# Earthquake forecast models for Italy based on the RI algorithm

Kazuyoshi Z. Nanjo[1]

[1] *Earthquake Research Institute, University of Tokyo, Tokyo, Japan*

## ABSTRACT

*This study provides an overview of relative-intensity (RI)-based earthquake forecast models that have been submitted for the 5-year and 10-year testing classes and the 3-month class of the Italian experiment within the Collaboratory for the Study of Earthquake Predictability (CSEP). The RI algorithm starts as a binary forecast system based on the working assumption that future large earthquakes are considered likely to occur at sites of higher seismic activity in the past. The measure of RI is the simply counting of the number of past earthquakes, which is known as the RI of seismicity. To improve the RI forecast performance, we first expand the RI algorithm to become part of a general class of smoothed seismicity models. We then convert the RI representation from a binary system into a testable CSEP model that forecasts the numbers of earthquakes for the predefined magnitudes. Our parameter tuning for the CSEP models is based on the past seismicity. The final submission is a set of two numerical data files that were created by tuned 5-year and 10-year models and an executable computer code of a tuned 3-month model, to examine which testing class is more meaningful in terms of the RI hypothesis. The main purpose of our participation is to better understand the importance (or lack of importance) of relative intensity of seismicity for earthquake forecastability.*

## 1. Introduction

The crust of the Earth is clearly extremely complex, and it is generally accepted that seismicity is a chaotic phenomenon [Turcotte 1997]. Thus, as in the case of weather forecasting, earthquake forecasting must be considered from a statistical stand point [Rundle et al. 2003]. The statistical properties of seismicity patterns can be used to forecast future earthquakes. Basic types of statistical seismicity precursors include foreshock, quiescence, swarms, activation and doughnuts [see, for example, Mogi 1985, Turcotte 1991, Scholz 2002, Kanamori 2003]. Several groups have systematically developed algorithms to search for premonitory seismicity patterns. A Russian group has studied premonitory seismic activation for some strong earthquakes that occurred in California and Nevada, using the «CN» algorithm, and for a magnitude $M > 8$ worldwide, using the «M8» algorithm [see, for example, Keilis-Borok 1990, Keilis-Borok and Rotwain 1990, Keilis-Borok and Kossobokov 1990, Keilis-Borok and Soleviev 2003]. Another group showed premonitory seismic quiescence for Armenian and Landers earthquakes using the parameter known as the Z-value [see, for example, Wyss 1997, Wyss and Martirosyan 1998, Wyss and Wiemer 2000]. A new approach to earthquake forecasting was proposed by Rundle and coworkers, known as the Pattern Informatics (PI) approach [Rundle et al. 2002, Rundle et al. 2003, Tiampo et al. 2002, Holliday et al. 2005, Holliday et al. 2007, Nanjo et al. 2006a, Nanjo et al. 2006b, Nanjo et al. 2006c]. This approach is based on the strong space-time correlations of seismicity, using the ideas of driven non-linear threshold dynamics, as well as of mean-field long-range theory. This PI technique has thus been used to detect precursory seismic activation and quiescence, and to provide earthquake forecasts for California and Japan, and also on a worldwide basis.

An alternative approach to earthquake forecasting is much simpler than the algorithms indicated above. This is the use of the Relative Intensity (RI) of past seismicity, which is based on counting the number of earthquakes that occurred in the past. RI suggests that future earthquakes are most likely to occur where historical seismicity has been relatively high [Tiampo et al. 2002]. RI belongs to a general class of smoothed seismicity models, which also includes proximity to past earthquakes [Rhoades and Evison 2004], cellular seismology [Kafka 2002], and other members [e.g., Kagan and Jackson 2000, Kossobokov 2004, Helmstetter et al. 2007] that essentially offer slightly different representations of the same basic hypothesis. Of note, seismicity changes are not taken into account with RI forecast generation. Although the RI idea is very simple, based on previous studies, its forecast performance is considerable [Rundle et al. 2002, Rundle et al. 2003, Tiampo et al. 2002, Holliday et al. 2005, Holliday et al. 2006a, Holliday et al. 2006b, Nanjo et al. 2006a, Nanjo et al. 2006b, Nanjo et al. 2006c, Zechar and Jordan 2008]. Our challenge is to move towards a systematic prospective for the testing of the RI hypothesis under a well-controlled environment.

Associated with resurgence in earthquake predictability research, the working group on Regional Earthquake Likelihood Models (RELM) has established a facility for prospective testing of scientific earthquake predictions in California [Field 2007]. One of the successors of RELM is an international partnership to develop a Collaboratory for the Study of Earthquake Predictability (CSEP), which is designed to support a global program of research on earthquake predictability [Jordan 2006]. Prospective testing experiments under the CSEP have been carried out for California and other regions (http://www.cseptesting.org/). One of the newly involved regions is Italy. The Italian territory is characterized by a relatively high seismic risk [Marzocchi 2008], with the strongest earthquakes ranging from medium to relatively large magnitudes (up to about $M$ 7). Using these circumstances, the CSEP Italian prospective testing experiment that is hosted by the European Testing Center was officially started on August 1, 2009. In this study, we present an overview of the RI models submitted for this experiment. Details of the "rules of the game" for this experiment are given on the website at: http://eu.cseptesting.org/. Three forecast models for the 5-year and 10-year classes and the 3-month class have been submitted to the testing center. Our aim here is to examine which of these classes is more meaningful for the RI hypothesis.

## 2. Evidence for using RI in forecasting seismicity

As described in the previous section, RI is a very simple algorithm. However, this provides us with a relatively strong approach to earthquake forecasting. Evidence in support of RI is based on three main lines, as summarized below:

*1. RI-based forecasting is better than random forecasting.* Together with the PI method, the RI method has been applied for prospective forecasting of a variety of tectonic regimes, including California and Japan, as well as on a worldwide basis [see, for example, Rundle et al. 2002, Rundle et al. 2003, Tiampo et al. 2002, Holliday et al. 2005, Nanjo et al. 2006a, Nanjo et al. 2006b, Nanjo et al. 2006c]. The output is a map of areas in a seismogenic region where earthquakes are forecast to occur over a future 10-year time span. Using Relative Operating Characteristics (ROC) diagrams [Mason 2003] and log-likelihood tests, both methods outperform random guessing (i.e., random forecast maps) under most circumstances.

*2. RI forecasting is associated with occurrence of large earthquakes.* Holliday et al. [2006a, 2006b] showed that PI and RI have comparable accuracies for spatial forecasting in retrospective testing of California data using ROC diagrams [Mason 2003]. In addition, they examined the relative forecast accuracies of PI and RI as functions of time, where they showed that the time-dependence is highly correlated with

the occurrence of large earthquakes: $M \geq 6$ earthquakes since 1960 have occurred during intervals of time where RI outperforms PI. These authors also found that their approach is applicable to Sumatra, to show that $M \geq 8$ earthquakes since 1980 have occurred over time intervals when RI provides better performance relative to PI.

*3. Significant gains of RI relative to the U.S. Geological Survey National Seismic Hazard Map.* Zechar and Jordan [2008] presented a method for testing alarm-based earthquake predictions, on the basis of Molchan diagrams [Molchan 1990]. These diagrams provide the plot of miss rate $\nu$ as a function of fraction of space-time occupied by alarm $\tau$. To illustrate the method, they considered a 10-year experiment by Rundle et al. [2002, 2003] to predict $M \geq 5$ earthquakes in California, and they tested forecasts from three models: RI, PI, and National Seismic Hazard Map (NSHM; a model that uses seismicity smoothed over regions of as much as 60 km, zones of background seismicity, and explicit fault information). These authors used the RI alarm function as the prior distribution, and they computed the Molchan trajectories for the PI and NSHM forecasts (Figure 1a). The Molchan trajectories for PI and NSHM are close to the descending diagonal. This indicates much smaller probability gains for PI and NSHM using the RI reference. Based on these results, the authors stated that neither PI nor NSHM provide significant performance gains relative to the RI reference model. Their Molchan diagram is a one-sided test with only one forecast model considered for setting up the null hypothesis: only that taking RI as a reference. We used the data and program code in their online article, and took NSHM as the reference. Figure 1b shows that PI and RI are far from the descending diagonal for $\nu \sim 0.0$-$0.5$ at $>95\%$ confidence (or significance $\alpha < 5\%$), indicating significant gains relative to NSHM at low miss rates $\nu$. These Molchan analyses indicated that both PI and RI can deliver performances that are statistically superior to NSHM. Thus, both PI and RI can be significantly better at forecasting future earthquakes than NSHM at the 95% confidence level, even though RI is the simplest algorithm among the three. We also took PI as the reference and found that neither RI nor NSHM provide significant performance gains relative to the PI reference model.

