Geophysical Journal International

Geophys. J. Int. (2023) **235,** 1541–1551
Advance Access publication 2023 August 04
GJI Seismology

https://doi.org/10.1093/gji/ggad312

Comparison between alarm-based and probability-based earthquake forecasting methods

Emanuele Biondini [1] and Paolo Gasperini [1,2]

[1] *Dipartimento di Fisica e Astronomia, Università di Bologna,* 40127 *Bologna, Italy. E-mail: emanuele.biondini2@unibo.it*
[2] *Istituto Nazionale di Geofisica e Vulcanologia, Sezione di Bologna,* 40127 *Bologna BO, Italy*

Accepted 2023 August 3. Received 2023 July 27; in original form 2023 May 5

SUMMARY

In a recent work, we applied the every earthquake a precursor according to scale (EEPAS) probabilistic model to the pseudo-prospective forecasting of shallow earthquakes with magnitude M 5.0 in the Italian region. We compared the forecasting performance of EEPAS with that of the epidemic type aftershock sequences (ETAS) forecasting model, using the most recent consistency tests developed within the collaboratory for the study of earthquake predictability (CSEP). The application of such models for the forecasting of Italian target earthquakes seems to show peculiar characteristics for each of them. In particular, the ETAS model showed higher performance for short-term forecasting, in contrast, the EEPAS model showed higher forecasting performance for the medium/long-term. In this work, we compare the performance of EEPAS and ETAS models with that obtained by a deterministic model based on the occurrence of strong foreshocks (FORE model) using an alarm-based approach. We apply the two rate-based models (ETAS and EEPAS) estimating the best probability threshold above which we issue an alarm. The model parameters and probability thresholds for issuing the alarms are calibrated on a learning data set from 1990 to 2011 during which 27 target earthquakes have occurred within the analysis region. The pseudo-prospective forecasting performance is assessed on a validation data set from 2012 to 2021, which also comprises 27 target earthquakes. Tests to assess the forecasting capability demonstrate that, even if all models outperform a purely random method, which trivially forecast earthquake proportionally to the space–time occupied by alarms, the EEPAS model exhibits lower forecasting performance than ETAS and FORE models. In addition, the relative performance comparison of the three models demonstrates that the forecasting capability of the FORE model appears slightly better than ETAS, but the difference is not statistically significant as it remains within the uncertainty level. However, truly prospective tests are necessary to validate such results, ideally using new testing procedures allowing the analysis of alarm-based models, not yet available within the CSEP.

Key words: Computational seismology; Earthquake interaction, forecasting and prediction; Statistical seismology.

INTRODUCTION

The occurrence of earthquakes in regions, cities and densely populated areas poses enormous risks that can threaten national prosperity and social welfare (Jordan 2009). Quantifying urban earthquake risk is a complex problem that requires detailed knowledge of the environment, infrastructure and understanding of earthquake phenomena, as well as human behaviour in hazardous situations (MacPherson-Krutsky *et al.* 2023). Empirical models based on earthquake clustering can provide methods for forecasting and predicting earthquakes that can be useful to the decision-makers to prepare seismic risk mitigation operations (Marzocchi & Lombardi 2009; Azarbakht *et al.* 2021).

We follow Jordan *et al.* (2011) to distinguish between a probabilistic forecast and a deterministic prediction using strict definitions: the former is the indication of the probability that one or more target event will occur in a given space–time domain, whereas the latter is the binary assertion that target events will occur or not in a given space–time alarm window.

Within the ambit of the collaboratory for the study of earthquake predictability (CSEP, Jordan 2006; Zechar *et al.* 2010b; Michael &

Werner 2018; Schorlemmer *et al.* 2018), to evaluate the forecasting performance of probabilistic models, a suite of statistical test has been developed (Schorlemmer *et al.* 2007; Werner *et al.* 2010; Bayona *et al.* 2022), some of which are included in the pyCSEP package (Savran *et al.* 2022a,b).

Tests appropriate for deterministic models were developed by Molchan (1990, 1991), Zechar & Jordan (2008, 2010) and Shebalin *et al.* (2011), but they have not yet integrated in the pyCSEP package. The two approaches are not directly compatible and hence a problem arises for comparing the performance of probabilistic versus deterministic models.

The implementation of the probabilistic approach, to assess the forecasting performance of deterministic models (for example using the CSEP tests by Savran *et al.* 2022a,b) is not trivial because it can be difficult to associate a rate of occurrence of target earthquakes with some kind of potentially precursory phenomenon. Conversely, probabilistic models can be easily converted to the deterministic approach by simply choosing a threshold rate above which issuing a binary alarm. A similar approach had been implemented by Murru *et al.* (2009) and Console *et al.* (2010) to evaluate probabilistic models applied to Italian and New Zealand seismicity respectively. Thus, to compare the forecasting performance of probabilistic and deterministic methods, the use of tests and procedure developed for deterministic methods is usually preferable.

In this work, the probabilistic forecasting models described in Biondini *et al.* (2023) are compared with a deterministic forecasting method based on the occurrence of strong (fore) shocks (hereafter FORE method) described by Gasperini *et al.* (2021). The latter sets an alarm of duration Δt every time an earthquake of magnitude $4.4 \leq M_w \leq 4.8$ ($M_w = 4.6 \pm 0.2$) occurs in a given circle of radius $R = 30$ km. The results obtained retrospectively by Gasperini *et al.* (2021) show that approximately 60 and 70 per cent of earthquakes with magnitude $M \geq 5.0$ that occurred between 1960 and 2020 were anticipated by foreshocks in the three months and one year respectively preceding them with space–time coverages of alarms less than 1 per cent. Such percentages increase by extending the alarm window, highlighting how the occurrence of strong shocks can contribute to improving stronger earthquake prediction. The foreshocks as precursors of strong earthquakes were previously studied by Jones (1984, 1985, 1994) and Agnew & Jones (1991) in California.

The probabilistic models described by Biondini *et al.* (2023) are the every earthquake a precursor according to scale (EEPAS) model (Rhoades & Evison 2004) and the epidemic type aftershock sequence (ETAS) model (Ogata 1988). The EEPAS model is based on the increase in rate and in magnitude of the minor seismicity observed before the occurrence of major earthquakes (ψ-phenomenon) and each earthquake is assumed to contribute to the transient increment of the rate density of future seismicity (Rhoades & Evison 2004). The ETAS model considers each aftershock as capable of triggering subsequent aftershocks and the occurrence rate for a certain tth instant of time is given by the superposition, weighted by the magnitude of the parent event, of the decays described by the time-shifted Omori–Utsu decay function (Ogata & Zhuang 2006).

Their implementations to various regions of the world (California, New Zealand, Italy and Japan) where reliable earthquake catalogues are available are widely described in the seismological literature of the last two decades (e.g. Console *et al.* 2006; Rhoades 2007, 2011; Falcone *et al.* 2010; Lombardi & Marzocchi 2010; Mizrahi *et al.* 2023).

SETTING UP THE DETERMINISTIC EXPERIMENT

The probabilistic models EEPAS and ETAS are fitted by Biondini *et al.* (2023) to forecast earthquakes with magnitude $M \geq 5.0$ in the Italian region, over a spatial grid consisting of 177 non-overlapped square cells of side $L = 30\sqrt{2}$ km showed in Fig. 1. Such grid is assumed as application region R even for the present forecasting experiment. Even the FORE method, based on the occurrence of potential foreshocks, was adapted to this region R of analysis. According to Biondini *et al.* (2023) such region only considers cells where a $M_w \geq 4$ earthquake occurred inland in the last four centuries. This is to avoid an overestimation of the performance of all forecasting methods caused by the inclusion of almost completely aseismic areas (like for example the Sardinia Island). To implement the alarm-based approach, the analysis and optimization of the rate thresholds for the probabilistic models and of the foreshock magnitude range for the FORE method was conducted for the period 1990–2011, while a pseudo-prospective alarm-based forecasting experiment is conducted for the period 2012–2021. Such periods are the same as those used for the learning and the pseudo-prospective application, respectively of the EEPAS and ETAS models by Biondini *et al.* (2023).

