

Pseudo-prospective earthquake forecasting experiment in Italy based on temporal variation of the b -value of the Gutenberg–Richter law

E. Biondini¹, F. D’Orazio¹, B. Lolli² and P. Gasperini^{1,2}

¹Dipartimento di Fisica e Astronomia, Università di Bologna, Bologna I-40127, Italy. E-mail: emanuele.biondini2@unibo.it

²Istituto Nazionale di Geofisica e Vulcanologia, Sezione di Bologna, Bologna I-40127, Italy

Accepted 2024 December 30. Received 2024 December 20; in original form 2024 October 10

SUMMARY

We present the BVAL method, designed to forecast potentially damaging earthquakes ($M_w \geq 5.0$) in Italy based on temporal variations of the b -value of the Gutenberg–Richter frequency–magnitude distribution. The b -value is used as an indicator of stress within the Earth’s crust, with lower b -values associated with higher stress levels and an increased likelihood of significant seismic events. This method issues alarms when the b -value falls below a critical threshold. It is optimized using the HOMogenized instrUMENTal Seismic catalogue data from 1990 to 2004 and validated pseudo-prospectively using data from 2005 to 2022. Our analysis uses the recently developed b -positive (b^+) method to compute the b -value from magnitude differences, providing resilience against data incompleteness. We compare the performance of the BVAL method with two established models: the Epidemic Type Aftershock Sequence (ETAS) model, which forecasts earthquake rates based on the epidemic principle that each shock triggers subsequent shocks, and the FORE model, which relies on the occurrence of strong foreshocks. Additionally, we evaluate two ensemble models that combine BVAL and FORE through additive (EADD) and multiplicative (EMUL) strategies to balance false alarms and missed events. The EADD model declares an alarm when either BVAL or FORE signals it, while the EMUL model triggers alarms only when both methods agree. We assess the predictive efficiency of these models using the area skill score, derived from Molchan diagrams, which plot the miss rate against the fraction of space-time occupied by alarms. Our results demonstrate that BVAL is less effective than FORE and ETAS at high space-time fractions, but it outperforms ETAS at low fractions ($\tau < 2$ –4 per cent), indicating its potential utility in scenarios where minimizing false alarms is critical. This comprehensive comparison highlights the strengths and limitations of each method, suggesting that integrating multiple forecasting strategies can enhance the reliability of earthquake preparedness and response efforts in Italy.

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

INTRODUCTION

The slope (b -value) of the frequency–magnitude distribution (FMD) of earthquakes (Gutenberg & Richter 1944) is indicated by some studies (Scholz 1968, 2015; Ambrano 2003) as a proxy of the state of stress within the Earth’s crust and then as an indicator of the state of preparation of a future strong earthquake (Gulia & Wiemer 2010, 2019, 2021; Gulia *et al.* 2018, 2024). The ratio behind such studies is that low b -values are likely to be associated with high stress levels and then with a larger probability of occurrence of strong shocks, and high b -values with low stress levels and a lower probability.

In the last two decades, several works studied the space-time variations of b -value before moderate and strong earthquakes in different parts of the world. Among the others, we can mention Murru *et al.* (2004) for the 1997 central Italy seismic sequence, Nuannin *et al.* (2005) for the great 2004 Sumatra earthquake, Tsukakoshi & Shimazaki (2008) for the 2003 Northern Miyagi earthquake in Japan, Chan *et al.* (2012, 2023) for large earthquakes in Taiwan, Varotsos *et al.* (2012) before large earthquakes in Greece, El-Isa & Eaton (2014) for Central California and Andaman–Sumatran regions, Montuori *et al.* (2016) and García-Hernández *et al.* (2021) for the 2016 Central Italy sequence, Lu (2017) for New Zealand, Öztürk (2020) for Central Anatolia, DeSalvio and Rudolph (2021)

Table 1. Summary of tested forecasting models.

Model	Main features	References
BVAL	Deterministic space/time-dependent model based on the observation that the b -value of the Gutenberg–Richter law inversely correlates with the accumulation of differential stress in crustal rocks. An alarm is triggered whenever the b -value drops below a certain threshold value.	(Scholz 1968, 2015; Amitrano 2003)
FORE	Deterministic space/time-dependent model based on the occurrence of potential foreshocks as precursor signals. An alarm is issued every time an earthquake within the potential foreshock magnitude range occurs.	(Gasperini <i>et al.</i> 2021)
ETAS	Epidemic-type aftershock sequence model based on the hypothesis that each earthquake can perturb the rate of subsequent earthquakes and generate its own Omori-like decay sequence. In this work, the SVP model (Console & Murru, 2001) is used to assess the background seismicity. The ETAS model is applied using an alarm-based deterministic approach, where an alarm is issued whenever the expected daily rate of target events exceeds a pre-defined threshold.	(Ogata 1988, 1989; Ogata & Zhuang 2006)

for western Nord America and Pastoressa *et al.* (2023) for Central Italy. In such works, there is a general consensus that strong earthquakes are preceded by low b -values but even the opposite is found in a few cases (e. g. Parsons 2007; Sharma *et al.* 2023).

In this work, we develop a method to forecast potentially damaging earthquakes ($M_w \geq 5.0$) in Italy based on the temporal behaviour of the b -value (called BVAL). Our algorithm issues an alarm every time the b -value goes below a given threshold coming from a higher value. The duration of the alarm is made to vary between a fraction of second to the total duration of the experiment.

We take advantage of the recently developed b -positive (or b^+) method (van der Elst 2021) to compute the b -value from the distribution of magnitude differences which was proven to be more resilient to incompleteness than methods based on the distribution of simple magnitudes (Tinti & Gasperini 2024).

We optimize the forecasting algorithm based on the analysis of the HOMogenized instrUMENTal Seismic (HORUS) catalogue (Lolli *et al.* 2020) from 1990 to 2004 (learning period). Using the same catalogue, we test the algorithm pseudo-prospectively on a successive interval from 2005 to 2022 (testing period) and compare its performance with that of other two forecasting methods already applied in Italy: the Epidemic Type Aftershock Sequence (ETAS; Ogata 1988; Murru *et al.* 2009; Lombardi & Marzocchi 2010; Biondini *et al.* 2023; Mizrahi *et al.* 2023), the FORE (Gasperini *et al.* 2021; Biondini & Gasperini 2023), and two ensemble models obtained, respectively, as addition and multiplication (Bayona *et al.* 2022) of BVAL and FORE methods.

In the additive model (EADD), an alarm is triggered if either of the two individual methods detects the precursor signal. Consequently, the prediction window of this ensemble model is the union of the prediction windows from both methods. This design increases the sensitivity of the model, aiming to minimize missed detections of the precursor signal. However, it may result in a higher number of false alarms due to its broader criteria for issuing alerts.

In contrast, the multiplicative model (EMUL) activates an alarm only when both simple methods independently detect the precursor signal. Therefore, the prediction window for this ensemble model is the intersection of the prediction windows of the two methods. This approach makes the EMUL model more prudential, focusing on reducing false alarms by requiring a higher level of confirmation before triggering an alert.

The choice between these models depends on the user's priority: the EMUL model seeks to limit false alarms by demanding agreement from both methods, while the EADD model aims to enhance

detection of the precursor signal, even at the risk of generating more false alarms.

See in Table 1, a summary of various forecasting models considered in this work.

ANALYSIS REGION AND EARTHQUAKE CATALOGUE

As application region, we consider the overlap of two tessellations (R_1 and R_2) of the Italian territory spanning from 7° E to 19° E in longitude and from 36° N to 47° N in latitude. Both tessellations consist of square cells with side length of $L = 30\sqrt{2}$ km. The tessellation R_2 is shifted by $L/2$ both in latitude and in longitude with respect to R_1 (Fig. 1). The purpose of this dual tessellation is to enhance the spatial alignment between the forecasting features detected by different algorithms with the actual earthquakes. This approach almost doubles the portion of space covered by the alarms, thus reducing the efficiency of the various algorithms, but avoids failing a correct prediction because the earthquake to be predicted, although close, is located slightly outside the alarmed cell as in the sketch of Fig. 2, in which the precursor seismicity (blue circles) is able to predict the target earthquake (red star) using the red grid but not the black grid.

