is accordingly known as "time-shift coefficient". However, when using this approach, one must notice to perform the prediction in such a way that the precursors would necessarily have an overlapped ensemble in the end of the 5-year time interval. For instance, if prediction of 5-year precursors has begun already as of today, then the prediction of 4-year precursors shall cover a time span from next year up until the next coming five years. In the same way, prediction of 3-year precursors must begin from two years later so that it will definitely cover the end of the specified 5-year time interval, and this procedure continues till the monthly intervals meet this criterion as well, i.e. the prediction time span of 3-month precursory cells will be the last three months of the specified five-year-old interval. The last three-month period of the prediction time interval for which all the several precursors would impact the warning factors indicating an imminent earthquake event, is called the "ensemble subinterval of the precursory alarms", that in general, equals to τ. Figure (2) gives a better look on this criterion and the overlapped ensemble sub-interval. In this figure, the short-term precursors implying an immediate warning are signified in dark colors, while the other precursors with relatively longer warning time spans are shown with light colors. Meanwhile, regarding this fact that, say, the 5-year long precursory alarm is supposed to benefit the entire time interval of the next five years, and given the probability distribution of the announced alarm data is uniform, then the likelihood for an event to occur within the last three months of the interval will, according to the 5-year precursory alarm, be scaled with a coefficient of 3/60=0.05. In this manner, the darker colors, corresponding to bigger alarm coefficients, will simply indicate more critical alarms within the last three months of the 5-year long interval. Generally speaking, the "scaling coefficient of precursory warning" is equal to $\beta = \tau / \tau = 1 / a$ where a is the pre-mentioned time-shift

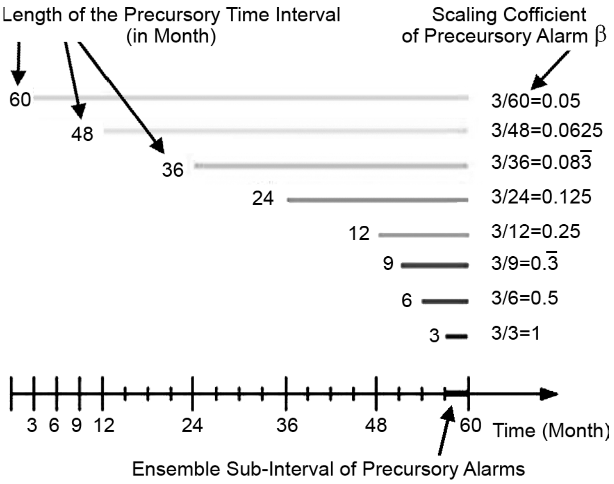

Figure 2. The target prediction time interval for different precursors. All predictions show to have definite overlapped ensemble in the last three months of the 5-year long interval. Short-term precursors that imply immediate warning are marked with dark colors and other precursors with relatively longer alarm intervals are shown in lighter hues. The darker colors are also indicating the bigger scaling coefficients β, which will be corresponding to the more efficient or more serious warnings as for the ensemble subinterval.

coefficient. Through this definition, we can now define the lower and upper bounds for β as $0 < \beta \leq 1$. As is well depicted in Figure (2), for the above example the scaling coefficient of precursory warning lies in the interval $0.05 < \beta \leq 1$.

While the effects of all different precursory cells are accumulated within the ensemble subinterval which lies in the end of the forecasting period, it will be quite fair to think of a unique number as the representative of that specific interval. Once this number is calculated, the earthquake researcher will be able to attribute a specific degree of seriousness or “degree of warning” to the particular time interval, e.g. the 5-year interval that is under consideration.

Given the ith precursory cells with a time shift of $a_i\tau$, we define D_i as the ratio of the number of dual cells with a history of earthquake occurrence within “precursory interval” to the total number of identified dual cells in the ith set. The “precursory interval” is defined as a period of time during which the occurrence of any earthquake event would be closely inspected as an alarm for an event in the target cell. For instance, suppose that there are three 5-year precursory cells to the target cell, and suppose that in all three cells and during the precursory intervals, an earthquake with a magnitude greater than the threshold level happened. Then, the amount of resultant warning degree for the 5-year precursory cells will be equal to $D_{(5\text{-}year)} = 3/3 = 1$.