Both the optimization and the pseudo-prospective forecasting experiment are conducted using the seismic data of the HORUS seismic catalogue (Lolli *et al.* 2020). As the two formulations of EEPAS (EEPAS-NW and EEPAS-W) and of the ETAS (ETAS-SUP and ETAS-SVP) described by Biondini *et al.* (2023) produced quite similar results and performances, for this experiment we only consider the EEPAS-NW and ETAS-SVP (now on simply EEPAS and ETAS) models, which are characterized by slightly better forecasting performances than the EEPAS-W and ETAS-SUP, respectively. A summary of the characteristics of each model analysed in this work is given in Table 1.

SETTING UP FORECAST ASSUMPTIONS AND TESTING PROCEDURES

The expected daily rate for the EEPAS and ETAS models is re-evaluated for each cell, every time an earthquake of magnitude $m \geq 2.5$ (depth < 40 km) occurs within the analysis polygon R (Fig. 1) according to the HORUS seismic catalogue. An alarm with duration Δt, is issued in a given cell, every time the expected rate estimated by such models exceeds some threshold value. Similarly, for the FORE method, an alarm is issued whenever a strong shocks of magnitude within $M \pm \Delta M$ occurs.

A target earthquake ($M \geq 5.0$) is considered successfully predicted if it occurs within one or more alarm windows of duration Δt. On the contrary, it is considered as a failure to predict if it occurs outside any active alarm window. According to Molchan (1990, 1991), the miss rate ν, that is the fraction of unpredicted earthquakes, is computed as

$$\nu = \frac{N - h}{N}, \tag{1}$$

where h is the number of target events successfully forecasted and N is the total number of target events.

As well, the overall space–time duration of the alarms is computed as

$$\tau = \frac{1}{K} \sum_{i=1}^{K} \left(\frac{d_{ci}}{T} \right), \tag{2}$$

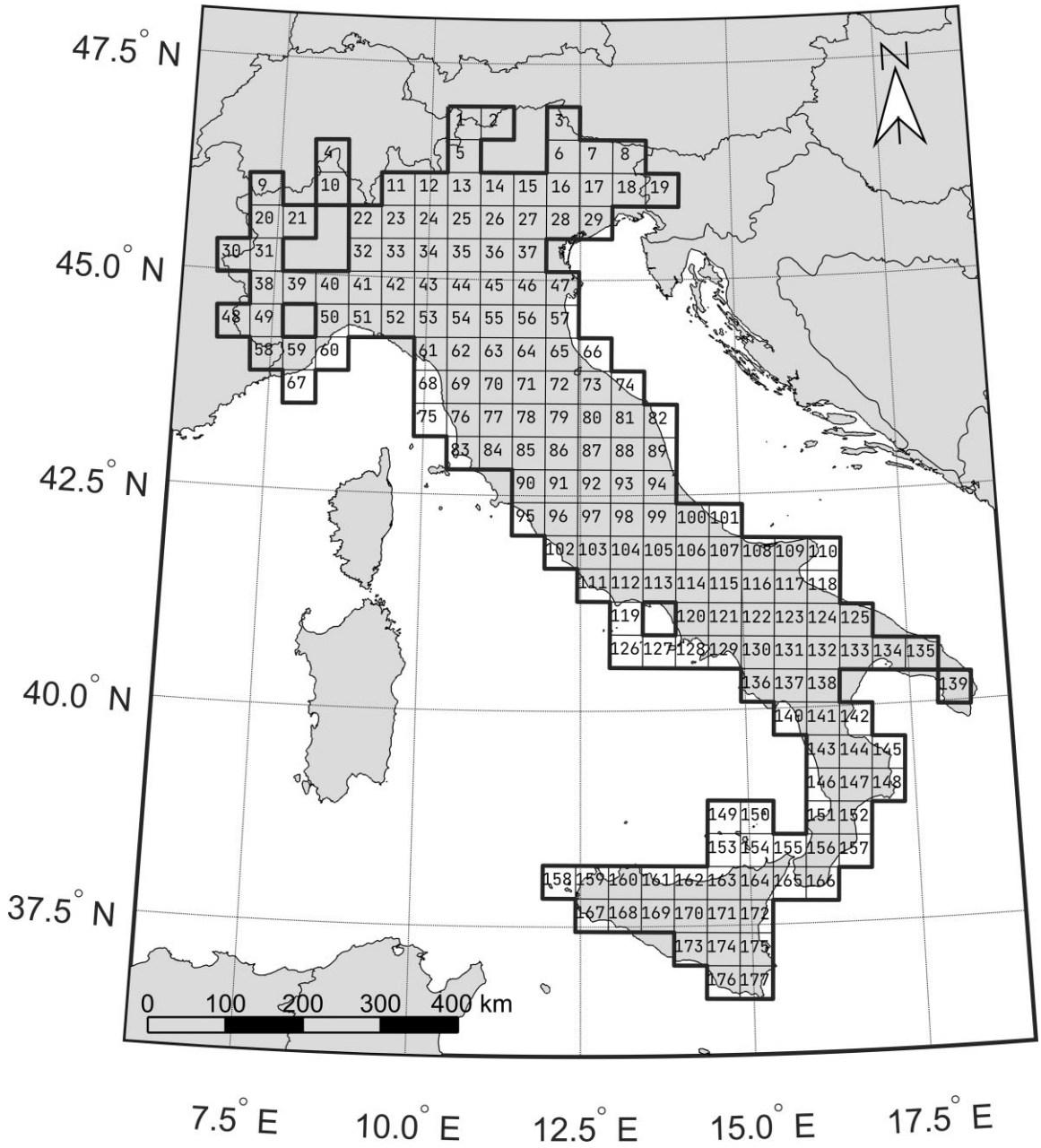

Figure 1. Tessellation of the Italian territory region used for the fitting of parameters and for the pseudo prospective experiment. The thick black line delimits the analysis region R. The cells that R comprises are only those within which at least one earthquake with $M \geq 4.0$ from 1600 to 2021 have occurred according to CPTI15 catalogue (Rovida *et al.* 2020) and have $30\sqrt{2}$ km of side.

where K is the total number of cells, T is the total duration of the forecasting experiment and d_{c_i} is the total time coverage of alarms within each ith spatial cell. d_{c_i} can be computed by multiplying the window length Δt by the number n of alarms casted and then subtracting the sum of intersections between alarm time windows. For more details see also Gasperini *et al.* (2021). The experiment is then repeated by varying the alarm time duration Δt from a few second to the total duration of the experiment (10 yr).

Following Shebalin *et al.* (2011), we calculated the fraction of the space–time occupied also by weighting each alarm with the long-term earthquake rate within each cell. We compute such rates using the historical earthquakes occurred in the time period 1620–1959 according to the CPTI15 V4.0 catalogue (Rovida *et al.* 2020, 2022) following the procedure described in Gasperini *et al.* (2021). Hence for each Δt of the considered forecasting models, we compute both weighted τ_w and unweighted τ_u space–time fractions occupied by alarms. Details of such computations for each cell are reported in Table S1.

The miss rate ν and the fraction of space–time occupied by alarms τ are used to draw the so-called Molchan diagram (Molchan 1990, 1991; Molchan & Kagan 1992) (e.g. Fig. 2). The line joining the points (τ, ν) obtained by varying Δt is called Molchan trajectory. In a paradoxical forecasting method that does not issue any alarm, $\tau = 0$ and it is impossible to predict any target event. This corresponds to the point $(\tau, \nu) = (0, 100$ per cent) in the upper left corner of the Molchan diagram. Conversely, a prediction method that issues

Table 1. Summary of tested forecasting models.