The challenge is to move from specific case studies to a systematic prospective testing under the well-controlled Italian CSEP experiment.

## 3. RI and its modification

Here, we first provide the basics of the RI algorithm, and then a possible modification that allows RI to belong to a general class of smoothed seismicity models. Note that this modification does not violate the RI hypothesis. Instead, our idea is to examine whether inclusion of smoothing into
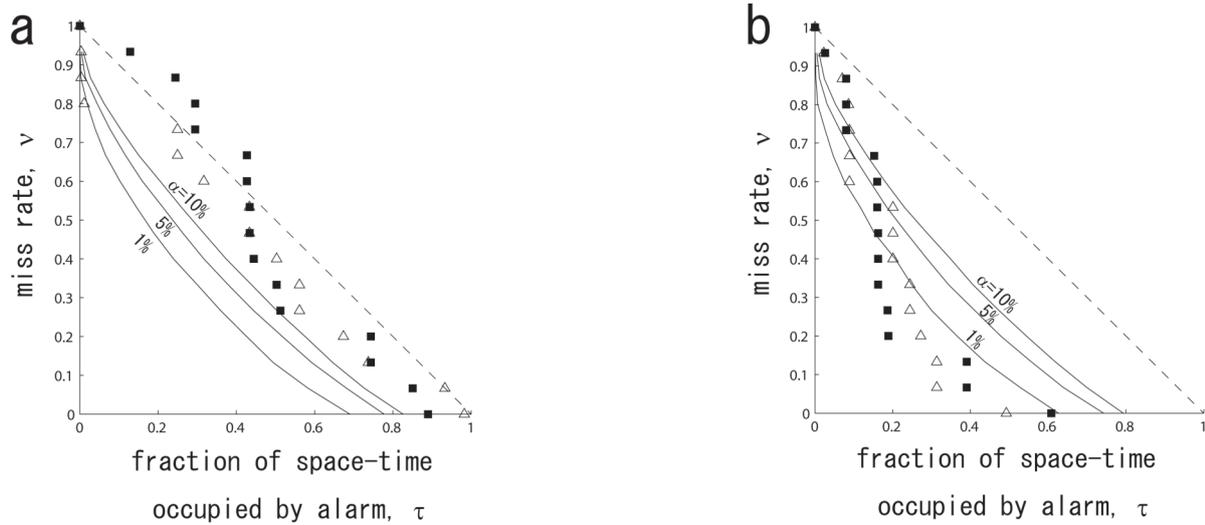
**Figure 1.** Molchan trajectory analysis. (a) PI (filled squares) and NSHM (open triangles) relative to the RI reference model. (b) RI (filled squares) and PI (open triangles) relative to the NSHM reference model. The Molchan trajectories are shown by the plots of the miss rate $\nu$ as functions of fraction of space-time occupied by alarm $\tau$. Each plot also shows the $\alpha = 1\%$, 5% and 10% significance boundaries. In the Molchan trajectory plots, the points below these boundaries reject the null hypothesis. In (a), neither PI nor NSHM provide significant gain relative to RI. In (b), both PI and RI can be significantly better at forecasting future earthquakes than the NSHM for $\nu \sim 0.0$-$0.5$ at the 95% confidence level (or $\alpha < 5\%$). (a) is a reproduction using the data and program code supplemented in the online version of Zechar and Jordan [2008], while the same data and program code were used to create (b).

construction of the RI model can improve the performance of earthquake forecasting. Most seismicity-based models use smoothing [Rhoades and Evison 2004, Kafka 2002, Kagan and Jackson 2000, Kossobokov 2004, Helmstetter et al. 2007]. Inclusion of smoothing into model construction reflects likelihood of an increase in a subsequent earthquake occurrence in neighboring areas to current and past earthquakes. The physical basis for this increase is the concept of fault interaction via Coulomb Failure Function (CFF): the location for subsequent earthquakes off the fault of the main shock can be explained by the spatial changes in CFF [e.g., Stein et al. 1992]. Of course, the action of each earthquake is neither isotropic nor uniform; however, we use uniform smoothing as the simplest solution to take theses effects into account.

Smoothing also has different supporting reasons. First, cataloged earthquake data that researchers use are always limited in time: for example, as described below, we have used data since 1985. On a high-resolution grid such as the case of the Italian CSEP experiment, even in areas of high seismic activity many bins can remain "empty", just on a statistical basis, and not because of physical reasons. Smoothing can thus compensate for this lack of data. Secondly, the accuracy of epicenter determination can be lower than the grid step, particularly in the older sections of earthquake catalogs. Smoothing can reflect the degree of trust that we have in the spatial resolution. Therefore, the use of smoothing for the binning process in the generation of a forecast can account for error in the event location. In any case, a fundamental question is the effect of smoothing on improvements in the performance of earthquake forecasting.

For RI formulation, we define two periods: a forecast period from $t_2$ to $t_3$ ($t_2 < t_3$) that is fixed under a predefined

rule, and a learning period from $t_0$ to $t_1$ ($t_0 < t_1 \le t_2 < t_3$) that can be defined by a modeler who generates a RI forecast. Using earthquakes in the learning period, we then forecast seismicity in the period from $t_2$ to $t_3$.

The first step of the original RI approach is to divide a region of interest into a grid of boxes. We next count the number of earthquakes with $M \ge$ a lower cut-off magnitude ($M_L$) for the $i$-th box during the period from $t_0$ to $t_1$, denoted by $n_i(t_0, t_1, M_L)$. This number is computed for all of the boxes. Then, the relative values of these numbers are given in the form $n_i(t_0, t_1, M_L)/\sum_j n_j(t_0, t_1, M_L)$, where the sum for $j$ is taken over all of the boxes. For the sake of simplicity, we define $P_i(t_0, t_1, M_L) \equiv n_i(t_0, t_1, M_L)/\sum_j n_j(t_0, t_1, M_L)$, where $\sum_i P_i(t_0, t_1, M_L) = 1$. Thus, $P_i(t_0, t_1, M_L)$ reflects the relative intensity of seismicity on the basis of counting the number of earthquakes. In the framework of RI, large earthquakes in the period from $t_2$ to $t_3$ are considered likely to occur for boxes with high $P_i(t_0, t_1, M_L)$ values. This type of approach is called alarm-based forecast. That is, the RI forecast is a binary: an earthquake is forecast to occur in the $i$-th box when $P_i(t_0, t_1, M_L)$ is larger than a given threshold, while it is forecast not to occur when $P_i(t_0, t_1, M_L)$ is smaller than the threshold. The standard approach to the evaluation of binary forecasts is use of Molchan [Molchan 1990] and ROC [Mason 2003] diagrams.

We modified the original RI approach for the process of binning the data. This idea was first introduced to improve PI forecasts [Holliday et al. 2007, Nanjo et al. 2006a]. We take a similar approach to RI model construction. The binning is implemented by the averaging seismicity over the $i$-th box and the nearest 8 surrounding boxes. Where there is an earthquake with $M \ge M_L$ in the $i$-th box during the period from $t_0$ to $t_1$, we assign $9^{-1}$ to the $i$-th box and 8 surrounding boxes, while unity is assigned only to the $i$-th box in the

original approach. Thus, implementation of this assignment smoothes the seismicity pattern in space. Above, we have considered the immediately adjacent boxes. Expansion includes the second nearest boxes: the number of $25^{-1}$ is assigned to the $i$-th box and the 24 surrounding boxes. This expansion to higher orders (third nearest boxes, fourth nearest, and so on) is possible: the number of $(2S - 1)^{-2}$ is assigned to the $i$-th box and the $(2S - 1)^{-2} - 1$ surrounding boxes, where $S = 1, 2, 3, \ldots$ and where the case of $S = 1$ corresponds to the original RI approach without the binning implementation. Hereinafter, we refer to $S$ as the smoothing parameter. The larger the $S$ is, the smoother the spatial pattern of $P_i(t_0, t_1, M_L)$ is. Note that the minimum value ($S = 1$) already implies smoothing because the RI approach without smoothing corresponds to an epicenter map of past events. The smoothing parameter $S$ is tuned because for each model a better value for $S$ is unknown.