To calculate the total degree of warning, D_t, we then add up different amounts of D_i, each of them scaled by the corresponding scaling coefficient $1/a_i$

$$D_t = 1/N\sum_{i=1}^{N} D_i / a_i = 1/N\sum_{i=1}^{N} iD_i \tag{4}$$

where N is the number of different precursory periods. Considering the variation interval for scaling coefficient β, (4) implies that the variation interval for the resultant warning degree D_t will be $0 \leq D_t \leq 1$.

Figure (3) depicts the details of an example for calculating D_t. In this figure, any square represents a precursory cell. However, if the square is red-filled, it means that an earthquake event has happened in that cell during the precursory interval. To make it more clear, if there are five cells as 3-year (36-month) long precursors to the target cell, and if during the precursory time it happens for only cells -out of five- to witness an earthquake event, then the amount of resultant warning degrees for the 3-year precursory cells will be equal to $D_{(3\text{-}year)} = 2/5 = 0.4$.

For example, in Figures (2) and (3), the amount of D_t equals to 0.177, that is the resultant degree of warning as for the 3-month ensemble time interval for this example. Regarding the fact that in this example, there are lots of precursors for the reference cell which are active, see Figure (3), and since the range of resultant (total) degree of

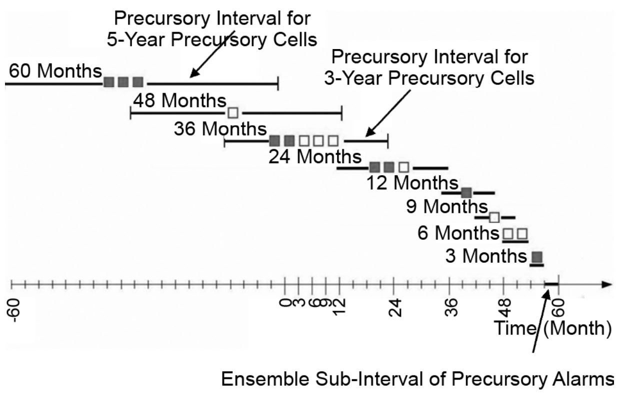

Figure 3. Calculation of D_t. In each precursory interval, the squares represent the total number of precursor cells, which are identified as a dual to target cell. The red-filled squares indicate the number of seismically active cells within that precursory time interval.

warning, D_t, is known to be: $0 \leq D_t \leq 1$, hence it can be simply concluded that big amounts of this parameter can imply drastically serious warnings.

If the location-based BZEFM algorithm is applied, then a highly precise prediction map can be outlined, which will serve the seismic forecasting procedure for significant points and cities around the world. Such a precise map will be continually brought up to new codes and data. Furthermore, when the related codes are efficiently implemented and some regular simplifications are done, then these algorithms may be used for all cells of the spatial-temporal matrix of an earthquake event. Currently, the typical algorithms of this kind make use of 2×2 degree cells in order to simplify the prediction problem. However, if the algorithms are implemented and applied efficiently, we can hope to reduce the size of cells down to 1×1 so that we will get better results as more efficient forecasting.

3.2. Time-Based Event Forecasting

Unlike the location-based forecasting approach which was already explained, in this section we are supposed to study the earthquake events which date back beyond the past 24 hours or even the last week (it might be hard to determine which period of time would be more efficient) and might have occurred anywhere around the world. Then, in the second step, by means of a database compiled in advance by BZEFM approach for dual zones, it will be determined that which event(s) would act out as precursor or alarm to which cell(s).

