



Long-Term Probabilistic Forecast for $M \geq 5.0$ Earthquakes in Iran

MOHAMMAD TALEBI,¹ MEHDI ZARE,¹ ANTONELLA PERESAN,^{2,3} and ANOOSHIRAVAN ANSARI¹

Abstract—In this study, a long-term forecasting model is proposed to evaluate the probabilities of forthcoming $M \geq 5.0$ earthquakes on a 0.2° grid for an area including the Iranian plateau. The model is built basically from smoothing the locations of preceding events, assuming a spatially heterogeneous and temporally homogeneous Poisson point process for seismicity. In order to calculate the expectations, the space distribution, from adaptively smoothed seismicity, has been scaled in time and magnitude by average number of events over a 5-year forecasting horizon and a tapered magnitude distribution, respectively. The model has been adjusted and applied considering two earthquake datasets: a regional unified catalog (MB14) and a global catalog (ISC). Only the events with $M \geq 4.5$ have been retained from the datasets, based on preliminary completeness data analysis. A set of experiments has been carried out, testing different options in the model application, and the average probability gains for target earthquakes have been estimated. By optimizing the model parameters, which leads to increase of the predictive power of the model, it is shown that a declustered catalog has an advantage over a non-declustered one, and a low-magnitude threshold of a learning catalog can be preferred to a larger one. In order to examine the significance of the model results at 95% confidence level, a set of retrospective tests, namely, the L test, the N test, the R test, and the error diagram test, has been performed considering 13 target time windows. The error diagram test shows that the forecast results, obtained for both the two input catalogs, mostly fall outside the 5% critical region that is related to results from a random guess. The L test and the N test could not reject the model for most of the time intervals (i.e. ~ 85 and $\sim 62\%$ of times for the ISC and MB14 forecasts, respectively). Furthermore, after backwards extending the time span of the learning catalogs and repeating the L test and N test for the new dataset as well as the R test, it is observed that a low-quality longer catalog does not essentially improve the predictive skill of the model than a high-quality shorter one. The performed retrospective tests suggest that the model yields some statistically acceptable efficiency for the studied area, at least with respect to the spatially uniform reference model. Thus, the

considered model may provide useful information as a reference for more refined time-independent models and also in combination with long-term indications from seismic hazard maps; this is particularly relevant in areas characterized by a high level of predicted ground shaking and high forecast rate.

Key words: Probabilistic forecasting, adaptively smoothed seismicity, retrospective tests, seismic hazard, Iranian plateau.

1. Introduction

The Iranian plateau from time to time experiences destructive earthquakes due to the convergence of Arabian and Eurasian plates. These earthquakes usually lead to a lot of property damage and human casualties. Hazard assessment is one of the key elements, which may permit to reduce such kind of earthquake effects. On the other hand, earthquake forecasts can be considered as a part of the seismic hazard studies (Panza et al. 2012; Zechar and Jordan 2010). Current approaches in operational earthquake forecasting, as well as their practical relevance in mitigating the impact from possible impending earthquakes, are matter of active scientific discussion (Jordan et al. 2014; Kossobokov et al. 2015; Peresan et al. 2012).

Several earthquake forecasting methods have been already applied to the Iranian plateau (e.g., Maybodan et al. 2014; Radan et al. 2013; Talebi et al. 2015), most of them involving a set of predictions with “Yes–No” responses. Another common type of earthquake forecast, performed in different parts of the world, is the long-term probabilistic forecast of earthquakes (e.g., Helmstetter et al. 2007; Helmstetter and Werner 2012; Kagan and Jackson 1994, 2000; Werner et al. 2010a, 2011). This type of studies includes some of the methods evaluated in the

¹ International Institute of Earthquake Engineering and Seismology, 21 Arghavan Street, Tehran 19395/3913, Iran. E-mail: m.talebi@iiees.ac.ir; mzare@iiees.ac.ir; a.ansari@iiees.ac.ir

² National Institute of Oceanography and Experimental Geophysics, Seismological Research Centre, Udine, Italy. E-mail: aperesan@units.it

³ Department of Mathematics and Geosciences, University of Trieste, Trieste, Italy.

framework of the Collaboratory for the Study of Earthquake Predictability (CSEP) (Schorlemmer et al. 2010; Werner et al. 2010b; Zechar et al. 2010b, 2013). Several of these researches are carried out as part of regional earthquake risk studies and can be factored in the global earthquake model (Zechar and Jordan 2010).

In this study, a similar long-term probabilistic forecasting model is presented for the Iranian plateau. The forecast is based on spatial smoothing of the earthquakes occurred in the study region. This investigation is similar to those performed by Helmstetter et al. (2007) and Werner et al. (2011) for California and by Werner et al. (2010a) for Italy. Several studies indicated that, according to selected statistical tests, the model by Helmstetter et al. (2007) outperforms other models; moreover, the tests failed to reject the model during the period under review. Thus, it is now being used in the USGS seismic hazard model for California (Field et al. 2014).

The present study is a kind of grid-based forecast in which the study area is divided into several cells and the expected number of earthquakes is determined for each cell. The possibility is explored to choose the model parameters according to regional data and following an optimization process. Moreover, in order to investigate the impacts of input data and find the more robust features of seismicity, two different earthquake datasets are implemented in the present investigation.

In this study, we suggest that the proposed model may show satisfactory performances in forecasting target $M \geq 5.0$ earthquakes. Specifically, the model is intended to provide an improved quantitative description of the space distribution of seismicity over a uniform model. Such an experiment can be important for developing forecast techniques in Iran. Given the important role that earthquake forecast would play in combination with earthquake hazard assessment, this investigation could be useful in identifying areas of high seismic risk in the Iranian plateau. Moreover, such kind of studies provides valuable information about the utilization and validation of smoothed intensity forecast methods in different seismotectonic zones.

2. Data

The studied area covers a region including the Iranian plateau limited to the coordinate ranges 26–39°N and 45–61°E. In order to consider the effect of seismicity outside the testing area on the seismicity of its inside, the area for collecting data is additionally extended by 3°. Earthquakes within this area can possibly affect the hazard of Iran.

In this study, dates, locations, and magnitudes of earthquakes are used. Instrumental seismicity of Iran and adjacent regions are reported in several regional and global catalogs. For all these catalogs three main periods can be identified, which are characterized by a different data accuracy, namely: 1900–1964, 1964–1996 (i.e. after the global-scale installation of modern seismological instruments till 1996), and after 1996 (i.e. after the development of seismic local networks in Iran). Accordingly, the values of uncertainties in the earthquake parameters decreased over time.

As the quality of the most recent part of the catalogs (i.e. from 1996 till now) is high compared to that of the others, this part of the data has been used as main part of our dataset. We have also extended the time span of the data during retrospective consistency tests; the advantage of this analysis is to see if data from the longer period can provide more predictive information or not.

Similar to data used for hazard analysis and most of other forecasting methods (Peresan et al. 2011), the data used for this study should be homogeneous in terms of magnitude; i.e., the magnitudes should be consistent over time as possible. In fact, the robustness of the applied method can be assessed with respect to uncertainties of randomly defective and incomplete input data (e.g., Peresan et al. 2002). Still, any sound method would fail (or it is expected to fail) if the input data provide a systematically inconsistent picture of reality, like in the case reported by Peresan et al. (2000).

Based on the Gutenberg–Richter relation, Maybodian et al. (2014) have discussed in detail the homogeneity and variation of earthquake magnitudes for seismicity of Iran. They studied two regional catalogs that provided by the Iranian Seismological

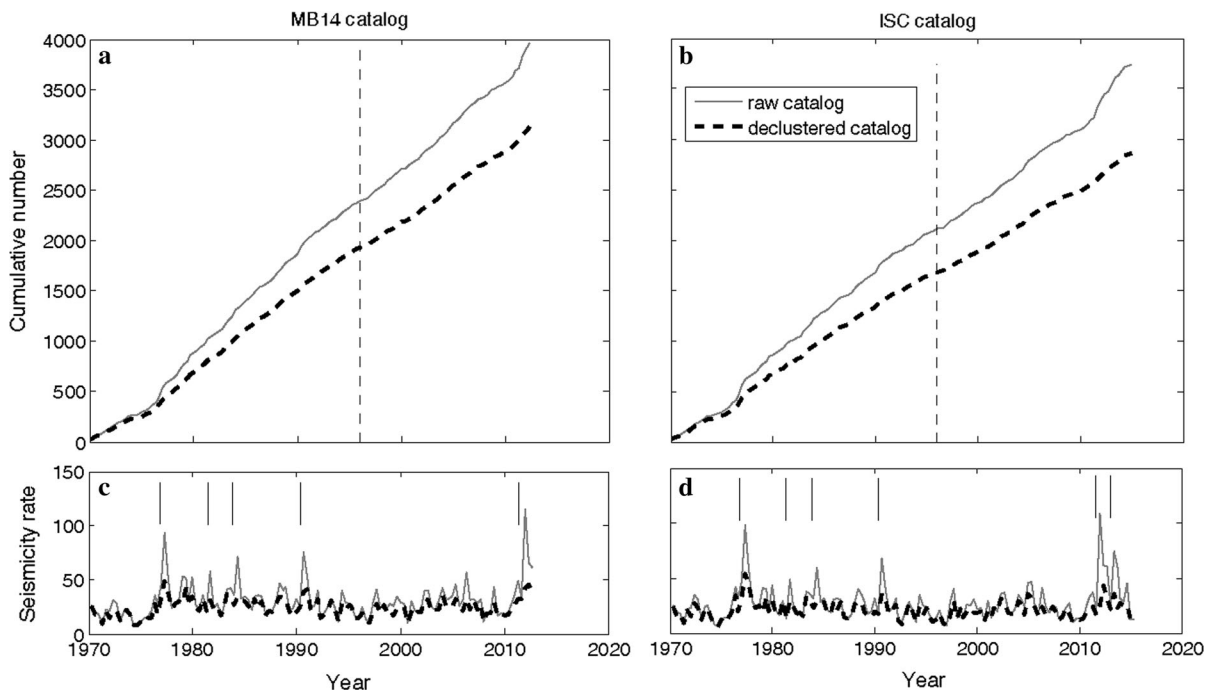


Figure 1

Temporal distributions of original and declustered seismicity with $M \geq 4.5$. **a, b** Display accumulated number of earthquakes in the MB14 and ISC catalogs, respectively. The *dash lines* separate the most recent part of the catalogs (i.e. from 1996 till now). **c, d** Seismicity rate per 4 months for the MB14 and ISC catalogs, respectively. The *vertical solid lines* indicate the times of occurrence of $M > 6.5$ events and sharp rises of seismicity. The *curves* show that the statistic trend of the seismicity is not subjected to change after ~ 1980 and most of the dependent seismicity causing the large fluctuations in time distribution has been eliminated in the declustered catalogs

Center (IGUT) and the International Institute of Earthquake Engineering and Seismology (IIEES), as well as two global catalogs collected by the National Earthquake Information Center (NEIC) and the International Seismological Centre (ISC). Since the global catalogs commonly provide several estimates for an earthquake, based on different magnitude scales, Maybodian et al. (2014) have chosen the largest magnitude among all of the reported values. According to their study, the ISC catalog outperformed the others from the point of view of magnitude completeness (M_c) and homogeneity. Accordingly, the ISC catalog is used in the present study by considering the maximum magnitude reported for each earthquake. This criterion is also employed by Kagan and Jackson (2010) to extract data of the PDE catalog. For moderate size earthquakes, the maximum magnitude is often reported in mb and Ms scales, while for large earthquakes the biggest one is usually in Mw scale.