Model	Main features	Reference
FORE	Deterministic space/time-dependent model based on the occurrence of potential foreshocks as precursor signal.	Gasperini *et al.* (2021)
EEPAS-NW	Space/time-dependent model based on the hypothesis that each earthquake ($M \geq m_c$) contributes to the transient increment of the future rate of $M \geq m_T$ in its vicinity according to $\psi-$predictive relations.	Rhoades & Evison (2004)
ETAS-SVP	Epidemic-type aftershock model based on the hypothesis that each earthquake can perturb the rate of earthquakes and generate its own Omori-like decay sequence. The space variable Poisson model (SVP, Console & Murru 2001) is used as background model.	Ogata (1988, 1989), Ogata & Zhuang (2006)

alarms at any time and at any location covers the entire space–time domain, ensuring that no target events are missed. This method is represented by the point $(\tau, \nu) = (100$ per cent, $0)$ in the lower right corner of the diagram. The diagonal line connecting such two points (e.g. thick black diagonal line in Fig. 2) corresponds to a purely random forecasting that simply predicts target earthquakes proportionally to the fraction of space–time occupied by alarms. Its equation is given by

$$\nu = 1 - \tau \tag{3}$$

such line divides the Molchan diagram in two regions, the skilled (below the line) and the unskilled (above) ones. The ratio between the success rate and the space–time fraction occupied by alarms given by

$$G = \frac{1-\nu}{\tau} \tag{4}$$

represents the 'probability gain' with respect to the random chance (Kagan 2009). A skilled forecasting method is characterized by a $G > 1$.

To assess how much better (or worse) the performance of a forecasting model is, compared to the random model, Zechar & Jordan (2008, 2010) proposed the area skill (AS) score statistic $a_f(\tau)$. The latter is defined as the integral of the success rate function $1 - \nu(\tau)$ normalized to the alarm space–time coverage τ and takes the form:

$$a_f(\tau) = \frac{1}{\tau} \int_0^\tau [1 - \nu(t)]\, dt. \tag{5}$$

The AS score is normalized so that its value ranges between 0 and 1 and corresponds to the area above the Molchan trajectory. The closer the AS score to 1 the better the performance of a model. The AS score of a purely random method, obtained by substituting eq. (3) in eq. (5), is

$$a_f(\tau) = \frac{1}{\tau} \int_0^\tau [1 - (1-t)]\, dt = \frac{\tau}{2}. \tag{6}$$

Zechar & Jordan (2008, 2010) also derived the Gaussian asymptotical estimate of the variance of $a_f(\tau)$ as $\sigma^2 = 1/(12\ N)$ (where N is the number of target shocks). Even if the AS score can be computed for any τ, Zechar & Jordan (2008, 2010) argued that the power of the test tends to increase with increasing τ and therefore it is the best to use $a_f(\tau = 1)$ for hypothesis testing.

OPTIMIZATION OF ALARM THRESHOLDS

To identify the optimal expected daily rate threshold for the probabilistic models EEPAS and ETAS, we varied it logarithmically from 10^{-6} to 0.5. For the FORE method, the optimal range of potential foreshocks has been chosen by considering various magnitude ranges with central values varying from 4.1 to 4.5 and ranges with respect to the central values varying from 0.1 to 0.5. Using the different thresholds, the various models were thus applied to retrospectively forecast the 27 target events occurred during the learning period (1990–2011).

In Gasperini *et al.* (2021), the maximum AS score and the minimum number of alarms were considered as criteria for selecting the optimal threshold value. However, in the present experiment, since the expected daily rate for EEPAS and ETAS models was estimated at each earthquake occurrence time, the alarms in periods of high seismicity are mostly overlapped. For this reason, the number of alarms may not be representative of the fraction of space–time occupied by the alarms. Hence, to choose the best threshold, we used here only the maximum AS score criterion.

In Figs 3–5 we show the behaviours of the AS scores of the EEPAS, ETAS and FORE models respectively as a function of the analysed alarm thresholds. Red and dark blue lines refer to the unweighted (τ_u) and weighted (τ_w) fractions of space–time occupied by alarms, respectively. In such figures we also report for reference with grey bars the unweighted fraction of space–time occupied by alarms using a $\Delta t = 3$ months ($\tau_{3\,\mathrm{months}}$). In the same figures, the black arrowheads indicate our choices of the best probability thresholds or ranges: $p = 3 \times 10^{-5}$ for EEPAS, $p = 2 \times 10^{-4}$ for ETAS and 4.5 ± 0.3 for FORE.

RESULTS OF PSEUDO-PROSPECTIVE TESTING

Using the thresholds optimized for the learning interval 1990–2011, the models were applied pseudo-prospectively to forecast the 27 target events with $M_w \geq 5.0$ that occurred in the test period 2012–2021. For the EEPAS model, alarms of duration Δt are declared when the expected daily rate in a cell is larger than the chosen threshold $p = 3 \times 10^{-5}$. In Fig. 2 the Molchan trajectories obtained by varying Δt from a few seconds to the total duration $T = 10$ yr of the test period are reported. Red and dark blue lines refer to the unweighted (τ_u) and weighted (τ_w) fractions of space–time occupied by alarms respectively (see in Table S2 the numerical values of plotted curves). Both the red and dark blue lines in Fig. 2 lie below the Molchan trajectory of the purely random model (black diagonal line), indicating a better forecasting performance than the pure chance, for all explored Δt. The model fails to predict the total number of target earthquakes even for $\Delta t = 10$ yr, for which only 14 over 27 (51.9 per cent) target events are predicted with a corresponding $\tau_u \approx 21$ per cent and $\tau_w \approx 48$ per cent. Such fraction of forecasted target events is reached starting from $\Delta t = 0.25$ d (6 hr) with a corresponding $\tau_u \approx 7.8$ per cent and $\tau_w \approx 19$ per cent. The overall AS scores $a_f(\tau = 1) = 0.71 \pm 0.06$ and $a_f(\tau_w = 1) = 0.63 \pm 0.06$, based on the Student's t-test, are larger than the expectance of a random method (0.5).

In Fig. 6 the same plot as in Fig. 2 is showed but for the ETAS model using $p = 2 \times 10^{-4}$ as daily rate threshold to declare the

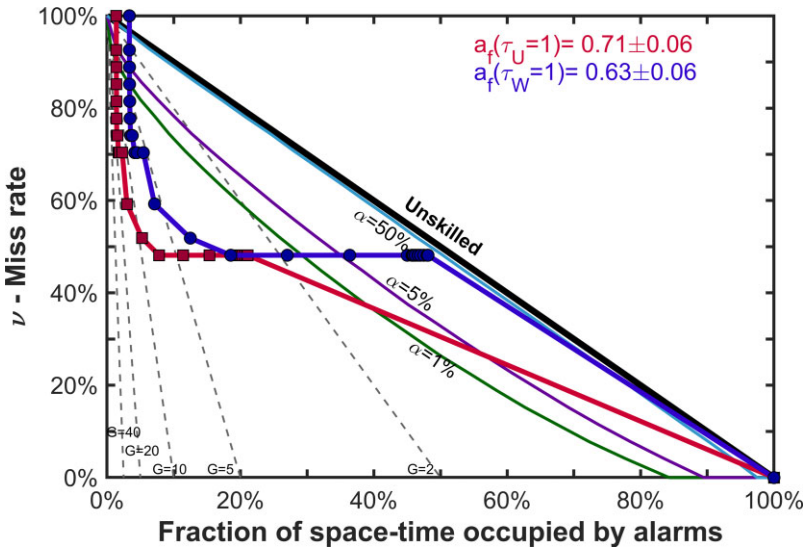

Figure 2. Molchan diagram and AS score of the EEPAS model for target shocks with $M_w \geq 5.0$ from 2012 to 2021. Red and dark blue lines indicate the forecasting performance of expected daily rate threshold 3×10^{-5} for unweighted (τ_u) and weighted (τ_w) fractions of space–time occupied by alarms respectively (see text). The black continuous line indicates a purely random forecasting method that separates skilled (below the line) from unskilled (above) forecasting methods. The light blue, violet and green lines indicate the confidence limits for $\alpha = 50$, 5 and 1 per cent, respectively. The black dashed lines indicate probability gains $G = 2$, 5, 10, 20 and 50.