The tessellation R_1 was developed starting from the initial point $x_1 = (7^\circ$ E, 47° N) and moving the centre from x_1 in steps of L km for each cell, until reaching $x_2 = (19^\circ$ E, 36° N). The centres of the cells in tessellation R_2 are shifted relative to R_1 by $-L/2$ km in longitude and $+L/2$ km in latitude. To ensure uniform cell dimensions, we converted the coordinates of points x_1 and x_2 from the WGS84 geographic reference system to kilometric coordinates in the RDN2008 Italy zone (E-N) CRS EPSG:7794 using the QGIS software (see the Data Availability section).

Following Gasperini *et al.* (2021), Biondini *et al.* (2023) and Biondini & Gasperini (2023) we only consider cells in which at least one earthquake of magnitude $M_w \geq 4.0$ ($z \leq 50$ km) occurred on the Italian mainland during the period 1600–1959 according to the seismic catalogue CPTI15 (Rovida *et al.* 2022). Isolated cells not contiguous with the two main tessellations are excluded. The resulting grids are composed of 167 and 170 cells for R_1 and R_2 , respectively (Fig. 1). This approach is consistent with the methodology defined by Gasperini *et al.* (2021) to prevent the overestimation of the forecasting performance by excluding predominantly aseismic areas, such as the island of Sardinia.

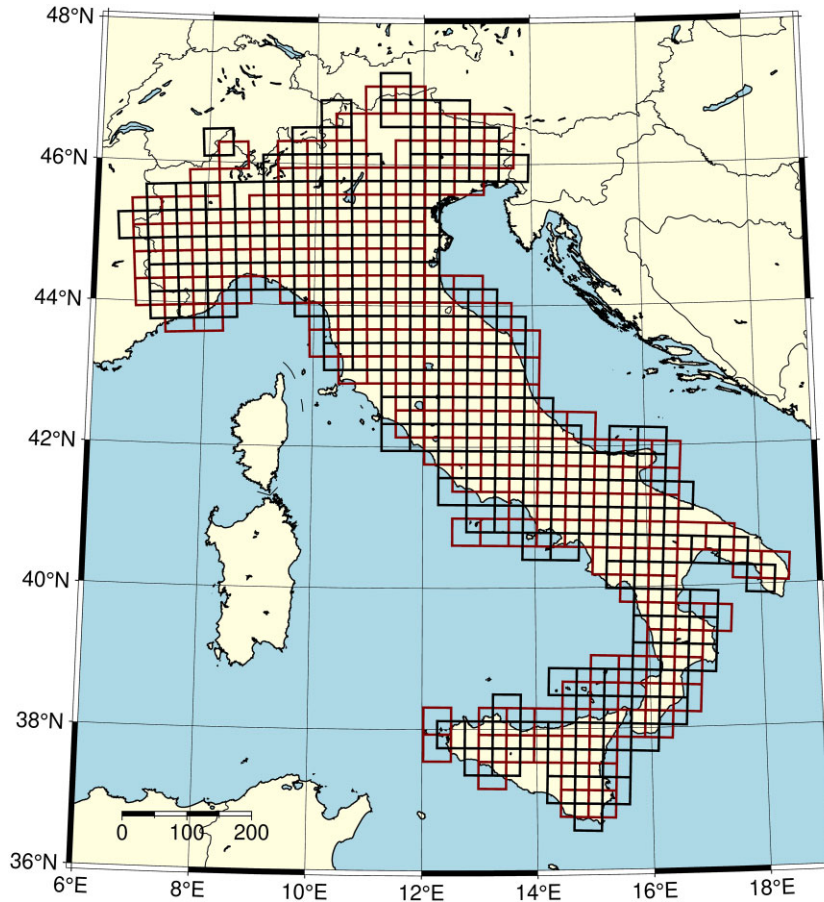


Figure 1. Analysis region used for the models' calibration, application and comparison. The red and black partially overlapped squares of side $L = 30\sqrt{2}$ km represent the tessellation R_1 and R_2 , respectively. Each cell contain at least one earthquake with $M_w \geq 4.0$ occurred inland from 1600 to 1959 according to the CPTI15 catalogue (Rovida *et al.* 2022).

To calibrate and pseudo-prospectively test the relative efficiency of various forecasting models, we use the HORUS catalogue (Lolli *et al.* 2020; see the Data Availability section), which collect earthquakes from 1960 to the present and is homogeneous in terms of magnitudes. To balance statistical resolution and the inherent uncertainties in magnitude measurements, we preliminarily bin the magnitudes to a single decimal digit. Hence, all magnitude thresholds mentioned below must be understood as reduced by half of the bin size (0.05).

During the learning period ranging from 1990 to 2004, the catalogue reports 18 target shocks with $M_w \geq 5.0$ and 8 with $M_w \geq 5.5$ within the double tessellation. In the testing period ranging from 2005 to 2022, it reports 34 target shocks with $M_w \geq 5.0$ and 12 with $M_w \geq 5.5$.

According to Lolli *et al.* (2020), the HORUS catalogue can be considered complete for the Italian mainland for $M_w \geq 4.0$ since 1960, for $M_w \geq 3.0$ since 1981, for $M_w \geq 2.5$ since 1990, for $M_w \geq 2.1$ since 2003 and for $M_w \geq 1.8$ since 2005.

However, for the purpose of determining the b -value, using the method of magnitude differences (van der Elst 2021; Tinti & Gasperini 2024), which is notably resilient to incompleteness, we re-estimated the completeness magnitude (M_c) using the maximum curvature method (Wiemer & Wyss 2000; Woessner & Wiemer 2005), which is less accurate but preserves more data. The seismicity distribution used for these calculations is illustrated in Fig. 3,

which shows the epicentres of earthquakes within the analysis region (R) and the CPTI15 polygon, highlighting the spatial coverage and density of seismic events in the study area. The re-estimated M_c are $M_w \geq 2.7$ since 1960, $M_w \geq 2.0$ since 1981, $M_w \geq 2.1$ since 1997, $M_w \geq 1.9$ since 2003 and $M_w \geq 1.2$ since 2005 (Fig. 4).

COMPUTATION OF b -VALUE

Up to a few years ago, the determination of the b -value was usually done through a formula, derived many decades ago by Utsu (1965) and Aki (1965), which assumes a continuous exponential distribution of magnitudes. If the magnitudes are binned to only one decimal digit, Utsu (1966) suggested an approximate correction to the formula consisting of subtracting one-half of the binning size (0.05) from the minimum magnitude of completeness of the data set. As the distribution of binned magnitudes is not continuous but discrete, Guttorp & Hopkins (1986) derived the exact formula, using the geometric distribution theory. Such formula, however, was almost totally neglected by the following literature (see Tinti & Gasperini 2024 for a thorough overview).