4. Experimental Results and Validity of BZEFM

In order to experiment with the BZEFM performance and to test its validity, first the world's seismicity catalogue was received for the time span 1.1.1973 through 6.30.2010. The data extracted from the interval 1.1.1973-6.3.2009 was then used to look for any dual zones, and the last year's period of time, i.e. the interval 6.30.2009-6.30.2010 was considered to serve as an "interval of evaluation" for measuring the accuracy of predictions based on the dual zones which were determined already. In this paper, we refer to these two periods as identification and evaluation periods respectively. Here, the world's map or the global surface is segregated into cells of 2×2 degrees in size, i.e. each cell covers, approximately, an area of 40000 square Kilometers. Selecting smaller areas makes the predictions more useful and important; however, it increases the complexity of the problem when the current algorithm is employed. At the next step, the events with magnitudes greater than the threshold level 5.5, which have occurred within the desired time interval in every specific cell have been taken into account. Regarding the 2×2 division of the map, the ensuing spatial-temporal matrix of earthquake events would convey 16200 rows (cells). Meanwhile, regarding the starting and ending dates of the catalogue, as well as the three monthly divisions, this matrix is expected to include 152 columns presenting the 152 seasons. The selection of the time period is again based upon a tradeoff. A too-long time period gives no sense of similarity of seismic activity of two cells. On the other hand, if we choose the time period too short, then there is very little chance for contemporary events in two cells, so the basic strategy of BZEFM does not apply anymore.

Once the event matrix is completed, to reduce the complexity of the algorithm, the rows of the matrix with less than three events, were removed. Via this strategy, the entire number of rows representing the earthquake events were largely reduced from 16200 into 398; a fact that in turn influenced the procedure with a significant reduction in the computational complexity for the algorithms.

Following this, values of parameters A, B, and C were calculated for each reference cell and then, the outcomes were sorted out in accord with the observed maximum probabilities. This way, some cells of the matrix happened to demonstrate noticeable spatial correlations. However, a more startling point appeared when these dual cells which were identified in time intervals prior to 6.30.2010 showed similar attitudes within the evaluation time interval, and this could further assert the accuracy of the approach introduced by BZEFM. Table (1) demonstrates the performance of the new method over dual zones of two reference regions A (*lat*: -57, *long*: -27) and B (*lat*: -15, *long*: -171). The positions of the reference cell A and its dual zones, i.e., dual_{1A}, and dual_{2A} are shown in Figure (4). Similarly, the positions of cell B and its dual zones, i.e. dual_{1B} and dual_{2B}, can be seen in Figure (5). The high percent of duality for the set of cells A and B, indicates that an earthquake event in each of the dual cells may be considered as an alarm for the corresponding reference cell. It is

Table 1. BZEFM's performance over identification and evaluation periods.

Reference Cell	***ref***$_A$ **(-57, -27)**	***ref***$_B$ **(-15, -171)**
First Best Cell in Duality	Dual1$_A$(-35, -103)	Dual1$_B$ (-17,-171)
Second Best Cell in Duality	Dual2$_A$ (13, -87)	Dual2$_B$ (29, -179)
Rate of Contemporary Events in the Three Cells Over the Identification Period	38%	38%
Rate of Contemporary Events in the Three Cells Over the Evaluation Period	75%	96%

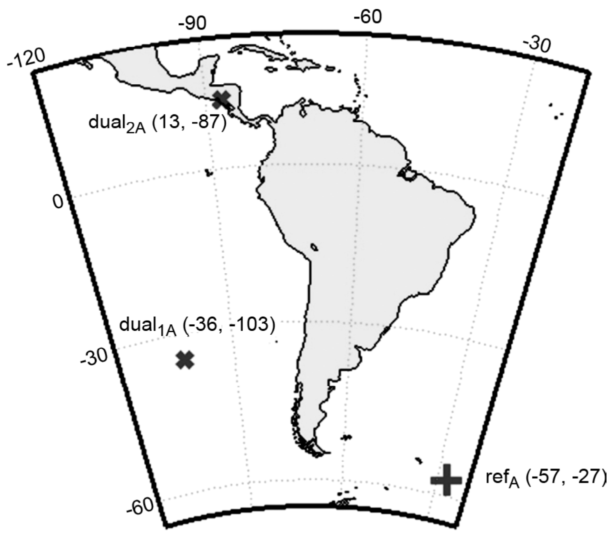

Figure 4. The positions of the reference region A (-57,-27), and its first best and second best dual zones.

