

Assessing performances of pattern informatics method: a retrospective analysis for Iran and Italy

M. Y. Radan · H. Hamzehloo · A. Peresan · M. Zare · H. Zafarani

Received: 10 August 2012 / Accepted: 21 March 2013 / Published online: 30 March 2013
© Springer Science+Business Media Dordrecht 2013

Abstract Pattern informatics (PI) algorithm, which was introduced at the beginning of past decade, uses instrumental earthquake catalogs to investigate the time-dependent rate of seismicity in the study area and, based on the information from past events, calculates the probabilities for the occurrence of future large earthquakes. The main measure in this method is the number of events above a specified magnitude threshold M_c that is counted over a gridded area. PI has been applied in several regions of the world and different variants of the method have been developed over the past decade. Hence, the problem of formally evaluating and comparing the performances of the different PI variants needs to be addressed from an operational perspective, in order to identify the preferred application scheme and as well as to improve the performances of the method. In this study, PI is applied for the first time to the retrospective analysis of the earthquake catalogs of Iran and Italy, so as to check whether this method could forecast the past large events in these two regions with different level of data completeness and complex seismotectonic setting. The original PI algorithm and one of its modified variants, as well as the relative intensity (RI) model, are used to check the stability and statistical significance of the obtained results. In order to assess and compare the obtained results, the performances of the different PI variants are analyzed considering different evaluation strategies, which turn out to provide significantly different scores even for the same algorithm variant. We show that a critical point in the assessment of the obtained results is related with the definition and quantification of the space uncertainty of the issued forecasts, that is, with the extent of the territory where large earthquakes are to be expected. Accordingly, we emphasize the need for an appropriate definition of the evaluation strategies, clearly and unambiguously indicating the area where a large earthquake has to be expected. The study shows that, with

M. Y. Radan · H. Hamzehloo (✉) · M. Zare · H. Zafarani
International Institute of Earthquake Engineering and Seismology, IIEES, Tehran, Iran
e-mail: hhamzehloo@yahoo.com

A. Peresan
Department of Mathematics and Geosciences, University of Trieste, Trieste, Italy

A. Peresan
The Abdus Salam International Centre for Theoretical Physics, SAND Group, Trieste, Italy

respect to application in Iran and Italy, the performances of PI algorithm (both original and modified variants) are highly dependent on the selected evaluation strategy and do not provide better information than the simple RI model, which does not account for temporal properties of seismicity evolution. The overall performances can be improved by introducing specific thresholds that discard the less active cells; however, being based on some posterior optimization, a rigorous prospective testing is required to assess the forecasting capability of the method. In this paper, we aim to set up the rules for such testing, including advance definition of the evaluation strategy.

Keywords Earthquake forecasting · Pattern informatics · Relative intensity · Evaluation strategy

1 Introduction

Assessing and comparing performances of forecasting models have been the subject of many studies in the last decade. The conclusions, however, may be quite dependent on the specific way the results are evaluated, which includes selection of the statistical tests and reference models, but also a number of possible options in the simple quantification of forecasting scores (e.g., definition of forecasted events and false alarms). All of these options provide opportunity for some tuning, particularly when dealing with retrospective results; in this study, we focus on the definition of the forecasted earthquakes, in connection with the definition of the area where such events have to be expected.

A number of methods aimed at earthquake prediction have been proposed to investigate whether large earthquakes can be forecasted before their occurrence. These efforts can be grouped in two main categories, although different classifications can be proposed (e.g., Jordan et al. 2011). One is based on the empirical observations of the dynamics of the system (e.g., GPS observations) and some other measurable variable physical phenomena that can be eventually related to earthquakes (e.g., geochemical, geo-electrical anomalies); this kind of methods is commonly referred as *precursory methods*. Another possibility, with low cost, is to analyze earthquake catalogs using statistical approaches in order to find models of seismic evolution and to make forecasts about the possible occurrence of large earthquakes; this second approach is referred as *statistical methods*. Also, seismicity patterns (e.g., activation, quiescence, Mogi doughnuts) that are derived from earthquake catalogs by the means of some statistical and mathematical procedures, can be categorized as statistical methods.

The pattern informatics (PI) approach to earthquake forecasting (described later on) has been proposed in the beginning of past decade (Rundle et al. 2000a; Tiampo et al. 2002a) and has experienced several modifications till now, which makes a statistical evaluation of its performances very difficult, if not impossible so far. The approach is based on the strong space–time correlations that are responsible for the cooperative behavior of driven threshold systems and arise both from threshold dynamics and from the mean field (long range) nature of the interactions.

Driven threshold systems are composed of interacting spatial networks of cells, each having one or more inputs, an internal state variable that evolves in time in response to inputs, and one or more outputs. Each cell is connected to other cells by means of a network of interactions and to an external driving source. Threshold dynamics arise when a cell is subjected to this persistent external forcing, increasing the value of the internal state variable through time until a predefined failure threshold is reached, where the cell fails,

reducing the internal state variable to a residual value. Thresholds, residual values, internal states, and the resulting dynamics may be affected by the presence of noise and disorder. Mean field threshold systems arise when the coupling between cells is long range but weak, leading to suppression of all but the longest wavelength fluctuations. The system dynamics often result in strong space–time correlations in cell failures over many scales (Klein et al. 1997; Tiampo et al. 2003). Tiampo et al. (2002a) suggested that seismicity can be described by pure phase dynamics (Mori and Kuramoto 1997; Rundle et al. 2000a, b), in which the main changes in seismicity are mainly associated with the rotations of the state vectors in a Hilbert space (Fukunaga 1970; Holmes et al. 1996; Mori and Kuramoto 1997; Rundle et al. 2000a). The length of the state vector represents the average temporal frequency of events in the region and is nearly related to the rate at which stress is dissipated. Following that the information about space–time fluctuations in the system state can be represented only by the phase angle of the state vector. Changes in the norm of the state vector represent only random fluctuations, most of which can be removed by requiring the system state vector to have a constant norm.

The PI method was first introduced by Rundle et al. (2000a) as an implication of the diffusive mean field nature of earthquake dynamics. By treating seismicity as a self-organizing threshold system, they created a forecast map for the occurrences of future large earthquakes in southern California. At that time, the method was known as phase dynamical probability change (PDPC). Tiampo et al. (2002a) described the PDPC method in mathematical terms and provided an explanation for each step of calculations. They also performed likelihood tests against various null hypotheses and showed that the PDPC method can forecast earthquakes better than the methods using other measures of past seismicity, like relative intensity (described later on) measures.

Seismic activation and quiescence have been widely studied for earthquake prediction purposes, and significant efforts are currently devoted to validate forecasting models based on seismic activity changes. The ultimate answer about predictive capability of any method, however, can only come from rigorous prospective testing. Although several authors have claimed that such rate changes cannot really be detected (Hardebeck et al. 2008; van Stiphout et al. 2011), some formally defined methods, which underwent rigorous real-time testing over the last decades, already demonstrated that specific patterns in seismic activity may provide statistically significant precursors for large earthquakes (Kossobokov 2012). PI method detects both activation and quiescence in a region by means of the analysis of the seismicity rate changes over time (i.e., Holliday et al. 2005; Chen et al. 2005; Wu et al. 2008a, b; Shcherbakov et al. 2010; Jiang and Wu 2010, 2011; Li 2008; Zhang et al. 2012). In this study, we aim to set up an experiment, which may permit to assess PI forecasting capability, as well as to compare its performances with other tested methods based on the analysis of seismicity patterns.

PI method has been applied in various regions worldwide, generally concluding that PI method has some predictive skill (mostly based on retrospective analysis), at least in the tested areas like California, Taiwan, Canada, and also on a worldwide basis. Specifically, Holliday et al. (2005) performed forecasts for California, Japan and worldwide using PI method in its original formulation. Chen et al. (2005) modified the PI method to investigate the binary forecasts for the 1999 Chi–Chi earthquake in Taiwan, using the Taiwanese catalogs. Later on, Holliday et al. (2006) performed a systematic procedural and sensitivity analysis of the method; they modified the ordering of the different steps of the procedure and the parameter values and found optimal choices for the southern California region. Nanjo et al. (2006) further modified the PI method for use with the Japanese catalogs. Their results showed that the epicenter of the October 23, 2004, $M = 6.8$ Niigata, Japan

earthquake, could be successfully forecasted by retrospective analysis. A common feature to all these experiments is a series of ad hoc adaptations of the procedure to the study area, all implying some sort of data fitting, which hamper the standard application and large-scale validation of the approach.

A Regional Earthquake Likelihood Models earthquake forecasting based on PI was set up by Holliday et al. (2007). In that paper, they prepared a composite forecast map by superimposing PI map on RI map for a 5-year period for California. Their results showed that PI outperformed the RI method in California. Wu et al. (2008a) applied PI to Taiwan seismicity for detecting precursory seismic activation of the Pingtung (Taiwan) offshore doublet events. Later on, Wu et al. (2008b) used PI to detect migration patterns in seismicity. Zechar and Jordan (2008) presented a method for testing alarm-based earthquake forecasts, based on Molchan errors diagram (described in Sect. 4). They tested forecasts from RI, PI, and NSHM models, and their results have shown that neither PI nor NSHM provides significant performance gain relative to the RI reference model in California. Toya et al. (2009) used PI approach to forecast earthquakes in 3D. Tiampo et al. (2010) investigated the ergodicity in seismic catalogs and its insights for PI applications. Shcherbakov et al. (2010) applied PI for forecasting large earthquakes on a worldwide basis. Jiang and Wu (2010) with a retrospective procedure investigated the PI applicability to Wenchuan earthquake and showed that PI forecast outperformed the RI forecast. Also, Jiang and Wu (2011) investigated the effects of aftershocks removal on the performance of PI algorithm. This study indicated that when an intense aftershock sequence is included in the “sliding time window,” the hotspot picture may vary, and the variation lasts for about 1 year. PI forecasts seem to be affected by the aftershock sequence included in the “change interval,” and the PI forecast using “background events” seems to have a better performance. Li and Chen (2011) used PI for investigating the characteristics of long-term regional seismicity before the Wenchuan event. Also, Zhang et al. (2012) performed a retrospective study on the prediction capabilities of PI considering the Wenchuan M8.0 and Yutian M7.3 earthquakes and showed that the PI method could forecast both events.

2 Methodology

The PI algorithm measures the seismicity in two time intervals, the so-called reference (long-term) and change (medium-term) periods. By comparison of these measurements, the relative activation and quiescence are defined in terms of intensity changes and the probabilities for the occurrence of large earthquakes are computed. In order to test the forecasting algorithm, a potential map for future large earthquakes is outlined. The PI method has been proposed as a technique for forecasting large events based on retrospective applications to earthquake data from southern California (e.g., Tiampo et al. 2002a, b; Rundle et al. 2002, 2003). Later on, several variants of the method have been proposed (Chen et al. 2005; Nanjo et al. 2006, Holliday et al. 2006) in order to improve its performances and to allow its application in different regions worldwide. In this work, we consider the versions described by Tiampo et al. (2002a) and Chen et al. (2005).

2.1 PI basic version

The basic version of PI that was introduced by Tiampo et al. (2002a) is a six-step process that creates a time-dependent system state vector in a real-valued Hilbert space and uses the phase angle to predict future states (Rundle et al. 2003). The method is based on the

idea that the future time evolution of seismicity can be described by pure phase dynamics (Rundle et al. 2000a, b; Tiampo et al. 2002a). Hence, a real-valued seismic phase function $\hat{I}(x_i, t_b, t)$ is constructed and allowed to rotate in its Hilbert space. To denoise the seismicity, temporal averages of seismic activity are utilized in the method.