More recently, van der Elst (2021) showed that the b -value can be determined using the distribution of positive differences between magnitudes of successive shocks and, in this case, the estimation is more resilient to the incompleteness of the data set, particularly if the differences lower than a given trimming threshold $\Delta M_c'$ are

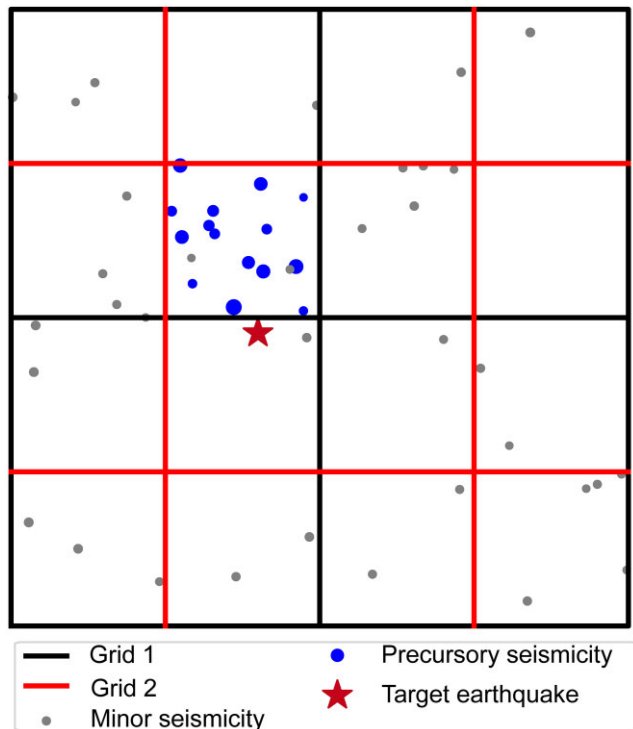


Figure 2. Sketch explaining the role of the double grid. The precursor seismicity (blue circles) is able to predict the target earthquake (red star) only using grid 2 (red lines) but not using grid 1 (black lines) because in the latter case it occurs close but outside the cell where the target occurs.

discarded. Tinti & Gasperini (2024) confirmed that the effect of incompleteness is mitigated by using trimmed magnitude differences but not that positive differences are significantly better than negative ones. However, as the so-called b^+ estimator proposed by van der Elst (2021) is largely used in the most recent literature, we adopt it even in this work. For computing the b -value, we use the expression proposed by Tinti & Gasperini (2024):

$$b = \frac{1}{2\delta \ln(10)} \ln \left[\frac{\overline{\Delta M} - \Delta M'_c + 2\delta}{\overline{\Delta M} - \Delta M'_c} \right], \quad (1)$$

where δ is equal to one half of the magnitude bin size (typically 0.05) and $\overline{\Delta M}$ is the average of the positive magnitude differences. Such expression is algebraically equivalent to that proposed by van der Elst (2021) in its eq. (A3) but can be preferred because it does not use the hyperbolic arc-cotangent function, which is not provided in some software environments. Although the b^+ estimator is more resilient than methods based on the distribution of magnitudes, it is not completely insensitive to incompleteness. Hence, a completeness analysis (Woessner & Wiemer 2005) is needed anyhow before using it.

van der Elst (2021), in analysing real sequences, considered two options: in the first one, he used all catalogued earthquakes and then applied the b^+ estimator above a minimum positive magnitude difference $\Delta M'_c$ ranging from 0.3 to 0.4 units, while in the second option, he used only the earthquakes with magnitude above the completeness magnitude M_c computed as that of maximum curvature of the FMD + 0.2 according to Woessner & Wiemer (2005) and applied the b^+ estimator to all positive magnitude differences. He obtained similar results in the two cases. However, as the true b -value of real sequences is not known, it is not possible to decide which strategy is the best. Both selection strategies

imply a reduction of the number of remaining data available for b -value determination but, according to the computation made by van der Elst (2021) for the 2019 Ridgecrest Sequence, the first strategy ($\Delta M'_c$ cut and not M_c cut) would seem to preserve more data.

Tinti & Gasperini (2024), by statistical analyses of simulated data sets, showed that the theoretical b -value is correctly determined when the trimming threshold is equal to the bin size (only $\Delta M = 0$ discarded), if the cutting magnitude M_c is assumed to be that of maximum curvature of the FMD (without + 0.2 correction), and when the data set is strongly incomplete data sets (without M_c cut), if using trimming thresholds four to five times the binning size. Such analyses also indicate that the number of remaining data is 20–30 per cent larger in the former case. Hence in the following, we will adopt the strategy of cutting magnitudes below the maximum curvature of the FMD and considering all positive differences ($\Delta M > 0$).

SETTING UP THE BVAL FORECASTING METHOD

We first remove all magnitudes below the point of maximum curvature of the FMD determined as reported in section analysis region and earthquake catalogue. The b -value is then calculated in each tessellation cell starting from the beginning of the catalogue, using a sampling window with a fixed number of events, $N_m = 150$ according to Gulia *et al.* (2024). For each sampling window, we compute the difference vector D obtained by subtracting the magnitude of each earthquake from the magnitude of the preceding one, that is, $D_i = (m_i - m_{i-1})$ for $i = 2, 3, \dots, N_m$. We then filter out any negative or zero difference and compute the average of the remaining ones denoted as $\overline{\Delta M}$. The latter is used in eq. (1) to calculate the b -value. We create a temporal series of the b -values by moving forward the sampling windows one event at time and recalculating the parameter for each new window until the end of the data set is reached. According to this forecasting model, alarms are triggered each time the b -value falls below a critical threshold value b_{th} , coming from a larger value. The alarm lasts for a duration Δt , starting from the time of the last earthquake in the sample of N_m events. If the b -value in the previous window was already below the threshold, no alarm is issued to avoid repeated warnings for the same low b -value condition. The choice of the optimal threshold b_{th} is determined in a specific calibration step where various thresholds are tested using the learning data set to find the value that best balances the sensitivity (detecting potential seismic events) and specificity (minimizing false alarms). Using the calibrated threshold, we evaluate pseudo-prospectively the efficiency of the forecasting method by applying it to the testing data set.

TESTING ALARM-BASED FORECASTING METHODS

A target earthquake is considered successfully predicted if it occurs within a cell where one or more alarm windows of duration Δt are active. Conversely, it is considered a failure to predict if it occurs outside any active alarm window. Thus, the prediction of a target earthquake is evaluated according to a dichotomy: predicted or not predicted.

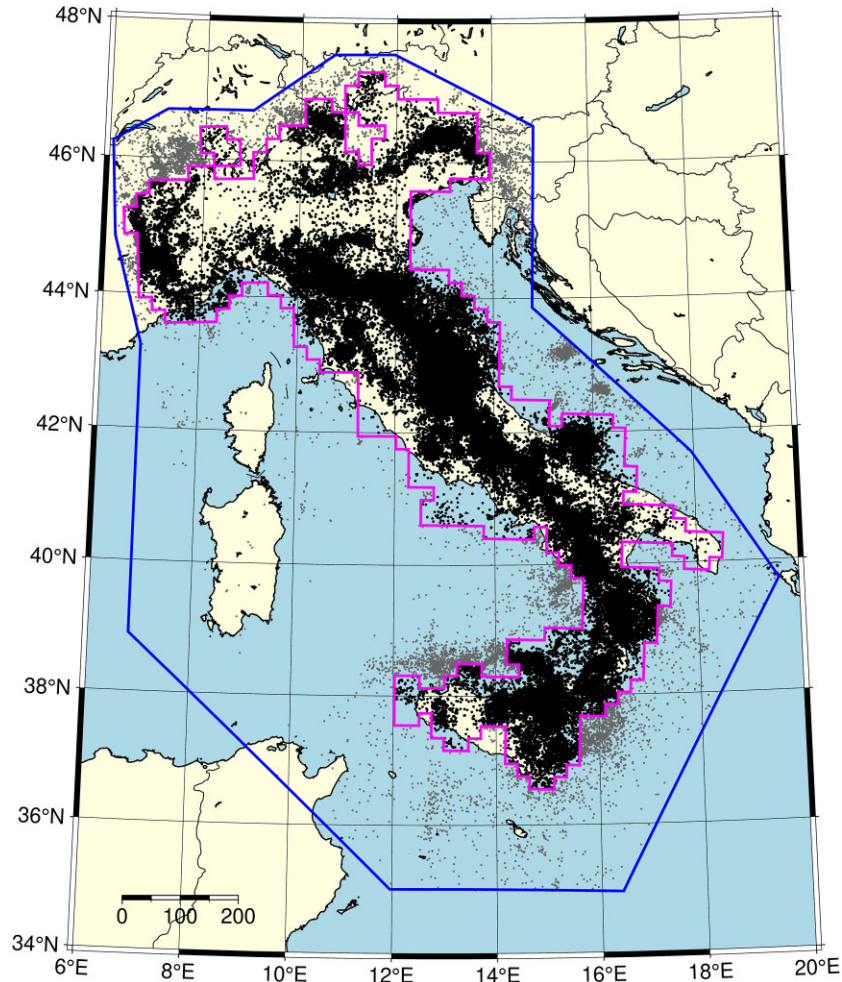


Figure 3. Seismicity map of Italy for the period 1990–2023 from HORUS catalogue (Lolli *et al.* 2020). Black dots indicate the epicentres of earthquakes within the analysis region (R) represented by the magenta shape, while grey dots indicate the epicentres of earthquakes within the CPTI15 polygon (Rovida *et al.* 2020) represented by the blue shape.