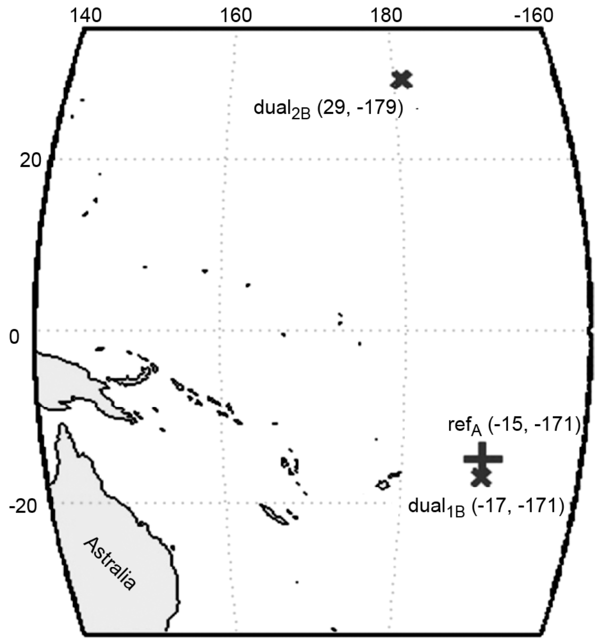

Figure 5. The positions of the reference region B (-15,-171) and its first best and second best dual zones.

also interesting that the dual zones of the reference cell *A* are located far from one another.

Following this step, and in order to make use of the event matrix as a tool conveying the precursory data for the future events, the reference row basically needs to get a one-column time shift (i.e. a three-month time interval) at a time, and the newly-shaped shifted matrix row will then be compared with the rest of the rows. For instance, according to the findings of BZEFM algorithm applied to the seismic catalogue's data gathered until June 30, 2009, it was revealed that the 2x2 cells with longitudinal and latitudinal degrees (*lat*: -23, *long*: -177) as for the center point, served well as a precursor for a second cell with coordinates (*lat*: 17, *long*: -101). Figure (6) shows how these two cells are taking a geographical stand. Other successful precursory cases as for the succeeding events of a three-month interval imminence (expected to occur within the target cell) are shown in Table (2). It is noteworthy that the target cell, here, is located on the Pacific Ring of Fire in the neighborhood of Mexico City, Mexico.

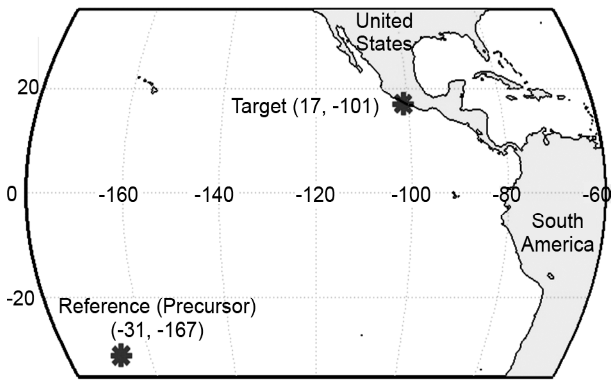

Figure 6. The positions of the reference (precursor) cell (-31, -167) and the target cell (17,-101).

Strictly speaking, these cases are the direct hints for calling the two mentioned areas as dual zones with a time difference of three months. In other words, the reference cell is fairly considered a precursor to the second cell. This duality is then examined over the evaluation time interval, i.e., the last year of the earthquake catalogue.

Table 2. An instance of a successful precursory case for the succeeding events of a three-month interval imminence in the target cell, found by BZEFM.

	Reference (Precursor) Cell: (-23,-177)	Target Cell: (17, -101)
Identification Period	4.8.1973 – M= 5.5	7.16.1973 – M= 6.2
	11.11.1974 – M= 5.6	2.22.1975 – M= 5.8
	2.14.1976 – M= 5.9	6.7.1976 – M=6.7
	1.13.1988 – M= 5.7	2.8.1988 – M= 5.8
	4.11.1998 – M= 6.2 4.11.1998 – M= 5.5 4.12.1998 – M= 5.5	7.11.1998 – M= 5.5 7.12.1998 – M= 5.5
Evaluation Period	2.22.2010 – M= 6 3.18.2010 – M= 5.6	4.14.2010 – M= 5.5

Based on this procedure and in regard to an earthquake event with magnitude of 6 (in Richter scale) dating back to Feb. 22, 2010, which has occurred in the reference cell, another earthquake event with a magnitude greater than 5.5 has simply been anticipated within a three-month interval as of March 1, 2010 to June 1, 2010. This forecasting did prove to be true since an earthquake of 5.5 Richter occurred on April 14, 2010 within the second cell. It should be noted that choosing crisp boundaries for time intervals is a drawback and avoids finding many dual zones worldwide. In the case of the examples for reference-target cells presented in this paper, some events fail to act as a precursor, because they were slightly out of the specific time interval. Therefore, if we choose the boundary of the cells fuzzy, it is expected to achieve promising results. This may be the subject of further research on BZEFM.