The region under study is binned into N boxes with dimension $dx \times dx$ centered at a point x_i . Within each box, a time series $n_{\text{obs}}(x_i, t)$ is defined by counting how many earthquakes with magnitude greater than the magnitude of completeness of seismic data (M_c) occurred during the time period from t to $t + dt$. Next, the activity rate function $I(x_i, t_b, t)$ is defined as the average rate of occurrence of earthquakes in box i over the period t_b to T :

$$I(x_i, t_b, T) = \frac{1}{T - t_b} \sum_{t=t_b}^T n_{\text{obs}}(x_i, t)$$

The activity rate function is then normalized by subtracting the spatial mean over all boxes and by scaling to give a unit norm:

$$\hat{I}(x_i, t_b, T) = \frac{I(x_i, t_b, T) - \frac{1}{N} \sum_{j=1}^N I(x_j, t_b, T)}{\sqrt{\sum_{j=1}^N [I(x_j, t_b, T) - \frac{1}{N} \sum_{k=1}^N I(x_k, t_b, T)]^2}}$$

The relevant changes in seismicity will be given by the change in the normalized activity rate function for the time period t_1 to t_2 :

$$\Delta \hat{I}(x_i, t_b, t_1, t_2) = \hat{I}(x_i, t_b, t_2) - \hat{I}(x_i, t_b, t_1)$$

In order to remove the last free parameter in the system, the choice of base year, and to further reduce random noise components, the changes in the normalized activity rate function are averaged over all possible base time periods:

$$\Delta \hat{I}(x_i, t_b, t_1, t_2) = \frac{1}{t_1 - t_0} \sum_{t_b=t_0}^{t_1} \Delta \hat{I}(x_i, t_b, t_1, t_2)$$

Finally, the probability of change of activity in a given box is deduced from the square of its base averaged, mean normalized change in activity rate:

$$P(x_i, t_0, t_1, t_2) = [\Delta \hat{I}(x_i, t_0, t_1, t_2)]^2$$

This probability function is often given relative to the background by subtracting its spatial mean:

$$P'(x_i, t_0, t_1, t_2) = P(x_i, t_0, t_1, t_2) - \frac{1}{N} \sum_{j=1}^N P(x_j, t_0, t_1, t_2)$$

where P' indicates the probability of change in activity (activation or quiescence) and is measured relative to the background.

2.2 PI-modified version

In this modified version, in addition to the use of Moore neighborhood in calculating intensities in each cell and use only the most active cells into calculations, Chen et al. (2005) considered both temporal and spatial normalization after computing the changes of

intensities in the cells. Accordingly, in order to compare the intensities from different time intervals, the intensities should have the same statistical properties. Therefore, each of the seismic intensities should be normalized both individually over all choices for t_b and aggregately at each choice for t_b . This normalization is performed by subtracting the mean seismic activity from the intensity and dividing the results by the standard deviation of the seismic activity. The statistically normalized seismic intensity of box i is thus defined by:

$$\tilde{\Delta I}_i(t_b, t_1, t_2) = \frac{\Delta I_i(t_b, t_1, t_2) - \langle \Delta I_i(t_b, t_1, t_2) \rangle_T}{\sigma_T}$$

$$\hat{\Delta I}_i(t_b, t_1, t_2) = \frac{\tilde{\Delta I}_i(t_b, t_1, t_2) - \langle \tilde{\Delta I}_i(t_b, t_1, t_2) \rangle_A}{\sigma_A}$$

where $\langle \Delta I_i(t_b, t_1, t_2) \rangle_T$ is the mean intensity difference of box i averaged over all choices of t_b , and $\langle \tilde{\Delta I}_i(t_b, t_1, t_2) \rangle_A$ is the time averaged mean intensity difference averaged over all the boxes at each choice of t_b , and then σ_T and σ_A are the respective standard deviations (Chen et al. 2005; Wu et al. 2008a, b).

2.3 Relative intensity (RI)

Holliday et al. (2005) used RI model, as a null hypothesis better than a random guess, for testing PI forecasts. RI is a non-clustered seismicity model, according to which future large earthquakes are considered more likely where higher seismic activity occurred in the past. To assess that, it uses the rate of occurrence of past earthquakes in the area under investigation. It was originally formulated as a binary forecast, although it has been modified in several ways since that time.

In RI application, the region under study is tiled into square boxes with dimension $dx \times dx$. Within each box, the number of earthquakes with magnitude $M \geq M_c$ is determined over the entire time period of the catalog, where M_c is the magnitude of completeness of the data. The RI score for each box then is computed as the total number of earthquakes in the box in that time period, divided by the largest value obtained in all of the considered boxes. A threshold value in the interval $[0, 1]$ is then selected, and all values above that are expected to have a large event over the forecast period of interest, resulting in a binary forecast. The remaining boxes with RI scores smaller than the threshold represent locations at which large earthquakes are not expected to occur. The result is a map of locations where large earthquakes are forecast to occur over some future intermediate-term time span (Holliday et al. 2005; Tiampo and Shcherbakov 2012).

3 Applications of PI method

Following the original version of PI as described in Tiampo et al. (2002a) and the modified version of PI as defined by Chen et al. (2005), the corresponding program codes have been developed. The new codes have been preliminary tested in California, where the hotspot map obtained by Tiampo et al. (2002a) was properly reproduced, with a few minor differences, which are apparently due to catalog revision. Similarly, applying Chen's modified PI, the test described by Holliday et al. (2005) was carried out for California. The results we obtained are pretty similar to what they mentioned, with marginal differences,

which could be also attributed to differences in the data used in this study and those used by Holliday et al. (2005). Once tested the codes, we applied these variants of PI to Iranian and Italian earthquake catalogs. These two areas were selected in order to investigate the performances of PI method in two regions having a quite complex tectonic settings and different level of data completeness. Further research is envisaged, to assess the possible dependence of PI performances on the tectonic setting of the study area.

3.1 Study regions

Iran is one of the seismically active areas of the world and is frequently affected by destructive earthquakes, causing heavy losses in human lives and widespread damage. The Iranian plateau is situated within the Alpine-Himalayan seismic belt, which is recognized as one of the seismically active areas of the world. Deformation and seismicity in the Iranian plateau are mainly due to the continental shortening between Eurasian and Arabian plates. The Iranian plateau can be divided into five major geological units based on tectonic history, magmatic events, or sedimentary features (Nabavi 1976). These units are well known as Zagros, Sanandaj–Sirjan, Central Iran, East and Southeast zones, and Alborz, each of which is subdivided into a number of subunits with specific characteristics. Its deformation is due to the continuous convergent movement between Arabian plate, to the southwest, and Turan platform to the northeast, with the north-northeast drift of African-Arabian plate against Eurasia.

Compared to Iran, the Italian peninsula and the whole Mediterranean area exhibit a considerable heterogeneity in the tectonic regime, revealed by the coexistence of fragmented seismogenic structures of greatly different kind (Cuffaro et al. 2011; Meletti et al. 2000, 2008). The recent geodynamics of the Central Mediterranean region is controlled by the Africa–Europe plate interaction and by the subduction of the southwestern margin of the Adria plate, along the Apennines (Cuffaro et al. 2010). Northern Italy is characterized by the presence of a main structure, the Alpine arc, which is generally uplifting (Cloetingh et al. 2006; D’Agostino et al. 2005) with some westerly strike-slip motion, and therefore, the majority of focal mechanisms are compressive or transpressive (Vannucci and Gasperini 2004; Guidarelli and Panza 2006; Basili et al. 2008). The central part of the Italian peninsula, along the Apennines, is characterized by a band with tensional seismotectonic behavior, with prevailing dip-slip focal mechanism (normal faulting). The southern part of the Italian peninsula is characterized by a seismotectonic regime controlled by the sinking of the ionic plate under the Calabrian Arc, toward the Tyrrhenian Sea (e.g., Brandmayr et al. 2010; Splendore et al. 2010).

3.2 Data sets and parameters

For the application of PI method in Iran, we consider the global composite catalog compiled by ISC (2012). This composite catalog reports the size of earthquakes by the means of various magnitude scales, whenever available, estimated by different agencies. In this study, we considered maximum reported magnitude (M_{\max}) for each individual event. We use all of the available data from ISC Bulletin (in CSV format) up to the end of 2011, including also unrevised data since July 1, 2010.

The area under investigation is between 44°E–64°E longitude and 25°N–40°N latitude. To cope with uncertainties in location, we choose a box size $dx = 0.2^\circ$ in longitude and latitude, which is larger than the standard box size ($dx = 0.1^\circ$) used in most of PI applications that is hardly compatible with accuracy in epicentral determinations (Cho and

Tiampo 2012). A preliminary analysis of magnitude of completeness as a function of time has been performed, within the area of investigation, by considering the distribution of the number of events versus time and magnitude (Romashkova and Peresan 2013), as well as using ZMAP software (Wiemer 2001). Accordingly, the cutoff magnitude here is set to $M_c = 4.5$, so that the resulting catalog is satisfactorily complete for the analysis since 1980 (Fig. 1a). The epicenters of all $M \geq 4.5$ earthquakes reported in *ISC* catalog since 1980 are shown in Fig. 1b.

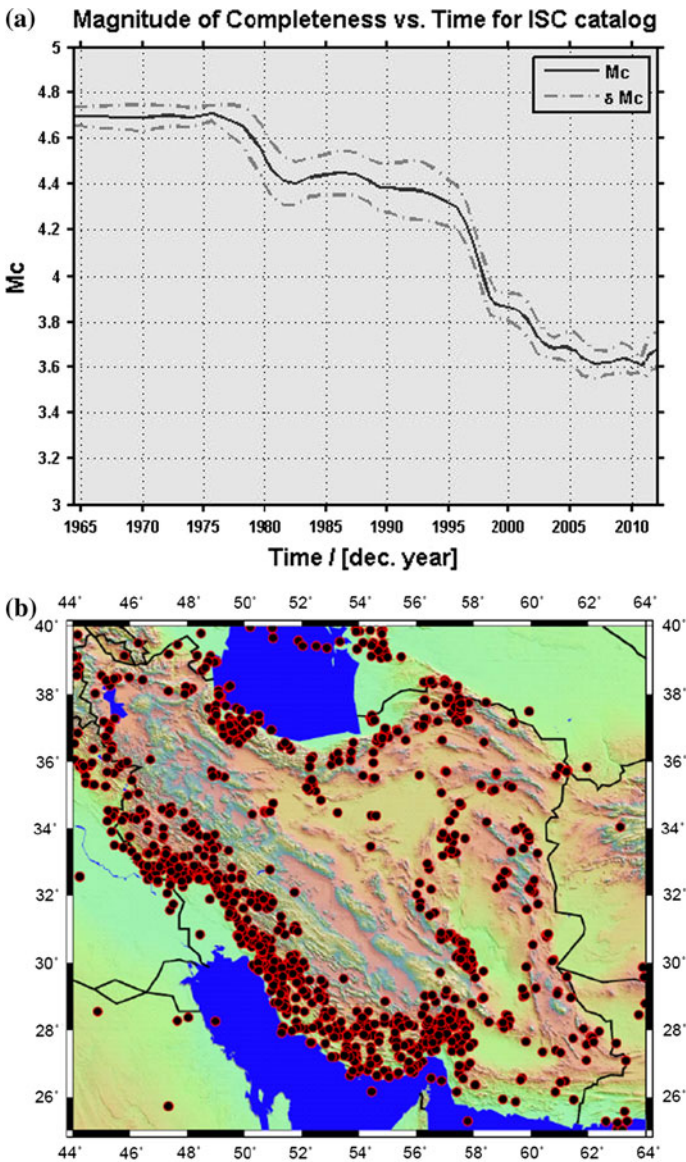


Fig. 1 **a** Analysis of M_c versus time for *ISC* catalog within the Iranian area, performed by ZMAP. **b** Distribution of all $M \geq 4.5$ events as reported in *ISC* catalog for 1980–2011

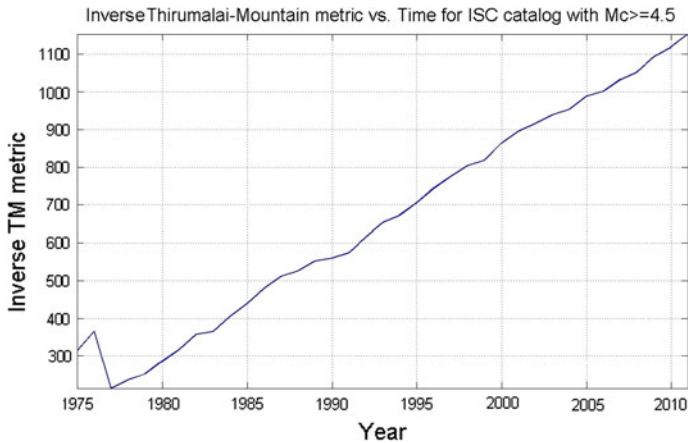


Fig. 2 Plot of inverse TM metric versus time, considering earthquakes with $M \geq 4.5$ reported in *ISC* catalog for the Iranian area (Fig. 1b)

Using insights from previous investigations of ergodicity in earthquake catalogs (Tiampo et al. 2007, 2010), a study of ergodicity of the catalogs was carried out. In these applications, Thirumalai-Mountain (TM) fluctuation metric (Thirumalai et al. 1989; Mountain and Thirumalai 1992; Thirumalai and Mountain 1993) over some periods of time was calculated, and the curves of inverse TM metric versus time were plotted. Effective ergodic periods are identified as the time windows when the inverse TM metric is linear in time with a positive slope. In these windows, the PI algorithm has been proposed to be able to estimate the time-dependent earthquake rates. Cho et al. (2010) proposed that the evolution of the cumulative number of events can be seen as a diffusive process during these time windows and deviations of the inverse TM metric from a linear trend occurs as a result of the clustering of events in time and/or location. As shown in Fig. 2, for *ISC* catalog over the mentioned time, the plot of inverse TM metric for $M_c = 4.5$ has a monotonic positive slope.