In addition to the data from global and regional scientific centers, in recent years, some unified earthquake catalogs have been compiled for the Iranian plateau (e.g., Karimiparidari et al. 2013; Mousavi-Bafrouei et al. 2014; Shahvar et al. 2013; Zare et al. 2014). The catalog provided by Mousavi-Bafrouei et al. (2014) is the most recent one, and the period of its instrumental data, extending to early 2013, is longer than other data sets; therefore, it is used as the second catalog in this investigation. Mousavi-Bafrouei et al. (2014) have applied general orthogonal regression (GOR) method to convert magnitudes of Mb, Ms, Ml, and Mn type into Mw. In the following, MB14 stands for the study of Mousavi-Bafrouei et al. (2014).

Using the maximum curvature method (Wiemer and Wyss 2000), we have estimated the completeness level of the considered datasets as $M_c = 4.5$ over the whole area. Figure 1a, b shows the accumulated number of $M \geq 4.5$ earthquakes reported in the MB14

and ISC catalogs, respectively. It appears that in both the MB14 and the ISC catalogs the statistic trend of seismicity is not subjected to change after 1996.

As we want to compare the results from the MB14 and ISC catalogs, and the period covered by the MB14 catalog is up to 2013 only, we used the data of the ISC catalog until the beginning of 2013 for optimizing the model and carrying out the retrospective tests. However, in order to compute the final forecast, the data between 2013 and 2015 in the ISC catalog are considered too.

3. Declustering

In order to estimate the time-independent earthquake rate, we removed the dependent events (i.e. the aftershocks), which are associated with large fluctuations of seismic activity in space and time. For this purpose the catalogs were declustered by Reasenber's algorithm (1985), as adapted by Helmstetter et al. (2007), setting the depth and lateral error in the earthquake location to 15 and 10 km, respectively.

Declustering process for the MB14 and ISC catalogs identified, respectively, 20 and 28% of the earthquakes as dependent ones. After declustering, the MB14 catalog for the study area includes 1240 earthquakes with magnitude $M \geq 4.5$ (from 1996 to 2013), and the ISC catalog contains 1178 earthquakes with magnitude $M \geq 4.5$ (from 1996 to 2015).

Figure 1c, d shows that the declustering process could remove the sharp rises, which are associated with the temporal clustering, in the seismicity rate. The cumulative number curves related to the declustered catalogs are also smoother than those of the raw catalogs (Fig. 1a, b). In general, it can be deduced that most of the foreshocks and aftershocks causing the large fluctuations in time have been removed from the final catalogs.

4. Methodology

4.1. Spatial Distribution

As the forecasting model used in this study is described in detail by Helmstetter et al. (2007), here

we just present a summary of the method. This model aims to predict the spatial distribution of seismicity. The main assumption of the model is that a future target earthquake is likely to occur in the vicinity of epicenters of prior events. In other words, the locations of past earthquakes provide important information about locations of the forthcoming ones. Earlier studies (e.g., Helmstetter et al. 2007) have shown that information about small earthquakes may increase the performance of the forecasting model.

In order to describe the spatial distribution of seismicity in the gridded geographical region, we estimated the seismicity rate in each cell of the grid, after having applied a kernel method to smooth the epicentral location of all earthquakes. Specifically, an isotropic adaptive kernel, centered on each earthquake epicenter, was applied for the spatial smoothing of the location of an earthquake i as follows:

$$K_{d_i}(\vec{r} - \vec{r}_i) = C(d_i) / \left(|\vec{r} - \vec{r}_i|^2 + d_i^2 \right)^{1.5}. \quad (1)$$

This function is a power-law kernel, where $|\vec{r} - \vec{r}_i|$ is the horizontal distance between the i th earthquake epicenter location \vec{r}_i and an arbitrary point \vec{r} (e.g., a node of the grid), d_i is the scale parameter or bandwidth, and $C(d_i)$ is a normalizing constant, such that the integral of the kernel $K_{d_i}(\vec{r} - \vec{r}_i)$ over an infinite area is equal to 1.

The scale parameter d_i is a function of the density of events around the i th earthquake. We calculated this parameter as the lateral distance between the i th event and its k th nearest neighbor (Helmstetter et al. 2007). The parameter k , which is a single parameter computed for the whole catalog, was adjusted at the stage of regional model optimization, so that d_i was small where the density of events was high and vice versa. This thus can improve the resolution of the forecast in places with higher density. The parameter d_i has the lowest values of ~ 3 km for the MB14 dataset and ~ 2.5 km for the ISC dataset.

The density λ at any point \vec{r} was then calculated by summing the contributions of all N past earthquakes, included in each learning catalog, as follows:

$$\lambda(\vec{r}) = \sum_{i=1}^N K_{d_i}(\vec{r} - \vec{r}_i). \quad (2)$$

However, the forecast provides an expected number of events within each cell. Hence, Eq. (2) was then integrated over each cell to obtain the seismicity rate in this cell.

4.2. Model Optimization

A powerful tool to assess the adaptation of real and predicted model is the likelihood function. The optimal value for the parameter k for each model was obtained by a criterion of the maximum likelihood estimation. To do this, several sets of non-overlapping catalogs, i.e. pairs of a learning catalog and a testing catalog, were extracted, so that earthquakes used to calculate the seismicity rate were not used to evaluate the results. To build each pair of learning and testing catalogs, we considered the same input catalog but with different start times, end times, and magnitude thresholds.

For each cell, the rate λ was computed and normalized according to Helmstetter et al. (2007). Assuming spatiotemporal independency of earthquakes, the log-likelihoods of the models (L) were calculated by the Poisson distribution (Schorlemmer et al. 2007, Appendix A). Finally, the best models were selected by calculating the differential probability gain per target earthquake (G) for each model as follows:

$$G = \exp\left(\frac{L - L_{\text{unif}}}{N_t}\right), \quad (3)$$

where N_t is the number of target earthquakes at each testing catalog, and L_{unif} is the log-likelihood of a spatially uniform reference model, which assumes that earthquake epicenters are distributed uniformly in space (i.e. over a regular grid, the number of epicenters is the same within each cell).

4.3. Magnitude Distribution

For the magnitude distribution, we assume a tapered Gutenberg–Richter distribution, proposed by Helmstetter et al. (2007) as

$$P(m) = 10^{-b(m-m_{\min})} \exp\left(10^{1.5(m_{\min}-m_c)} - 10^{1.5(m-m_c)}\right), \quad (4)$$

where b is the b value parameter, m_{\min} is the minimum magnitude of target earthquakes and m_c is the upper corner magnitude. For Iran, we have used $m_c = 7.56$ as suggested by Bird and Kagan (2004) for the type of active continent; this estimate is compatible with records from historical and instrumental catalogs of Iran, where earthquakes with magnitude around 7.5 are reported in Zagros, Alborz, Kopetdagh, and Makran regions. Using maximum likelihood algorithm as implemented in seismic tool ZMAP 6.0 (Wiemer 2001), we estimated the b value as 1.14 ± 0.02 for both the ISC and the MB14 catalogs.

Figure 2 depicts the cumulative frequency–magnitude distributions of $M \geq 4.5$ events, as reported in the ISC and MB14 declustered catalogs, and the line fitted by Eq. (4). It can be observed that the difference between the two distributions is low for small and moderate earthquakes, up to $M = 6.0$. The fitted line, which is applied for estimating the space–magnitude rate, fits the rate of $M \leq 6$ events fairly well, whereas it tends to underestimate the rate for the largest target magnitudes. However, since the earthquakes with $5 \leq M \leq 6$ naturally make the most part of the target catalogs, the influence of such b value estimate on the obtained results is negligible.

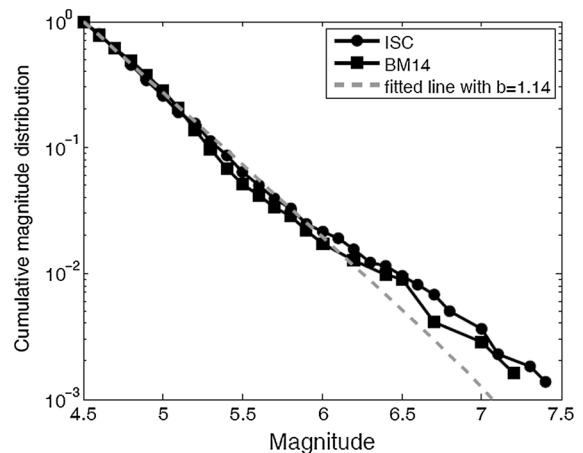


Figure 2

Cumulative frequency–magnitude distribution of $M \geq 4.5$ events, reported in the MB14 and ISC declustered catalogs since 1996. The dashed line is fitted according to Eq. (5) with $b = 1.14 \pm 0.02$. The fit appropriately describes the rate of $M \leq 6$ earthquakes, whereas it tends to underestimate it for larger magnitudes

4.4. Space–Time–Magnitude Distribution

The expectation (E) for each coupled bin of space (i_x, i_y), and magnitude (i_m) is estimated by

$$E(i_x, i_y, i_m) = \lambda'(i_x, i_y) N_h P(i_m), \quad (5)$$

where λ' is the normalized spatial density λ , such that the sum of λ' over all the cells is equal to 1, and N_h is measured as the average number of target earthquakes in the whole region over a time span corresponding to the forecast horizon h . In order to estimate the parameter N_h for each catalog, the total number of observed target events is divided by the catalog duration, and then it is multiplied by the duration of the forecast horizon. Using the declustered catalogs, we estimated this parameter as ~ 67.53 and ~ 81.01 earthquakes of $M \geq 5.0$ per 5 years for the ISC and MB14 datasets, respectively. $P(i_m)$ is the cumulative probability of a target earthquake occurring within the magnitude bin i_m and is calculated from Eq. (4).

5. Results

Table 1 reports the results of the model optimization for a number of model variants. Based on these results we investigate some of the factors affecting the differential probability gain of a forecast, which leads to choose the model parameters. Having varied the smoothing parameter k between 1 and 50, we selected the value of k that corresponds to the maximum log-likelihood score of the model. In fact, higher gain values indicate relatively higher model efficiency, with the gain of the reference homogeneous model corresponding to 1.0.

Figure 3a, b illustrates how the results change depending on the size (large or small) of the cells used for gridding the study territory, considering two different 5-year target periods (2003–2007, 2008–2012) with target magnitudes $M \geq 5$.

In general, for these two time periods the average curves of the differential probability gain display larger values for models with smaller cell sizes, for all the learning magnitude thresholds of 4.5, 5.0, 5.5, and 6.0. As expected, when only the largest

earthquakes (above the target magnitude) are used to train the model, the results get very poor.