In another experiment, we reduced the size of cells to 1×1 degrees and obtained an event matrix with 64800 rows and 152 columns. To challenge the complexity problem again, we removed those rows of the matrix with less than five events $M \geq 4$. This way, we reduced the rows of the matrix to 4714. We then searched through the matrix to find the dualities. We found at least three dual cells for 4648 target cells out of 4714 (98.6%) over the identification period from January 1, 1973 to June 3, 2009. Using these dual cells, we managed to do successful forecast over the evaluation period, June 30, 2009 to June 30, 2010. This period includes 591 earthquakes with $M \geq 5.5$, which are happened in 272 days. Considering the 1×1 geographical divisions and the three-month time divisions, BZEFM may raise an alarm in 360×180×(12/3) occasions, corresponding to the elements of the event matrix on the validation period. This sparse matrix only has 438 non-zero elements.

Here, the performance of the algorithm is presented using the well-known matrix of confusion [29-30]. This matrix provides the rate of True/False forecasts for 192600 Positive/Negative occasions. BZEFM correctly forecasts d = 283 earthquakes with $M \geq 5.5$, and $a = 191348$ non-active elements of the event matrix. The model also incorrectly raises an earthquake alarm for $c = 1252$ cells, and also misses $b = 155$ events. The resulting confusion matrix is shown in Table (3), from which some standard statistical measures may be concluded. The BZEFM accuracy defined as the proportion of the total number of predictions that were correct is $(a + d)/(a + b + c + d)$=0.9926. The recall or true positive rate (TP) as the proportion of positive cases that are correctly identified is $d/(c + d) = 0.1844$. Similarly, the true negative rate (TN) is defined as the proportion of negatives cases that are classified correctly: $a/(a + b) = 0.9992$. Finally, precision (P) is the proportion of the predicted positive cases that are correct: $d/(b + d) = 0.6461$.

Table 3. Confusion matrix for forecasting earthquakes with M ≥ 5.5 over the evaluation period, 6.30.2009-6.30.2010, Using BZEFM.

		Prediction Outcome	
		Negative	Positive
Actual Value	Negative	True Negative: a = 190910	False Negative: b = 155
	Positive	False Positive: c = 1252	True Positive: d = 283

The BZEFM seems to have great accuracy; however, it may not be an adequate performance measure [31], because the number of negative cases, in this experiment, is much greater than the number of positive cases. In fact, there are 191065 negative cases out of 192600 elements of the event matrix. If the model classifies them all as negative, the accuracy would be 99.77%, even though the classifier missed all positive cases. Therefore, other performance measures should be employed, e.g., geometric mean g_{mean} [31], as defined as:

$$g_{mean1} = \sqrt{(TP.P)} \quad , \quad g_{mean2} = \sqrt{(TP.TN)} \tag{5}$$

and $F_{measure}$ [32]

$$F = \frac{(\beta^2 + 1)P \times TP}{(\beta^2 + 1)P + TP} \qquad (6)$$

where β is a value from 0 to infinity and is used to control the weight assigned to *TP* and *P*.

Any classifier evaluated using (5) or (6) will have a measure value of 0, if all positive cases are classified incorrectly. Here, $g_{mean1} = 0.3452$, and $g_{mean2} = 0.4292$. For $\beta = 1$, which assigns equal weights to precision and recall, $F_{measure} = 2P.TP / (P + TP) = 0.2869$.

The above statistical measures are listed in Table (4). These values indicate the acceptable forecasting power of BZEFM in our experiment.

Table 4. Standard statistical measures for BZEFM results, forecasting earthquakes M ≥ 5.5 over the evaluation period, 6.30.2009-6.30.2010.