## 4. Requirement for CSEP testing

The first requirement for CSEP testing is that the N-, L- and R-Tests used by the CSEP [Schorlemmer et al. 2007] cannot handle a forecast that consists of boxes with zero values if a target earthquake occurs at a zero-value box. To put any RI forecast model under such a test, the smallest value among $P_i(t_0, t_1, M_L) > 0$ is assigned to every box with $P_i(t_0, t_1, M_L) = 0$ in the process of model construction.

A second requirement is that any model must forecast the number of earthquakes for a given magnitude bin $M$ at the $i$-th box, $\lambda_{iM}$. Each bin expressed by $M$ covers from $M - 0.05$ to $M + 0.05$. For the 3-month class, the range of magnitude bins is $M = 4.0, 4.1, \ldots, 9.0$. Similarly, the magnitude bins for the 5-year and 10-year classes are $5.0, 5.1, \ldots, 9.0$. To meet these requirements, our challenge is to move RI from the original alarm-based forecast to the CSEP forecast. Below we give a brief explanation of this conversion of $P_i(t_0, t_1, M_L)$ to $\lambda_{iM}$.

We first estimate the total number of events with $M \geq M_L$ in the learning period $t_0$ to $t_1$, and we call this $N_T$. We next compute the number of events with $M \geq M_L$ in a time window with a length equal to $\Delta t \equiv t_3 - t_2$, $N_T \Delta t (t_1 - t_0)^{-1}$. We then multiply $P_i(t_0, t_1, M_L)$ by $N_T \Delta t (t_1 - t_0)^{-1}$ for the $i$-th box, and assume that this gives the number of events with $M \geq M_L$ for the forecast period of $\Delta t$.

To extrapolate our forecast into any magnitude bin expressed by $M_1 \leq M < M_2$, we use the Gutenberg-Richter (GR) frequency-magnitude law:

$$\log N = A - bm , \qquad (1)$$

where the parameters $A$ and $b$ are constants, and $N$ is the number of earthquakes with $M \geq m$. Substituting $m = M_L$ and $N = N_T \Delta t (t_1 - t_0)^{-1} P_i(t_0, t_1, M_L)$ into Equation (1), we obtain $A = \log_{10}[N_T \Delta t (t_1 - t_0)^{-1} P_i(t_0, t_1, M_L)] + bM_L$ for the $i$-th box. Substitution of this equation into Equation (1) gives $N = 10^{\beta(m)}$

with $\beta(m) = \log_{10}[N_T \Delta t (t_1 - t_0)^{-1} P_i(t_0, t_1, M_L)] - b(m - M_L)$. This is the equation to compute the forecast number of events with $M \geq m$ for the $i$-th box in the period from $t_2$ to $t_3$. Finally, the difference in $N$ between $m = M_1$ and $m = M_2$ gives the forecast number ($\lambda_{iM}$) of events for the bin $M_1 \leq M < M_2$ at the $i$-th box in the period from $t_2$ to $t_3$: $\lambda_{iM} \equiv 10^{\beta(M_1)} - 10^{\beta(M_2)}$. Thus, the final output is a set of $\lambda_{iM}$ values.

## 5. Data

According to the "rules of the game", the testing center has provided modelers with three catalogs for forecast-model development: two historical catalogs, known as the CPTI (Catalogo Parametrico dei Terremoti Italiani; http://emidius.mi.ingv.it/CPTI/), from 1901 to 2006, and the CSI 1.1 (Catalogo della Sismicità Italiana; http://csi.rm.ingv.it/), from 1981 to 2002; plus the Italian seismic bulletin (Bollettino Sismico Italiano, BSI; http://bollettinosismico.rm.ingv.it/) that currently provides data since April 16, 2005. The bulletin data prior to this date are still being revised. The testing center provides modelers with the BSI that include data until March 31, 2009, for model development and forecast generation. That is, we cannot use data from April 1, 2009, to the model submission deadline (June 31, 2009).

For our model development, we used the CSI 1.1 and the BSI because the magnitude scale used for these catalogs is the local magnitude scale and both include the microseismicity data. On the other hand, the CPTI includes macroseismicity. Furthermore, each event cataloged in CPTI has information on the magnitude in one or multiple scales: the scales used are body-wave magnitude, moment magnitude, surface magnitude, and local magnitude. There is no obvious authorized conversion equation between the local magnitude scale and the moment magnitude scale for Italy, and we would like to use microearthquake information to construct RI forecasts. Thus, we did not use the CPTI.

We also did not use CSI 1.1 data prior to 1985, because the number of reported earthquakes in the period of 1981-1984 was quite small. This was associated with the many network changes that occurred in the early 1980s.

The gap between the periods over which the CSI 1.1 and the BSI cover ranges from $t_{GS} =$ January 1, 2003, to $t_{GE} =$ April 15, 2005. In our forecast generation, the effect of this gap is taken into consideration where there is the need to use the two catalogs for the generation.

## 6. Application to the Italian testing experiment

Previous studies have used RI for 10-year experiments to predict earthquakes in California and Japan, and also on a worldwide basis [Rundle et al. 2002, Rundle et al. 2003, Tiampo et al. 2002, Holliday et al. 2005, Holliday et al. 2007, Nanjo et al. 2006a, Nanjo et al. 2006b, Nanjo et al. 2006c]. Following theses studies, our first application was for the long-term (5-year and 10-year) models. Our further
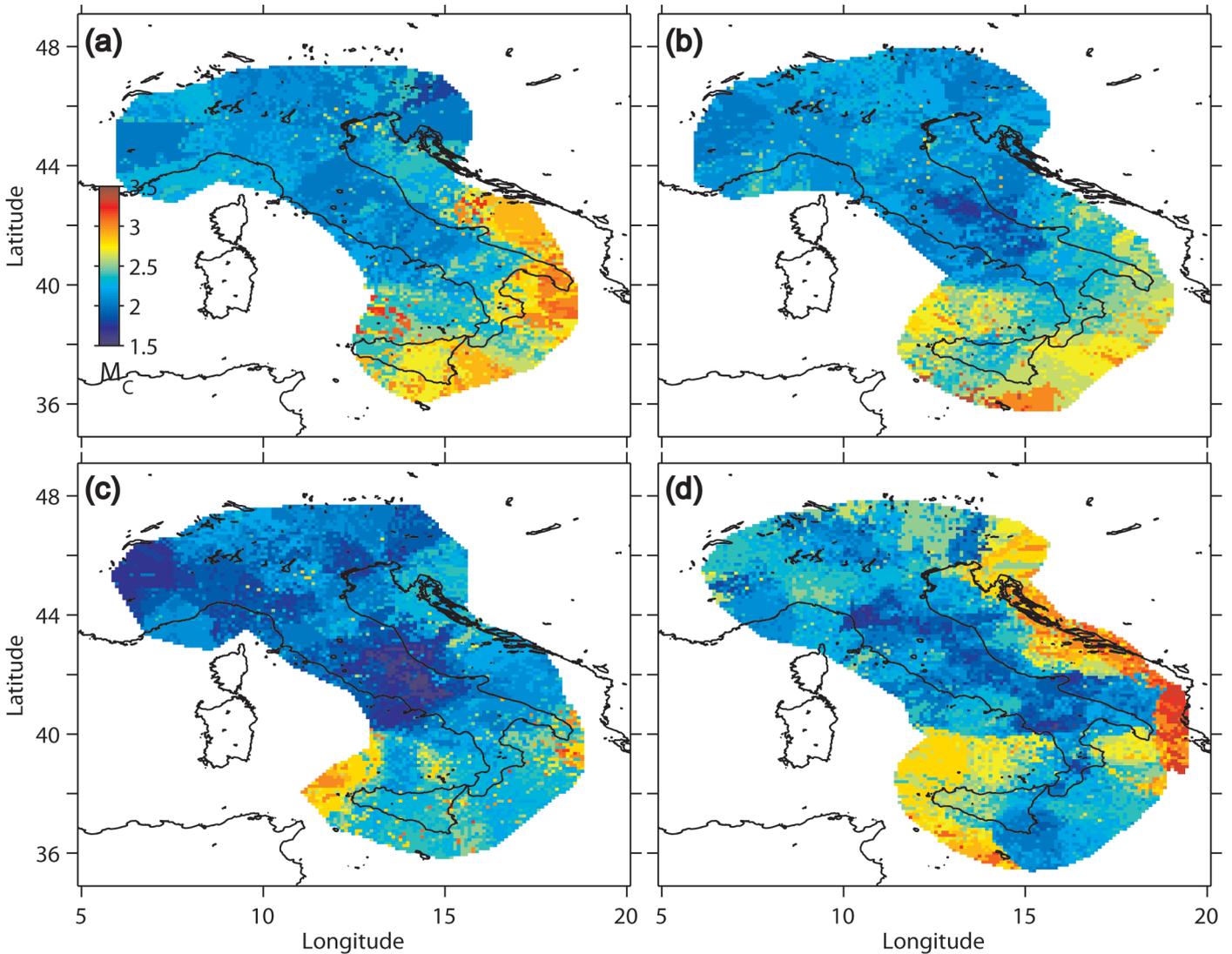
**Figure 2.** $M_C$ maps based on events in the four periods of: (a) January 1, 1985 - December 31, 1989; (b) January 1, 1990 - December 31, 1996; (c) January 1, 1997 - December 31, 2002; and (d) April 16, 2005 - March 31, 2009.