For PI application in Italy, we consider the *UCI* catalog (Peresan and Panza 2002 and its updates), which is currently used for real-time prediction experiment, based on the identification of precursory seismicity patterns (Peresan et al. 2005). This choice is due to certain discontinuity and heterogeneity of the Italian instrumental earthquake catalogs available (Nanjo 2010; Romashkova and Peresan 2013), which may affect calibration and testing results. Moreover, it permits comparing the results of PI algorithm with those provided by CN algorithm (Peresan et al. 2005) using the same input data. The area under investigation is between 6.8°E – 18.3°E longitude and 36°N – 47°N latitude (Fig. 4). For each individual event, we considered the magnitude estimate selected according to the priority order (M_{prio}) as defined in Peresan et al. (2005). Accordingly, the *UCI* catalog can be considered sufficiently complete, within the Italian territory, with a cutoff magnitude of $M_c = 3$, starting on January 1, 1955. Also, we set $dx = 0.2^\circ$ to cope with uncertainties in the location of earthquakes reported in the input catalog.

Similar to the analysis performed for Iran, we construct TM plots for Italy for different time windows. The application of TM metric to *UCI* catalog, with $dx = 0.2^\circ$ and for the period 1955–2011, is shown in Fig. 3. Although the catalog is fairly complete for $M_c = 3$ since 1955, according to the criteria by Jiang and Wu (2011), based on TM plots, it seems

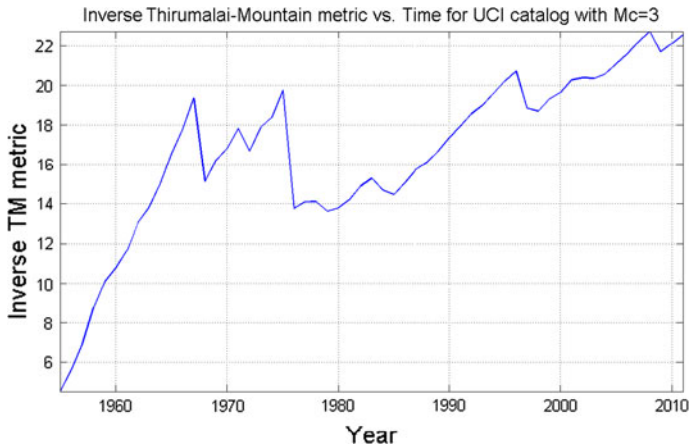
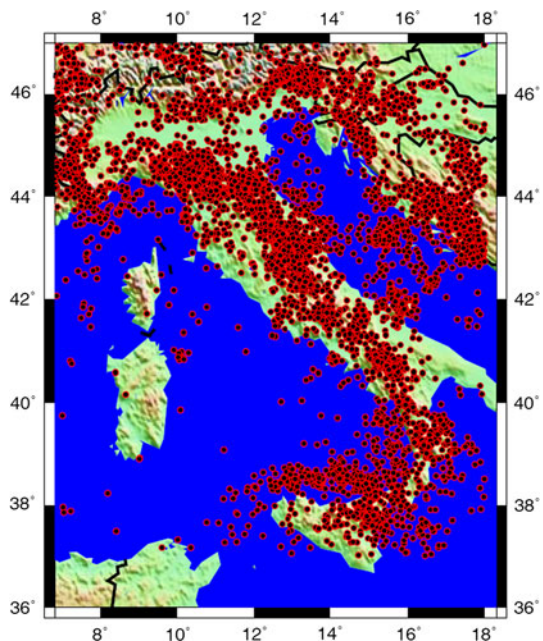


Fig. 3 Inverse TM metric versus time, considering the earthquakes with $M \geq 3.0$ reported in *UCI* catalog for the period 1955–2011 and for the area shown in Fig. 4

preferable to set 1975 as the beginning of data analysis. The epicenters of the $M \geq 3$ earthquakes, as reported in *UCI* catalog since 1975, are shown in Fig. 4.

In PI application to Iran and Italy, we wish to focus on retrospective forecasting of large earthquakes in the period 1996–2011. However, according to TM plots, in Iran, the data are complete for $M_c = 4.5$ only after 1980, while in Italy the data are adequate for $M_c = 3.0$ only after 1975. Since we wish to consider the same forecasting time intervals in Iran and Italy, we select $t_1 = 1990$, which guarantees a long-term period of 10 years in Iran and 15 years in Italy, respectively. This choice determines the hotspot maps to be valid from 1996, accounting for a change period of 5 years.

Fig. 4 Distribution of epicenters for all $M \geq 3$ events within the Italian area, as reported in *UCI* catalog for 1975–2011



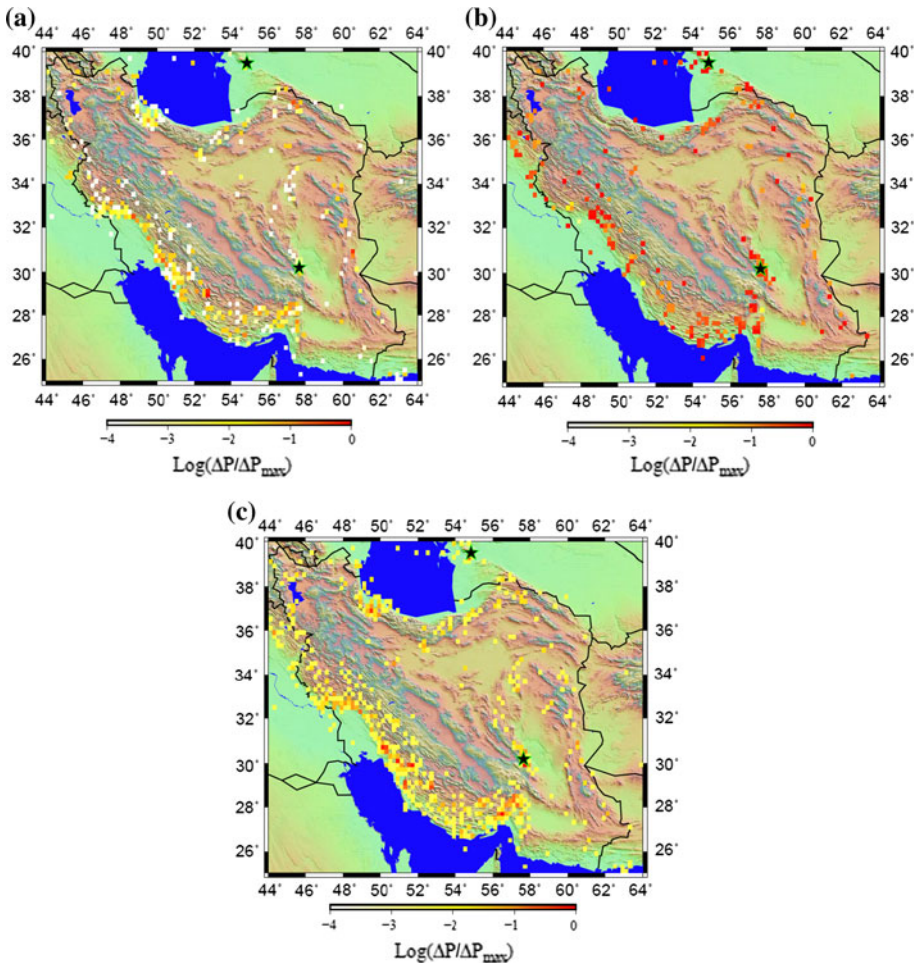


Fig. 5 Hotspot maps for $M \geq 6.5$ earthquakes within Iran territory, referred to the period from January 1, 1998, to January 31, 2002: **a** PI-T version. **b** PI-C version with $N \geq 1$. **c** RI. The *black stars* are the target events that occurred in the forecast time window

3.3 Results for Iran

A number of tests is carried out using the original version of PI algorithm (hereinafter, T) and its modified version (hereinafter, C), as well as the (RI) model. A new application scheme, based on yearly updating of hotspot maps with $dx = 0.2$, is proposed. Here, we apply the method in both versions, targeting sixteen 5-year forecast sliding time intervals (shifted by 1 year), beginning from January 1996 and up to January 2011. It is notable that in application of C version, we test different thresholds for selecting active cells to be included in the computation, that is, cells having at least $N = 1$, $N = 2$, $N = 3$ events inside, although only the case with $N = 1$ is shown in the map (Fig. 5b). At this stage, all the cells having positive PI index are considered as possible locations for future large earthquakes; that is, no threshold is selected to identify the alarm area. The earthquakes that occurred in Iran during the forecast time span are listed in detail in Table 2, along with the outcomes from different tests. As an

example, Fig. 5 shows the corresponding hotspot maps for both T and C versions of PI and also RI map, referred to the forecast interval from January 1, 1998, to December 31, 2002.

3.4 Results for Italy

Similarly to Iran, a number of tests have been carried out for the Italian territory, using both original version (T) and modified version (C) of PI algorithm. In the application of C version, we set some thresholds and, in the calculations, we consider only the most active cells. Here, for fixed $dx = 0.2$ and $M_c = 3.0$, we apply these versions targeting sixteen 5-year forecast sliding time intervals (shifted by 1 year), beginning from 1996. The earthquakes occurred in Italy during these forecast windows are listed in Table 3, including the prediction results from different tests. As an example, we present here the results for the forecast time window from January 1, 2005, to December 31, 2009, which includes the M6.3 L'Aquila earthquake and its largest early aftershocks. Figure 6 shows the corresponding hotspot maps for both T and C versions of PI and also RI map.

4 Evaluation of forecasts

Evaluation of forecast results plays a key role in determining the quality and validity of any forecast method. The most well-known methods for evaluating binary forecasts, like PI, are Molchan diagram (Molchan 1997, 2003, 2010) and ROC diagram (Jolliffe and Stephenson 2003; Holliday et al. 2005).

For binary forecasts like PI, four quantities can be involved in the evaluation process. In Molchan error diagram, there are two ratios that can be used to represent a specific result (i.e., a point) in the diagram:

1. *Fraction of alarms* the number of alerted cells in the study area divided by the number of all cells in the area.
2. *Failure to predict* the number of events that are not predicted divided by the number of all occurred events.

In Molchan diagram, the diagonal line from the upper left corner to the lower right corner corresponds to the results that can be obtained by random guess. When the point takes further distance from the diagonal line toward the origin of the axes (i.e., for low fraction of both alarms and failures), the forecast shows better performance. Also, when the distance of the point from the axis of *failures to predict* is small, it indicates a high probability gain (maximum success with minimum alarm rate).

For the ROC diagram, we have two ratios as well:

1. *Hit rate (H)* the number of predicted events divided by the number of all occurred events.
2. *False alarm rate (F)* the number of alarmed cells where no target events did occur divided by the number of all alarmed cells in the study area.

The upper left corner of the ROC diagram represents a perfect forecast system (no false alarms, only hits). The closer any forecast result is to this upper left corner, the higher the skill. It should be noted that, differently from Molchan error diagrams, in ROC the time of confirmed alarms (i.e., successful hits) is not added to the time of false alarms, so that their sum would represent the rate of successful hits by chance.