In similar studies, carried out in the framework of RELM (Regional Earthquake Likelihood Models test of earthquake forecasts in California) (Schorlemmer and Gerstenberger 2007), the distance between grid points was set to 0.1° .

Although the location error for earthquakes of the Iranian plateau is greater than that of California (e.g., Engdahl et al. 2006; Richards-Dinger and Shearer 2000), our results show that the models of cell size $0.1^\circ \times 0.1^\circ$ could be appropriate also for the Iranian region. The reason may be the fact that the smoothing technique has smoothed the impacts of errors in the earthquake location too. Indeed, the obtained results indicate an almost negligible difference between the values related to the cell sizes of $0.1^\circ \times 0.1^\circ$ and $0.2^\circ \times 0.2^\circ$. In this study, however, we prefer using the cell size $0.2^\circ \times 0.2^\circ$, which better copes with the error of earthquake location and finite earthquake source dimensions. On the other hand, earthquakes are not points and $0.1^\circ \times 0.1^\circ$ cell size is even less than the approximate source size for the largest target earthquakes (e.g., $M \geq 5.0$); accordingly, the fault of a real target earthquake may exceed the cell size.

In order to determine the effects of the declustering process, we have studied the results of a model with cell size $0.2^\circ \times 0.2^\circ$ using events from the non-declustered and declustered catalogs. For the two previously mentioned target time periods, the gain values have been calculated for target magnitudes $M \geq 5$. Figure 3c, d shows the average results obtained for each magnitude threshold of the learning catalogs.

For both the ISC and the MB14 catalogs, models based on the declustered catalogs provide higher gains than those obtained from the raw catalogs. These results indicate that non-declustered data are too localized in space for application in a time-independent long-term forecasting model. Without the declustering procedure, more complicated models are possibly required (Werner et al. 2010a). Thus, in what follows, we have selected the declustered catalogs as reference database.

We have also studied the effects of using different magnitude thresholds of the learning catalogs on the

Table 1

Results from the model optimization analysis, obtained using the MB14 and ISC catalogs

Model			Learning catalog				Target catalog				Result			
No	Cell size (°)	Catalog	t1	t2	m	N	t1	t2	m	N	L	G	k	b_w (km)
1	0.1	MB14	1996	2007	4.5	655	2008	2012	5	111	-601.8	2.4	3	40.8
2	0.1	MB14	1996	2002	4.5	347	2003	2007	5	62	-386.5	1.9	3	53.6
3	0.1	MB14	1996	2007	5	164	2008	2012	5	111	-605.0	2.3	2	64.1
4	0.1	MB14	1996	2002	5	102	2003	2007	5	62	-386.1	1.9	4	146.2
5	0.1	MB14	1996	2007	5.5	31	2008	2012	5	111	-660.5	1.4	3	265.4
6	0.1	MB14	1996	2002	5.5	17	2003	2007	5	62	-413.3	1.2	2	290.0
7	0.1	MB14	1996	2007	6	12	2008	2012	5	111	-680.6	1.2	4	486.7
8	0.1	MB14	1996	2002	6	7	2003	2007	5	62	-421.8	1.1	3	716.1
9	0.2	MB14	1996	2007	4.5	655	2008	2012	5	111	-458.0	2.3	3	40.8
10	0.2	MB14	1996	2002	4.5	347	2003	2007	5	62	-306.6	1.9	3	53.6
11	0.2	MB14	1996	2007	5	164	2008	2012	5	111	-459.5	2.3	2	64.1
12	0.2	MB14	1996	2002	5	102	2003	2007	5	62	-307.2	1.9	4	146.2
13	0.2	MB14	1996	2007	5.5	31	2008	2012	5	111	-514.5	1.4	3	265.4
14	0.2	MB14	1996	2002	5.5	17	2003	2007	5	62	-335.5	1.2	2	290.0
15	0.2	MB14	1996	2007	6	12	2008	2012	5	111	-533.8	1.2	4	486.7
16	0.2	MB14	1996	2002	6	7	2003	2007	5	62	-343.4	1.1	3	716.1
17	0.5	MB14	1996	2007	4.5	655	2008	2012	5	111	-296.6	1.8	3	40.8
18	0.5	MB14	1996	2002	4.5	347	2003	2007	5	62	-203.9	1.9	4	66.0
19	0.5	MB14	1996	2007	5	164	2008	2012	5	111	-292.1	1.9	2	64.1
20	0.5	MB14	1996	2002	5	102	2003	2007	5	62	-207.2	1.8	4	146.2
21	0.5	MB14	1996	2007	5.5	31	2008	2012	5	111	-331.6	1.3	2	218.2
22	0.5	MB14	1996	2002	5.5	17	2003	2007	5	62	-234.3	1.2	2	290.0
23	0.5	MB14	1996	2007	6	12	2008	2012	5	111	-350.6	1.1	4	486.7
24	0.5	MB14	1996	2002	6	7	2003	2007	5	62	-241.4	1.1	3	716.1
25	1	MB14	1996	2007	4.5	655	2008	2012	5	111	-202.1	1.6	13	103.6
26	1	MB14	1996	2002	4.5	347	2003	2007	5	62	-135.5	1.8	2	41.2
27	1	MB14	1996	2007	5	164	2008	2012	5	111	-199.4	1.6	2	64.1
28	1	MB14	1996	2002	5	102	2003	2007	5	62	-142.5	1.6	4	146.2
29	1	MB14	1996	2007	5.5	31	2008	2012	5	111	-226.8	1.3	5	381.2
30	1	MB14	1996	2002	5.5	17	2003	2007	5	62	-166.8	1.1	3	388.2
31	1	MB14	1996	2007	6	12	2008	2012	5	111	-241.8	1.1	6	738.9
32	1	MB14	1996	2002	6	7	2003	2007	5	62	-168.5	1.0	3	716.1
33	0.1	ISC	1996	2007	4.5	563	2008	2012	5	75	-427.5	2.7	2	33.1
34	0.1	ISC	1996	2002	4.5	301	2003	2007	5	70	-424.1	2.0	3	62.8
35	0.1	ISC	1996	2007	5	154	2008	2012	5	75	-432.3	2.6	2	65.0
36	0.1	ISC	1996	2002	5	84	2003	2007	5	70	-431.3	1.8	2	90.6
37	0.1	ISC	1996	2007	5.5	36	2008	2012	5	75	-468.8	1.6	2	190.6
38	0.1	ISC	1996	2002	5.5	20	2003	2007	5	70	-461.5	1.2	2	261.4
39	0.1	ISC	1996	2007	6	16	2008	2012	5	75	-479.7	1.4	4	439.3
40	0.1	ISC	1996	2002	6	7	2003	2007	5	70	-471.3	1.0	3	714.4
41	0.2	ISC	1996	2007	4.5	563	2008	2012	5	75	-332.0	2.6	2	33.1
42	0.2	ISC	1996	2002	4.5	301	2003	2007	5	70	-331.9	2.0	3	62.8
43	0.2	ISC	1996	2007	5	154	2008	2012	5	75	-337.1	2.5	2	65.0
44	0.2	ISC	1996	2002	5	84	2003	2007	5	70	-338.4	1.9	2	90.6
45	0.2	ISC	1996	2007	5.5	36	2008	2012	5	75	-369.4	1.6	2	190.6
46	0.2	ISC	1996	2002	5.5	20	2003	2007	5	70	-371.2	1.2	2	261.4
47	0.2	ISC	1996	2007	6	16	2008	2012	5	75	-381.4	1.4	4	439.3
48	0.2	ISC	1996	2002	6	7	2003	2007	5	70	-380.8	1.0	3	714.4
49	0.5	ISC	1996	2007	4.5	563	2008	2012	5	75	-225.2	2.1	2	33.1
50	0.5	ISC	1996	2002	4.5	301	2003	2007	5	70	-216.0	2.1	3	62.8
51	0.5	ISC	1996	2007	5	154	2008	2012	5	75	-226.5	2.1	2	65.0
52	0.5	ISC	1996	2002	5	84	2003	2007	5	70	-221.6	1.9	2	90.6
53	0.5	ISC	1996	2007	5.5	36	2008	2012	5	75	-247.6	1.5	2	190.6

Table 1 continued

Model			Learning catalog				Target catalog				Result			
No	Cell size (°)	Catalog	t1	t2	m	N	t1	t2	m	N	L	G	k	b_w (km)
54	0.5	ISC	1996	2002	5.5	20	2003	2007	5	70	-257.4	1.1	2	261.4
55	0.5	ISC	1996	2007	6	16	2008	2012	5	75	-258.7	1.3	4	439.3
56	0.5	ISC	1996	2002	6	7	2003	2007	5	70	-266.2	1.0	3	714.4
57	1	ISC	1996	2007	4.5	563	2008	2012	5	75	-156.7	1.7	9	86.7
58	1	ISC	1996	2002	4.5	301	2003	2007	5	70	-141.5	1.8	4	74.6
59	1	ISC	1996	2007	5	154	2008	2012	5	75	-158.9	1.7	7	163.4
60	1	ISC	1996	2002	5	84	2003	2007	5	70	-146.0	1.7	2	90.6
61	1	ISC	1996	2007	5.5	36	2008	2012	5	75	-174.5	1.4	2	190.6
62	1	ISC	1996	2002	5.5	20	2003	2007	5	70	-179.3	1.1	2	261.4
63	1	ISC	1996	2007	6	16	2008	2012	5	75	-179.9	1.3	4	439.3
64	1	ISC	1996	2002	6	7	2003	2007	5	70	-184.1	1.0	3	714.4

t1 and t2 are first year and last year used to select the data, respectively; m and N are the minimum magnitude threshold and the corresponding number of events in the catalog, respectively; G is the differential probability gain per earthquake (Eq. 3); b_w represent the average adaptive bandwidth. Having varied the smoothing parameter k between 1 and 50, we selected the one, which corresponds to the maximum log-likelihood score (L)

predictive power of the model. The diagrams in Fig. 3e show that the gain reduces gradually by increasing the magnitude threshold of the learning catalogs, although this is less evident for the MB14 catalog. These results are in agreement with the results of other studies, such as Helmstetter et al. (2007) and Werner et al. (2011) for California, and show that the information carried by earthquakes with small magnitudes can lead to increase the model resolution; this could be eventually due to the increased number of data used to compute the forecasting model. As discussed in Maybodian et al. (2014), the magnitude completeness of the Iranian regional catalogs may well be lower in recent years, i.e. $M_c \approx 2.5$ since 2006, although a much longer time span is needed for the purposes of intermediate- or long-term forecasting algorithms, which require a preliminary calibration or learning stage.

In contrast to the similar study on CPTI catalog of Italy by Werner et al. (2010a), the current results do not display large fluctuations and turn out fairly stable, which could be explained by the sufficient number of target events reported in the catalogs (on average about 74 earthquakes of $M \geq 5.0$ per 5 years).

In Fig. 3e, the average curves of the differential probability gains obtained for the two target periods (2003–2007 and 2008–2012) are shown for both the ISC and the MB14 catalogs. The results indicate a

slight difference between the values. However, the ISC catalog obtained higher gain values than the MB14 catalog, for all the learning magnitude thresholds of 4.5, 5.0, 5.5, and 6.0. Thus, the adequate number of target events and the robustness of the method apparently preserve similarity of results, in spite of the differences between these catalogs.