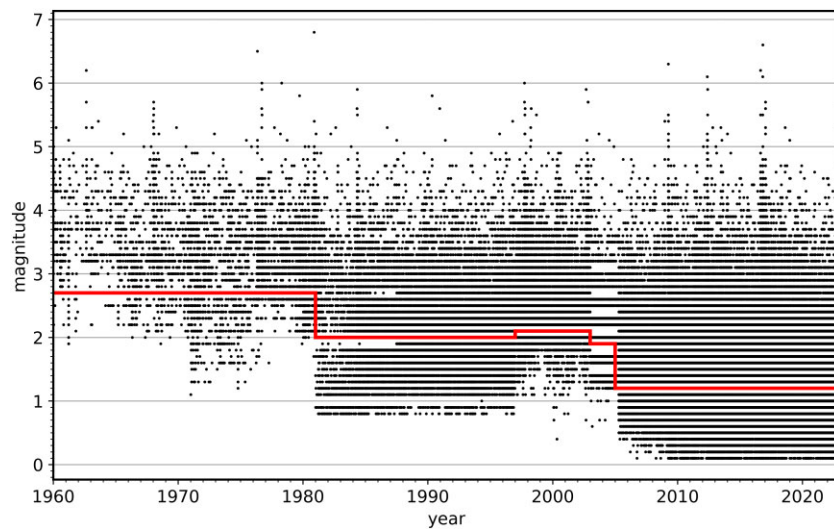


Figure 4. Time versus magnitude distribution of earthquakes (black dots) occurred within the analysis regions from 1960 to 2023 ($z \leq 50$ km) according to the HORUS catalogue (Lolli *et al.* 2020) and used for the models' calibration, testing and optimization. The red line represents the magnitude of completeness estimated using the maximum curvature method (Wiemer & Wyss 2000).

According to Molchan (1990, 1991) and Molchan & Kagan (1992), the miss rate ν , that is the fraction of unpredicted earthquakes, is computed as:

$$\nu = \frac{N - h}{N}, \quad (2)$$

where N is the total number of target events and h is the number of target events that were successfully predicted. Additionally, the overall fraction of space-time occupied by alarms is computed as:

$$\tau = \frac{1}{K} \sum_{i=1}^K \left(\frac{dc_i}{T} \right), \quad (3)$$

where K is the total number of cells, T is the total duration of the forecasting experiment and dc_i is the total time coverage of alarms within each i -th spatial cell. dc_i can be calculated by multiplying the alarm time Δt by the number n of defined alarms and then subtracting the sum of the intersections between the different alarm windows $\cap t_s$ as follows:

$$dc_i = (\Delta t \times n) - \sum \cap t_s. \quad (4)$$

Following Shebalin *et al.* (2011), we also calculate a weighted fraction of space-time occupied by alarms, which accounts for the long-term earthquake rate within each cell. Such rates are computed using the Italian historical earthquake data in the period 1620–1959, according to the CPTI15 V4.0 catalogue (Rovida *et al.* 2020, 2022; see the Data Availability section), following the procedure described in Gasperini *et al.* (2021). The weighted fraction of space-time occupied by alarms is then computed as:

$$\tau_w = \frac{\sum_{i=1}^K \bar{\lambda}_i \left(\frac{dc_i}{T} \right)}{\sum_{i=1}^K \bar{\lambda}_i}, \quad (5)$$

where $\bar{\lambda}_i$ is the long-term average rate of earthquakes $M_w \geq 4.0$ for each i -th cell. Details of weight computations are reported in Table S1 (Supporting Information). This weighting procedure penalizes forecasts in highly seismic areas, where predictions are inherently easier, by increasing the space-time fraction occupied by alarms proportionally to the historical earthquake rate.

The miss rate ν and the fractions of space-time occupied by alarms τ_u (unweighted) or τ_w (weighted) are used to construct the so-called Molchan diagram (Molchan 1990, 1991, 1997; Molchan & Kagan 1992; Molchan & Keilis-Borok 2008). The latter is a graphical tool commonly used to assess the forecasting performance of alarm-based models (see for example the one reported in Fig. 5).

In a paradoxical forecasting method that does not issue any alarm, we have $\tau = 0$ and it is impossible to forecast any target event. This corresponds to the point $(\tau, \nu) = (0 \text{ per cent}, 100 \text{ per cent})$ in the upper left corner of the diagram of Fig. 5. On the other hand, if the method issues a set of alarms that cover the entire space-time volume, all target earthquakes are certainly predicted. This corresponds to the point $(\tau, \nu) = (100 \text{ per cent}, 0 \text{ per cent})$. The diagonal line connecting these two points of equation $\nu = 1 - \tau$ represents the prediction performance of a purely random method that simply forecast earthquakes proportionally to the space-time fraction occupied by alarms.

By analysing different alarm durations Δt , we calculated the corresponding values of τ and ν for each duration, allowing us to plot a set of points on the Molchan diagram forming a path known as the ‘Molchan trajectory’. Models with trajectories that lie below the diagonal line and are close to the lower left-hand corner of coordinates $(\tau, \nu) = (0 \text{ per cent}, 0 \text{ per cent})$ show higher predictive capabilities than the random prediction model. In contrast, models

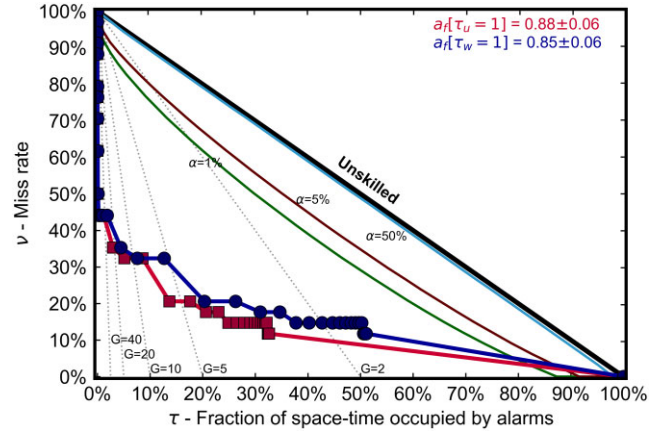


Figure 5. Molchan diagram and AS scores of the BVAL model for target earthquake with $M_w \geq 5.0$ occurred from 2005 to 2022. Red and blue lines indicate the Molchan trajectories for $b_{th} = 0.9$, using unweighted (red) and weighted (blue) space-time occupied by alarms (τ_u and τ_w), respectively (see the main text). The solid black diagonal line represents the Molchan trajectory of a purely random method, which forecasts earthquakes in proportion to the space-time occupied by alarms. This line divides the diagram into two parts: the area above the diagonal, indicating unskilled performance, and the area below the diagonal, indicating skilled performance. The light blue, violet and green lines represent the confidence limits for $\alpha = 50$, 5 and 1 per cent, respectively. The grey dotted lines indicate the probability gains $G = 2, 5, 10, 20$ and 40.

with trajectories extending above the diagonal line show lower predictive performance than the random model (see Fig. 5). The ratio between the success rate and the space-time fraction occupied by alarms,

$$G = \frac{1 - \nu}{\tau} \quad (6)$$

indicates the probability gain compared to the random chance (Kagan 2009). A value of $G = 1$ aligns with the random chance diagonal, while $G > 1$ characterises skilled forecasting methods.