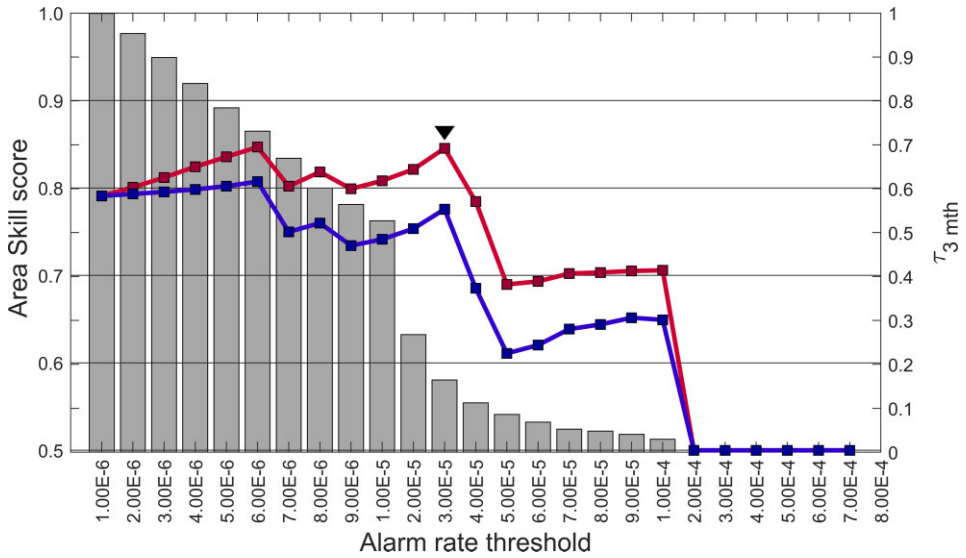

Figure 3. For EEPAS forecasting model, area skill (AS) score computed for targets with $M_w \geq 5.0$, using unweighted (red) and weighted (dark blue) fractions of space–time occupied by alarms, and fractions of space–time occupied by of alarms considering $\Delta t = 1$ yr (bars), as a function of the expected daily rate threshold. The chosen threshold is indicated by the black arrowhead (3×10^{-5}).

alarm in a cell (see Table S3 for numerical values). The performance is definitely better than EEPAS. In particular, the Molchan trajectories remain well below the diagonal line and, for τ_u and τ_w lower than 70–80 per cent, they are also below the $\alpha = 1$ per cent confidence curve (green). As for the EEPAS, ETAS fails to predict the total number of target earthquakes even for $\Delta t = 10$ years, for which only 25 over 27 (92.5 per cent) target events are predicted with a corresponding $\tau_u \approx 53$ per cent and $\tau_w \approx 73$ per cent. Such percentage of predicted events is reached starting from $\Delta t = 2$ yr with the corresponding $\tau_u \approx 37$ per cent and $\tau_w \approx 58$ per cent. The overall AS scores $a_f(\tau = 1) = 0.92 \pm 0.06$ and $a_f(\tau_w = 1) = 0.90 \pm 0.06$, are higher than those obtained by EEPAS.

In Fig. 7 it is showed the Molchan trajectory obtained by the FORE model using $M_w = 4.5 \pm 0.3$ as magnitude range for the foreshock events (see Table S4Table S4 for numerical values). The forecasting performance is better than EEPAS and slightly lower but substantially comparable to that of ETAS. In particular, the Molchan trajectories remain well below the diagonal line and, for τ_u and τ_w lower than 70–80 per cent, they are also below the $\alpha = 1$ per cent confidence curve (green line). As for the two previous models, FORE fails to predict the total number of target earthquakes even for $\Delta t = 10$ yr, for which only 22 over 27 (81.4 per cent) target events are predicted with corresponding $\tau_u \approx 14$ per cent and $\tau_w \approx 20$ per cent. For all Δt analysed, the FORE model is characterized

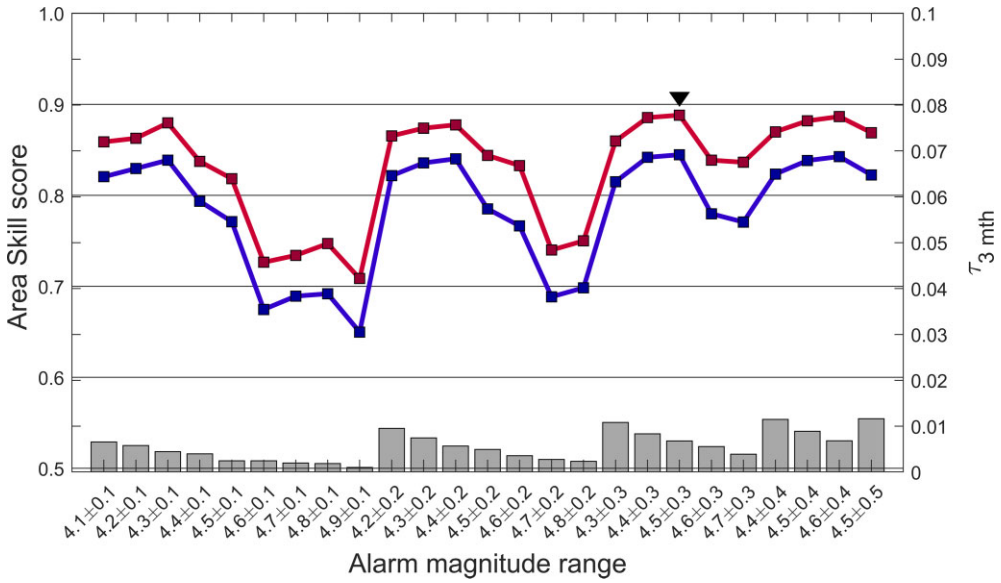

Figure 5. Same as Fig. 2 for the FORE model. The AS score and the fraction of space–time occupied by alarms are computed as a function of the foreshock magnitude range. The arrowhead indicates the chosen range ($M_w = 4.5 \pm 0.3$).

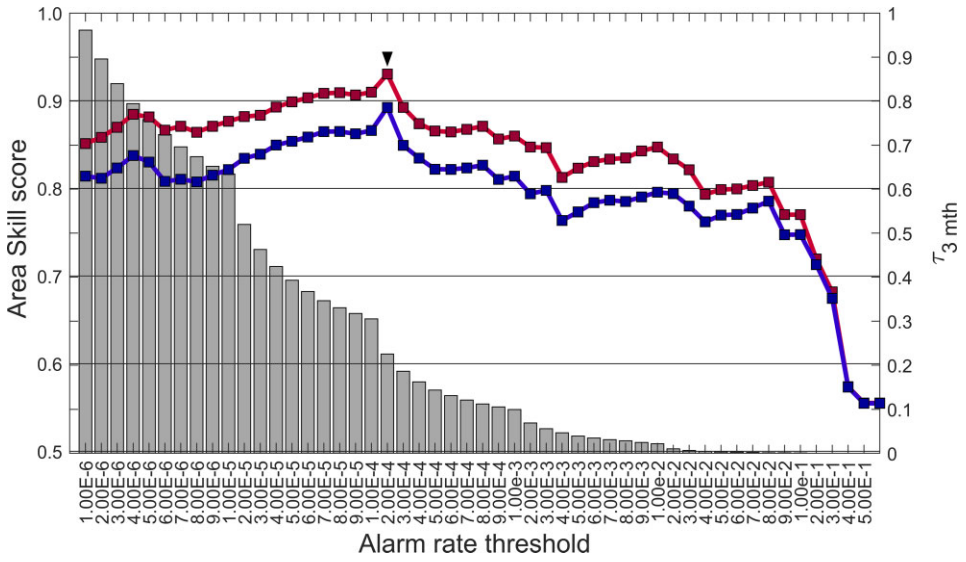

Figure 4. Same as Fig. 2 for ETAS. The chosen threshold is indicated by the black arrowhead (3×10^{-4}).

by lower τ_u and τ_w than those of the other models (Fig. 8 and Tables S2, S3 and S4). The overall AS scores $a_f(\tau_u = 1) = 0.91 \pm 0.06$ and $a_f(\tau_w = 1) = 0.89 \pm 0.06$, are higher than that obtained by EEPAS and slightly lower than that obtained by ETAS.