Figure 4 shows the differential probability gain versus the average optimized bandwidth of the smoothing kernel. It can be seen that increasing the bandwidth generally leads to decreasing the gain value, so that it is going to converge to 1.0. A reason for this phenomenon is that increasing bandwidth makes the forecasting model go to a spatially uniform model.

It is found that the optimal value of the parameter k for the territory of Iran does not vary exceedingly in all the models, and ranges from 2 to 13. Its average value, used for the final forecasts, is equal to 3 for both the ISC and the MB14 catalogs. According to Werner et al. (2010a), the smoothing parameter k for the CPTI catalog of Italy varied in a much wider range, from 1 to 35. The difference in the value of the parameter k for Iran and Italy can be explained mainly by the different number of target events; in Italy the small number of target earthquakes makes the forecast model quite unstable. However, the low variability of the smoothing parameter might also reflect the influence of seismotectonic conditions on a

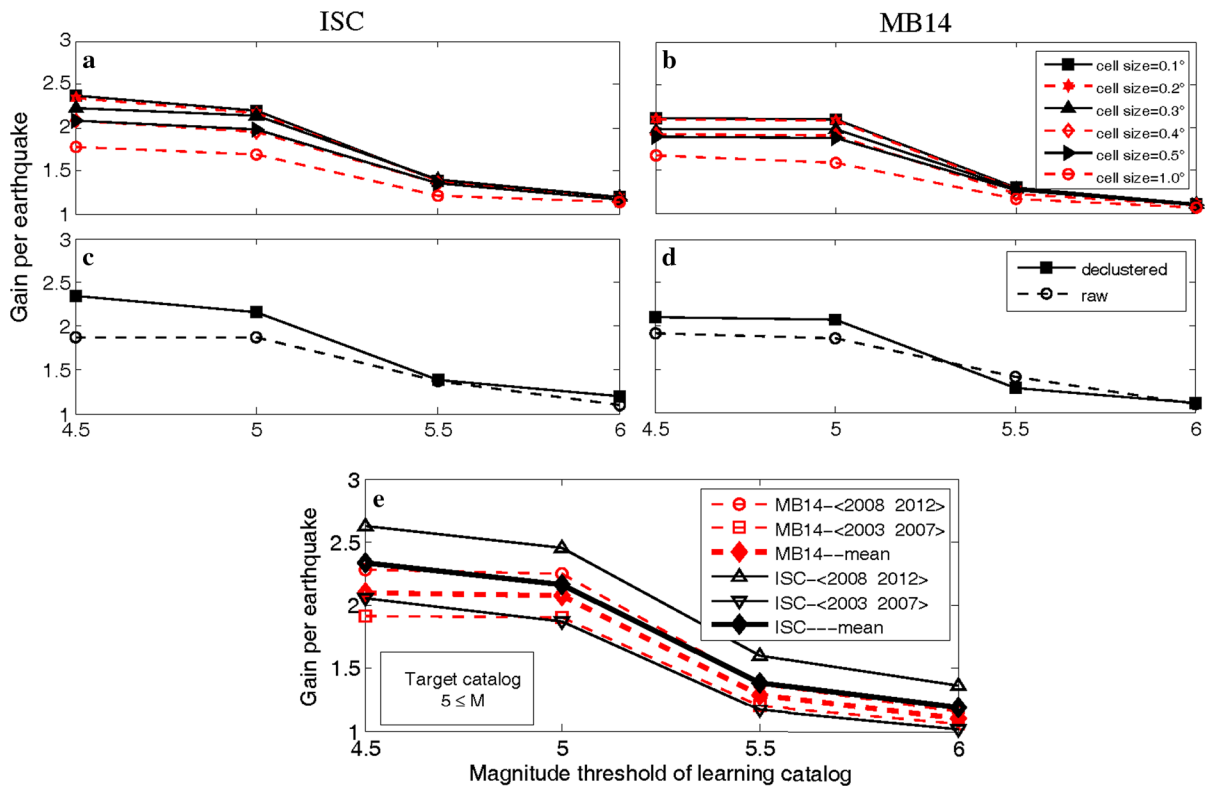


Figure 3

Differential probability gain per earthquake against magnitude threshold of the learning catalogs for two different 5-year target periods (2003–2007, 2008–2012) with target magnitudes $M \geq 5.0$. **a, b** Depict the effects of selecting small and big cell sizes on the results for the ISC and MB14 catalogs, respectively. The effects of declustering in estimation of the differential probability gain are shown in **c** for the ISC catalog and in **d** for the MB14 catalog. **e** Displays the effects of using each learning magnitude threshold on the predictive power of the model for the catalogs. **Bold curves** in **e** as well as ones in **a, b, c,** and **d** present the average values of the differential probability gains obtained for the target periods

smoothed intensity earthquake forecast; in fact, when the target events occur within formerly active areas (i.e. tend to repeat in the same locations), the value of the parameter k is generally small (Werner et al. 2010a)_ENREF_34. This suggests that the assumptions of the smoothed intensity forecast method are more compatible with seismotectonic setting of Iran, than with that of Italy, which is very heterogeneous (e.g., Radan et al. 2013).

Figures 5 and 6 present the final 5-year earthquake forecasts for the target $M \geq 5.0$ earthquakes, as obtained from ISC and MB14 catalogs, respectively. For both the maps, the higher seismicity rates of moderate earthquakes are predicted in the Zagros zone, extended in the western and southern parts of Iran. By a visual assessment, it seems that the maps

are approximately similar, and most of the observed target events that occurred after the learning time period, i.e. from 1 January 2013 to 2016 (white squares), are consistent with the space forecast.

6. Discussion

6.1. Retrospective Tests

In order to check the consistency of observational data with the forecasts, the models have been evaluated by different types of retrospective tests, namely, the L test, the N test, the R test (Schorlemmer et al. 2007; Werner et al. 2011; Zechar 2010; Zechar et al. 2010a), and the error diagram test (Molchan

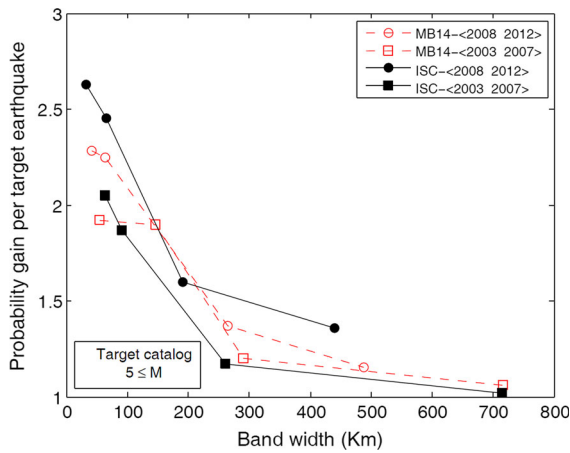


Figure 4

Differential probability gain per earthquake versus the mean kernel bandwidth, over the two target periods 2003–2007 and 2008–2012. Increasing the bandwidth generally leads to decreasing the gain value, so that it is going to converge to 1.0

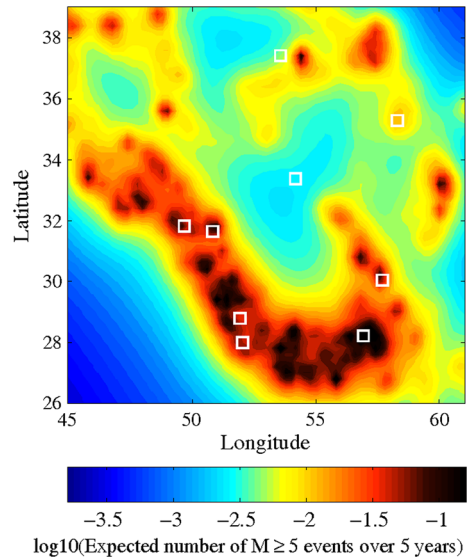


Figure 6

Expected seismicity rate of the target $M \geq 5.0$ earthquakes over a 5-year forecast horizon from 2015 to 2019, based on the ISC catalog. *White squares* are $M \geq 5.0$ declustered events that happened from 1 January 2015 to 2016, based on the ISC catalog

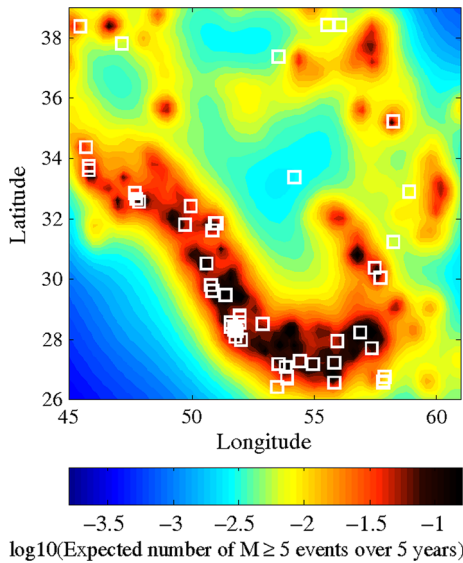


Figure 5

Expected seismicity rate of the target $M \geq 5.0$ earthquakes over a 5-year forecast horizon from 2013 to 2017, based on the MB14 catalog. *White squares* are $M \geq 5.0$ declustered events that happened from 1 January 2013 to 2016, as reported in the available ISC catalog. We may observe that the locations of observed $M \geq 5.0$ earthquakes well correlate with the hotspots in the forecasted map

1991; Molchan and Kagan 1992). It is worth noting that in order to build the forecasts examined in this section, the events of target periods have been excluded from learning catalogs.

In the L test, the observed joint log-likelihood (that is the likelihood of observing the target events given the forecast, i.e. the joint likelihood of each bin's observation given each bin's forecast) is calculated, and the consistency of forecasts is examined. In order to assess how well the observed catalog matches the forecasts, about 10,000 simulated catalogs consistent with each forecast were generated, and then the observed joint log-likelihood was compared with the distribution of the simulated joint log-likelihoods. In these simulated catalogs, locations and number of earthquakes were randomly assigned, using a Poisson distribution; the procedure for generating the synthetic catalogs is discussed in detail in Zechar et al. (2010a) and Zechar (2010). Similarly, in the N test, the consistency of the number of observed earthquakes with the number of those forecasted is examined.

In Fig. 7, the results of the L and N tests are shown for a set of 5-year target periods, partially overlapping and shifted by 1 year, beginning from 1996. For each model, rejected and passed forecasts are, respectively, shown by squares and circles. The 95% confidence limits from the Poisson distribution of the results are also presented with bars.