interest was to apply this to a 3-month testing model, to examine which testing class is more meaningful for RI forecasting.

In computing $\lambda_{iM}$ for an Italy-wide forecast based on the RI hypothesis, the parameters ($t_0$, $t_1$, $M_L$, $b$, $S$, depth, and box size) were tuned, while $t_2$ and $t_3$ (giving $\Delta t$) were predefined following the "rules of the game". Our tuning was based on the available past seismicity. While a model with a well-tuned set of parameters would be used retrospectively to forecast seismicity, its ability to prospectively forecast may be disappointing. Therefore, our preferred approach was that all of the parameters were determined independent of retrospective forecasting. However, an appropriate value of $S$ is unknown, although we explained the physical basis and other important reasons for this parameter in Section 3. To find a better $S$ value for each model, we conducted parameter searches based on retrospective forecasting. Below we briefly describe our rationale behind each decision, after a brief explanation of depth and box size.

### 6.1. Depth and box size

We considered all of the events down to a 30-km depth for forecast generation, because the depths of forecast earthquakes are less than 30 km, according to the "rules of the game".

The testing center provided modelers with a list of 8993 nodes (longitude and latitude pairs), where each $\lambda_{iM}$ must be computed. The ruled node spacing is 0.1° in latitude and longitude, and the set of nodes covers all of the Italian territory. We assumed that each node was the center of a box considered for RI model construction, and that the box size is 0.1° in latitude and longitude.

### 6.2. Completeness magnitude and the $t_0$ value

A small $M_L$ and a long period from $t_0$ to $t_1$ are desirable, because the ideal construction of any RI forecast map is based on including smaller events and longer seismic history, based on previous studies [e.g., Rundle et al. 2002, Rundle et al. 2003, Tiampo et al. 2002, Holliday et al. 2005, Nanjo et al. 2006a, Nanjo et al. 2006b, Nanjo et al. 2006c]. On the other
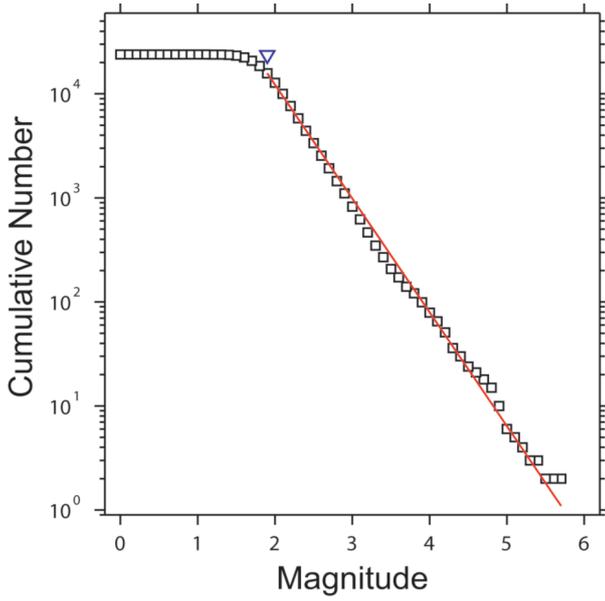
**Figure 3.** Cumulative number of earthquakes as a function of *m*. Based on events in the period from April 16, 2005, to March 31, 2009. The straight line shows the GR law in Equation (1), with *b* = 1.19 and *A* = 6.28. The inverted triangle indicates $M_C$ = 1.9.

hand, small events are in general more frequently missed by seismograph networks than large ones, and seismic networks have been generally modernized over the past to improve their earthquake detection capabilities. Thus, the work necessary is the estimation of the completeness magnitude ($M_C$), above which all of the events are interpreted as recorded. Various techniques to compute $M_C$ and their applications were reviewed by Wiemer and Wyss [2002] and Schorlemmer and Woessner [2008]. A frequently-used method is based on detecting the point of deviation from the GR distribution in Equation (1). The magnitude of the deviation point is defined as $M_C$. Below $M_C$, a fraction of events is not detected by the seismic network. We adopted a GR-based method known as the entire-magnitude-range (EMR) method [Woessner and Wiemer 2005]. We used the «ZMAP» seismicity-analysis software [Wiemer 2001], which routinely maps the EMR-based $M_C$.

To monitor the completeness in time and space, we created $M_C$ maps (Figure 2) for the four periods: (a) January 1, 1985 - December 31, 1989; (b) January 1, 1990 - December 31, 1996; (c) January 1, 1997 - December 31, 2002; and (d) April 16, 2005 - March 31, 2009. Data from the CSI 1.1 and the BSI were used for (a-c) and (d), respectively. We set the sample size of earthquakes for each node as 200, and the node spacing as 0.1˚. We excluded low- or no-seismicity areas from the mapping nodes. For (a-d), the $M_C$ values are mainly 1.5-2.5. $M_C$ = 3.0-3.5 can be seen mainly in offshore regions for (a-d), and in the western Balkan peninsula for (d): both of these regions are on and around the periphery of the testing region. Although the $M_C$ map in (a) was based on events for the oldest period, the maximum $M_C$ was about 3.5. Based on the visual inspection of $M_C$, we take $M_L$ to be 3.5 and $t_0$ to be January 1, 1985.

### 6.3. The b-value

We used earthquakes with depths of less than 30 km for the period April 16, 2005 - March 31, 2009, for the cumulative frequency of earthquakes *N* as a function of *m* in Figure 3. To find the best set of *A* and *b*, we applied the EMR method to the same data. We obtained *b* = 1.19 ± 0.01, and *A* = 6.28 for $M_C$ = 1.9. Using the GR fitting in Equation (1) with these values, we got good agreement with the observation for *m* ≥ 1.9. Thus, we assumed *b* = 1.20 for our forecast generation.

If the *b*-value change led to significant changes in the results in the parameter searches given below, the use of *b* = 1.2 for prospective forecasting would be questionable. To address this point, we also took two *b*-values in retrospective forecasting: *b* = 1.1 and 1.3. However, below we see no significant influences, supporting the use of *b* = 1.2 for formal prospective testing.