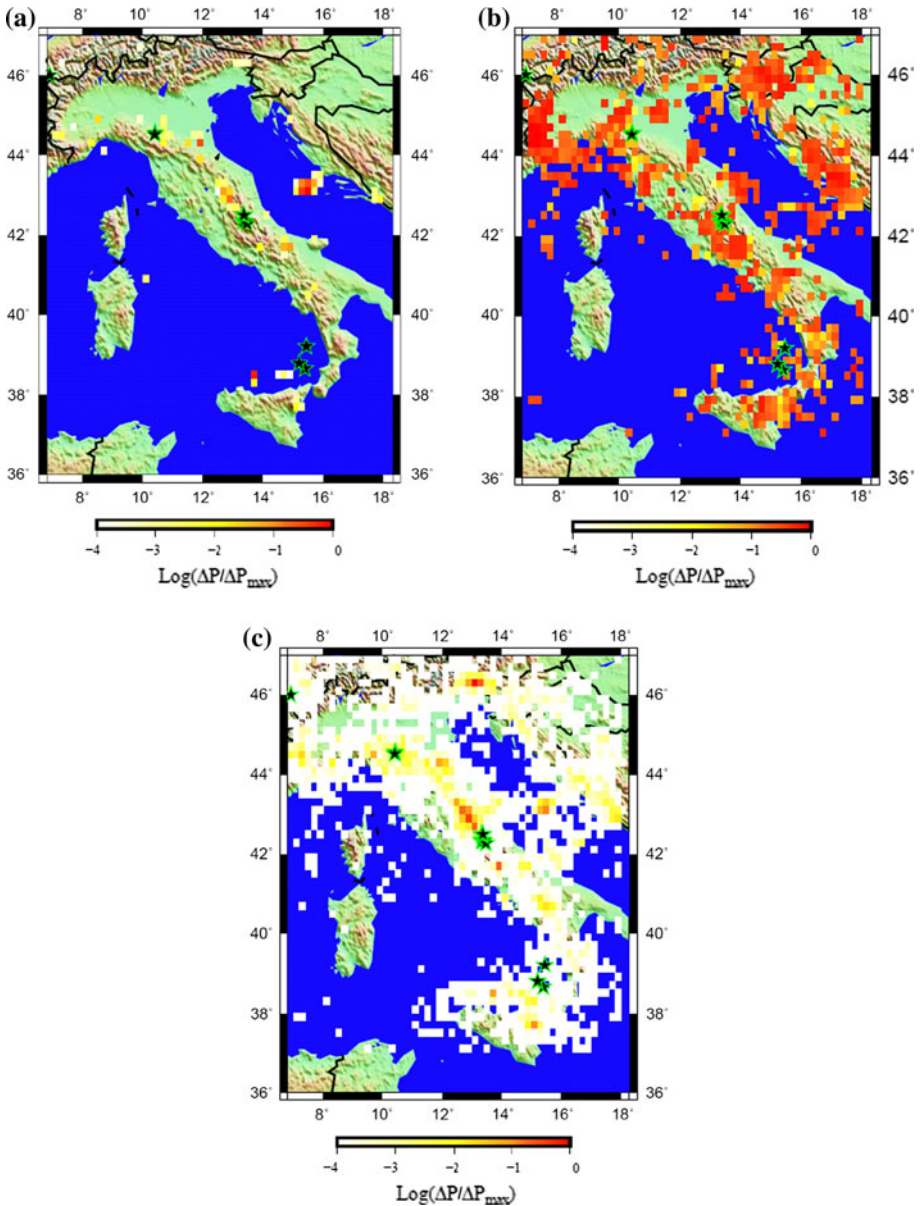


Fig. 6 Hotspot maps for $M \geq 5.0$ earthquakes within Italian territory, referred to the period from January 1, 2005, to December 31, 2009: **a** PI-T version. **b** PI-C version with $N \geq 1$. **c** RI. The black stars are the target events that occurred in the forecast time window

There are three strategies to define the mentioned ratios when constructing the error and ROC diagrams. In the first strategy, if a target earthquake occurs inside an alarmed cell, that cell is counted as a successful alarm. In the second strategy, proposed by Holliday et al. (2005), Moore neighbors are also involved when counting the successes or failures to predict. In this strategy, if an earthquake occurs inside an alarmed cell or even inside each

of its eight adjacent cells (i.e., its Moore neighbors, being also alarmed cells or not), that earthquake is counted as a successful hit. The second strategy has obviously a better performance than the first, if we account for the Moore neighbors only when estimating the hit rate, but not including them in the false alarm rate. But in fact, if we want to consider the Moore neighbors as possible locations of impending events (due to errors in location and some other uncertainties), we should count all actual (declared hotspots) and virtual (Moore neighbors) alarmed cells in the construction of the ratios for error diagrams. This is the third strategy. We denote the first strategy as O, the second as M, and the third as MM, respectively.

The above-mentioned measure of the alarm rate is clearly a simplistic one, being based only on the number of alarmed cells and not accounting for their actual seismic activity. Therefore, to assess the significance of a given method, the results should be compared to those provided by the RI method, which can be considered as a null hypothesis.

In order to draw the Molchan and ROC diagrams for both Iran and Italy, we run a set of sixteen applications of PI, each of them having a 5-year forecast time window (t_2 - t_3) and sliding forward every 1 year, from 1996 to 2011. According to this application scheme, forecasts are updated on a yearly basis and the score of each forecast map is examined against the earthquakes, which occur during the 1-year period (i.e., before next updating). Cumulative scores of these applications are used to construct the Molchan and ROC diagrams. Also, as a null hypothesis, we applied RI method with the same conditions as PI, both for Iran and for Italy, and added the results in the Molchan and ROC diagrams.

As mentioned above, the output of PI calculation is a map of hotspots that have positive probability. According to previous works, that is, Holliday et al. (2005), the declared hotspots can be decreased by introducing a threshold for hotspot activation. In this work, we apply ten thresholds to define hotspot activation level for both Iran and Italy. We use these thresholds to construct Molchan and ROC diagrams and to analyze the temporal variability of the fraction of alarmed area for Iran and Italy, as are shown in Figs. 7 and 8. The random guess outlines at 95 and/or 99 % confidence levels in Figs. 7d and 8d show explicitly which results are significantly different from random guess of alarm time, based on the alarm rate and on the corresponding rate of failures to predict.

With regard to these 16 years retrospective application, the corresponding hit rate and alarm rate for each strategy are summarized in Table 1. The details are listed in Tables 2 and 3 for Iran and Italy, respectively. It is inferred from the error diagrams, presented above, that none of the basic (T) and modified (C) versions of PI method can outperform the RI method in Iran and Italy. The alarmed area for RI map is too large, however, unless a threshold is introduced to select only the most active cells.

5 Discussion and conclusion

We applied the PI approach in retrospective tests for the identification of the areas of anomalous seismic activity, targeting past events in Iran and Italy. These regions were selected due to interest in how the PI method works in areas characterized by a complex seismotectonic setting and also on account of data availability. PI technique is not an earthquake prediction tool (in its specific definition of prediction, to tell exactly when and where an earthquake will occur), but it is a forecast that tells where the future large earthquakes are expected to occur during a relatively mid- or long-time window (~ 5 – 10 years). The objective of PI is to reduce the extent of alarmed areas relative to long-term earthquake hazard assessments (Holliday et al. 2005). At the stage of evaluating

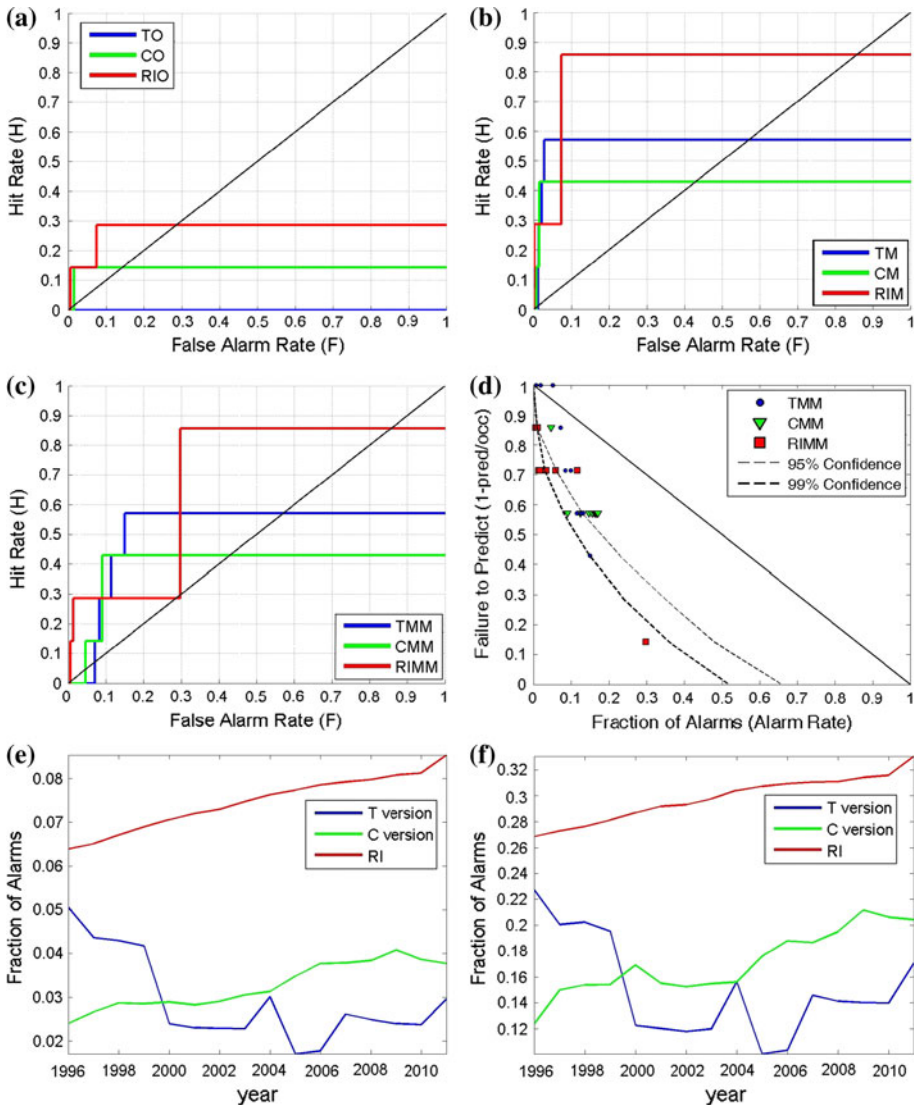


Fig. 7 a–c ROC diagrams for Iran for O, M, and MM strategies, respectively. **d** Molchan diagram for MM strategy. **e** Fraction of alarms versus time for O and M strategies. **f** Fraction of alarms versus time for MM strategy. In Molchan diagram, **d** the diagonal line corresponds to the results of a random guess and the 95 and 99 % confidence curves identify the results that are significantly different from a random identification of alarmed cells

the forecasts, according to both Molchan and ROC diagrams, it is obvious that smaller hotspots are preferable, since they may provide a high probability gain; however, larger fractions of alarms have naturally lesser failures to predict. Moreover, the size of the cells should account for the uncertainty in the location of earthquake epicenters (both for past and for future events), including the finite source dimensions for the largest events, which might exceed the cell size.