Following Shebalin *et al.* (2011), the miss rates (ν) of the analysed models can also be compared to another skilled reference model, characterized by its space–time fractions occupied by alarms (τ_{ref}) and miss rates (ν_{ref}), instead of the purely random model. In such comparison, the miss rates ν_{ref} of the reference model are plotted on the diagonal line of the Molchan diagram were $\tau = 1 - \nu_{ref}$. Even for the other compared models, the expected miss rates (ν) must be plotted versus $\tau = 1 - \nu_{ref}$, but as they are computed at different values of τ than the τ_{ref} of the reference model, a linear interpolation is required. Hence for each τ_{ref} of the reference ETAS model, we compute the interpolated miss-rates, for both EEPAS and FORE models, as

$$\nu_{int} = \frac{\nu_a - \nu_b}{\tau_a - \tau_b}(\tau_{ref} - \tau_b) + \nu_b, \tag{7}$$

where τ_a and τ_b are the fractions occupied by alarms, of the compared models, immediately larger and smaller than τ_{ref}, respectively, and ν_a and ν_b the corresponding miss rates, respectively.

As for the Molchan diagrams showed above (Figs 2, 6 and 7), if the miss rates of a compared model (ν_{int} for the corresponding $\tau = 1 - \nu_{ref}$ are lower than ν_{ref}, it has a better predictive ability than the reference model. In Figs 9 and 10 the Molchan (or better Molchan–Shebalin) diagrams and the AS scores for unweighted (τ_u) and weighted (τ_w) fractions of space–time occupied by alarms are reported, respectively (see Tables 2 and 3 for numerical values). Where the ETAS model is reported as the diagonal line (orange)

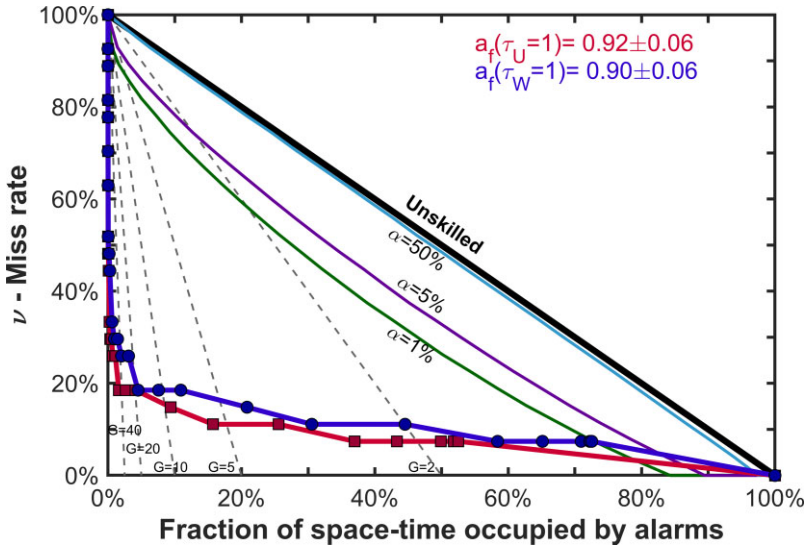

Figure 6. Same as Fig. 2 for the ETAS model.

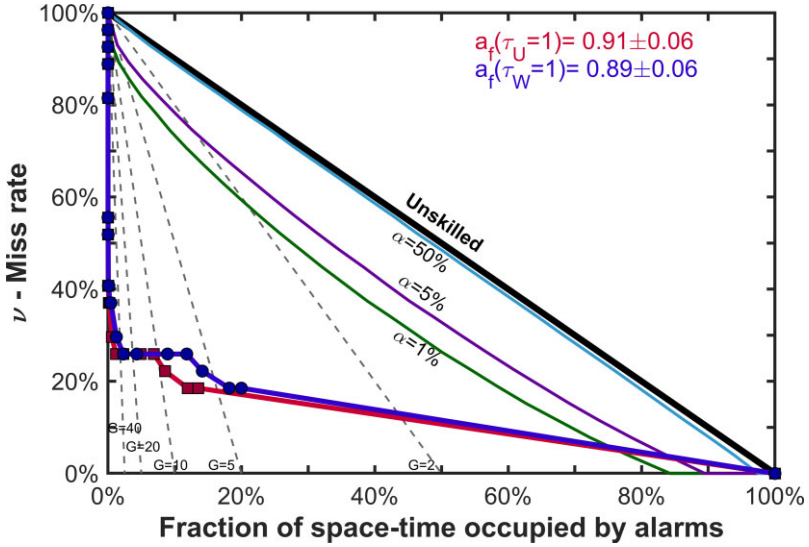

Figure 7. Same as Fig. 2 for the FORE model.

joining the coordinate points $(1-\nu_{\rm ref},\ \nu_{\rm ref})$. The Molchan trajectory for the EEPAS (blue) and FORE models (red) join the points $(1-\nu_{\rm ref},\ \nu_{\rm int})$. In both Figs 9 and 10 the EEPAS model is characterized by a Molchan trajectory well above the reference diagonal line showing a worse predictive performance than ETAS since for the same $\tau = 1-\nu_{\rm ref}$, the $\nu_{\rm int}$ of the EEPAS is always larger than that of ETAS. Such worst forecasting performance is also confirmed by the AS scores $a_{\rm f}(\tau_{\rm u}=1)$ and $a_{\rm f}(\tau_{\rm w}=1)$, which both equal to 0.13 ± 0.06, that is smaller than that (0.5) of the reference ETAS model. On the contrary, the FORE model is characterized by Molchan trajectories lower or close to the reference model for $\tau < 60$ per cent. For $\tau > 60$ per cent the FORE Molchan trajectory lies for short stretches slightly above the diagonal line. Overall, the predictive performance is (slightly) better than the reference model because the AS scores for the unweighted and weighted trajectories, $a_{\rm f}(\tau_{\rm u}=1)=0.540.06$ and $a_{\rm f}(\tau_{\rm w}=1)=0.550.06$, respectively, are larger than that of the reference model ETAS. However, the AS score difference between FORE and ETAS is well within the confidence bounds and hence cannot be considered significant in the statistical sense.