### 6.4. Smoothing parameters

*Log-likelihood tests*: Our statistical test to find a better *S* value was based on the log-likelihood (*LL*), which is used to evaluate the consistency of a forecast (a set of $\lambda_{iM}$) with an observation (a set of target earthquakes). Our *LL* was defined as follows:

$$LL = \sum_i \left( \theta_i + \omega_i \log\theta_i - \log\omega_i! \right) \qquad (2)$$

where $\theta_i$ is the sum of $\lambda_{iM}$ over all of the target magnitude bins ($\theta_i \equiv \sum_M \lambda_{iM}$), and $\omega_i$ is the number of target earthquakes. This is a simplified version of the CSEP test [Schorlemmer et al. 2007]. Our test compared in *LL* among forecasts: the higher the *LL* value is, the better the agreement between forecast and observation is. A useful measure for this comparison is $\Delta LL \equiv LL_{MAX} - LL(S)$, where $LL_{MAX}$ is the maximum for *LL* among forecasts generated using different *S* values, and $LL(S)$ is the *LL* obtained for *S*. $\Delta LL$ = 0 indicates the best among the candidates. Thus the corresponding *S* value is considered to be optimal to forecast the observed targets. On the other hand, $\Delta LL$ > 0 indicates worse performing forecast than does $\Delta LL$ = 0.

*5-year and 10-year classes*: Our target was a set of 31 observed 4 ≤ *M* ≤ 9 earthquakes that occurred in the testing region during the period from $t_2$ = April 16, 2005, to $t_3$ = March 31, 2009 ($\Delta t$ = 1446 days). This is the period over which the BSI covers. For the formal prospective testing in the 5-year and 10-year classes, the target magnitude range was 5 ≤ *M* ≤ 9. However, the number of earthquakes in this range during the same period was 1: this was not enough for statistical testing. Thus, our lowest magnitude was reduced to *M* = 4. Also note that our target period was about 4 years: this period is shorter than the periods for the 5-year and 10-year classes. It is possible to take $t_2$ to be a date before the start day of the catalog gap ($t_{GS}$), in order to set the lengths of the retrospective forecast periods as 5 and 10 years.

However, such cases are believed to be far from 5-year and 10-year seismicity clustering patterns. Thus, we assumed that the use of our target period can capture the essence of such long-term testing classes.

Our forecast for $4 \leq M \leq 9$ was based on 857 events with $M \geq M_L = 3.5$, for the period from $t_0 =$ January 1, 1985, to $t_1 =$ December 31, 2002. For the forecast generation, we used $b = 1.2$, $\Delta t = 1446$ days, and different $S$-values: 1, 2, 3, ..., 7.

$\Delta LL$ is given in Figure 4a as a function of $S$ for $b = 1.2$ (red diamonds). The optimal forecast was obtained for $S = 5$. It was also seen that the case of the minimal smoothing ($S = 1$) did not show a better performance than for the optimal case ($S = 5$). This shows that the inclusion of smoothing into constructing RI can improve the performance of earthquake forecasting. To indicate that the $LL$-Test did not show different results when various $b$-values were used, the curves for $b = 1.1$ and $1.3$ are given in Figure 4; these provide support for the result for $b = 1.2$. Using the parameters and the optimal $S = 5$, we mapped the cumulative numbers ($\theta_i$) of forecast earthquakes over $4 \leq M \leq 9$ in Figure 5a. We then plotted the 31 targets. Our prospective forecasts for the 5-year and 10-year testing classes were generated using $S = 5$.

*3-month class*: For our parameter search, we first assumed the target period $t_2 =$ January 1, 2009, to $t_3 =$ March 31, 2009 ($\Delta t = 89$ days). The lowest target magnitude was down to $M = 3.5$. This is smaller than the $M = 4.0$ predefined for the CSEP 3-month class, for the same reason as described for the 5-year and 10-year classes. The highest target magnitude was $M = 9.0$. The number of target earthquakes with $3.5 \leq M \leq 9$ was six.

To generate a forecast, we used 930 events with $M \geq M_L = 3.5$ that occurred in the period since $t_0 =$ January 1, 1985. We again used a default $b = 1.2$. We then took $\Delta t = 89$ days and $t_1 =$ December 31, 2008. Note that the period from $t_0$ to $t_1$ included the catalog gap period ($t_{GS}$ to $t_{GE}$); thus, the period considered for the forecast generation consisted of two periods from $t_0$ to $t_{GS}$ and from $t_{GE}$ to $t_1$. Finally, taking various $S$ values (1, 2, ..., 7), we generated the forecasts. Using the $LL$-Test, the result is given in Figure 4b. The forecast with $S = 3$ gaves the best performance. As with the long-term case, the forecast with the minimal smoothing ($S = 1$) did not show a better performance than that with the optimal smoothing ($S = 3$).

The use of 6 targets appears to be insufficient for statistical testing. To increase the number of target earthquakes, taking a value smaller than $M = 3.5$ would be possible. However, this is not the case here, because as shown in the completeness study (Section 6.2), the Italian seismic network can only reliably detect $M \geq 3.5$ earthquakes that occur in the testing region. Thus, a meaningful alternative was to carry out the same testing for a different target period. We took the forecast period from $t_2 =$ October 1, 2008, to $t_3 =$ December 31, 2008 ($\Delta t = 91$ days), and the targets were 9 earthquakes. The number of earthquakes in the learning period until $t_2 =$ September 30, 2008 was 921. The result in Figure 4b shows $\Delta LL = 0$ at $S = 5$, but very low $\Delta LL$ values are also seen at $S = 3$ and 4. Thus, better forecasts for the second period were given at $S = 3$-5. As for the long-term case, we checked the influence of the $b$-value change on the $\Delta LL$-$S$ behavior, using $b = 1.1$ and $1.3$: no significant influence was seen (data not shown).

Considering both cases of the different target periods, $S = 3$ was our choice for the 3-month class. The 3-month forecast maps using $S = 3$ for the first and second
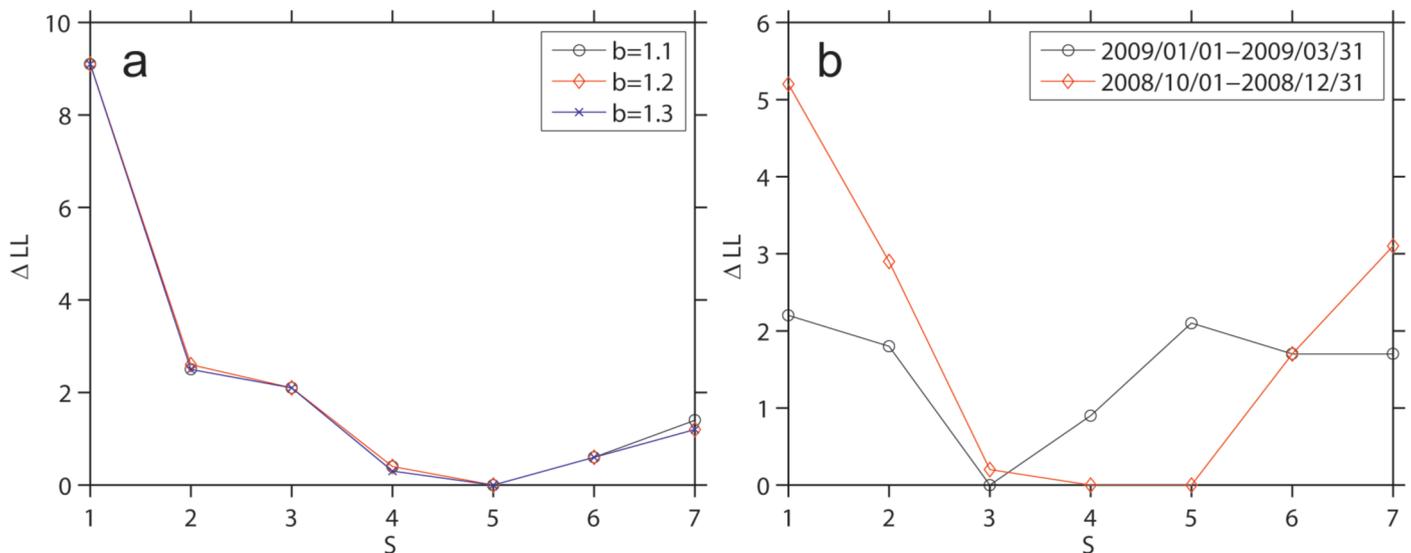