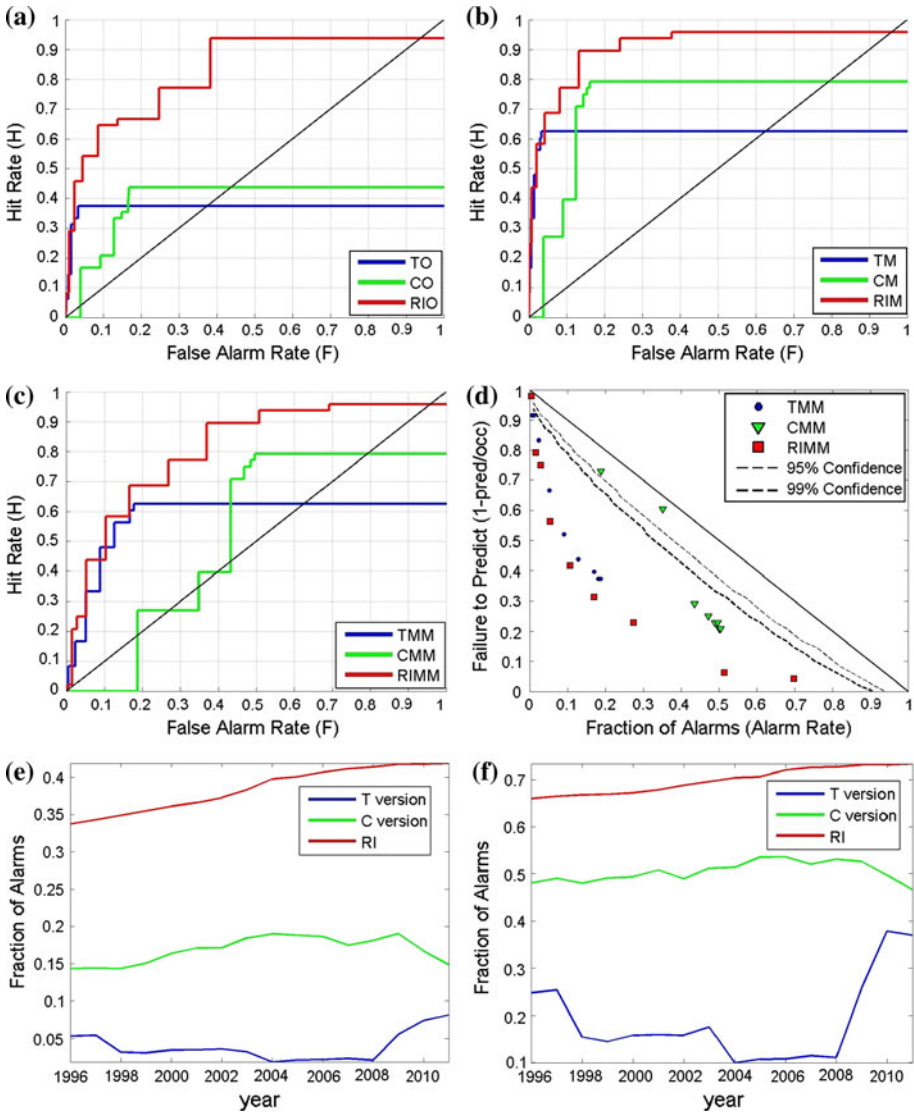


Fig. 8 a–c ROC diagrams for Italy for O, M, and MM strategies, respectively. **d** Molchan diagram for MM strategy. **e** Fraction of alarms versus time for O and M strategies. **f** Fraction of alarms versus time for MM strategy. In Molchan diagram, **d** the diagonal line corresponds to the results of a random guess and the 95 and 99 % confidence curves identify the results that are significantly different from a random identification of alarmed cells

Applications of the original and modified versions of this method in California, Canada, Japan, and Taiwan have shown some success in retrospective forecasting of the locations of past large earthquakes and also in real (prospective) forecasting. In this study, the problem of formally evaluating and comparing the performances of the different PI variants has been addressed considering different evaluation strategies, in order to identify the preferred application scheme and as well as to improve the performances of the method.

Table 1 The hit rate, alarm rate, miss rate, and space–time volume of PI application for different versions and evaluation strategies

Strategy	TO	TM	TMM	C1O	C1M	C1MM	C2O	C2M	C2MM	C3O	C3M	C3MM
<i>Region: Iran (total number of target events in 1996–2011 = 7)</i>												
Hit rate	0	0.571	0.571	0.142	0.428	0.428	0.142	0.285	0.285	0.142	0.285	0.285
Alarm rate	0.028	0.028	0.149	0.032	0.032	0.170	0.009	0.009	0.060	0.004	0.004	0.028
Miss rate	1	0.428	0.428	0.857	0.571	0.571	0.857	0.714	0.714	0.857	0.714	0.714
Space–time	0.492	0.492	0.738	0.436	0.436	0.649	0.249	0.249	0.439	0.151	0.151	0.299
Strategy	TO	TM	TMM	C1O	C1M	C1MM	C3O	C3M	C3MM	C5O	C5M	C5MM
<i>Region: Italy (total number of target events in 1996–2011 = 48)</i>												
Hit rate	0.325	0.625	0.625	0.437	0.791	0.791	0.312	0.604	0.604	0.375	0.604	0.604
Alarm rate	0.038	0.038	0.187	0.168	0.168	0.504	0.077	0.077	0.270	0.046	0.046	0.189
Miss rate	0.675	0.375	0.375	0.562	0.208	0.208	0.687	0.395	0.395	0.625	0.395	0.395
Space–time	0.282	0.282	0.619	0.405	0.405	0.545	0.322	0.322	0.654	0.265	0.265	0.545

Table 2 Target events ($M \geq 6.5$) occurred in Iran in the period 1996–2011, as reported in ISC catalog

Date	Time	Long.	Lat.	Mag.	Depth	Forecast status
2/4/1997	10:37 a.m.	57.2892	37.7385	6.6	35.7	TM
5/10/1997	7:57 a.m.	59.823	33.8784	7	6.7	Not forecasted!
3/14/1998	7:40 p.m.	57.6123	30.1606	6.7	43.5	TM, C1, C2, C3
12/6/2000	5:11 p.m.	54.811	39.493	7.4	33	TM, C1M, C2M, C3M
12/26/2003	1:56 a.m.	58.3044	28.9653	6.8	15	Not forecasted!
12/20/2010	6:41 p.m.	59.17	28.42	6.7	12	Not forecasted!
1/18/2011	8:23 p.m.	63.97	28.78	7.3	0	TM, C1M

The last column indicates the status of forecast with PI variants for the corresponding event: *T* Tiampo original alarmed region, *TM* Tiampo–Moore region, *C* Chen original alarmed region, *CM* Chen–Moore region. The number after *C* represents the threshold that defined the most active cells

The results shown in the error diagrams in Sect. 4 (Figs. 7, 8) indicate that, for each evaluating approach (i.e., O, M and MM), T version of PI outperforms the C version in Iran, while in Italy T version has much better gain, although C version has better forecast skills. However, both T and C versions do not provide better forecasts than the simple model RI, in agreement with results obtained by Zechar and Jordan (2008) for California. Also, it is inferred from the ROC diagrams that for M strategy, in accordance with Holliday et al. (2005), the T version has the best score in Iran, while in Italy the C version provides better performances. The fraction of alarms versus time clearly shows that the T version provides always a significantly lower rate of alarms in Italy (Fig. 8e, f), while in Iran the alarmed territory (Fig. 7e, f) is quite comparable with T and C versions.

The predictive capability of RI method appears quite satisfactory when turned into a binary forecast, as mentioned in Tiampo and Shcherbakov (2012); namely, a threshold value can be introduced, so that only the cells having RI above the specified value are expected to experience a large event, over the forecast period of interest. By considering, for example, a threshold that selects about 30 % of the monitored territory, RI predicts

Table 3 Target events ($M \geq 5.0$) occurred in Italy in the period 1996–2011, as reported in UCI catalog

Date	Time	Long.	Lat.	Mag.	Depth	Forecast status
2/20/1996	2:12 p.m.	17.23	44.51	5	10	T, C1, C3
9/5/1996	8:44 p.m.	17.94	42.8	6	10	T
9/9/1996	3:57 p.m.	17.87	42.77	5.2	10	TM
10/15/1996	9:55 a.m.	10.78	44.79	5.4	10	TM
9/26/1997	12:33 a.m.	12.88	43.05	5.7	10	T, C1, C3, C5
9/26/1997	9:40 a.m.	12.81	43.08	6.1	10	T, C1, C3, C5
10/3/1997	8:55 a.m.	12.79	43.08	5.3	10	T, C1, C3, C5M
10/6/1997	11:24 p.m.	12.84	43.04	5.5	10	T, C1, C3, C5
10/12/1997	11:08 a.m.	12.9	42.9	5.3	10	T, C1, C3, C5
10/14/1997	3:23 p.m.	12.89	42.96	5.6	10	T, C1, C3, C5
11/9/1997	7:07 p.m.	12.95	42.9	5	10	T, C1, C3, C5
3/13/1998	3:14 p.m.	14.2	45.59	5	10	C1M
3/26/1998	4:26 p.m.	12.97	43.26	5.4	10	TM, C1, C3, C5
4/3/1998	7:26 a.m.	12.7	43.16	5.2	10	T, C1M, C3, C5
4/5/1998	3:52 p.m.	12.72	43.19	5	10	T, C1M, C3, C5
4/12/1998	10:55 a.m.	13.65	46.24	5.9	10	C1, C3, C5M
5/18/1998	5:19 p.m.	15.11	39.25	5.4	279	C1M
9/9/1998	11:27 a.m.	15.98	40.03	5.5	10	C1, C3, C5M
12/29/1999	8:42 p.m.	10.31	46.6	5.2	10	C1, C3, C5M
5/10/2000	4:52 p.m.	12	44.32	5.1	10	T
7/17/2001	3:06 p.m.	11.2	46.74	5.6	10	Not forecasted!
11/26/2001	12:56 a.m.	12.07	43.6	5	10	TM, C1M, C3M
2/14/2002	3:18 a.m.	13.17	46.37	5.2	10	T, C1, C3, C5
9/6/2002	1:21 a.m.	13.7	38.38	5.9	5	C1M
10/31/2002	10:32 a.m.	14.87	41.79	5.7	10	C1M, C3M, C5M
11/1/2002	3:09 p.m.	14.88	41.73	5.7	10	C1M, C3M, C5M
3/29/2003	5:42 p.m.	15.46	43.11	5.4	10	C1
4/11/2003	9:26 a.m.	8.89	44.79	5.1	4	C1, C3, C5M
9/14/2003	9:42 p.m.	11.45	44.33	5.6	10	TM, C1M, C3M, C5M
5/5/2004	1:39 p.m.	14.81	38.51	5.5	228	C1, C3, C5
5/23/2004	3:19 p.m.	17.45	43.41	5.1	10	TM, C1, C3M, C5M
7/12/2004	1:04 p.m.	13.64	46.3	5.7	7	C1M, C3M, C5M
11/24/2004	10:59 p.m.	10.56	45.63	5.5	17	C3M
11/25/2004	6:21 a.m.	15.36	43.17	5.2	21	T, C3, C5
9/8/2005	11:27 a.m.	6.9	46	5.3	7	C1M, C3M, C5M
10/26/2006	2:28 p.m.	15.39	38.65	5.7	212	C1M, C3M, C5M
7/4/2007	11:55 p.m.	15.2	38.79	5.2	279	C1M, C3M, C5M
12/17/2008	9:57 p.m.	15.44	39.21	5.2	257	TM, C1, C3M, C5M
12/23/2008	3:24 p.m.	10.41	44.56	5.3	28	TM
12/23/2008	9:58 p.m.	10.39	44.52	5.1	31	C1M, C3M
4/6/2009	1:32 a.m.	13.33	42.33	6.3	8	TM, C1M, C3M, C5M
4/6/2009	2:37 a.m.	13.34	42.37	5.1	10	TM, C1M, C3M, C5M
4/7/2009	9:26 a.m.	13.39	42.34	5	10	TM, C1M, C3M, C5M

Table 3 continued

Date	Time	Long.	Lat.	Mag.	Depth	Forecast status
4/7/2009	5:47 p.m.	13.46	42.28	5.5	15	TM, C1M, C3M, C5M
4/9/2009	12:52 a.m.	13.34	42.48	5.3	15	T, C1, C3, C5
4/9/2009	7:38 p.m.	13.33	42.51	5	2	T, C1, C3, C5
4/13/2009	9:14 p.m.	13.36	42.5	5.1	7	T, C1, C3, C5
11/3/2010	6:13 p.m.	13.25	40.04	5.3	468	Not forecasted!

The last column indicates the status of forecast with PI variants for the corresponding event: *T* Tiampo original alarmed region, *TM* Tiampo–Moore region, *C* Chen original alarmed region, *CM* Chen–Moore region. The number after C represents the threshold that defined the most active cells

about 70 % of the earthquakes in Italy, whereas in Iran, RI predicts more than 80 % of the large earthquakes, with a threshold selecting 30 % most active cells.

Similarly, in order to reduce the rate of alarms, when considering C version of PI, it is possible to introduce a threshold to discard the less active cells, that is, the cells including a number of epicenters lower than a threshold *N*. A set of experiments has been performed considering the thresholds *N* = 1, 2, 3 for Iran and *N* = 1, 3, 5 for Italy. The ROC diagrams, comparing the performances of thresholds in C version for Iran and for Italy, are presented in Fig. 9. Accordingly, the alarm territory significantly reduces for the increasing thresholds; the performances improve in Italy, while the improvement is modest in Iran.