The Molchan diagram provides insights into how the miss rate fluctuates with changes in the fraction of space-time covered by alarms, highlighting the essential trade-off between alarm coverage and minimizing failure rates. However, it remains primarily a visual tool. To quantitatively assess the forecasting performance of models, Zechar & Jordan (2008, 2010) introduced the area skill (AS) score $a_f(\tau)$. The latter quantifies how much better or worse a model performs compared to a random prediction baseline and is defined as the integral of the success rate $1 - \nu(\tau)$ normalized to the alarm space-time coverage τ :

$$a_f(\tau) = \frac{1}{\tau} \int_0^{\tau} [1 - \nu(t)] dt. \quad (7)$$

The overall AS score, calculated across the entire Molchan trajectory ranging from $\tau = 0$ per cent to $\tau = 100$ per cent, quantifies the area above the Molchan trajectory. The AS score is normalized such that its value ranges from 0 to 1, with values closer to 1 indicating better forecasting performance than the reference random model, represented by the diagonal line (Fig. 5), which is characterized by an overall AS score of 0.5.

Zechar & Jordan (2008, 2010) also derived a Gaussian asymptotical estimate of the variance of $a_f(\tau)$ as $\sigma^2 = \frac{1}{12} N$, where N indicates the number of target shocks. While the AS score can be computed for any τ , Zechar & Jordan (2008, 2010) suggested that the statistical power of the test generally improves with larger values

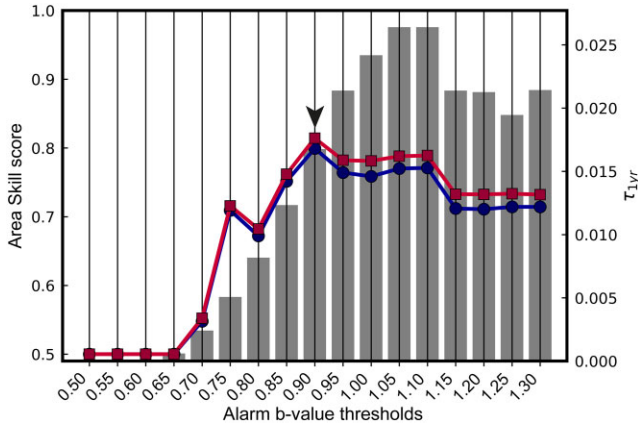


Figure 6. AS score trend for the BVAL model computed for targets earthquakes with magnitude $M_w \geq 5.0$, using the unweighted (red line) and weighted (blue line) fractions of space-time occupied by alarms. The grey bars represent the fractions of space-time occupied by alarms of duration equal to 1 yr ($\tau_{1\text{yr}}$). The chosen b -value threshold ($b_{\text{th}} = 0.9$) is characterized by the highest AS score and is indicated by the black arrowhead. The AS is represented on the left vertical axis, and the fractions of space-time occupied by alarms are reported on the right vertical axis.

of τ . Therefore, they recommended to use $a_f(\tau = 100 \text{ per cent})$ for hypothesis testing to maximize the sensitivity of the analysis.

CALIBRATION AND OPTIMIZATION OF FORECASTING MODELS

BVAL, FORE and ETAS models are applied to forecast target earthquakes with $M_w \geq 5.0$, which occurred inland in Italy during the learning period 1990–2004. As possible precursory seismicity, also the data in the preceding 15 yr interval from 1975 to 1989 is considered.

Following the approach of Biondini & Gasperini (2023), we used the maximum AS score as the criterion for selecting the optimal forecasting thresholds for each analysed model. Additionally, we also provide information on the fraction of space-time ($\tau_{1\text{yr}}$) occupied by alarms of one year duration ($\Delta t = 1 \text{ yr}$) for each analysed threshold. This metric indicates the proportion of time and geographic area where alarms were issued. Even if not used for the optimal threshold selection, this additional information provides a comprehensive view of the operational impact of the chosen thresholds. For example, a high fraction of space-time occupied by alarms may indicate a high number of false alarms, potentially reducing the model practical utility. Conversely, a too-low fraction may suggest that the chosen precursor is overly prudential and may miss target events.

For the BVAL model, in Fig. 6 (with numerical values reported in Table S2 of Supporting Information) we show the behaviour of the AS scores (lines) and of $\tau_{1\text{yr}}$ (grey bars) with the b -value threshold b_{th} varying from 0.50 to 1.30. The red and blue lines indicate the AS scores as a function of the unweighted and weighted fraction of space-time occupied by alarms, respectively. For both curves the highest overall scores ($a_f(\tau_u = 100 \text{ per cent}) = 0.81$ and $a_f(\tau_w = 100 \text{ per cent}) = 0.79$, respectively) are obtained for $b_{\text{th}} = 0.90$. For such threshold the fraction of space-time covered by alarms of duration $\Delta t = 1 \text{ yr}$ is $\tau_{1\text{yr}} = 0.017$ that is 1.7 per cent of the total duration of the learning time interval.

For the FORE model, an alarm is issued in a spatial cell every time a potential foreshock with $M_w = M \pm \Delta M$ occurs within

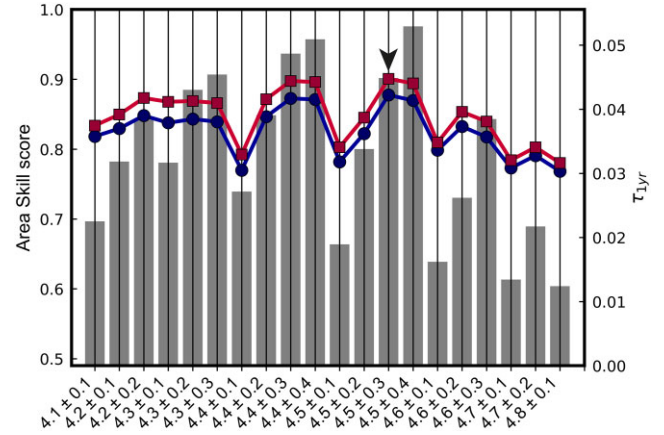


Figure 7. Same as Fig. 6 for the FORE model. The AS score and the fraction of space-time occupied by alarms are calculated based on the potential foreshock magnitude range. The arrowhead indicates the selected optimal magnitude range ($M_w = 4.5 \pm 0.3$).

that cell (Gasperini *et al.* 2021). The optimal range of potential foreshocks was determined by comparing the AS scores obtained in the learning interval by using magnitude ranges with central values $4.1 \leq M \leq 4.8$, and deviations from the central value $0.1 \leq \Delta M \leq 0.4$ but ensuring that the upper bound is not larger than 4.9 ($M + \Delta M \leq 4.9$), so that the magnitude of the foreshocks is lower than that of any target earthquakes (Fig. 7; with values in Table S3 of Supporting Information). The maximum overall AS scores ($a_f(\tau_u = 100 \text{ per cent}) = 0.90$ and $a_f(\tau_w = 100 \text{ per cent}) = 0.88$) are obtained for $M_w = 4.5 \pm 0.3$. For such ranges, the fraction of space-time covered by alarms of duration $\Delta t = 1 \text{ yr}$ is $\tau_{1\text{yr}} = 0.045$ that is 4.5 per cent of the total duration of the learning time interval.