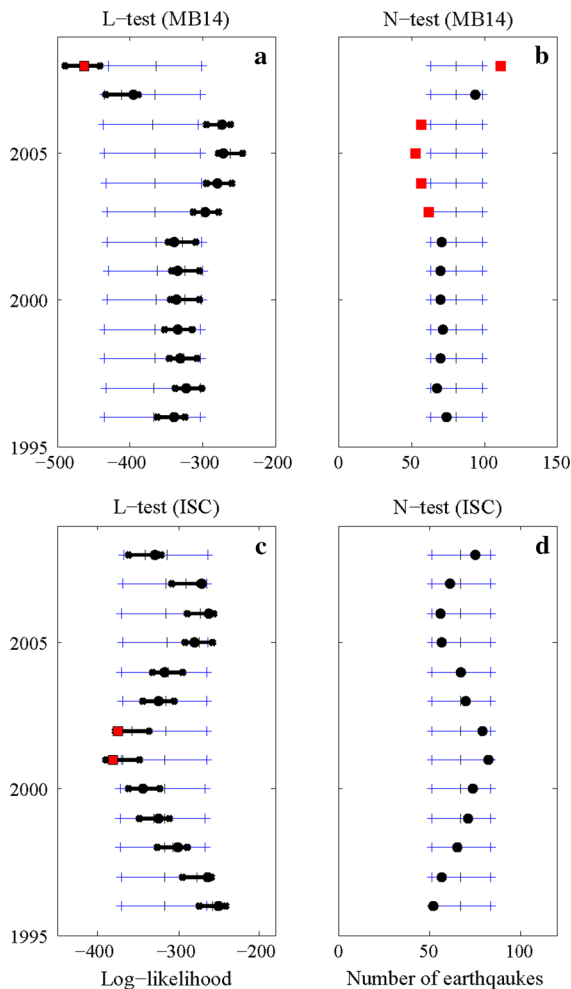


Figure 7

Results from the retrospective tests (unconditional L test, conditional L test, and N test) for a set of 5-year target periods, partially overlapping and shifted by 1 year, starting on 1996. **a, b** Display the results for the forecasts based on the MB14 catalog. **c, d** Depict the results for the forecasts based on the ISC catalog. The 95% confidence limits from the Poisson distribution of the corresponding results of the unconditional L test and N test are presented with *thin bars*, and the associated bounds for the expected conditional likelihood score are shown with *bold bars* (panels **a** and **c**). For each model, *symbols* denote the values of observed log-likelihoods and the number of observed events; a *black circle* is used for a passed forecast, while a *red square* is applied for a rejected forecast, i.e. when the model is inconsistent with the observations within the considered time period

The L test is a one-sided test and rejects a forecast when the observed log-likelihood locates at the lower limits of distribution of the simulated log-likelihoods. Accordingly, the models based on the MB14 and ISC

catalogs have been rejected 1 and 2 times among 13 periods, respectively (Fig. 7a, c).

The N test is a two-sided test and rejects a forecast if the number of earthquakes observed in the target period locates either in the lower or upper limits of the distribution of the number of events in the simulated catalogs. According to Fig. 7b, d, this test did not reject any of the forecasts related to the ISC catalog, whereas for the MB14 catalog the model results have been rejected 5 times.

Figure 7 shows that an increased number of target earthquakes results in a reduced log-likelihood score; in other words, they are correlated, with opposite trend. Since the expected likelihood score decreases when the expected number of events increases, and given that the L test is a one-sided test, a model that over-predicts the number of earthquakes might pass the L test trivially. A remedy for this issue is to perform the conditional L test (Werner et al. 2011), which conditions the likelihood range expected under the model to the number of observed earthquakes. As shown in Fig. 7, the corresponding confidence limits of the conditional L test, illustrated by bold bars, are narrower than those of the unconditional/original L test; still, the conditional L test does not reject any of the models during any of the periods.

We observed that in a number of periods the unconditional L test (hereafter, the L test) and the N test provided an opposite response, namely the L test did not reject the model, while the N test rejected it and vice versa. Only in one target period, i.e. 2008–2012, the model based on the MB14 catalog was rejected by both the tests. Compared with Werner et al. (2011), these tests seem to provide some additional information about the computed models. In other words, the N test cannot reject the models based on the ISC data, whereas the L test appears more stringent than the N test. On the contrary, the N test is more effective than the L test in discarding the models based on the MB14 data.

The same tests just described have been also performed after extending backward the duration of the learning catalogs, which increases the number of events and eventually decreases the statistical variability of results. For this purpose, we included in the datasets the events which occurred in the time interval from 1980 to 1996, as the occurrence rate

Table 2
 Summary results of the retrospective tests (*R* test, *N* test, and unconditional *L* test)

	<i>R</i> test				<i>N</i> test	<i>L</i> test
	<i>A</i> = MB14-m80	<i>A</i> = MB14-m96	<i>A</i> = ISC-m80	<i>A</i> = ISC-m96		
B = MB14-m80	–	4	7	7	5	1
B = MB14-m96	4	–	7	7	5	1
B = ISC-m80	3	3	–	5	2	3
B = ISC-m96	3	3	6	–	0	2

In each row, the numbers of the target periods, out of the 13 considered periods, in which the model B is rejected by the tests is reported. In the case of the *R* test, a null hypothesis that model A (in columns) is correct is assumed. The prefixes “-m80” and “-m96” indicate the models based on the longer and shorter catalogs, respectively

of events is approximately constant since ~ 1980 (Fig. 1). The results obtained based on the longer datasets show that the outcomes of the tests based on the extended MB14 catalog (Table 2; Fig. 8a, b) remain very similar to those from the shorter one, while the number of rejected forecasts increased for the longer ISC catalog (Table 2; Fig. 8c, d).

In order to quantitatively cross-compare the relative forecast skill of each of the models, which were built based on the longer and shorter catalogs, respectively, the *R* test has been used (Schorlemmer et al. 2007). Under a null hypothesis that model A is correct, the quantile score α_{AB} returns the portion of synthetic likelihood ratios, given the model A, that is less than the observed likelihood ratio for the forecasts A and B. Large values of α_{AB} support the model A over the model B. The same procedure leads to provide α_{BA} , assuming a null hypothesis that the model B is correct.

Table 2 shows the numbers of the target periods, out of the 13 considered ones, in which the models are rejected by the tests. For the *R* test, the results present the number of periods when the model B (in rows) should be rejected compared to the model A (in columns), using the critical significant level of 0.05. The models based on the longer and shorter catalogs are denoted by “-m80” and “-m96” prefixes, respectively, and the periods in which the preferred models have been rejected by the *L* test or *N* test have been excluded for the evaluations by the *R* test. The results show that there is not a strong evidence to make a conclusion about improving the predictive skill of the models by a longer learning period, as there are approximately the same number of periods

that corresponding models of the longer and shorter catalogs were not rejected. Comparing the corresponding forecasts obtained from the MB14 and ISC catalogs, the ISC-based models can be favored over the models based on the MB14 catalog.

However, the *L* test measures a combined score of rate and space components of a forecast, and the *N* test only considers impacts of the rate forecast (Zechar 2010). In order to focus on the space component of the forecasts, the results have been evaluated also by the error diagram analysis (Molchan 1991).

As this method tests basically alarm-based forecasts, firstly the water level method (Kossobokov 2006; Zechar and Jordan 2008) has been used to obtain alarm-based outputs of the models, assuming a uniform spatial density distribution of seismicity. In practice, having set some thresholds on the expectation E , for each threshold we calculated the space volume T of alarms by summing up the numbers of alarmed cells in which the expectation was higher than this threshold. Accordingly, the space rate of alarms is simply given by the number of alarmed cells versus the total number of analyzed cells. The miss rate ν was also computed as the number of target events that happened outside the alarmed cells.

Figure 9a, b presents the average error diagrams (the percentage of T versus the percentage of ν) of the forecasts based on the MB14 and ISC catalogs in two non-overlapping periods (2003–2007 and 2008–2012). The diagonal line displays the long-term trend of the efficiency of random predictions (i.e. the set of results which can be obtained by a random guess). The curve corresponding to a

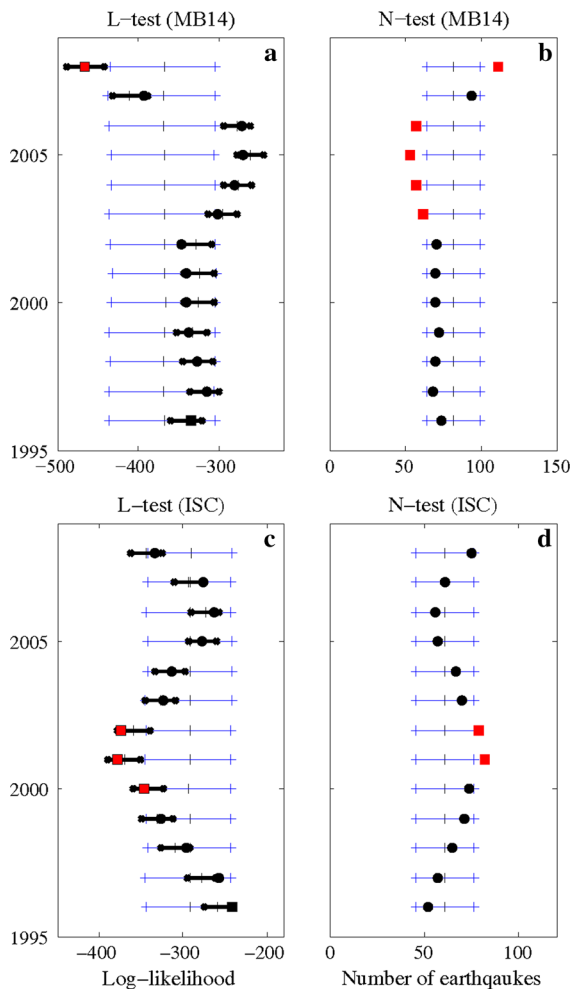


Figure 8

Results from retrospective tests (unconditional L test, conditional L test, and N test) for a set of 5-year target periods, partially overlapping and shifted by 1 year, starting on 1996. **a, b** Display the results for the forecasts based on the MB14 catalog (from 1980 onwards). **c, d** Depict the results for the forecasts based on the ISC catalog (from 1980 onwards). The 95% confidence limits from the Poisson distribution of the corresponding results of the unconditional L test and N test are presented with *thin bars*, and the associated bounds for the expected conditional likelihood score are shown with *bold bars* (panels **a** and **c**). For each model, *symbols* denote the values of observed log-likelihoods and the number of observed events; a *black circle* is used for a passed forecast, while a *red square* is applied for a rejected forecast, i.e. when the model is inconsistent with the observations within the considered time period

confidence level of 95% (Molchan 1991; Zechar and Jordan 2008) is also provided in the diagram to show the statistical significance of the results. Each confidence bound is calculated based on the corresponding

average number of target earthquakes occurred within the considered time periods. It is seen that the models cannot be rejected by this test, as the points do not exceed the 95% confidence bound.

However, the assumption of a uniform spatial density distribution of seismicity provides a simplistic reference model for the test. In order to supply a more conservative estimation of T , the empty cells, namely non-active cells with no $M \geq 4.5$ earthquakes, were excluded from the analysis (Kossobokov and Carlson 1995). Indeed, in this way we have made a reference model of past seismicity, which grossly permits to account for the non-uniform density distribution of earthquake epicenters. Accordingly, the results (black open circles in Fig. 9) indicate that the model cannot be rejected by this test, at least at 95% confidence level. Only part of the results, corresponding to large values of T , falls in the critical region of the diagram, between the curve of confidence bound and the diagonal line of random guessing. However, given the reference model of past seismicity, the efficiency of the forecasting model is evidently lower than that estimated considering a spatially uniform reference model of seismicity. This suggests that the forecasting model provides rather limited additional information, with respect to the location of past earthquakes, and might need more complexities to outperform stricter reference models.