**Figure 4.** $\Delta LL$ as a function of $S$. (a) Retrospective forecast period from $t_2 =$ April 16, 2005, to $t_3 =$ March 31, 2009. Targets are 31 earthquakes of $4 \leq M \leq 9$ that occurred during this period. We considered $b = 1.2$ (red diamonds). Also included is the $\Delta LL$-$S$ dependence for $b = 1.1$ (black circles) and $1.3$ (blue crosses). (b) Retrospective forecast periods from $t_2 =$ January 1, 2009, to $t_3 =$ March 31, 2009 (black circles), and from $t_2 =$ October 1, 2008, to $t_3 =$ December 31, 2008 (red diamonds). Targets are 6 and 9 earthquakes with $3.5 \leq M \leq 9$ for the former and latter periods, respectively.
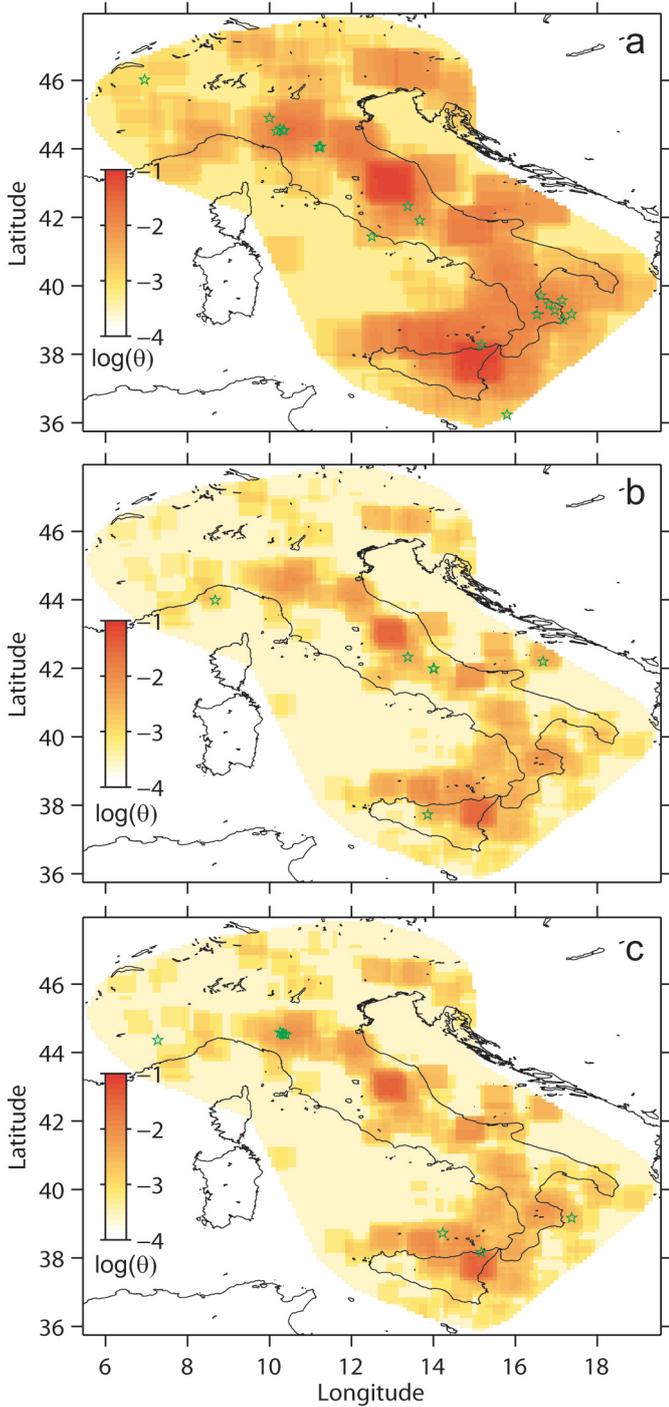
**Figure 5.** RI-based retrospective forecast maps. (a) Forecast period from $t_2 =$ April 16, 2005, to $t_3 =$ March 31, 2009 for $4 \leq M \leq 9$ events. The smoothing parameter used was the optimal $S = 5$, based on the search shown in Figure 4a. The logarithm of the cumulative forecast number ($\theta_i$) is given in the color code. The targets (stars) show 31 $4 \leq M \leq 9$ events that occurred during the forecast period. (b, c) Forecast periods from $t_2 =$ January 1, 2009, to $t_3 =$ March 31, 2009 (b), and from $t_2 =$ October 1, 2008, to $t_3 =$ December 31, 2008 (c) for $3.5 \leq M \leq 9$ earthquakes. Smoothing parameter used both was the appropriate value ($S = 3$) based on the search in Figure 4b. Stars indicate the 6 and 9 target earthquakes that occurred in the forecast periods, respectively.

retrospective tests are given in Figure 5b,c, respectively. Figure 5b,c also shows the respective plots for the 6 and 9 earthquakes that were used as the targets in the retrospective testing.

## 7. Conclusion and discussion

Based on the RI hypothesis, we prospectively forecast the numbers ($\lambda_{iM}$) of earthquakes for each box for each of the magnitude bins $4 \leq M \leq 9$ for the predefined consecutive 3-month periods (starting at $t_2 =$ August 1, 2009, November 1, 2009, February 1, 2010, and May 1, 2010). Similarly, we forecast this number for each of the magnitude bins $5 \leq M \leq 9$ for the predefined 5-year period ($t_2 =$ August 1, 2009, to $t_3 =$ August 1, 2014) and the 10-year period ($t_2 =$ August 1, 2009, to $t_3 =$ August 1, 2019). An executable computer code for the 3-month model and numerical tables for the 5-year and 10-year models have been submitted to the Italian prospective testing at the CSEP European Testing Center. The main purpose of the submission of the RI-based models to different classes is to determine which class is more meaningful for the RI hypothesis.

For the 5-year-class model submitted, we chose the set of $M_L = 3.5$, $t_0 =$ January 1, 1985, and $t_1 =$ March 31, 2009. The number of earthquakes with depths less than 30 km for the generation of a forecast was 936. Note that the data until $t_1$ were only available before the model submission deadline. We took the catalog gap period ($t_{GS}$ to $t_{GE}$) into account. We then used $S = 5$ and $b = 1.2$. The map of the cumulative number ($\theta_i$) for $5 \leq M \leq 9$ is shown in Figure 6a, using the numerical data in the submitted table. Our model predicts $\sum_i \theta_i = 2.8$ events in total for the testing region.

We did the same for the 10-year model, with $t_3$ as August 1, 2019. As for the 5-year model, the resultant map is shown in Figure 6b. The cumulative number ($\theta_i$) for the 10-year class for each box is twice that for the 5-year class, as the forecast period for the former class is twice that for the latter. The predicted total number of earthquakes is $\sum_i \theta_i = 5.6$, which is again twice that for the 5-year forecast model.

In the computer code of our 3-month model, we fixed $M_L = 3.5$, $S = 3$, $t_0 =$ January 1, 1985, and $b = 1.2$. Similar to the long-term models, the code takes the catalog gap period ($t_{GS}$ to $t_{GE}$) into account in the forecast generation. For each forecast implementation, given $t_2$ and $t_3$, we asked the testing center to equate $t_1$ to $t_2$ and to provide the BSI that includes the earthquakes until $t_1$ (i.e., $t_2$) as an input data catalog. This is because the data until $t_1$ (i.e., $t_2$) must be available for forecast generation. Using the submitted code with the fixed parameter set and the currently available input data until $t_1 =$ March 31, 2009, Figure 6c shows the map of the cumulative number ($\theta_i$) of events over the magnitudes $4 \leq M \leq 9$.