We evidence in this study that, when estimating the hit rates, alarm rates and failures to predict, the Moore neighbors should be treated consistently. In fact, according to Holliday et al. (2005) and Chen et al. (2005), when calculating the fractions of alarms and false alarms rates, only hotspots are counted as alarmed areas (*M* strategy), whereas Moore neighbors are also considered as possible locations for target events. This implies that earthquakes that occurred in a wider territory than the alerted one (i.e., outside the hotspots and inside Moore neighbors) are considered as successful predictions. Following Holliday et al. (2005) and Chen et al. (2005), supplementary authors have also applied *M* strategy. From an operational perspective, if we consider Moore neighbors as possible locations for a strong earthquake, the Moore neighbors should also be considered as alarmed cells and the fraction of alarms would increase significantly (*MM* strategy). According to this

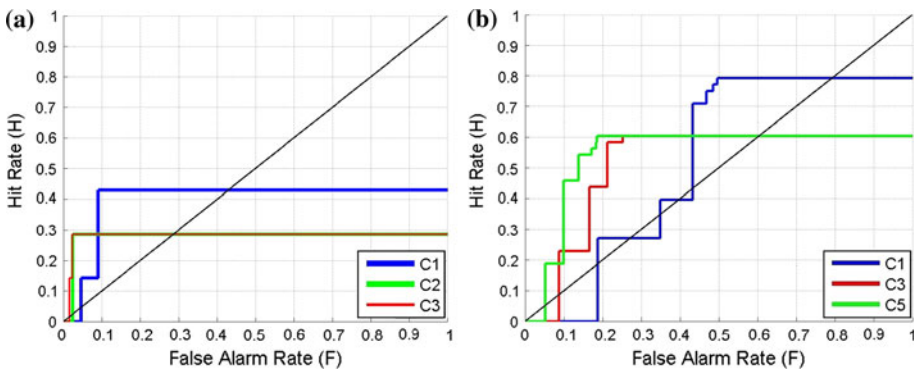


Fig. 9 ROC curves for C version of PI algorithm, obtained according to MM strategy, for: **a** Iran, with *N* = 1, 2, 3; **b** Italy, with *N* = 1, 3, 5

criterion, where both hotspots and their Moore neighbors are “alerted,” the hotspot maps, shown in Fig. 5 for Iran and in Fig. 6 for Italy, should be redrawn as shown in Figs. 10 and 11, respectively, assigning the same weight to all of the hotspots and also considering the Moore neighbors as alarmed cells.

The results from this study emphasize the need for an appropriate definition of the evaluation strategies, clearly and unambiguously indicating the area where a large earthquake has to be expected. An appropriate criterion to assess and compare the prediction results from different variants of the method is, in fact, a necessary prerequisite in order to improve the performances of PI or RI methods. The obtained results from PI application in Iran and in Italy indicate that, in their original form, both T and C versions of PI do not provide better information than RI maps, which do not account for temporal properties of seismicity evolution. The poor performances of PI become especially evident when considering a more appropriate measure of the space–time volume of alarms, which accounts for the seismicity rate associated with the alarmed areas (Fig. 12): The results are very close to the diagonal line corresponding to a random guess. These findings are in line with those obtained for Italy by Nanjo (2010), based on different input data. Some marginal

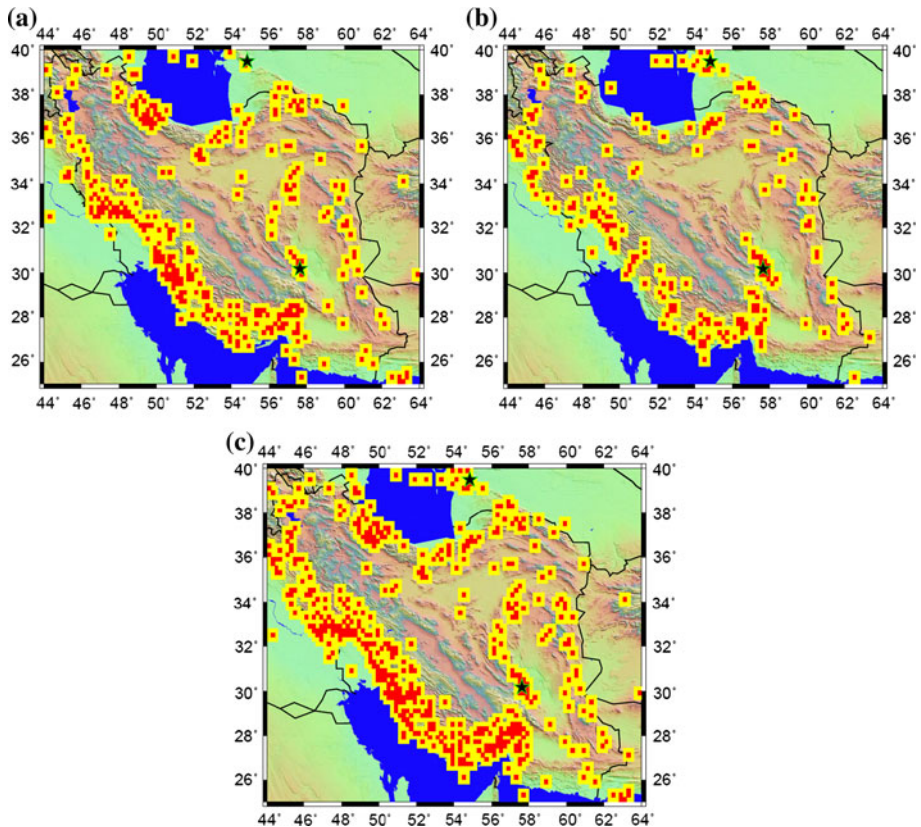


Fig. 10 Hotspot maps for $M \geq 6.5$ earthquakes within the territory of Iran, obtained according to MM strategy and referred to the period from January 1, 1998, to January 31, 2002, for: **a** PI-T version; **b** PI-C version with $N \geq 1$; **c** RI. All cells having positive PI or RI index are considered. The *black stars* are target events that occurred in the forecast time window

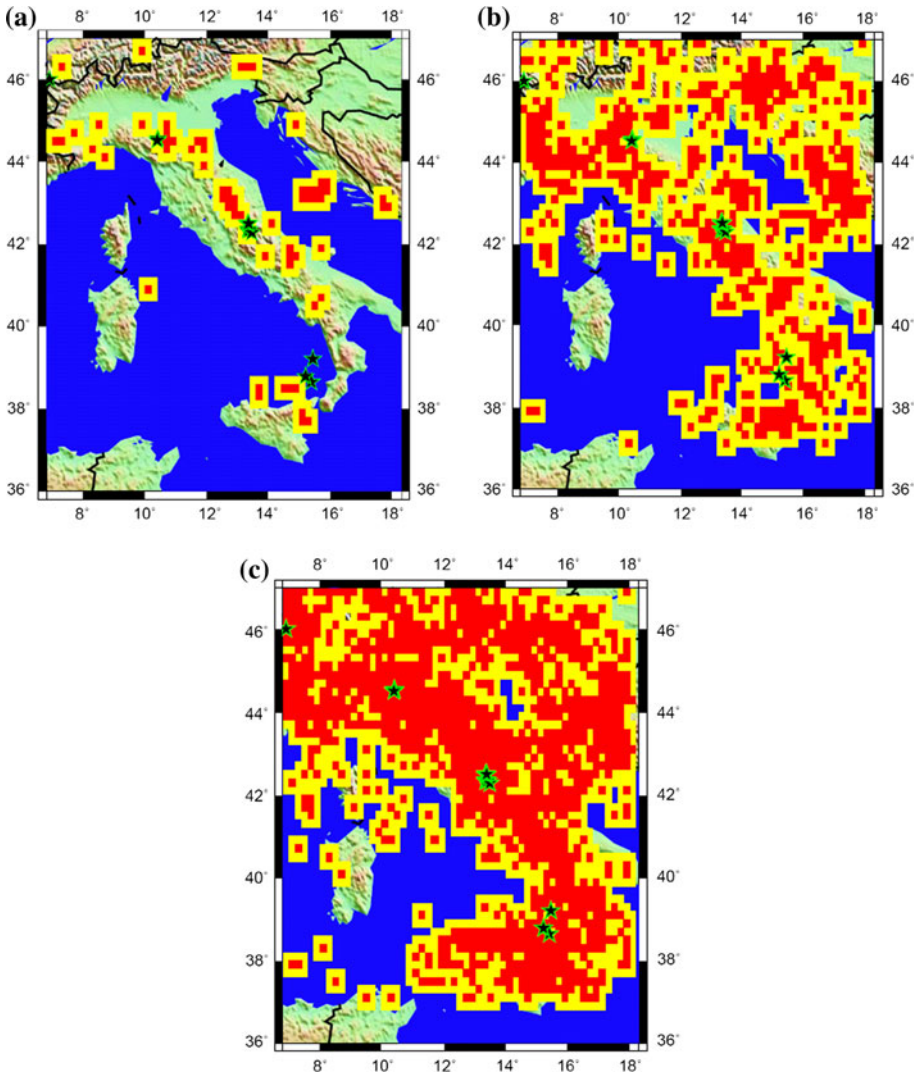


Fig. 11 Hotspot maps for $M \geq 5.0$ earthquakes within the territory of Italy, obtained according to MM strategy and referred to the period from January 1, 2005, to January 31, 2009, for: **a** PI-T version; **b** PI-C version with $N \geq 1$; **c** RI. All cells having positive PI or RI index are considered. The *black stars* are target events that occurred in the forecast time window

predictive capability for PI, both T and C versions, could be attributed only by the M evaluation strategy, which is evidently a biased strategy that maximizes the hit rate while minimizing the alarmed area.

Thus, the results from standard applications of PI method, although providing a seemingly high spatial accuracy, do not appear preferable compared with other methods based on formally defined patterns of seismicity, like CN or M8 (Peresan et al. 2005 and references therein). It is evidenced that some further considerations should be taken to possibly improve the performances of PI in Iran and Italy, eventually accounting for the

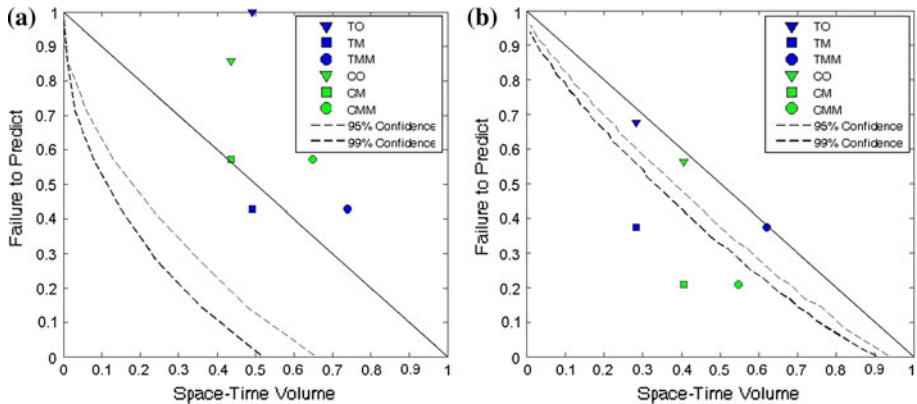


Fig. 12 Molchan diagrams for **a** Iran and **b** Italy, using the measure of alarms rate which accounts for the seismicity, as described in Sect. 4. Only results below the 95 or 99 % confidence curves are significantly different from a random guess

specific seismotectonic setting of the study region (e.g., Peresan et al. 1999) and taking into account possible long-range interactions. For instance, while PI identification of hotspots is mainly based on local variations of seismic activity, Latchman et al. (2008) suggested that it is possible to have elevated activity distant from the eventual location of a significant magnitude event. The comparative analysis of results from prospective testing of different methodologies will eventually provide new insights into the relevance of long-range correlations to earthquake forecasts.

The selection of the optimal thresholds, based on a trade-off between the rate of alarms and the rate of failures, is a decision-making problem that goes beyond the scope of this paper (see Davis 2012). However, based on the results described in previous sections and those shown in Fig. 9, we suggest using $N \geq 1$ and $N \geq 5$ for C application in Iran and Italy, respectively. Also, with the aim to test the RI method, we suggest alerting only 30 % of the cells in Iran and Italy, selecting those characterized by the largest values of the RI. To allow assessing the statistical significance of forward forecasts based on the mentioned application scheme of PI and RI methods, we compute the forecast maps starting from January 1, 2012 (Figs. 13, 14); these maps, according to the procedure proposed in this study, will be subject to update on January 2013.