In order to make a quantitative assessment of the results, we used two scalar quantities, namely the S -value ($S = T + v$) and the probability gain or predictive ratio value ($\text{Pr} = [1 - v]/T$) (Kagan and Jackson 2006; Kagan and Knopoff 1977; Kossobokov 2006). Higher S -values and lower Pr -values relate to lower forecast efficiency and vice versa. In the case of the reference model of a uniform spatial density distribution, the model based on the ISC data has the lowest S -value of 46% and the highest Pr -value of 14.6 (Fig. 10).

It is worth mentioning that while the forecasting model is compared to a spatially uniform model of seismicity, the difference between the maximum probability gain value based on Eq. 3 ($G = 2.6$ in Table 1) and that from error diagram ($\text{Pr} = 14.6$) arose due to the different definitions of the gain score by these two approaches. In fact, by setting thresholds

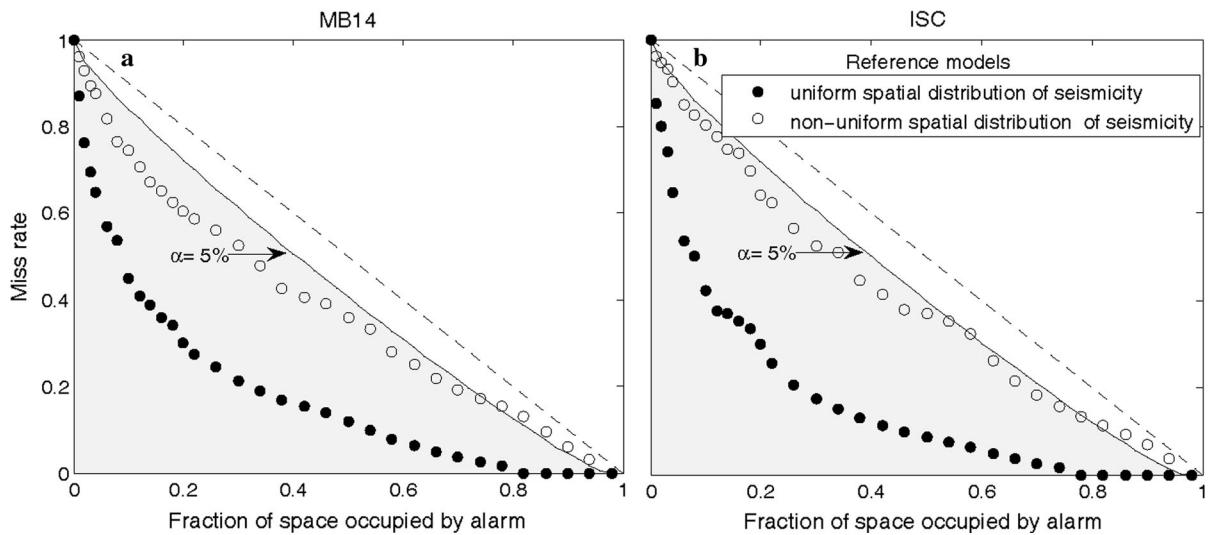


Figure 9

Average results of the error diagram test, as obtained from the models application in two non-overlapping periods (2003–2007 and 2008–2012), based on a reference model of a uniform spatial density distribution of seismicity (*black circles*) and a reference model of past seismicity (*open circles*). **a, b** Display the error diagrams obtained from the ISC and MB14 catalogs, respectively. The *dashed lines* show the long-term trend of the values corresponding to random guessing. The *filled areas* show acceptable scores at above 95% confidence level (below the confidence bound of $\alpha = 5\%$)

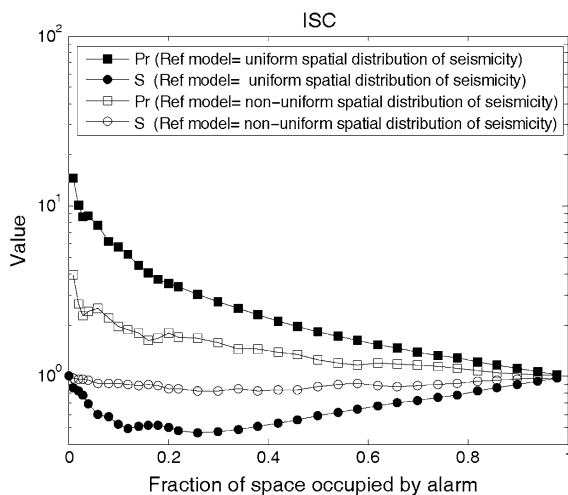


Figure 10

Illustration of the *S*-values (*circles*) and the predictive ratio values (*squares*) of the ISC forecasting model relative to a reference model (stated briefly in the legend as Ref model) of a uniform spatial density distribution of seismicity (*bold symbols*) and a reference model of past seismicity (*open symbols*). Higher *S*-values and lower Pr-values relate to lower forecast efficiency and vice versa

on the expectation *E*, we eventually reduce the space volume *T* of alarms. When the parameter *T* is close to zero, the probability gain defined by error diagram

(Pr) accounts only for the cells with high rates and gives its maximum value. In this point of view, however, the differential probability gain defined by the likelihood approach (*G*) gives the same weight to all the cells, i.e. it accounts for all the cells with high and low rates; in this condition, positive scores at high rates and negative scores at low rates cancel each other out (Shebalin et al. 2014). Although it seems that the values of the probability gain obtained by these different techniques may not be compared directly, their difference may show that good forecasts for low seismicity rates are not as important as for high rates.

When considering the second reference model, i.e. the model of past seismicity, the results sum up to $S = 82\%$ and $Pr = 3.9$. In terms of forecast strategy (Molchan 2003), $S = 46\%$ and $Pr = 14.6$ can be interpreted as a rather effective strategy. On the contrary, when the forecasting model is compared with a reference model of past seismicity, it turns out to be a strategy with a rather low predictive ratio $Pr = 3.9$, although the achieved statics passed the test at 95% confidence level for most of the values of *T*. Anyway as mentioned before, due to a fundamental property of probability gain scoring (Molchan 2003),

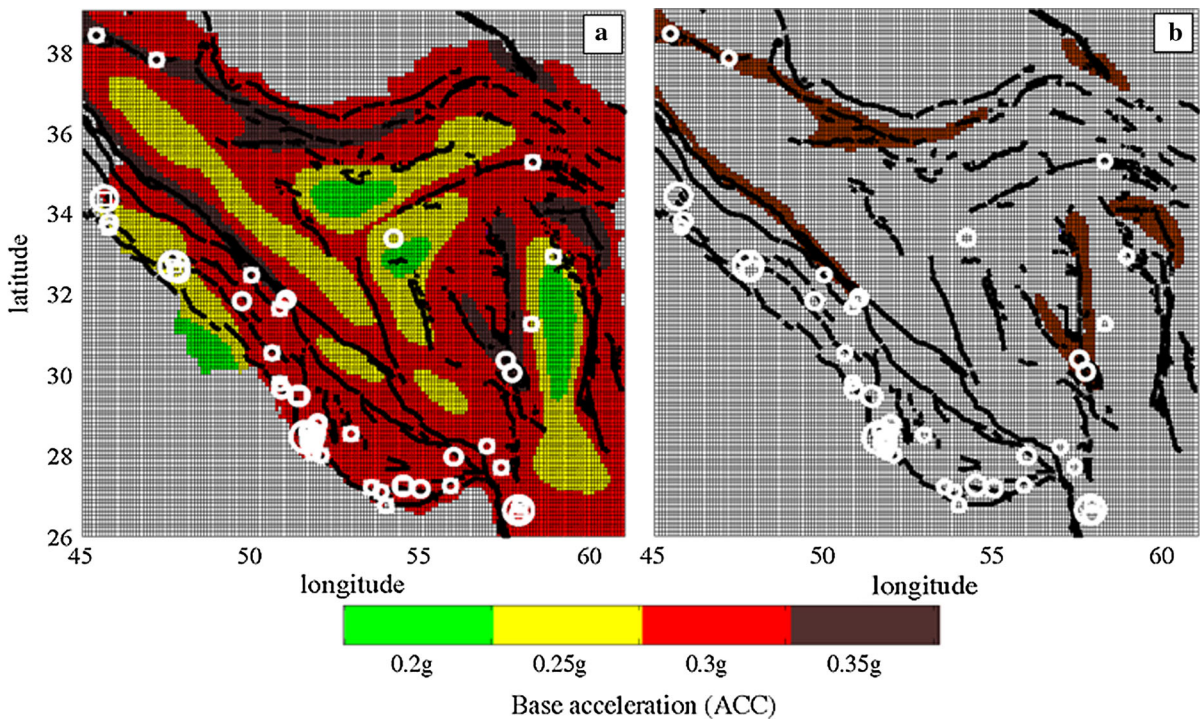


Figure 11

a The national probabilistic seismic hazard map (PSHA) of Iran, which is discretized by $0.1^\circ \times 0.1^\circ$ cells. It has been categorized into low, moderate, high, and very high seismic hazardous zones, based on the base accelerations (ACC), i.e. 0.2, 0.25, 0.3, and 0.35 g, respectively. **b** Same map, showing only cells with very high hazard (i.e. above 90th percentile). *White symbols* are $M \geq 5.0$ declustered events that happened from 1 January 2013 to 2016, as reported in the available ISC catalog, and the *solid lines* show major faults of the region

the maximum gain is reached when the T tends to zero; as a result, a single success obtained by randomly declaring few sporadic alarms may lead to very high probability gain estimates.

The forecasting model based on the MB14 catalog provided scores quite similar to those from the ISC data; thus the above-mentioned considerations are pretty general and do not depend on the specific catalog.

6.2. Comparison Between the Model and the National Seismic Hazard Map of Iran

The forecasting model developed in this study aims to provide some time-independent information about the probability of earthquake occurrence in the region. According to this model, the areas where the expected rate/probability is relatively high might be identified as the most hazardous areas. In this section, we wish to explore to what extent the hazard

estimates provided by the forecasting map, based on smoothed seismicity, differ from the national probabilistic seismic hazard map (PSHA) of Iran, Fig. 11a, which is provided by permanent committee for revising the Iranian code of practice for seismic resistant design of building (www.bhrc.ac.ir).