The present study has at least two very important independent aspects. First, the RI model is most commonly considered as a reference (or null hypothesis) to test other forecast models, for at least those that are alarm based. We extended the RI-type models, which are originally alarm-based, to a more general class, to be probability based. Thus the RI-type models become natural references for joint tests
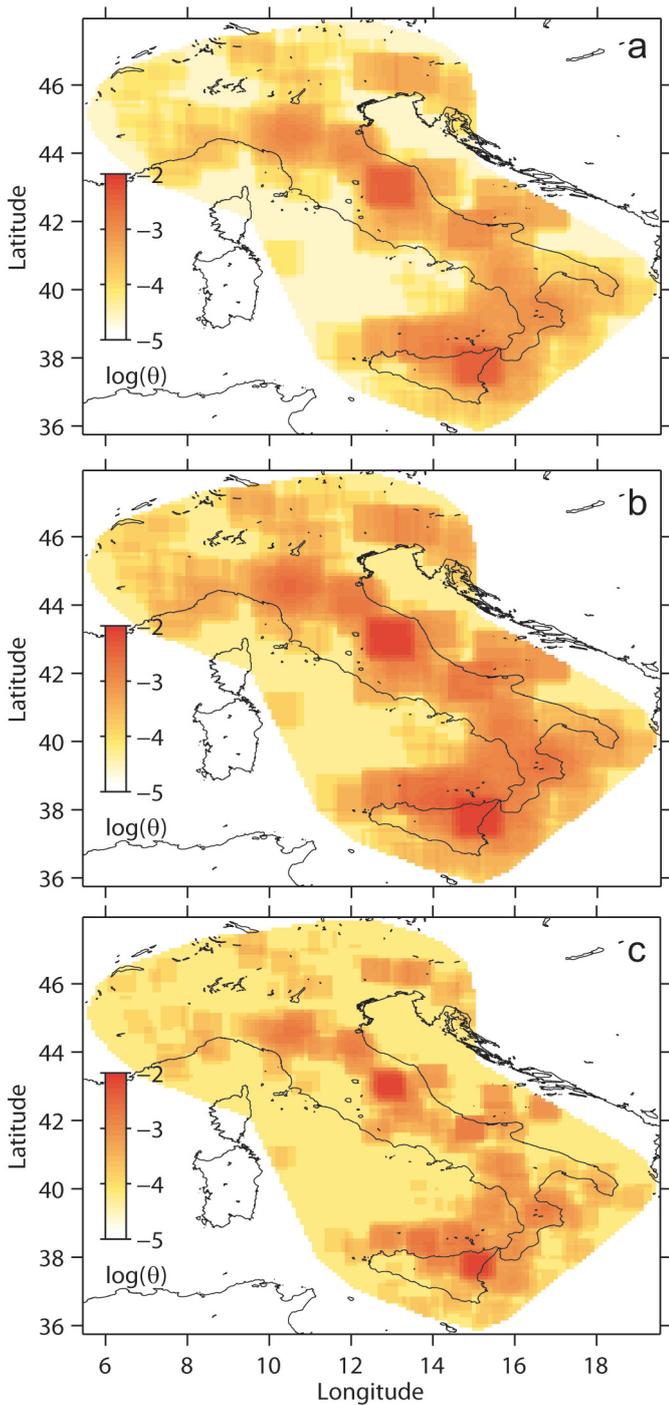
**Figure 6.** RI-based prospective forecast maps. (a, b) Forecast maps for the 5-year period from $t_2$ = August 1, 2009, to $t_3$ = August 1, 2014 (a) and the 10-year period from $t_2$ = August 1, 2009, to $t_3$ = August 1, 2019 (b), for $5 \leq M \leq 9$ events. The logarithm of the cumulative forecast number ($\theta_i$) is given in the color code. The tables of the numerical data used for these maps have been submitted to the testing center. (c) Forecast map for the period from $t_0$ = January 1, 1985, to $t_1$ = March 31, 2009, for the 3-month forecast for $4 \leq M \leq 9$ earthquakes. An executable computer code that can compute the forecast numbers $\lambda_{iM}$ used for (c) has been submitted to the testing center.

spatial smoothing of earthquake frequencies. We introduced a wider smoothing in the modified version than in the original one, and we studied different sizes (changing the smoothing factor $S$) in retrospective forecast tests. Our goal was to understand how wide the smoothing should be to provide a better forecasting performance.

However, the smoothing also had a negative effect. The seismicity is usually strongly clustered to the faults. The smoothing "disperses" the earthquake frequencies, assigning predatory values to the less active places between and outside the faults. Thus, for evaluating some forecast models, it is not appropriate to use a reference model with significantly wide smoothing. Accordingly, the significant gain of RI relative to NSHM might result from a similar effect: the PI and RI models are just much less smoothed in comparison to that of NSHM (see Figure 1b). This is similar to the evaluation of any forecast model *versus* a uniform reference model: practically any model gives results that are far from the diagonal in this case. To examine the effects of this smoothing on prospective forecasting, we have already submitted a variety of improved RI-based models with wide smoothing ranges to the Japanese experiment [Hirata et al. 2009] within CSEP.

Although we generated the RI-based forecasts on the basis of a non-declustered (original) catalog, the use of a declustering algorithm [e.g., Gardner and Knopoff 1974, Reasenberg 1985] would be of interest. A usual RI-type model based on a non-declustered catalog might overestimate forecast frequencies near the epicenters of past large earthquakes, because aftershocks that dominated the seismicity near these epicenters are included in the forecast generation. However, for purely statistical reasons, the application of any declustering algorithm might unfortunately prevent this hypothesis from being exploited for the Italian testing experiment. For example, for the long-term retrospective testing in Section 6.4, the forecasts were generated using events with $M \geq M_L = 3.5$ for the period from $t_0$ = January 1, 1985, to $t_1$ = December 31, 2002. However, the number of events used for the forecast generation was 857, a quite small number compared with the number of grid nodes (8993). In addition to this, it can be recalled that the number of earthquakes after declustering is, in general, approximately fewer than half the number of earthquakes before declustering, although the details of the declustered data depend on its desclustering algorithm. Without a large smoothing factor, target earthquakes might very likely fall into "empty" boxes, to which the minimal frequency value is assigned. Thus, declustering-based implementation needs to overcome these testing circumstances. According to the "rules of the game", some masking of the original testing region can be applied. If the center approves a proposal of masking off the areas of larger $M_C$ values, a testing region characterized by small $M_C$ values can be defined. This allows

with any forecast models. These tests will provide an important answer to the question: does a specific forecast model give more information about future earthquakes than just the known spatial distribution of the earthquake frequencies in the past? Secondly, the RI-type models use

the inclusion of small events into the dataset available for model development and forecast generation, resulting in avoiding targets falling into "empty" boxes. If this testing region is set in the second-round experiment, we would like to submit both non-declustered (original) and declustered versions based on the RI algorithm, and to examine whether the former significantly outperforms the latter in forecasting future seismicity.

In summary, our submitted RI models allow for a systematic test of the hypothesis that the relative intensity of past seismicity (on the basis of counting the number of events) can give information for forecasting future moderate-to-large earthquakes in Italy. Future modifications to RI model construction can include regional variations in the $b$-value [Wiemer and Schorlemmer 2007], while here we used a uniform constant, $b = 1.2$. Another example includes the use of a declustering algorithm [e.g., Gardner and Knopoff 1974, Reasenberg 1985] to remove aftershocks for forecast generation, while we used the non-declustered (original) catalog here, as discussed in some detail in the previous paragraphs. It might be of interest for future studies to compare the currently implemented version and the proposed modified one to determine whether these modifications are significantly fruitful for improving future seismicity forecasts. Despite leaving such possible modifications for later, we believe that the results that we will obtain through the Italian prospective experiment will be valuable for our better understanding of the importance (or lack of importance) of the simple relative intensity of seismicity for earthquake forecastability.

## References

Field, E.D. (2007). Overview of the working group for the development of Regional Earthquake Likelihood Models (RELM), Seismol. Res. Lett., 78, 7-16.