Accuracy	99.26%
True Positive Rate	18.44%
True Negative Rate	99.92%
Precision	64.61%
g_{mean1}	0.3452
g_{mean2}	0.4292
$F_{measure}$	0.2869

5. Conclusion

The main purpose of this survey, which elaborated BZEFM, was basically to introduce a new approach for predictions of the future earthquakes with different precursory intervals. In the experiments, we provided three cases of duality founded by the method, The many contemporary events in these dual zones, which were all away from each other, support the main idea presented by BZEFM; however, it should be mentioned that there is not always a dual for every cell, and so we cannot provide forecasts for any given region in the world.

Two main ideas that go along with this survey are as follows:

Those areas which follow one another as precursors and postcursors are referred to as seismic dual zones. Most of the zones are located not nearby but fairly distant from one another. Explaining the results in the framework of plate tectonics theory may be considered as the objective of another research paper. In this paper, we have only represented the dualities, which are achieved through an exhaustive search on the earthquake catalogue. In fact, the concept of “duality” introduced in this paper, is not based on a theory. Instead, it is the result of extracting information or reality from the raw data, using data mining algorithms.

In some cases, it happens for a particular cell to initially be a precursor to one specific zone within a specific time interval, and as time passes, it eventually turns out to be a precursor to another zone within a different period of time, i.e. there are some evidences that duality among two or multiple cells has changed in the time. Changes of this kind will need to be scrutinized.

Finally, it should be noted that reducing the size (in latitudinal/longitudinal degrees) of the cells and also reducing the threshold level of magnitude are of high significance. Both cases will be paid due attention during the complementary phase of this project, and the same procedure of BZEFM will be applied to them in that new phase of research.

References

1. Plagianakos, V.P. and Tzanaki, E. (2001). Chaotic Analysis of Seismic Time Series and Short Term Forecasting Using Neural Networks, The IEEE, International Joint Conference on Neural Networks, *In Proc.*, **3**, 1598-1602.

2. Geller, R.J., Jackson, D.D., Kagan, Y.Y., and Mulargia, F. (1997). Earthquakes Cannot Be Predicted, *Science*, **275**, 1616-1617.

3. Kossobokov, V.G., Keilis-Borok, V.I., Turcotte, D.L., and Malamud, B.D. (2000). Implication of a Statistical Physics Approach for Earthquake Hazard Assessment and Forecasting, *Pure App. Geophys.*, **157**, 2323-2349.

4. Kagan, Y.Y. and Jackson, D.D. (2000). Probabilistic Forecasting of Earthquakes, *Geophysical Journal International*, **143**(2), 438-453.

5. Bleier, T. and Freund, F. (2005). Earthquake Alarm, *IEEE Spectrum*, **42**(12), 16-21.

6. Johnston, M.J.S. (1997). Review of Electric and Magnetic Fields Accompanying Seismic and Volcanic Activity, *Surv. Geophys.*, **18**, 441-475.

7. Magro-Campero, A., Fleischer, R.L., and Likes,

R.S. (1980). Changes in Subsurface Radon Concentration Associated With Earthquakes, *Journal of Geophysical Research*, **85**, 3053-3057.

8. Singh, S., Kumar, A., Singh Bajwa, B., Mahajan, S., Kumar, V., and Dhar, S. (2010). Radon Monitoring in Soil Gas and Ground Water for Earthquake Prediction Studies in North West Himalayas, India, *Journal Terrestrial Atmospheric and Oceanic Sciences Cited Citing*, **21**(4), 685-695.

9. Zhiping, J. and Zanjun, W. (1995). Application of Electromagnetic Radiation in Short-Impending Earthquake Prediction, Platean Earthquake Research, 4.

10. Guo, G. and Wang, B. (2008). Cloud Anomaly Before Iran Earthquake, *International Journal of Remote Sensing*, **29**(7), 1921-1928.

11. Bhargava, N., Katiyar, V.K., Sharma, M.L., and Pradhan, P. (2009). Earthquake Prediction through Animal Behavior: a Review, *Indian J. of Biomechanics, Special Issue*, 159-165.

12. Turcotte, D.L. (1991). Earthquake Prediction, An. Rev. Earth Planet. Sci., **19**, 263-281.

13. Kanamori, H. (2003). Earthquake Prediction: an Overview, International Handbook of Earthquake and Engineering Seismology, Academic Press, Amsterdam, Netherland.