Although, the basic assumption is Poissonian earthquake occurrence in both the forecast model and the PSHA estimates, the quantities to be compared are necessarily different. The PSHA seismic hazard map provides a physically different description of the hazard, namely the expected Peak Ground Acceleration (PGA) that accounts for both seismicity rates and seismic waves propagation. The forecasting map based on smoothed seismicity model, instead, accounts for earthquake occurrence rate only. Another basic difference is the time window of the forecast: in fact, the forecast model estimates the rates over a 5-year window, whereas the hazard map refers to a longer time window. Therefore, in order to

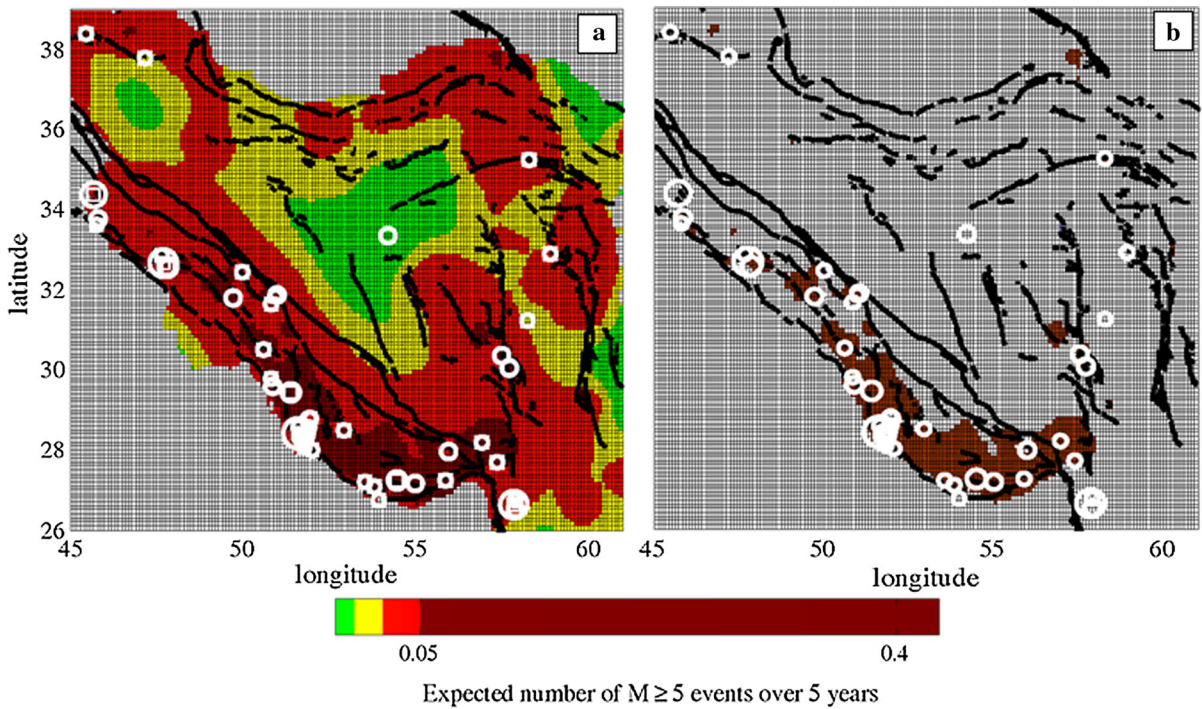


Figure 12

a Expected seismicity rate of the target $M \geq 5.0$ earthquakes over a 5-year forecast horizon from 2013 to 2017, based on Fig. 5, which is discretized by $0.1^\circ \times 0.1^\circ$ cells. It has been categorized into low/moderate, high, and very high seismic hazardous zones, based on the quantiles 0.05, 0.26, and 0.9 of the rate values. **b** Same map, showing only cells with very high hazard (i.e. above 90th percentile). *White symbols* are $M \geq 5.0$ declustered events that happened from 1 January 2013 to 2016, as reported in the available ISC catalog, and the *solid lines* show major faults of the region

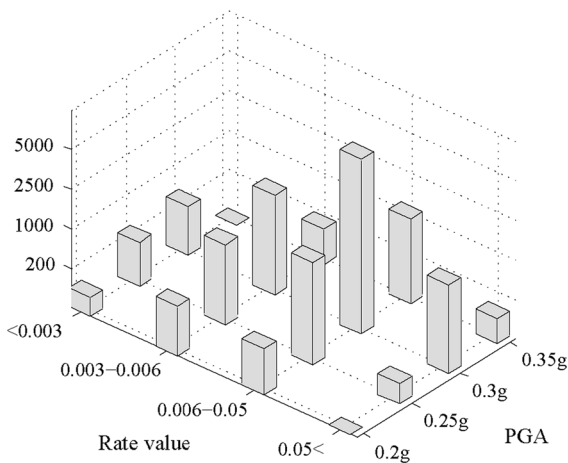


Figure 13

The bivariate histogram of the values as a function of PGA from the hazard map and rate from the forecast map. The quantiles 0.05, 0.26, and 0.9 of the rate values correspond to ~ 0.003 , ~ 0.006 , and ~ 0.05 , respectively

assess how well these substantially different hazard maps correlate each other, we need to use robust techniques (e.g., non-parametric statistics), applied to coarsely discretized values.

The PSHA map has been categorized into the zones of 4 types, according to the base accelerations (ACC). Accordingly, the zones with $ACC = 0.2, 0.25, 0.3,$ and 0.35 g have been interpreted as low, moderate, high, and very high seismic hazard zones, respectively. These acceleration thresholds correspond to the following quantiles: 0.05 (5th percentile), 0.26 (26th percentile) and 0.9 (90th percentile); that is, 5% of the territory is assigned low hazard, 21% moderate hazard, 64% high hazard, and 10% very high hazard.

For the purpose of comparison, the forecast results, Fig. 12a, have been discretized using for the rate values the same quantiles associated with PSHA

acceleration map, namely 0.05, 0.26, and 0.9. These quantiles were considered to categorize the studied area into zones with low, moderate, high, and very high probability of earthquake occurrence, respectively. In this way, we aim to address the following question: given that the same portion of the territory is assigned a given hazard category (i.e. low, moderate, high or very high), how it compares the classification map based on forecasts with that from PSHA?

A visual comparison of the two maps, carried out using the same color palette for the two maps of categorized hazard and forecast estimates, permits to see how well these values correlate each other. It can be observed in Fig. 11b that the PSHA map assigns higher hazard values in northern and eastern parts of Iran territory (e.g., Alborz region), with the highest values following the major faults. In the forecast map, instead, very high hazard (Fig. 12b) is assigned to the southern part of territory (i.e. Zagros region), an area characterized by higher level of seismic activity.

Figure 13 shows the bivariate histogram of the values as a function of PGA from the hazard map and the rate from the forecast map. The cells with larger forecast rate are mostly related with rather high PGA, i.e. 0.3 or 0.35 g. The correlation of the values provided at different sites by the two maps is positive (e.g., Spearman rank correlation coefficient ~ 0.25) and significant (p value < 0.01), but not strong. The rate from forecast appears very low in most of the territory, but that is natural within 5 years. Although high forecast rate is generally associated with areas of high seismic hazard (i.e. high ACC), there are a number of cells where a low forecast rate is assigned to areas of high expected ground shaking (e.g., Alborz region). Thus, we observed that time-independent forecasting maps based on smoothed seismicity provide a different picture of the seismic hazard, compared to PSHA. In fact, forecasting maps are intended to supply information about moderate-to-large earthquake occurrence at the intermediate-term scale (5 years), and thus may highlight areas where earthquakes are more likely to occur within a relatively short time span. On the other side, these maps may not be able to capture hazardous areas where the rate of seismic activity is low, but sporadic strong earthquakes are still possible.

7. Conclusions

In this study, a modified variant of the time-independent earthquake forecasting model by Helmstetter et al. (2007) is applied to calculate the probabilities of forthcoming $M \geq 5.0$ earthquakes within Iran, based on $M \geq 4.5$ events since 1996.

The results show that the outputs of the two used catalogs (the ISC and MB14 catalogs) are generally similar, and the most hazardous area for moderate earthquakes is the Zagros region. Moreover, the ISC catalog has a tendency to provide higher gains than the MB14 catalog, during both the model optimization and the retrospective tests, whose comparison can provide an idea about unavoidable uncertainties related with the input data. Accordingly and as discussed by Maybodan et al. (2014), the ISC catalog can serve well for developing forecast techniques in Iran. The experiments performed as part of the optimization process have shown that choosing the declustered catalogs with the lowest magnitude threshold can provide higher gains than other possible choices of the learning catalogs. We also suggested that a cell size of at least $0.2^\circ \times 0.2^\circ$ would be preferable for model application in the studied region.

Compared with the results of Werner et al. (2010a)_ENREF_33_ENREF_34 for Italy, the stability of the results discussed in Sect. 5, suggests the effects of seismotectonic conditions on an earthquake forecast that can manifest itself in the number of target events, which are on average ~ 74 and ~ 8 earthquakes per 5 years for Iran and Italy, respectively.

The L test has rejected the MB14 and ISC forecasts 1 and 2 times, respectively, out of the 13 considered periods. The N test has passed the model based on the ISC catalog in all the periods, while rejecting the MB14 forecasts 5 times. In the case of the ISC dataset, it is implied that the Poisson distribution can suitably describe the number of target events in the forecast horizon. In total, the ISC and MB14 forecasts have been rejected for ~ 15 and $\sim 38\%$ of times, respectively. The results also show that there is a negative correlation between the number of target events and the observed log-likelihoods.

After extending backward the duration of the learning catalogs (i.e. since 1980), the L test, N test, and R test have also been used to study a link between the durations of the datasets, their quality, and the predictive power of the model. Accordingly, it turns out that a lower-quality but longer catalog does not essentially provide a better forecast performance, than a shorter better-quality one. The general considerations about performances of the forecasting model, however, might be affected by some evidenced shortcomings of the standard testing methods applied for the analysis (e.g., Molchan 2012; Rhoades et al. 2011). Being aware of those limits, an additional testing scheme has been applied, based on the error diagram (Molchan 1991).

By the error diagram test, the forecasting models have been compared with two reference models. The first was the simple reference model of a uniform spatial distribution, which led to conclude that the forecasting models of smoothed seismicity may provide a rather good forecast strategy. The second reference model was a model of past seismicity. In this case, the forecasting models turn out to be a less informative forecast strategy. Still the achieved statistics could not be rejected at 95% confidence level through most parts of space volume of alarms; hence, forecasts seem to provide some statistically significant information, in addition to simple space distribution of past seismicity.

Overall, the two testing approaches considered here provide different assessments of the model results and account for the forecast rate and space. The retrospective tests illustrate that the model assumptions provide some limited, but statistically acceptable performances for the studied area.

Accordingly, the considered model may provide useful information in combination with long-term indications from seismic hazard maps. In fact, comparing the forecast rate map for target $M \geq 5.0$ earthquakes and the PSHA map available for the territory of Iran, as shown in Fig. 11, we may observe that the two maps turn out quite different, although some statistically significant correlation can be detected. Quite naturally, sites with large forecast rates are related with rather high values of acceleration (i.e. ACC 0.3 or 0.35 g). While there are still a huge number of sites where the hazard takes large

values, the forecast rate for a 5-year forecast horizon is very small. Thus we may conclude that, although high forecast rate is generally associated with seismically active areas, characterized by the presence of major faults and large earthquake occurrence (i.e. high expected ground base acceleration), the probability of their occurrence within a short time window keeps very low in most of the areas, given the assumption of Poissonian earthquake occurrence.

Thus, it is proposed that such kind of studies could be useful for recognizing regions most prone to seismic activation and high seismic hazard in Iran, specifically areas characterized by a high level of predicted ground shaking and high forecast rates. Still, areas where predicted rate is low, but ground shaking is potentially high, must be considered with caution: in fact, it cannot be excluded that sporadic (very low-rate) strong earthquakes may occur at any time, and this aspect should be factored in the decision-making process.