For the ETAS model, an alarm is issued, in a spatial cell, whenever the expected daily rate (λ_{daily}) exceeds a given threshold. ETAS parameters were fitted using the maximum likelihood approach (see details in Appendix A) for the learning period (1990–2004), separately for the two tessellations, R_1 and R_2 (see Table A1). Following Biondini & Gasperini (2023), the expected rate of earthquakes with $M_w \geq 5.0$ is recalculated for each cell whenever an earthquake with $M_w \geq 2.5$ occurs within the entire observation region $R = (R_1 + R_2)$. Each cell is thus characterised by its own time-series of expected λ_{daily} .

The optimal alarm threshold λ_{th} for the expected daily earthquake rate is determined by varying it logarithmically from 5×10^{-7} to 9×10^{-3} (Fig. 8; with values in Table S4 in the Supporting Information). The maximum overall AS scores ($a_f(\tau_u = 100 \text{ per cent}) = 0.88$ and $a_f(\tau_w = 100 \text{ per cent}) = 0.83$) are obtained for $\lambda_{\text{th}} = 1 \times 10^{-5}$. For such threshold the fraction of space-time covered by alarms of duration $\Delta t = 1 \text{ yr}$ is $\tau_{1\text{yr}} = 0.337$ that is 33.7 per cent of the total duration of the learning time interval. Such $\tau_{1\text{yr}}$, definitely larger than the one obtained by optimizing other two methods, highlights, right from this optimization stage, a different behaviour of the ETAS model compared to the other two forecasting models, which might condition its practical applicability.

RESULTS OF THE PSEUDO-PROSPECTIVE TESTING

We used the optimal thresholds and ranges determined in the previous step for applying the simple model BVAL, FORE and ETAS to pseudo-prospectively predict 34 earthquakes with magnitudes

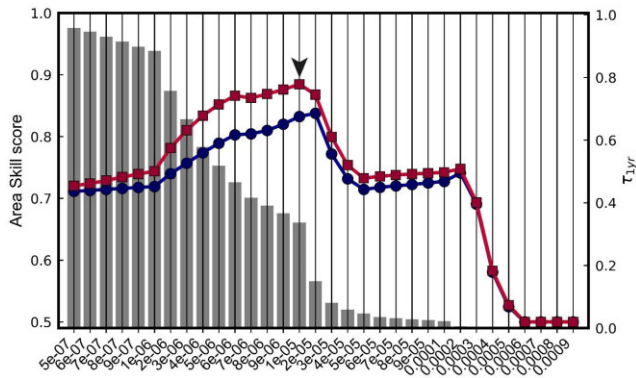


Figure 8. Same as Fig. 6 for the ETAS model. The AS score and the fraction of space-time occupied by alarms are calculated based on the daily expected rates thresholds. The arrowhead indicates the selected optimal alarm rate threshold $\lambda_{th} = 1.00E - 5$.

$M_w \geq 5.0$ and 12 earthquakes with magnitudes $M_w \geq 5.5$ occurred during the 18 yr of the pseudo-prospective test period 2005–2022. As possible precursor seismicity of targets events in the testing period 2005–2022, we used also the data in the preceding 18 yr interval from 1986 to 2004.

We use the same optimal thresholds even for applying the two ensemble models EADD and EMUL, obtained by summing and intersecting, respectively, the alarm windows of the two simple BVAL and FORE models.

In Fig 5, the Molchan trajectories of the BVAL model applied to the prediction of target earthquakes $M \geq 5.0$ are shown. They are obtained by varying the alarm time (Δt) from fractions of seconds to the total duration of the pseudo-prospective prediction experiment ($T = 18$ yr). The red and blue lines correspond to unweighted (τ_u) and weighted (τ_w) fractions of space-time occupied by alarms, respectively (see in Table S5 of the Supporting Information, the numerical values of the plotted trajectories). Both the red and blue Molchan trajectories lie well below the diagonal line representing the predictive performance of a purely random method and even below the 1 per cent confidence level threshold, highlighting their higher predictive ability. The BVAL model fails to predict the total number of target earthquakes with $M_w \geq 5.0$ even with $\Delta t = 18$ yr, for which only 30/34 (88.2 per cent) are predicted with time-space coverages of $\tau_u \approx 33$ per cent and $\tau_w \approx 51$ per cent. With $\Delta t = 1$ yr, approximately 68 per cent of target earthquakes are successfully forecasted, achieving space-time coverages of $\tau_u = 8.6$ per cent and $\tau_w = 12.7$ per cent. For a $\Delta t = 3$ months (0.25 yr), 64.7 per cent of targets earthquakes are successfully forecasted with $\tau_u = 2.96$ per cent and $\tau_w = 4.45$ per cent. Using $\Delta t = 1$ d, 50 per cent of target earthquakes are successfully forecasted, with $\tau_u = 0.05$ per cent and $\tau_w = 0.09$ per cent. The overall AS scores $a_f(\tau_u = 100 \text{ per cent}) = 0.88$ and $a_f(\tau_w = 100 \text{ per cent}) = 0.85$ are significantly larger of the expected score of the random model (0.5), based on the Student's t -test. The Molchan diagram for the BVAL model applied to predict target earthquakes with $M_w \geq 5.5$ is shown in Fig. S1 of Supporting Information. It shows high prediction skills, with overall AS scores of $a_f(\tau_u = 100 \text{ per cent}) = 0.92$ and $a_f(\tau_w = 100 \text{ per cent}) = 0.90$ for the unweighted and weighted fraction of space-time occupied by alarms, respectively. Even with $\Delta t = 18$ yr, the model forecasts only 11 of the 12 target events (92 per cent), with τ_u and τ_w remaining unchanged from the previous case due to consistent alarm issuance. For $\Delta t = 1$ yr and $\Delta t = 1$ d, 83 per cent (10/12) and 67 per cent (8/12) of target shocks are predicted, respectively. The

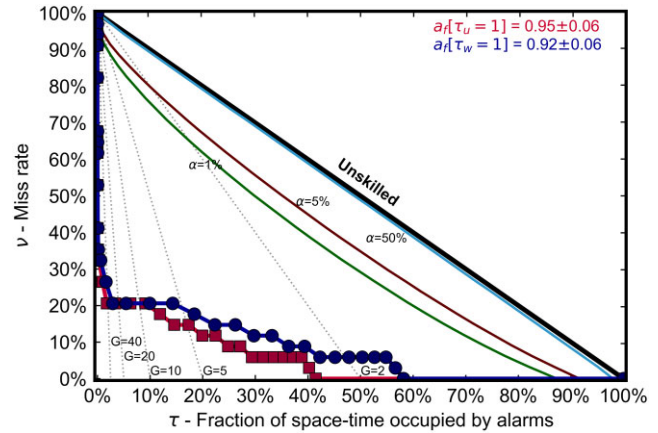


Figure 9. Same as Fig. 4 but for the FORE model applied to predict target shocks with $M_w \geq 5.0$.

numerical values of the plotted miss rates ν and of the AS scores as a function τ_u and τ_w for targets with $M_w \geq 5.0$ and $M_w \geq 5.5$ are reported in Tables S5 and S6 (Supporting Information), respectively.