CONCLUSIONS

In this work, the deterministic forecasting method based on the occurrence of strong foreshocks described by Gasperini *et al.* (2021) (called FORE) was compared with the probabilistic EEPAS and ETAS models described by Biondini *et al.* (2023) using the deterministic alarm-based approach. The forecasting models were applied to pseudo-prospectively forecast of shallow earthquakes (depth < 40 km) with magnitude $M \geq 5.0$ occurred in Italy from 2012 to 2021. All models were calibrated through a learning data set from 1990 to 2011. Both the learning and testing data sets are extracted from the HORUS seismic catalogue (Lolli *et al.* 2020). Based on the learning data set, various expected daily rate thresholds for the EEPAS and ETAS models and various magnitude ranges of foreshocks for the FORE model were considered to identify the

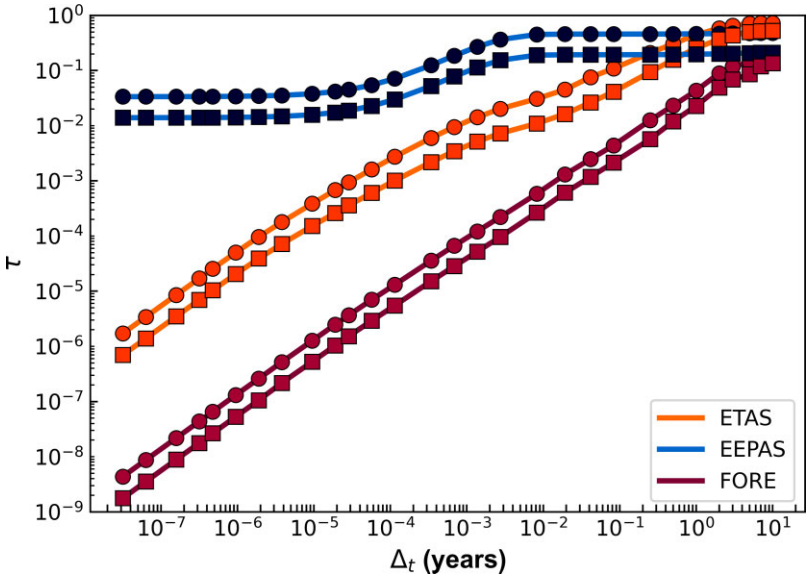

Figure 8. Behaviour of τ_u (squares) and τ_w (circles) as a function of Δt for the three forecasting models.

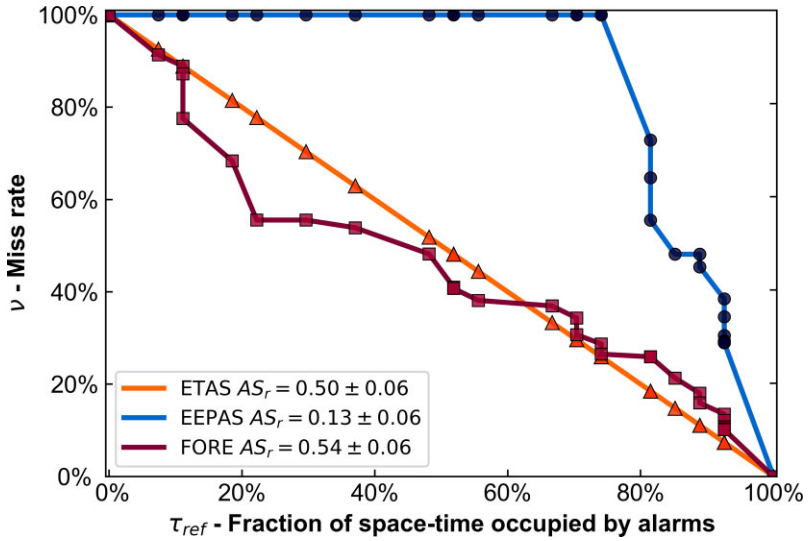

Figure 9. Molchan diagram and AS score of the EEPAS and FORE models considering the ETAS model as reference and unweighted fraction of space–time occupied by alarms (τ_u). The Molchan trajectory of EEPAS and FORE models are indicated by the blue and red curves, respectively. The diagonal continuous line (orange) indicates the miss rates of the ETAS model and separates models with higher performance (below the line) from with those with lower performance (above).

optimal alarm thresholds required to implement the deterministic approach. For each analysed threshold, the miss rates and the fractions of space–time occupied by the alarms were estimated by varying the temporal extension of the alarm window Δt from a fraction of a few seconds to the total duration of the learning data set (10 yr). Following the method described by Gasperini *et al.* (2021), such fractions were also estimated by considering the different levels of seismic activity in the various areas of Italy, by weighting more the alarm times in cells where the average seismicity rate, calculated from the CPTI15 seismic catalogue (Rovida *et al*. 2016, 2020) from 1600 to 1990, is higher. The choice of the optimal alarm thresholds was made by considering the Area Skill score (Zechar & Jordan 2008).

The tests of performance conducted using the Molchan diagram (Molchan 1990, 1991; Molchan & Kagan 1992) and the Area Skill score indicate that all the three models outperform a purely random method. In addition, following the approach suggested by Shebalin *et al.* (2011), the forecasting performances of the models were compared taking the ETAS as reference model. These tests show that for the same fractions of space–time occupied by the alarms, the EEPAS model presents a worse forecasting performance compared to the ETAS and FORE models. The FORE model shows a slightly better forecast performance than ETAS but the difference between the AS scores lies within the error bounds. Anyway, a more careful analysis should be done prospectively, maybe using new method developed within the CSEP.

Overall, this study highlights the potential of deterministic forecasting approaches in improving earthquake forecasting in Italy. Our findings contribute to ongoing efforts to develop more accurate and reliable earthquake forecasting methods, which can ultimately help mitigate the impact of seismic events on communities and infrastructure.

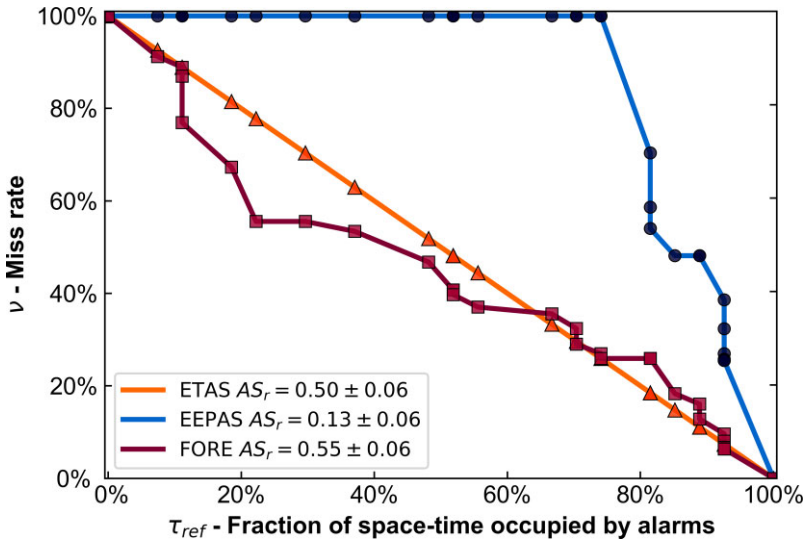

Figure 10. Same as Fig. 9 for weighted fraction of space–time occupied by alarms (τ_w).

Table 2. Values of variables in Molchan–Shebalin plot of Fig. 8 for unweighted fractions of space–time occupied by alarms (τ_u).