An alternative way to evaluate the forecasting skills and predictive capability of the different methods is to consider a different measure of the space–time volume occupied by alarms, based on the seismicity rate associated with alarmed areas (Kossobokov et al. 1999). Accordingly, each cell is assigned a “weight” corresponding to the number of earthquakes occurred inside it. For this purpose, we consider a “sample catalog” representative of the seismic activity of the territory under study. At a given time, we define the spatial percentage of alarm as the ratio of the number of epicenters from the sample catalog, which fall inside the alarmed cells, to the total number of epicenters, which fall inside the area under investigation. The space–time volume of alarm is then computed as the average spatial percentage of alarm over the total period of forecasts.

In the case of Iran, we use, as a sample catalog, all earthquakes of magnitude 4.5 or more contained in the ISC catalog, for the period 1980–2000. In Italy, the sample catalog is composed by the earthquakes with magnitude larger or equal to 4.0 as reported in the UCI catalog, for the period 1950–2000. The corresponding error diagrams are shown in

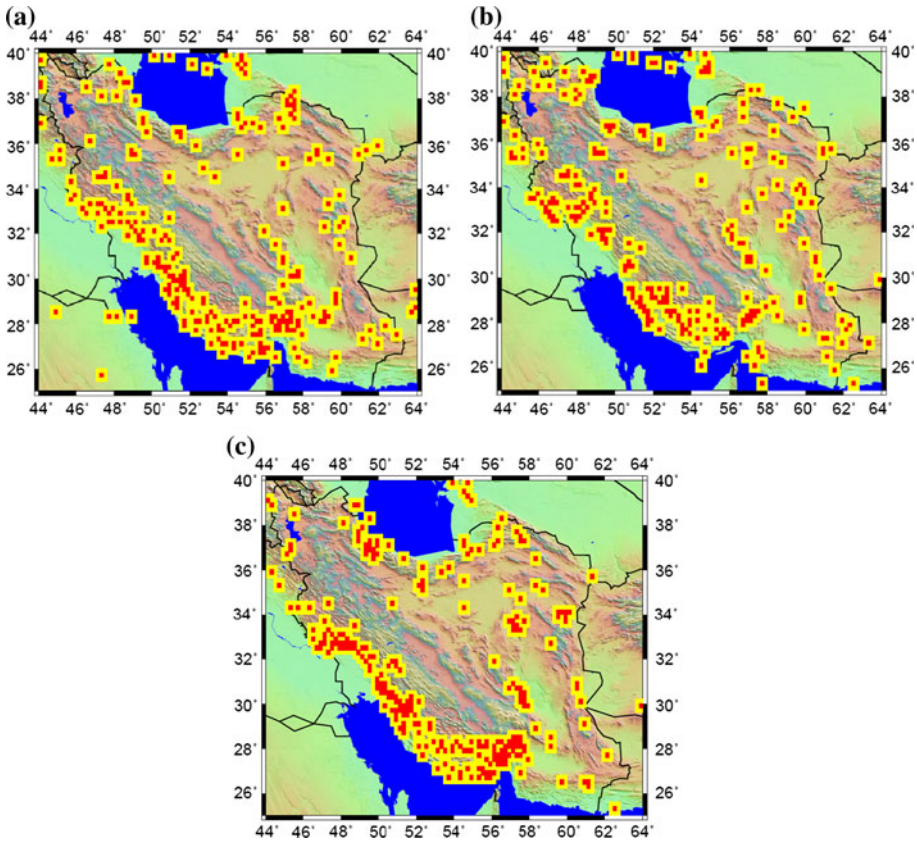


Fig. 13 Forward forecast hotspot maps for $M \geq 6.5$ earthquakes in Iran, based on MM strategy, as on January 2012: **a** PI-T version; **b** PI-C version, with $N \geq 1$; **c** RI, including only top 30 % of the cells characterized by the largest values of the RI. No target earthquakes occurred so far in the forecast time window. The map is subject to update on January 2013

Fig. 12a, b for Iran and Italy, respectively. The 95 and 99 % confidence curves show which results are significantly different from a random identification of alarmed cells, given the information on seismic density in space (Kossobokov 2006). All of the applications in Iran disclose a major failure, while in the case of Italy, the TM, CM, and CMM show some forecasting skills (with the M strategy providing biased estimates of the hit rate).

Since the beginning of the prospective testing for the proposed scheme of PI application, on January 2012, no $M \geq 6.5$ earthquakes occurred in Iran, whereas 8 events with $M \geq 5.0$ occurred in Italy. The forecast maps for Italy, valid for the year 2012, and the epicenters of the target earthquakes, including aftershocks, are shown in Fig. 14. The M6.1 Emilia earthquake that struck Northern Italy on May 20, 2012, as well as other target events, is located in the hotspots previously identified by PI method. It is possible to observe that although the earthquakes were correctly predicted in all of the three maps (i.e., two variants of PI application and RI map), the alarmed cells cover a significant part of the land, ranging from 11.3 % of the whole monitored territory for PI-T version to 22 % for PI-C version, and about 36.1 % for RI map. A detailed statistic of forecast results is provided in Table 4.

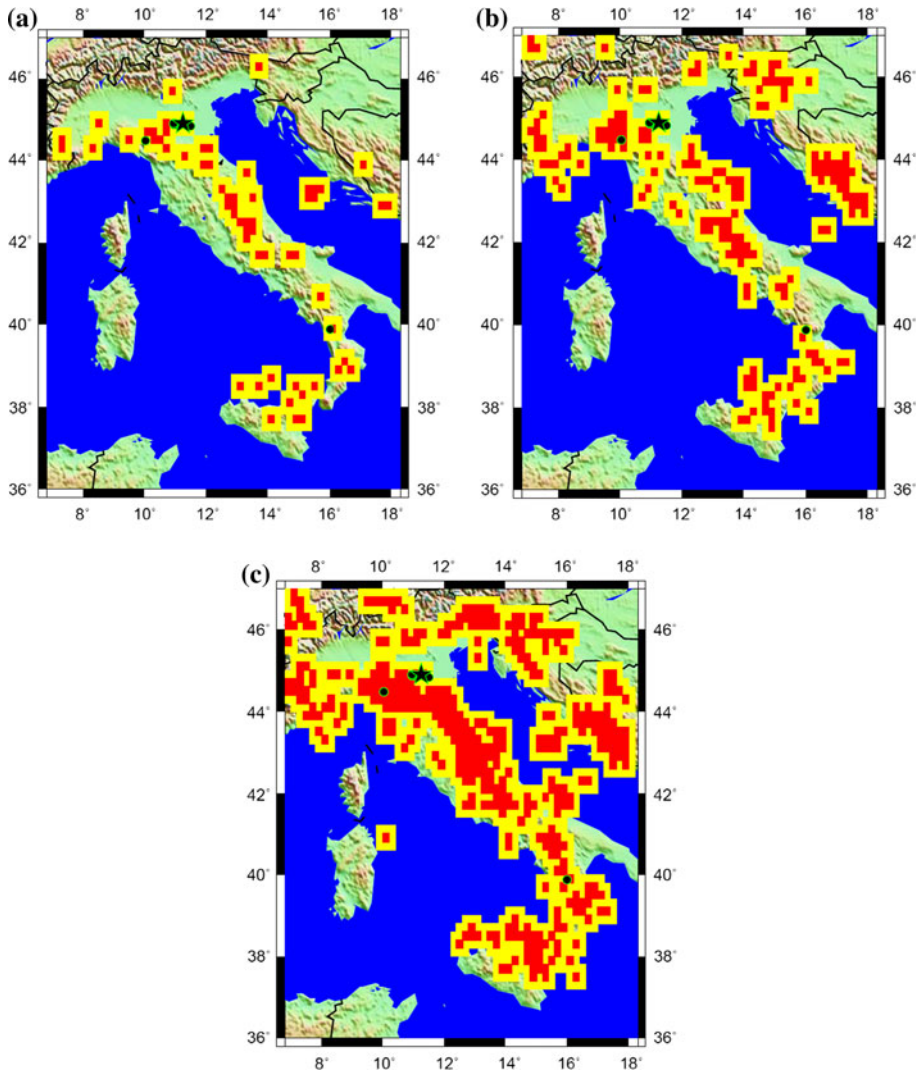


Fig. 14 Forward forecast hotspot maps for $M \geq 5$ earthquakes in Italy, based on MM strategy, as on January 2012: **a** PI-T version; **b** PI-C version, with $N \geq 5$; **c** RI, including only top 30 % of the cells characterized by the largest values of the RI. The *black circles* indicate the target events that occurred in the forecast time window; the $M_w = 6.1$ Emilia earthquake that occurred on May 20, 2012, is indicated by a *star*. The map is subject to update on January 2013

Although all of the earthquakes, occurred since the beginning of the experiment on January 2012, are correctly forecasted by these maps, large part of the predicted events occurred in Moore neighbors (yellow cells), hence restricting the alarmed territory to hotspots (red cells) would significantly reduce the predictive capability of the method. Accounting for Moore neighbors, on the other side, considerably increases the alarmed area, as shown in Table 4.

Table 4 Results from prospective testing of $M \geq 5.0$ earthquakes forecasts in the Italian region (period: January 1–December 1, 2012)

Experiment	NP	NPH	RA (%)	H/A (%)	RL (%)
PI-T	8	2	11.3	17.2	26
PI-C	8	4	22.0	26.7	50.7
RI-30 %	8	5	36.1	36.6	83.2

The number of target events within the monitored region is $N = 8$. *NP* Number of predicted events, *NPH* number of predicted events in hotspots, *RA* rate of alarmed cells versus all cells in the monitored region, *H/A* A rate of hotspot cells versus all alarmed cells, *RL* rate of alarmed cells versus all cells, considering only inland monitored territory

Acknowledgments We are thankful to the International Institute of Earthquake Engineering and Seismology (IIEES) and the Abdus Salam International Center for Theoretical Physics (ICTP) for providing the facilities for this study. We are indebted with G. F. Panza for his valuable comments and discussions during the preparation of the manuscript. We appreciate K. F. Tiampo and C. C. Chen for sharing some codes and comments. We are also grateful to Z. Wu, V. Kossobokov and an anonymous reviewer for their constructive comments. A. Peresan benefited by financial support from INGV-DPC Project S3 (2012–2013). Several images were plotted with the help of GMT software developed by Paul Wessel and Walter H. F. Smith.

References

- Basili R, Valensise G, Vannoli P, Burrato P, Fracassi U, Mariano S, Tiberti MM, Boschi E (2008) The database of individual seismogenic sources (DISS), version 3: summarizing 20 years of research on Italy's earthquake geology. *Tectonophysics*. doi:[10.1016/j.tecto.2007.04.014](https://doi.org/10.1016/j.tecto.2007.04.014)
- Brandmayr E, Raykova R, Zuri M, Romanelli F, Doglioni C, Panza GF (2010) The lithosphere in Italy: structure and seismicity. *J Virtual Explor, Electronic Edition*, ISSN 1441-8142, v. 36, paper 1
- Chen C-C, Rundle JB, Holliday JR, Nanjo KZ, Turcotte DL, Li S-C, Tiampo KF (2005) The 1999 Chi–Chi, Taiwan, earthquake as a typical example of seismic activation and quiescence. *Geophys Res Lett* 32:L22315. doi:[10.1029/2005GL023991](https://doi.org/10.1029/2005GL023991)
- Cho NF, Tiampo KF (2012) Effects of location errors in pattern informatics. *Pure Appl Geophys*. doi:[10.1007/s00024-011-0448-2](https://doi.org/10.1007/s00024-011-0448-2)
- Cho NF, Tiampo KF, McKinnon SD, Vallejos JA, Klein W, Dominguez R (2010) A simple metric to quantify seismicity clustering. *Nonlinear Proc Geophys* 17:293–302
- Cloetingh S, Tornu T, Ziedler PA, Beekman F (2006) Neotectonics and intraplate topography of the northern Alpine Foreland. *Earth Sci Rev* 74:127–196
- Cuffaro M, Riguzzi F, Scrocca D, Antonioli F, Carminati E, Livani M, Doglioni C (2010) On the geodynamics of the northern Adriatic plate. *Rend Fis Acc Lincei*. doi:[10.1007/s12210-010-0098-9](https://doi.org/10.1007/s12210-010-0098-9)
- Cuffaro M, Riguzzi F, Scrocca D, Doglioni C (2011) Coexisting tectonic settings: the example of the southern Tyrrhenian Sea. *Int J Earth Sci (Geol Rundsch)*. doi:[10.1007/s00531-010-0625-z](https://doi.org/10.1007/s00531-010-0625-z)
- D'Agostino N, Cheloni D, Mantenuto S, Selvaggi G, Michelini A, Zuliani D (2005) Strain accumulation in the southern Alps (NE Italy) and deformation at the northeastern boundary of Adria observed by CGPS measurements. *Geophys Res Lett* 32:L19306. doi:[10.1029/2005GL024266](https://doi.org/10.1029/2005GL024266)
- Davis CA (2012) Loss functions for temporal and spatial optimizing of earthquake prediction and disaster preparedness. *Pure Appl Geophys* 169:1989–2010. doi:[10.1007/s00024-012-0502-8](https://doi.org/10.1007/s00024-012-0502-8)
- Fukunaga K (1970) Introduction to statistical pattern recognition. Academic Press, New York
- Guidarelli M, Panza GF (2006) INPAR, CMT and RCMT seismic moment solutions compared for the strongest damaging events ($M_{4.8}$) occurred in the Italian region in the last decade. *Rend Acc Naz delle Scienze detta dei XL Mem Di Scienze Fisiche e Naturali*, 30: 81–98
- Hardebeck JL, Felzer KR, Michael AJ (2008) Improved tests reveal that the accelerating moment release hypothesis is statistically insignificant. *J Geophys Res* 113:B08310. doi:[10.1029/2007JB005410](https://doi.org/10.1029/2007JB005410)
- Holliday JR, Nanjo KZ, Tiampo KF, Rundle JB, Turcotte DL (2005) Earthquake forecasting and its verification. *Nonlinear Proc Geophys*. 12:965–977
- Holliday JR, Rundle JB, Tiampo KF, Klein W, Donnellan A (2006) Systematic procedural and sensitivity analysis of the pattern informatics method for forecasting large ($M > 5$) earthquake events in southern California. *Pure Appl Geophys* 163:2433–2454