For the FORE model, Fig. 9 shows the Molchan trajectories for target earthquakes $M_w \geq 5.0$. As for the BVAL model, also the FORE model overperform the prediction performance of the random method since the Molchan trajectories for both unweighted and weighted fraction of space-time occupied by alarms lies well below the diagonal line and the 1 per cent confidence level threshold. Unlike the BVAL model, the FORE model predicts all the target earthquakes with $M_w \geq 5.0$ for $\Delta t = 18$ yr, for which $\tau_u = 41.5$ per cent and $\tau_w = 58.2$ per cent. For $\Delta t = 1$ yr, approximately 79 per cent (27/34) of targets are predicted with $\tau_u = 3.4$ per cent and $\tau_w = 5.5$ per cent. With $\Delta t = 3$ months, 74 per cent (25/34) of target earthquakes are predicted with $\tau_u = 0.95$ per cent and $\tau_w = 1.58$ per cent. For $\Delta t = 1$ d, 47 per cent (16/34) of target earthquakes are predicted with $\tau_u = 0.01$ per cent and $\tau_w = 0.03$ per cent. The overall AS scores $a_f(\tau_u = 100 \text{ per cent}) = 0.95$ and $a_f(\tau_w = 100 \text{ per cent}) = 0.92$ are significantly higher than the random model's expected AS score (0.5). The Molchan diagram of the FORE model applied to $M_w \geq 5.5$ earthquakes (Fig. S2, Supporting Information) show very high prediction skills, with $a_f(\tau_u = 100 \text{ per cent}) = 0.98$ and $a_f(\tau_w = 100 \text{ per cent}) = 0.97$. The total number of target events (12) are predicted starting from $\Delta t = 9$ yr with $\tau_u = 24.8$ per cent and $\tau_w = 36.3$ per cent. For $\Delta t = 1$ yr and $\Delta t = 1$ d the 92 per cent (11/12) and the 58 per cent (7/12) of target shocks are predicted. Detailed numerical values and AS scores are provided in Tables S7 and S8 (Supporting Information).

For the ETAS model, Fig. 10 displays the Molchan trajectories for target earthquakes with $M_w \geq 5.0$. Like the two previous models, ETAS outperforms the random prediction method, as both the unweighted and weighted Molchan trajectories lie well below the diagonal line and even below the 1 per cent confidence level threshold. It predicts all target events with $\Delta t = 3$ d or larger, with $\tau_u = 30.1$ per cent and $\tau_w = 48.1$ per cent. With $\Delta t = 1$ yr, the fraction of space-time occupied by alarm increase to $\tau_u = 33.4$ per cent and $\tau_w = 53.2$ per cent. For $\Delta t = 1$ d, 97 per cent (33/34) of target earthquakes are predicted with $\tau_u = 20.8$ per cent and $\tau_w = 33.9$ per cent. The overall AS scores are $a_f(\tau_u = 100 \text{ per cent}) = 0.97$ and $a_f(\tau_w = 100 \text{ per cent}) = 0.95$. The Molchan diagram for the ETAS model applied to $M_w \geq 5.5$ earthquakes (Fig. S3, Supporting Information) show very high prediction skills, with $a_f(\tau_u = 100 \text{ per cent})$

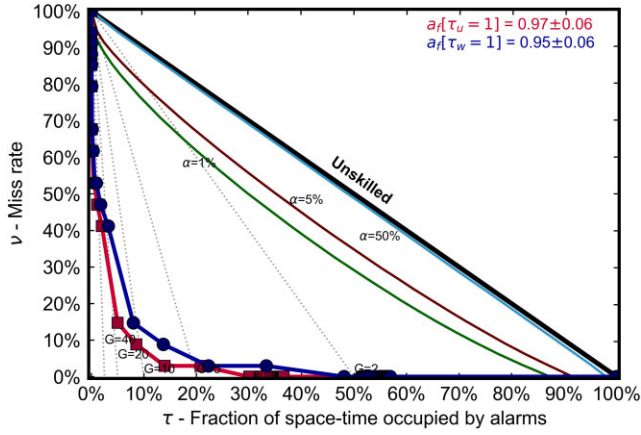


Figure 10. Same as Fig. 4 but for the ETAS model applied to predict target shocks with $M_w \geq 5.0$.

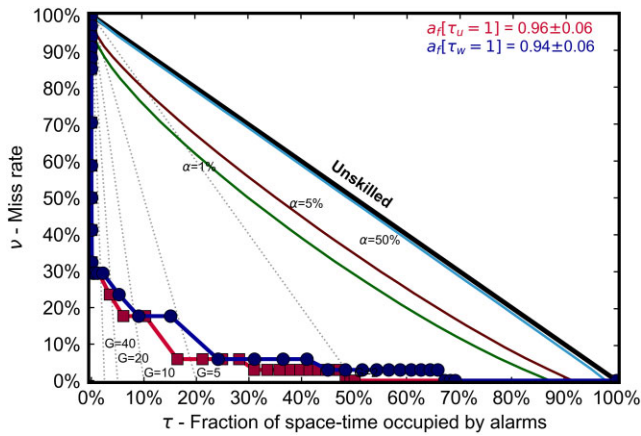


Figure 11. Same as the Fig. 5 but for the EADD ensemble model applied to predict target shocks with $M_w \geq 5.0$.

$= 0.99$ and $a_f(\tau_w = 100 \text{ per cent}) = 0.98$. All 12 target events are predicted starting from $\Delta t = 0.125$ d (3 hr), with corresponding $\tau_u = 5.0$ per cent and $\tau_w = 8.0$ per cent. Detailed numerical values and AS scores are provided in Tables S9 and S10 (Supporting Information).

For the ensemble EADD model, Fig. 11 shows the Molchan trajectories for target earthquakes $M_w \geq 5.0$. As for other models, it overperforms the prediction performance of the random method. It predicts all the 34 target earthquakes with $M_w \geq 5.0$ from $\Delta t = 16$ yr for which $\tau_u = 48.4$ per cent and $\tau_w = 67.3$ per cent. For $\Delta t = 1$ yr, approximately 82 per cent (28/34) are predicted with $\tau_u = 10.2$ per cent and $\tau_w = 15.1$ per cent. For $\Delta t = 3$ months, it predicts 76 per cent (26/34) of target earthquakes with $\tau_u = 3.5$ per cent and $\tau_w = 5.3$ per cent. For $\Delta t = 1$ d, 68 per cent (23/34) of target earthquakes are predicted with $\tau_u = 0.06$ per cent and $\tau_w = 0.11$ per cent. The overall AS scores $a_f(\tau_u = 100 \text{ per cent}) = 0.96$ and $a_f(\tau_w = 100 \text{ per cent}) = 0.94$ are significantly higher than the random model expected AS score (0.5). The Molchan diagram of the EADD model applied to $M_w \geq 5.5$ earthquakes (Fig. S4, Supporting Information) show very high prediction skills, with $a_f(\tau_u = 100 \text{ per cent}) = 0.97$ and $a_f(\tau_w = 100 \text{ per cent}) = 0.96$. The total number of target events (12) are predicted starting from $\Delta t = 6$ yr with $\tau_u = 31.0$ per cent and $\tau_w = 45.0$ per cent. For $\Delta t = 1$ yr and $\Delta t = 1$ d, the 92 per cent (11/12) and the 83 per cent (10/12) of target shocks are predicted, respectively. Detailed numerical values

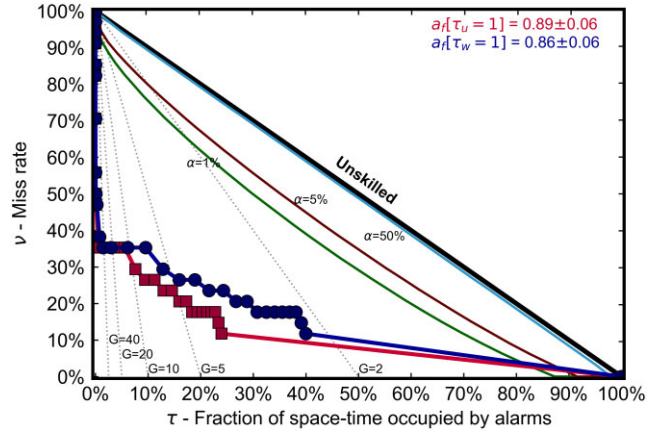


Figure 12. Same as Fig. 5 but for the EMUL ensemble model applied to predict target shocks with $M_w \geq 5.0$.

and AS scores are provided in Tables S11 and S12 (Supporting Information).