Acknowledgements

We would like to thank numerous colleagues, namely, Elham Malek-Mohammadi, Ehsan Noroozinejad, Meysam Mahmud-Abadi, and Mehdi Ahmadi-Borji for sharing their points of view on the manuscript. The authors would also like to acknowledge International Institute of Earthquake Engineering and Seismology (IIEES) for its help in providing research documents and methodological aspects of the job.

REFERENCES

- Bird, P., & Kagan, Y. Y. (2004). Plate-tectonic analysis of shallow seismicity: Apparent boundary width, beta, corner magnitude, coupled lithosphere thickness, and coupling in seven tectonic settings. *Bulletin of the Seismological Society of America*, 94(6), 2380–2399.
- Engdahl, E. R., Jackson, J. A., Myers, S. C., Bergman, E. A., & Priestley, K. (2006). Relocation and assessment of seismicity in the Iran region. *Geophysical Journal International*, 167(2), 761–778.
- Field, E. H., Arrowsmith, R. J., Biasi, G. P., Bird, P., Dawson, T. E., Felzer, K. R., et al. (2014). Uniform California Earthquake Rupture Forecast, Version 3 (UCERF3): The time-independent model. *Bulletin of the Seismological Society of America*, 104(3), 1122–1180.

- Helmstetter, A., Kagan, Y. Y., & Jackson, D. D. (2007). High-resolution time-independent grid-based forecast for $M \geq 5$ earthquakes in California. *Seismological Research Letters*, 78(1), 78–86.
- Helmstetter, A., & Werner, M. J. (2012). Adaptive spatiotemporal smoothing of seismicity for long-term earthquake forecasts in California. *Bulletin of the Seismological Society of America*, 102(6), 2518–2529.
- Jordan, T. H., Marzocchi, W., Michael, A. J., & Gerstenberger, M. C. (2014). Operational earthquake forecasting can enhance earthquake preparedness. *Seismological Research Letters*, 85(5), 955–959.
- Kagan, Y. Y., & Jackson, D. D. (1994). Long-term probabilistic forecasting of earthquakes. *Journal of Geophysical Research*, 99, 13685–13700.
- Kagan, Y. Y., & Jackson, D. D. (2000). Probabilistic forecasting of earthquakes. *Geophysical Journal International*, 143(2), 438–453.
- Kagan, Y. Y., & Jackson, D. D. (2006). Comment on ‘Testing earthquake prediction methods: “The West Pacific short-term forecast of earthquakes with magnitude M_w HRV ≥ 5.8 ” by VG Kossobokov. *Tectonophysics*, 413(1), 33–38.
- Kagan, Y. Y., & Jackson, D. D. (2010). Short-and long-term earthquake forecasts for California and Nevada. *Pure and Applied Geophysics*, 167(6–7), 685–692.
- Kagan, Y. Y., & Knopoff, L. (1977). Earthquake risk prediction as a stochastic process. *Physics of the Earth and Planetary Interiors*, 14(2), 97–108.
- Karimiparidari, S., Zaré, M., Memarian, H., & Kijko, A. (2013). Iranian earthquakes, a uniform catalog with moment magnitudes. *Journal of Seismology*, 17(3), 897–911.
- Kossobokov, V. G. (2006). Testing earthquake prediction methods: « The West Pacific short-term forecast of earthquakes with magnitude M_w HRV ≥ 5.8 ». *Tectonophysics*, 413(1), 25–31.
- Kossobokov, V. G., & Carlson, J. M. (1995). Active zone size versus activity: A study of different seismicity patterns in the context of the prediction algorithm M8. *Journal of Geophysical Research*, 100(B4), 6431–6441.
- Kossobokov, V. G., Peresan, A., & Panza, G. F. (2015). On operational earthquake forecast and prediction problems. *Seismological Research Letters*, 86(2A), 287–290.
- Maybodan, M., Zare, M., Hamzehloo, H., Peresan, A., Ansari, A., & Panza, G. F. (2014). Analysis of precursory seismicity patterns in Zagros (Iran) by CN algorithm. *Turkish Journal of Earth Sciences*, 23(1), 91–99.
- Molchan, G. M. (1991). Structure of optimal strategies in earthquake prediction. *Tectonophysics*, 193(4), 267–276.
- Molchan, G. M. (2003). Earthquake prediction strategies: a theoretical analysis. In: *Nonlinear dynamics of the lithosphere and earthquake prediction* (pp. 209–237), Springer, New York
- Molchan, G. M. (2012). On the testing of seismicity models. *Acta Geophysica*, 60(3), 624–637.
- Molchan, G. M., & Kagan, Y. Y. (1992). Earthquake prediction and its optimization. *Journal of Geophysical Research*, 97(B4), 4823–4838.
- Mousavi-Bafrouei, S. H., Mirzaei, N., & Shabani, E. (2014). A declustered earthquake catalog for the Iranian Plateau. *Annals of Geophysics*, 57(6), S0653–1–25.
- Panza, G. F., La Mura, C., Peresan, A., Romanelli, F., & Vaccari, F. (2012). Chapter three-seismic hazard scenarios as preventive tools for a disaster resilient society. *Advances in Geophysics*, 53, 93–165.
- Peresan, A., Kossobokov, V. G., & Panza, G. F. (2012). Operational earthquake forecast/prediction. *Rendiconti Lincei*, 23(2), 131–138.
- Peresan, A., Panza, G. F., & Costa, G. (2000). CN algorithm and long-lasting changes in reported magnitudes: The case of Italy. *Geophysical Journal International*, 141(2), 425–437.
- Peresan, A., Rotwain, I., Zaliapin, I., & Panza, G. F. (2002). Stability of intermediate-term earthquake predictions with respect to random errors in magnitude: The case of central Italy. *Physics of the Earth and Planetary Interiors*, 130(1), 117–127.
- Peresan, A., Zuccolo, E., Vaccari, F., Gorshkov, A., & Panza, G. F. (2011). Neo-deterministic seismic hazard and pattern recognition techniques: Time-dependent scenarios for North-Eastern Italy. *Pure and Applied Geophysics*, 168(3–4), 583–607.
- Radan, M. Y., Hamzehloo, H., Peresan, A., Zare, M., & Zafarani, H. (2013). Assessing performances of pattern informatics method: A retrospective analysis for Iran and Italy. *Natural Hazards*, 68(2), 855–881.
- Reasenber, P. (1985). Second-order moment of central California seismicity, 1969–1982. *Journal of Geophysical Research*, 90(B7), 5479–5495.
- Rhoades, D. A., Schorlemmer, D., Gerstenberger, M. C., Christophersen, A., Zechar, J. D., & Imoto, M. (2011). Efficient testing of earthquake forecasting models. *Acta Geophysica*, 59(4), 728–747.
- Richards-Dinger, K. B., & Shearer, P. M. (2000). Earthquake locations in southern California obtained using source-specific station terms. *Journal of Geophysical Research*, 105(B5), 10939–10960.
- Schorlemmer, D., & Gerstenberger, M. C. (2007). RELM testing center. *Seismological Research Letters*, 78(1), 30–36.
- Schorlemmer, D., Gerstenberger, M. C., Wiemer, S., Jackson, D. D., & Rhoades, D. A. (2007). Earthquake likelihood model testing. *Seismological Research Letters*, 78(1), 17–29.
- Schorlemmer, D., Zechar, J. D., Werner, M. J., Field, E. H., Jackson, D. D., Jordan, T. H., et al. (2010). First results of the regional earthquake likelihood models experiment. *Pure and Applied Geophysics*, 167(8–9), 859–876.
- Shahvar, M. P., Zare, M., & Castellaro, S. (2013). A unified seismic catalog for the Iranian plateau (1900–2011). *Seismological Research Letters*, 84(2), 233–249.
- Shebalin, P. N., Narteau, C., Zechar, J. D., & Holschneider, M. (2014). Combining earthquake forecasts using differential probability gains. *Earth, Planets and Space*, 66(1), 1–14.
- Talebi, M., Zare, M., Madahi-Zadeh, R., & Bali-Lashak, A. (2015). Spatial-temporal analysis of seismicity before the 2012 Varzeghan, Iran, M_w 6.5 earthquake. *Turkish Journal of Earth Sciences*, 24(3), 289–301.
- Werner, M. J., Helmstetter, A., Jackson, D. D., & Kagan, Y. Y. (2011). High-resolution long-term and short-term earthquake forecasts for California. *Bulletin of the Seismological Society of America*, 101(4), 1630–1648.
- Werner, M. J., Helmstetter, A., Jackson, D. D., Kagan, Y. Y., & Wiemer, S. (2010a). Adaptively smoothed seismicity earthquake forecasts for Italy. [arXiv:1003.4374](https://arxiv.org/abs/1003.4374).
- Werner, M. J., Zechar, J. D., Marzocchi, W., Wiemer, S., & Group, C.-I. W. (2010b). Retrospective evaluation of the five-year and ten-year CSEP-Italy earthquake forecasts. *Annals of Geophysics*, 53(3), 11–30

- Wiemer, S. (2001). A software package to analyze seismicity: ZMAP. *Seismological Research Letters*, 72(3), 373–382.
- Wiemer, S., & Wyss, M. (2000). Minimum magnitude of completeness in earthquake catalogs: Examples from Alaska, the western United States, and Japan. *Bulletin of the Seismological Society of America*, 90(4), 859–869.
- Zare, M., Amini, H., Yazdi, P., Sesetyan, K., Demircioglu, M. B., Kalafat, D., et al. (2014). Recent developments of the Middle East catalog. *Journal of Seismology*, 18(4), 749–772.
- Zechar, J. D. (2010). Evaluating earthquake predictions and earthquake forecasts: A guide for students and new researchers. *Community Online Resource for Statistical Seismicity Analysis*, 1–26.
- Zechar, J. D., Gerstenberger, M. C., & Rhoades, D. A. (2010a). Likelihood-based tests for evaluating space-rate-magnitude earthquake forecasts. *Bulletin of the Seismological Society of America*, 100(3), 1184–1195.
- Zechar, J. D., & Jordan, T. H. (2008). Testing alarm-based earthquake predictions. *Geophysical Journal International*, 172(2), 715–724.
- Zechar, J. D., & Jordan, T. H. (2010). Simple smoothed seismicity earthquake forecasts for Italy. *Annals of Geophysics*, 53(3), 99–105.
- Zechar, J. D., Schorlemmer, D., Liukis, M., Yu, J., Euchner, F., Maechling, P. J., et al. (2010b). The collaboratory for the study of earthquake predictability perspective on computational earthquake science. *Concurrency and Computation: Practice and Experience*, 22(12), 1836–1847.
- Zechar, J. D., Schorlemmer, D., Werner, M. J., Gerstenberger, M. C., Rhoades, D. A., & Jordan, T. H. (2013). Regional earthquake likelihood models I: First-order results. *Bulletin of the Seismological Society of America*, 103(2A), 787–798.

(Received July 2, 2016, revised January 30, 2017, accepted March 4, 2017, Published online March 11, 2017)