- Holliday JR, Chen C-C, Tiampo KF, Rundle JB, Turcotte DL, Donnellan A (2007) A RELM earthquake forecast based on pattern informatics. *Seismol Res Lett* 78(1):87–93
- Holmes P, Lumley JL, Berkooz G (1996) Turbulence, coherent structures, dynamical systems and symmetry. Cambridge University Press, Cambridge, UK
- ISC Catalogue (2012) <http://www.isc.ac.uk>, Int Seis Cent, Thatcham, UK
- Jiang CS, Wu ZL (2011) PI forecast with or without de-clustering: an experiment for the Sichuan-Yunnan region. *Nat Hazards Earth Syst Sci* 11:697–706. doi:10.5194/nhess-11-697-2011
- Jiang CS, Wu ZL, Forecast PI (2010) PI forecast for the Sichuan-Yunnan region: retrospective test after the May 12, 2008, Wenchuan earthquake. *PAGEOPH* 167(6/7):751–761. doi:10.1007/s00024-010-0070-8
- Joliffe LT, Stephenson DB (eds) (2003) Forecast verification: a practitioner's guide in atmospheric science. Wiley, Hoboken
- Jordan T, Chen Y, Gasparini P, Madariaga R, Main I, Marzocchi W, Papadopoulos G, Sobolev G, Yamaoka K, Zschau, J. (2011). Operational earthquake forecasting: state of knowledge and guidelines for utilization. *Ann Geophys* 54(4). doi:10.4401/ag-5350
- Klein W, Rundle JB, Ferguson CD (1997) Scaling and nucleation in models of earthquake faults. *Phys Rev Lett* 78:3793–3796
- Kossobokov VG (2006) Testing earthquake prediction methods: «The West Pacific short-term forecast of earthquakes with magnitude $M_{wHRV} = 5.8$ ». *Tectonophysics* 413:25–31
- Kossobokov VG (2012) Earthquake prediction: 20 years of global experiment. *Nat Hazards*. doi:10.1007/s11069-012-0198-1
- Kossobokov VG, Romashkova LL, Keilis-Borok VI, Healy JH (1999) Testing earthquake prediction algorithms: statistically significant advance prediction of the largest earthquakes in the Circum-Pacific, 1992–1997. *Phys Earth Planet Inter* 111:187–196
- Latchman JL, Morgan FDO, Aspinall WP (2008) Temporal changes in the cumulative piecewise gradient of a variant of the Gutenberg-Richter relationship, and the imminence of extreme events. *Earth Sci Rev* 87:94–112
- Li H-C, Chen C-C (2011) Characteristics of long-term regional seismicity before the 2008 Wen-Chuan, China, earthquake using pattern informatics and genetic algorithms. *Nat Hazards Earth Syst Sci* 11:1003–1009. doi:10.5194/nhess-11-1003-2011
- Meletti C, Patacca E, Scandone P (2000) Construction of a seismotectonic model: the case of Italy. *Pure Appl Geophys* 157:11–35
- Meletti C, Galadini F, Valensise G, Stucchi M, Basili R, Barba S, Vannucci G, Boschi E (2008) A seismic source zone model for the seismic hazard assessment of the Italian territory. *Tectonophysics* 450:85–108. doi:10.1016/j.tecto.2008.01.003
- Molchan GM (1997) Earthquake prediction as a decision-making problem. *Pure Appl Geophys* 149: 233–247
- Molchan GM (2003) Earthquake prediction strategies: a theoretical analysis. In: Keilis-Borok VI, Soloviev AA (eds) *Nonlinear dynamics of the lithosphere and earthquake prediction*. Springer, Heidelberg, pp 209–237
- Molchan GM (2010) Space-time earthquake prediction: the error diagrams. *Pure Appl Geophys* 167:8–9. doi:10.1007/s00024-010-0087-z
- Mori H, Kuramoto Y (1997) *Dissipative structures and chaos*. Springer, Berlin
- Mountain RM, Thirumalai D (1992) Ergodicity and loss of dynamics in supercooled liquids. *Phys Rev A* 45:3380–3383
- Nabavi MH (1976) An introduction to the geology of Iran. *Geol Surv Iran (in Farsi)*, 110 pp
- Nanjo KZ (2010) Earthquake forecast models for Italy based on the RI algorithm. *Ann Geophys*. 53(3). doi: 10.4401/ag-4810
- Nanjo K, Holliday J, Chen C-C, Rundle JB, Turcotte DL (2006) Application of a modified pattern informatics method to forecasting the locations of future large earthquakes in the central Japan. *Tectonophysics* 424(3):351–366
- Peresan A, Panza GF (2002) UCI2001: The updated catalogue of Italy, The Abdus Salam International Centre for Theoretical Physics. ICTP, Miramare, Trieste, Italy, Internal report IC/IR/2002/3
- Peresan A, Costa G, Panza GF (1999) Seismotectonic model and CN earthquake prediction in Italy. *Pure and Appl. Geophys* 154:281–306
- Peresan A, Kossobokov V, Romashkova L, Panza GF (2005) Intermediate-term middle-range earthquake predictions in Italy: a review. *Earth-Sci Rev* 69(1–2):97–132
- Romashkova L, Peresan A (2013) Analysis of Italian earthquake catalogs in the context of intermediate-term prediction problem. *Acta Geophys* 61:583–610
- Rundle JB, Klein W, Tiampo K, Gross S (2000a) Linear pattern dynamics in nonlinear threshold systems. *Phys Rev E* 61:2418–2432

- Rundle JB, Klein W, Gross S, Tiampo KF (2000b) Dynamics of seismicity patterns in systems of earthquake faults. In: Rundle JB, Turcotte DL, Klein W Geocomplexity and the physics of earthquakes. Geophysics Monograph Series, vol. 120, AGU, Washington, DC, pp 127–146
- Rundle JB, Tiampo KF, Klein W, Martins JSS (2002) Self-organization in leaky threshold systems: the influence of near mean field dynamics and its implications for earthquakes, neurobiology, and forecasting. *Proc Natl Acad Sci USA* 99(Suppl 1):2514–2521
- Rundle JB, Turcotte DL, Shcherbakov R, Klein W, Sammis C (2003) Statistical physics approach to understanding the multi scale dynamics of earthquake fault systems. *Rev Geophys* 41(4):1019. doi: [10.1029/2003RG000135](https://doi.org/10.1029/2003RG000135)
- Shcherbakov R, Turcotte DL, Rundle JB, Tiampo KF, Holliday JR (2010) Forecasting the locations of future large earthquakes: an analysis and verification. *Pure Appl Geophys* 167:743–749. doi:[10.1007/s00024-010-0069-1](https://doi.org/10.1007/s00024-010-0069-1)
- Splendore R, Marotta AM, Barzaghi R, Borghi A, Cannizzaro L (2010) Block model versus thermo mechanical model: new insights on the present-day regional deformation in the surroundings of the Calabrian Arc. *Geol Soc Lond Special Publ* 332:129–147. doi:[10.1144/SP332.9](https://doi.org/10.1144/SP332.9)
- Thirumalai D, Mountain RD (1993) Activated dynamics, loss of ergodicity, and transport in supercooled liquids. *Phys Rev E* 47:479–489
- Thirumalai D, Mountain RD, Kirkpatrick TR (1989) Ergodic behavior in supercooled liquids and in glasses. *Phys Rev A* 39:3563–3574
- Tiampo KF, Shcherbakov R (2012) Seismicity-based earthquake forecasting techniques: ten years of progress. *Tectonophysics* 522(523):89–121. doi:[10.1016/j.tecto.2011.08.019](https://doi.org/10.1016/j.tecto.2011.08.019)
- Tiampo KF, Rundle JB, McGinnis S, Klein W (2002a) Pattern dynamics and forecast methods in seismically active regions. *Pure Appl Geophys* 159:2429–2467
- Tiampo KF, Rundle JB, McGinnis S, Gross SJ, Klein W (2002b) Mean field threshold systems and phase dynamics: an application to earthquake fault systems. *Europhys Lett* 60:481–487
- Tiampo KF, Rundle JB, Klein W, Sá Martins JS, Ferguson CD (2003) Ergodic dynamics in a natural threshold system. *Phys Rev* 91(23):238501
- Tiampo KF, Rundle JB, Klein W, Holliday JR, Sá Martins JS, Ferguson CD (2007) Ergodicity in natural earthquake fault networks. *Phys Rev E* 75:066107
- Tiampo KF, Klein W, Li H-C, Mignan A, Toya Y, Kohen-Kadosh SZL, Rundle JB, Chen C-C (2010) Ergodicity and earthquake catalogs: forecast testing and resulting implications. *Pure Appl Geophys*. doi:[10.1007/s00024-010-0076-2](https://doi.org/10.1007/s00024-010-0076-2)
- Toya Y, Tiampo KF, Rundle JB, Chen C-C, Li H-C, Klein W (2009) Pattern informatics approach to earthquake forecasting in 3D. *Concurr Comput Pract Exp*. doi:[10.1002/cpe.1531](https://doi.org/10.1002/cpe.1531)
- Van Stiphout T, Schorlemmer D, Wiemer S (2011) The effect of uncertainties on estimates of background seismicity rate. *Bull Seismol Soc Am* 101(2):482–494. doi:[10.1785/0120090143](https://doi.org/10.1785/0120090143)
- Vannucci G, Gasperini P (2004) The new release of the database of earthquake mechanisms of the Mediterranean area (EMMA version 2). *Ann Geophys* 47(suppl 1):307–334
- Wiemer S (2001) A software package to analyze seismicity: ZMAP. *Seismol Res Lett* 72:373–382
- Wu Y-H, Chen C-C, Rundle JB (2008a) Precursory seismic activation of the Pingtung (Taiwan) offshore doublet earthquakes on December 26, 2006: a pattern informatics analysis. *Terr Atmos Ocean Sci* 19(6):743–749
- Wu Y-H, Chen C-C, Rundle JB (2008b) Detecting precursory earthquake migration patterns using the pattern informatics method. *Geophys Res Lett* 35:L19304. doi:[10.1029/2008GL035215](https://doi.org/10.1029/2008GL035215)
- Zechar JD, Jordan TH (2008) Testing alarm-based earthquake predictions. *Geophys J Int* 172(2):715–724. doi:[10.1111/j.1365-246X.2007.03676.x](https://doi.org/10.1111/j.1365-246X.2007.03676.x)
- Zhang Y, Zhang X, Wu Y, Yin X (2012) Retrospective study on the predictability of pattern informatics to the Wenchuan M8.0 and Yutian M7.3 earthquakes. *Pure Appl Geophys (Online First)*. doi:[10.1007/s00024-011-0444-6](https://doi.org/10.1007/s00024-011-0444-6)