Gardner, J.K. and L. Knopoff (1974). Is the sequence of earthquakes in southern California, with aftershocks removed, Poissonian?, Bull. Seismol. Soc. Am., 64, 1363-1367.

Helmstetter, A., Y.Y. Kagan and D.D. Jackson (2007). High-resolution time-independent grid-based forecast for M ≥ 5 earthquakes in California, Seismol. Res. Lett., 78, 1, 78-86.

Hirata, N., T.H. Jordan, H. Tsuruoka and K.Z. Nanjo (2009). Call for paper: Special issue of Earth, Planets, and Space (EPS), Earthquake Forecast Testing Experiment for Japan, Earth Planet Space, 61 (6), 805.

Holliday, J.R., K.Z. Nanjo, K.F. Tiampo, J.B. Rundle and D.L. Turcotte (2005). Earthquake forecasting and its verification, Nonlin. Proc. Geophys., 12, 965-977.

Holliday, J.R., J.B. Rundle, K.F. Tiampo and D.L. Turcotte (2006a). Using earthquake intensities to forecast earthquake occurrence times, Nonlin. Proc. Geophys., 13, 585-593.

Holliday, J.R., J.B. Rundle, D.L. Turcotte, W. Klein, K.F. Tiampo and A. Donnellan (2006b). Space-time clustering and correlations of major earthquakes, Phys. Rev. Lett., 97, 238501.

Holliday, J.R., C. Chen, K.F. Tiampo, J.B. Rundle, D.L. Turcotte and A. Donnellan (2007). A RELM earthquake forecast based on Pattern Informatics, Seismol. Res. Lett., 78 (1), 87-93.

Jordan, T.H. (2006). Earthquake predictability, brick by brick, Seismol. Res. Lett., 77 (1), 3-6.

Kafka, A.L. (2002). Statistical analysis of the hypothesis that seismicity delineates areas where future large earthquakes are likely to occur in the Central and Eastern United States, Seismol. Res. Lett., 73, 990-1001.

Kagan, Y.Y. and D. Jackson (2000). Probabilistic forecasting of earthquakes, Geophys. J. Int., 143, 438-453.

Kanamori, H. (2003). Earthquake prediction: an overview, in International Handbook of Earthquake and Engineering Seismology, edited by W.H.K. Lee, H. Kanamori, P.C. Jennings and C. Kisslinger, Amsterdam, 1205-1216.

Keilis-Borok, V.I. (1990). The lithosphere of the Earth as a nonlinear system with implications for earthquake prediction, Rev. Geophys., 28, 19-34.

Keilis-Borok, V.I. and V.G. Kossobokov (1990). Premonitory activation of earthquake flow: algorithm M8, Phys. Earth Planet. Inter., 61, 73-83.

Keilis-Borok, V.I. and I.M. Rotwain (1990). Diagnosis of time of increased probability of strong earthquakes in different regions of the world, Phys. Earth Planet. Inter., 61, 57-72.

Keilis-Borok, V.I. and A.A. Soleviev, eds. (2003). Nonlinear Dynamics of the Lithosphere and Earthquake Prediction, New York, 341 pp.

Kossobokov, V. (2004). Earthquake prediction: basics, achievements, perspectives, Acta Geod. Geoph. Hung., 39 (2-3), 205-221.

Marzocchi, W. (2008). Earthquake forecasting in Italy, before and after the Umbria-Marche seismic sequence 1997. A review of earthquake occurrence modeling at difference spatio-temporal-magnitude scales, Annals of Geophysics, 51 (2-3), 405-416.

Mason, I.B. (2003). Binary events, in Forecast Verification, edited by I.T. Jolliffe and D.B. Stephenson, Hoboken, 37-76.

Mogi, K. (1985). Earthquake Prediction, Academic Press, Tokyo, 355 pp.

Molchan, G.M. (1990). Strategies in strong earthquake prediction, Phys. Earth Planet. Int., 61, 84-98.

Nanjo, K.Z., J.R. Holliday, C.-C. Chen, J.B. Rundle and D.L. Turcotte (2006a). Application of a modified pattern informatics method to forecasting the locations of future large earthquakes in the central Japan, Tectonophysics, 424, 351-366.

Nanjo, K.Z., J.R. Holliday, C.-C. Chen, J.B. Rundle and D.L. Turcotte (2006b). Forecasting locations of future large earthquakes, using Pattern Informatics: A review, Proc. Inst. Stat. Math., 54 (2), 281-297.

Nanjo, K.Z., J.R. Holliday, C.-C. Chen, J.B. Rundle and D.L. Turcotte (2006c). Pattern informatics and its application for optimal forecasting of large earthquakes in Japan, Pure Appl. Geophys., 163 (11-12), 2417-2432.

Reasenberg, P. (1985). Second-order moment of central California seismicity, 1969-1982, J. Geophys. Res., 90 (B7), 5479-5495.

Rhoades, D.A. and F.F. Evison (2004). Long-range earthquake forecasting with every earthquake a precursor according to scale, Pure Appl. Geophys., 161, 47-72.

Rundle, J.B., K.F. Tiampo, W. Klein and J.S.S. Martins (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. U. S. A., 99, 2514-2521.

Rundle, J.B., D.L. Turcotte, R. Shcherbakov, W. Klein and C. Sammis (2003). Statistical physics approach to understanding the multiscale dynamics of earthquake fault systems, Rev. Geophys., 41 (4), 1019.

Scholz , C.H. (2002). The Mechanics of Earthquakes and Faulting, New York, 2nd ed., 471 pp.

Schorlemmer, D., M. Gerstenberger, S. Wiemer and D.D. Jackson (2007). Earthquake likelihood model testing, Seismol. Res. Lett., 78 (1), 17-29.

Schorlemmer, D. and J. Woessner (2008). Probability of detecting an earthquake, Bull. Seismol. Soc. Am., 98 (5), 2103-2117.

Stein, R.S., G.C.P. King and J. Lin (1992). Change in failure stress on the southern San Andreas fault system caused by the 1992 magnitude = 7.4 Landers earthquake, Science, 258, 1328-1332.

Tiampo, K.F., J.B. Rundle, S. McGinnis, S.J. Gross and W. Klein (2002). Mean field threshold systems and phase dynamics: an application to earthquake fault systems, Europhys. Lett., 60, 481-487.

Turcotte, D.L. (1991). Earthquake prediction, Annu. Rev. Earth Planet. Sci., 19, 263-281.

Turcotte, D.L. (1997). Fractals and Chaos in Geology and Geophysics, New York, 398 pp.

Wiemer, S. (2001). A software package to analyze seismicity: ZMAP, Seismol. Res. Lett., 72, 373-382.

Wiemer, S. and D. Schorlemmer (2007). ALM: An asperity-based likelihood model for California, Seismol. Res. Lett., 78 (1), 134-140.

Wiemer, S. and M. Wyss (2002). Mapping spatial variability of the frequency-magnitude distribution of earthquakes, Adv. Geophys., 45, 259-302.

Woessner, J. and S. Wiemer (2005). Assessing the quality of earthquake catalogs: Estimating the magnitude of completeness and its uncertainties, Bull. Seismol. Soc. Am., 95 (2), 684-698.

Wyss, M. (1997). Nomination of precursory seismic quiescence as a significant precursor, Pure Appl. Geophys., 149, 79-114.

Wyss, M. and A.H. Martirosyan (1998). Seismic quiescence before the M7, 1988, Spitak earthquake, Armenia, Geophys. J. Int., 134, 329-340.

Wyss, M. and S. Wiemer (2000). Change in the probability for earthquakes in southern California due to the Landers magnitude 7.3 earthquake, Science, 290, 1334-1338.

Zechar, J.D. and T.H. Jordan (2008). Testing alarm-based earthquake predictions, Geophys. J. Int., 172, 715-724.

Corresponding author: Kazuyoshi Z. Nanjo,
Earthquake Research Institute, University of Tokyo, Tokyo, Japan;
e-mail: nanjo@eri.u-tokyo.ac.jp