τ_{ref}	ν_{ref}	$a_f(\tau_{ref})$	$\nu_{int_{EEPAS}}$	$a_{EEPAS}(\tau_{ref})$	$\nu_{int_{FORE}}$	$a_{FORE}(\tau_{ref})$
0.000	1.000	0.000	1.000	0.000	1.000	0.000
0.074	0.926	0.037	1.000	0.000	0.913	0.043
0.111	0.889	0.056	1.000	0.000	0.889	0.062
0.111	0.889	0.056	1.000	0.000	0.873	0.062
0.111	0.889	0.056	1.000	0.000	0.776	0.062
0.185	0.815	0.093	1.000	0.000	0.684	0.145
0.222	0.778	0.111	1.000	0.000	0.556	0.184
0.296	0.704	0.148	1.000	0.000	0.556	0.249
0.370	0.630	0.185	1.000	0.000	0.539	0.290
0.481	0.519	0.241	1.000	0.000	0.482	0.336
0.519	0.481	0.259	1.000	0.000	0.410	0.352
0.519	0.481	0.259	1.000	0.000	0.407	0.352
0.519	0.481	0.259	1.000	0.000	0.407	0.352
0.556	0.444	0.278	1.000	0.000	0.381	0.369
0.667	0.333	0.333	1.000	0.000	0.369	0.411
0.704	0.296	0.352	1.000	0.000	0.343	0.423
0.704	0.296	0.352	1.000	0.000	0.307	0.423
0.741	0.259	0.370	1.000	0.000	0.287	0.437
0.741	0.259	0.370	1.000	0.000	0.265	0.437
0.815	0.185	0.407	0.729	0.012	0.259	0.465
0.815	0.185	0.407	0.647	0.012	0.259	0.465
0.815	0.185	0.407	0.555	0.012	0.259	0.465
0.852	0.148	0.426	0.481	0.033	0.213	0.478
0.889	0.111	0.444	0.481	0.053	0.180	0.491
0.889	0.111	0.444	0.454	0.053	0.159	0.491
0.926	0.074	0.463	0.385	0.074	0.135	0.506
0.926	0.074	0.463	0.346	0.074	0.121	0.506
0.926	0.074	0.463	0.306	0.074	0.107	0.506
0.926	0.074	0.463	0.294	0.074	0.103	0.506
0.926	0.074	0.463	0.290	0.074	0.102	0.506
0.926	0.074	0.463	0.290	0.074	0.102	0.506
0.926	0.074	0.463	0.290	0.074	0.102	0.506
0.926	0.074	0.463	0.290	0.074	0.102	0.506
1.000	0.000	0.500	0.000	0.132	0.000	0.539

$\boldsymbol{\nu_{ref}}$ is the miss rate of the ETAS model (reference), τ_{re} the fractions of space–time occupied by alarms used to plot the Molchan diagram and given by $1 - \nu_{ref}$, ν_{ref} the reference miss rate, ν_{ref} the interpolated miss rate, $a_f(\tau_{ref})$, $a_{EEPAS}(\tau_{ref})$ and $a_{FORE}(\tau_{ref})$ the area skill scores computed for the reference, EEPAS and FORE models, respectively.

Table 3. Values of variables in Molchan–Shebalin plot of Fig. 9 for weighted fractions of space–time occupied by alarms (τ_w).

τ_{ref}	ν_{ref}	$a_f(\tau_{ref})$	$\nu_{int_{EEPAS}}$	$a_{EEPAS}(\tau_{ref})$	$\nu_{int_{FORE}}$	$a_{FORE}(\tau_{ref})$
0.000	1.000	0.000	1.000	0.000	1.000	0.000
0.074	0.926	0.037	1.000	0.000	0.912	0.044
0.111	0.889	0.056	1.000	0.000	0.889	0.062
0.111	0.889	0.056	1.000	0.000	0.870	0.062
0.111	0.889	0.056	1.000	0.000	0.769	0.062
0.185	0.815	0.093	1.000	0.000	0.673	0.149
0.222	0.778	0.111	1.000	0.000	0.556	0.188
0.296	0.704	0.148	1.000	0.000	0.556	0.252
0.370	0.630	0.185	1.000	0.000	0.534	0.293
0.481	0.519	0.241	1.000	0.000	0.468	0.340
0.519	0.481	0.259	1.000	0.000	0.407	0.356
0.519	0.481	0.259	1.000	0.000	0.407	0.356
0.519	0.481	0.259	1.000	0.000	0.397	0.356
0.556	0.444	0.278	1.000	0.000	0.370	0.374
0.667	0.333	0.333	1.000	0.000	0.356	0.418
0.704	0.296	0.352	1.000	0.000	0.324	0.430
0.704	0.296	0.352	1.000	0.000	0.290	0.430
0.741	0.259	0.370	1.000	0.000	0.270	0.445
0.741	0.259	0.370	1.000	0.000	0.259	0.445
0.815	0.185	0.407	0.704	0.013	0.259	0.472
0.815	0.185	0.407	0.586	0.013	0.259	0.472
0.815	0.185	0.407	0.540	0.013	0.259	0.472
0.852	0.148	0.426	0.481	0.034	0.183	0.485
0.889	0.111	0.444	0.481	0.054	0.161	0.499
0.889	0.111	0.444	0.481	0.054	0.128	0.499
0.926	0.074	0.463	0.386	0.075	0.096	0.515
0.926	0.074	0.463	0.324	0.075	0.081	0.515
0.926	0.074	0.463	0.270	0.075	0.067	0.515
0.926	0.074	0.463	0.258	0.075	0.064	0.515
0.926	0.074	0.463	0.255	0.075	0.064	0.515
0.926	0.074	0.463	0.255	0.075	0.064	0.515
0.926	0.074	0.463	0.255	0.075	0.064	0.515
0.926	0.074	0.463	0.255	0.075	0.064	0.515
1.000	0.000	0.500	0.000	0.134	0.000	0.548

SUPPORTING INFORMATION

Supplementary data are available at *GJI* online.

Table S1. List of centre coordinates of cells with side $30\sqrt{2}$ km. $N_{4.5}$, $N_{5.0}$, $N_{5.5}$, $N_{6.0}$: numbers of earthquakes occurred within cell, according to the CPTI15 V4.0 catalogue (Rovida *et al.* 2020, 2022) up to year 1959, with $M_w \geq 4.5$, 5.0, 5.5 and 6.0, respectively in respective time intervals of completeness. $\lambda_{4.5}$, $\lambda_{5.0}$, $\lambda_{5.5}$, $\lambda_{6.0}$: average rates (events per year) of earthquakes with $M_w \geq 4.0$ within each cell computed from observed earthquakes with $M_w \geq 4.5$, 5.0, 5.5 and 6.0, respectively, and assuming a b-value = 1 (see text). λ_{ave}: average of non-null rates $\lambda_{4.5}$, $\lambda_{5.0}$, $\lambda_{5.5}$, $\lambda_{6.0}$. $w = \lambda_{ave} / \sum \lambda_{ave}$: overall weight of each cell ($\sum w = 1$).

Table S2. Values of variables in Molchan plot of Fig. 2 for the EEPAS model.

Table S3. Values of variables in Molchan plot of Fig. 6 for the ETAS model.

Table S4. Values of variables in Molchan plot of Fig. 7 for the FORE model.

Please note: Oxford University Press is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the paper.

ACKNOWLEDGMENTS

We thank the Editor Margarita Segou, and the reviewers Max Werner and Andy Michael for thoughful comments and suggestions. This paper benefited from funding provided by the European Union within the ambit of the H2020 project RISE (no. 821115), which in particular fully financed the PhD grant of one of the authors (EB).

DATA AVAILABILITY

The sources codes used for fitting the models and for computing and plotting the Molchan diagrams are provided in the supplementary material. The data of the HORUS and CPTI15 earthquake catalogues are available from public providers: http://horus.bo.ingv.it and https://emidius.mi.ingv.it/CPTI15-DBMI15/, respectively.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest regarding the publication of this manuscript.

REFERENCES

Agnew, D.C. & Jones, L.M., 1991. Prediction probabilities from foreshocks, *J. geophys. Res.*, **96**(B7), 11 959–11 971.

Azarbakht, A., Rudman, A. & Douglas, J., 2021. A decision-making approach for operational earthquake forecasting, *Int. J. Disaster Risk Reduct.,* **66,** doi:10.1016/j.ijdrr.2021.102591.

Bayona, J.A., Savran, W.H., Rhoades, D.A. & Werner, M.J., 2022. Prospective evaluation of multiplicative hybrid earthquake forecasting models in California, *Geophys. J. Int.,* **229,** 1736–1753.