14. Akasheh, B. and Kossobokov, V. (1989). Premonitory Clustering before Strong Earthquakes in Iran-Afghan Region, Bollettino di Geofisica Teorica ed Applicata, **31**(122), 159-162.

15. Keilis-Borok, V.I. and Kossobokov, V.G. (1987). Periods of High Probability of Occurrence of the World's Strongest Earthquakes, *Computational Seismology*, **19**, Allerton Press Inc., 45-53.

16. Keilis-Borok, V.I. and Kossobokov, V.G. (1990). Premonitory Activation of Seismic Flow: Algorithm M8, *Phys. Earth and Planet. Inter.*, **61**(1-2), 73-83.

17. Romashkova, L.L., Kossobokov, V.G., Peresan, A., and Panza, G.F. (2001). Intermediate-Term Medium-Range Earthquake Prediction Algorithm M8: A New Spatially Stabilized Application in Italy, ICTP, Trieste, Italy, Internal Report.

18. Romashkova, L.L., Kossobokov, V.G., Peresan, A., and Panza, G.F. (2002). The Spatially Stabilized Variant of M8 Algorithm in Application to Prediction of Earthquakes From Consequent Magnitude Ranges: Italy, ICTP, Trieste, Italy, Internal Report.

19. Kossobokov, V.G., Keilis-Borok, V.I., and Smith, S.W. (1990). Localization of Intermediate-Term Earthquake Prediction, *J. Geophys. Res.*, **95**, 19763-19772.

20. Vorobieva, I.A. (1999). Prediction of Subsequent Large Earthquake, *Phys. Earth Planet. Inter.*, **111**(3-4), 187-196.

21. Keilis-Borok, V.I., and Rotwain, I.M. (1990). Diagnosis of Time of Increased Probability of Strong Earthquakes in Different Regions of the World: Algorithm CN, *Physics of the Earth and Planetary Interiors*, **61**, 57-72.

22. Tiampo, K.F., Rundle, J.B., McGinnis, S., and Klein, W. (2002). Pattern Dynamics and Forecast Methods in Seismically Active Regions, *Pure App. Geophys.*, **159**, 2429-2467.

23. Nanjo, K.Z., Holliday, J.R., Chen, C.C., Rundle, J.B., and Turcotte, D.L. (2006). Application of a Modified Pattern Informatics Method to Forecasting the Location of Future Large Earthquakes in the Central Japan, *Tectonophysics*, **424**, 351-366.

24. Keilis-Borok, V.I. and Soloviev, A.A. (Eds.), (2003). Nonlinear Dynamics of the Lithosphere and Earthquake Prediction, Springer-Verlag, Heidelberg, ch. 1.

25. Kushida, Y. (2000). Method for Detecting Diastrophism by Detecting VHF Radio Waves Reflected by the Ionosphere, U.S. Patent, No. 6018244.

26. Kushida, Y. and Kushida R. (2002). Possibility of Earthquake Forecast by Radio Observations in the VHF Band, *Journal of Atmospheric Electricity*, **22**, 239-255.

27. Pulinets, S.A. (2004). Ionospheric Precursors of Earthquakes; Recent Advances in Theory and Practical Applications, *Terrestrial, Atmospheric and Oceanic Sciences*, **15**(3), 445-467.

28. Bali-Lashak, A. (2010). Probabilistic Model of Earthquake Forecasting Based on Earthquake Catalogue of Iran, Ph.D. Thesis, International Institute of Earthquake Engineering and Seismology, Tehran, Iran.

29. Provost, F. and Fawcett, T. (2001). Robust Classification for Imprecise Environments, *Machine Learning*, **42**(3), 203-231.

30. Provost, F., Fawcett, T., and Kohavi, R. (1998). "The Case Against Accuracy Estimation for Comparing Induction Algorithms", *Proceedings of the 15th International Conference on Machine Learning*, 445-453, Madison, WI. Morgan Kauffmann.

31. Kubat, M. and Matwin, S. (1997). "Addressing the Curse of Imbalanced Training Sets: One Sided Selection", *Proceedings of the 14th International Conference on Machine Learning*, 179-186, Nashville, Tennesse, Morgan Kaufmann.

32. Lewis, D. and Gale, W. (1994). "A Sequential Algorithm for Training Text Classifiers", Proceedings of ACM-SIGIR 1994.