Biondini, E., Rhoades, D.A. & Gasperini, P., 2023. Application of the EEPAS earthquake forecasting model to Italy, *Geophys. J. Int.,* **234**(3), 1681–1700.

Console, R. & Murru, M. 2001. A simple and testable model for earthquake clustering, *J. geophys. Res.,* **106**(), 8699–8711.

Console, R., Murru, M. & Catalli, F., 2006. Physical and stochastic models of earthquake clustering, *Tectonophysics,* **417,** 141–153.

Console, R., Murru, M. & Falcone, G., 2010. Retrospective forecasting of M $\geq$ 4.0 earthquakes in New Zealand, *Pure appl. Geophys.,* **167,** 693–707.

Falcone, G., Console, R. & Murru, M., 2010. Short-term and long-term earthquake occurrence models for Italy: ETES, ERS and LTST, *Ann. Geophys.,* **53**(3), 41–50.

Gasperini, P., Biondini, E., Lolli, B., Petruccelli, A. & Vannucci, G., 2021. Retrospective short-term forecasting experiment in Italy based on the occurrence of strong (fore) shocks, *Geophys. J. Int.,* **225,** 1192–1206.

Jones, L.M., 1984. Foreshocks (1966-1980) in the San Andreas system, California, *Bull. seism. Soc. Am.,* **74,** 1361–1380.

Jones, L.M., 1985. Foreshocks and time-dependent earthquake hazard assessment in Southern California, *Bull. seism. Soc. Am.,* **75,** 1669–1679.

Jones, L.M., 1994. Foreshocks, aftershocks, and earthquake probabilities: accounting for the Landers earthquake, *Bull. seism. Soc. Am.,* **84,** 892–899.

Jordan, T.H. *et al.*, 2011. Operational earthquake forecasting: state of knowledge and guidelines for utilization, *Ann. Geophys.,* **54**(4), doi:10.4401/ag-5350.

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

Jordan, T.H., 2009. Earthquake system science: potential for seismic risk reduction, *Sci. Iran.,* **16,** 351–366.

Kagan, Y.Y., 2009. Testing long-term earthquake forecasts: likelihood methods and error diagrams, *Geophys. J. Int.,* **177,** 532–542. doi: 10.1111/j.1365-246X.2008.04064.x.

Lolli, B., Randazzo, D., Vannucci, G. & Gasperini, P., 2020. The Homogenized instrumental Seismic Catalog (HORUS) of Italy from 1960 to present, *Seismol. Res. Lett.,* **91,** 3208–3222.

Lombardi, A.M. & Marzocchi, W., 2010. The ETAS model for daily forecasting of Italian seismicity in the CSEP experiment, *Ann. Geophys.,* **53**(3), doi:10.4401/ag-4848.

MacPherson-Krutsky, C., Lindell, M.K. & D. Brand, B., 2023. Residents' information seeking behavior and protective action for earthquake hazards in the Portland Oregon Metropolitan Area, *Risk Anal.,* **43,** 372–390.

Marzocchi, W. & Lombardi, A.M., 2009. Real-time forecasting following a damaging earthquake, *Geophys. Res. Lett.,* **36**(21), doi:10.1029/2009GL040233.

Michael, A.J. & Werner, M.J., 2018. Preface to the focus section on the Collaboratory for the Study of Earthquake Predictability (CSEP): new results and future directions, *Seismol. Res. Lett.,* **89**(4), 1226–1228.

Mizrahi, L., Nandan, S., Savran, W., Wiemer, S. & Ben-Zion, Y., 2023. Question-driven ensembles of flexible ETAS models, *Seismol. Res. Lett.,* **94,** 829–843.

Molchan, G.M., 1990. Strategies in strong earthquake prediction, *Phys. Earth planet. Inter.,* **61,** 84–98.

Molchan, G.M., 1991. Structure of optimal strategies in earthquake prediction, *Tectonophysics,* **193,** 267–276.

Molchan, G.M. & Kagan, Y.Y., 1992. Earthquake prediction and its optimization, *J. geophys. Res.,* **97,** 4823.

Murru, M., Console, R. & Falcone, G., 2009. Real time earthquake forecasting in Italy, *Tectonophysics,* **470,** 214–223.

Ogata, Y., 1988. Statistical models for earthquake occurrences and residual analysis for point processes, *J. Am. Stat. Assoc.,* **83,** 9–27.

Ogata, Y. & Zhuang, J., 2006. Space–time ETAS models and an improved extension, *Tectonophysics,* **413,** 13–23.

Rhoades, D.A., 2007. Application of the EEPAS model to forecasting earthquakes of moderate magnitude in Southern California, *Seismol. Res. Lett.,* **78,** 110–115.

Rhoades, D.A., 2011. Application of a long-range forecasting model to earthquakes in the Japan mainland testing region, *Earth Planets Space,* **63,** 197–206.

Rhoades, D.A. & Evison, F.F., 2004. Long-range earthquake forecasting with every earthquake a precursor according to scale, *Pure appl. Geophys.,* **161,** 47–72.

Rovida, A., Locati, M., Camassi, R., Lolli, B. & Gasperini, P., 2020. The Italian earthquake catalogue CPTI15, *Bull. Earthq. Eng.,* **18,** 2953–2984.

Rovida, A., Locati, M., Camassi, R., Lolli, B., Gasperini, P., Antonucci, A., *et al.* 2022. Italian Parametric Catalogue of Italian Earthquakes (CPTI15), version 4.0, Istituto Nazionale di Geofisica e Vulcanologia (INGV), doi:10.13127/cpti/cpti15.4.

Rovida, A., Locati, M., Camassi, R., Lolli, B & Gasperini, P.(eds), 2016. CPTI15, the 2015 version of the Parametric Catalogue of Italian Earthquakes, *Istituto Nazionale di Geofisica e Vulcanologia,* doi: 10.6092/INGV.IT-CPTI15.

Savran, W., Werner, M., Schorlemmer, D. & Maechling, P., 2022a. pyCSEP: a python package for earthquake forecast developers, *J. Open Source Software,* **7,** doi:10.21105/joss.03658.

Savran, W.H. *et al.*, 2022b. pycsep: a python toolkit for earthquake forecast developers, *Seismol. Res. Lett.,* **93**(5), 2858–2870.

Schorlemmer, D. *et al.*, 2018. The collaboratory for the study of earthquake predictability: achievements and priorities, *Seismol. Res. Lett.,* **89**(4), 1305–1313.

Schorlemmer, D., Gerstenberger, M.C., Wiemer, S., Jackson, D.D. & Rhoades, D.A., 2007. Earthquake likelihood model testing, *Seismol. Res. Lett.,* **78,** 17–29.

Shebalin, P., Narteau, C., Holschneider, M. & Schorlemmer, D., 2011. Short-term earthquake forecasting using early aftershock statistics, *Bull. seism. Soc. Am.,* **101,** 297–312.

Werner, M.J., Zechar, J.D., Marzocchi, W. & Wiemer, S., 2010. Retrospective evaluation of the five-year and ten-year CSEP-Italy earthquake forecasts, *Ann. Geophys.,* **53**(3), doi:10.4401/ag-4840.

Zechar, J.D. & Jordan, T.H., 2008. Testing alarm-based earthquake predictions, *Geophys. J. Int.,* **172,** 715–724.

Zechar, J.D. & Jordan, T.H., 2010. The area skill score statistic for evaluating earthquake predictability experiments, *Pure appl. Geophys.,* **167,** 893–906.

Zechar, J.D., Schorlemmer, D., Liukis, M., Yu, J., Euchner, F., Maechling, P.J. & Jordan, T.H., 2010b. The collaboratory for the study of earthquake predictability perspective on computational earthquake science, *Concur. Comp. Pract. Exp.,* **22**(12), 1836–1847.