For the ensemble EMUL model, Fig. 12 shows the Molchan trajectories for target earthquakes $M_w \geq 5.0$. As for other models, it overperforms the prediction performance of the random method since both the Molchan trajectories for unweighted and weighted fraction of space-time occupied by alarms lies well below the diagonal line. Even with $\Delta t = 18$ yr, it predicts only 30 of 34 target earthquakes (88 per cent) with $M_w \geq 5.0$. For $\Delta t = 1$ yr, approximately 65 per cent (22/34) are predicted with $\tau_u = 1.7$ per cent and $\tau_w = 3.0$ per cent. For $\Delta t = 3$ months predicts 62 per cent (21/34) of target earthquakes with $\tau_u = 0.4$ per cent and $\tau_w = 0.8$ per cent. For $\Delta t = 1$ d, 29 per cent (10/34) of target earthquakes are predicted with $\tau_u = 0.003$ per cent and $\tau_w = 0.01$ per cent. The overall AS scores $a_f(\tau_u = 100 \text{ per cent}) = 0.89$ and $a_f(\tau_w = 100 \text{ per cent}) = 0.86$ are significantly higher than the random model's expected AS score (0.5). The Molchan diagram of the EMUL model applied to $M_w \geq 5.5$ earthquakes (Fig. S5, Supporting Information) show high prediction skills, with $a_f(\tau_u = 100 \text{ per cent}) = 0.94$ and $a_f(\tau_w = 100 \text{ per cent}) = 0.92$. Even with $\Delta t = 18$ yr, it predicts only 11 of 12 target earthquakes (92 per cent) with $M_w \geq 5.5$. For $\Delta t = 1$ yr and $\Delta t = 1$ d the 83 per cent (10/12) and the 42 per cent (5/12) of target shocks are predicted, respectively. Detailed numerical values and AS scores are provided in Tables S13 and S14 (Supporting Information).

ANALYSIS OF PERFORMANCE OF MODELS BASED ON ALARM DISTRIBUTION

Based on the overall AS score (computed for $\tau = 100$ per cent), the ETAS model appears to be the best performing model for predicting target earthquake with both $M_w \geq 5.0$ and $M_w \geq 5.5$. However, a closer examination reveals significant issues associated with its implementation using the alarm-based approach. During the 18-yr testing period, the ETAS model casts more than 1600 000 alarms. These alarms were concentrated in only 134 out of 337 cells, leading to an average of approximately 1.8 alarms per day in each cell. While this high frequency of alarms explain why ETAS can predict all target earthquakes even with very short alarm times (a few days for $M_w \geq 5.0$ and a few hours for $M_w \geq 5.5$) it also results in an overwhelming number of alerts. This could be impractical for a real

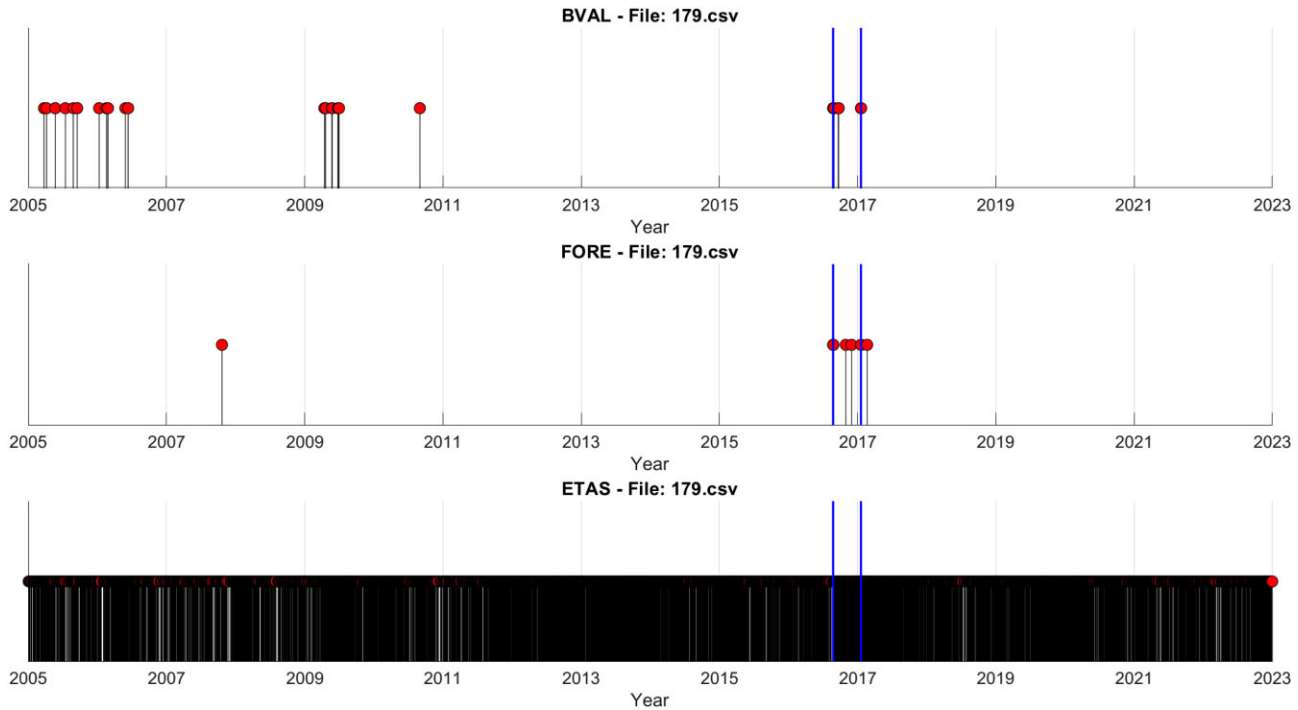


Figure 13. Time distribution of alarms for BVAL, FORE and ETAS forecasting methods in the cell nr. 179 where the 2016 August 24, Amatrice (M_w 6.2) main shock and other two strong aftershocks of 2017 January 18 (indicated with blue bars) occurred.

application, where the large number of alarms may complicate the decision-making process and strain the public resources.

Conversely, the BVAL and FORE models issues significantly fewer alarms during the same period, with the BVAL model generating 1796 alarms (less than one each year on average in each cell) and the FORE 782 alarms (about one every three years on average per cell).

The differences between these models are illustrated by their behaviour in Fig. 13 for cell 179, which includes three targets (blue vertical bars): the 2016 August 24 Amatrice earthquake ($M_w = 6.2$) and two significant aftershocks occurred on 2017 January 18 ($M_w = 5.3$ and 5.7). All the three methods are able to forecast the three target shocks with a minimum Δt of 6 hr for BVAL and FORE, and of about 30 s for ETAS. ETAS alarms (pins) are almost continuous throughout the 18-yr testing period, even when considering short alarm times Δt of just a few hours or minutes. This nearly constant state of alert reflects ETAS's high sensitivity but also highlights its potential for generating 'alarm fatigue', where the frequency of alarms diminishes their credibility (Bliss *et al.* 1995; Graham & Cvach 2010).

By comparison, the alarms issued by the BVAL and FORE models in the same cell are much sparser. While these models may not capture every possible precursor signal, they offer a more balanced approach, issuing alarms selectively and potentially providing more reliable warnings. This contrast highlights the fundamental trade-off in earthquake forecasting between the comprehensive coverage of ETAS and the more targeted and practical utility of BVAL and FORE. This comparison offers useful insights into the practical effects of each forecasting method, helping to determine which model is the most suitable, based on the desired balance between sensitivity and reliability.

As noted by Zechar & Jordan (2008, 2010), the overall AS Score (computed for $\tau = 100$ per cent) is a valuable criterion to compare

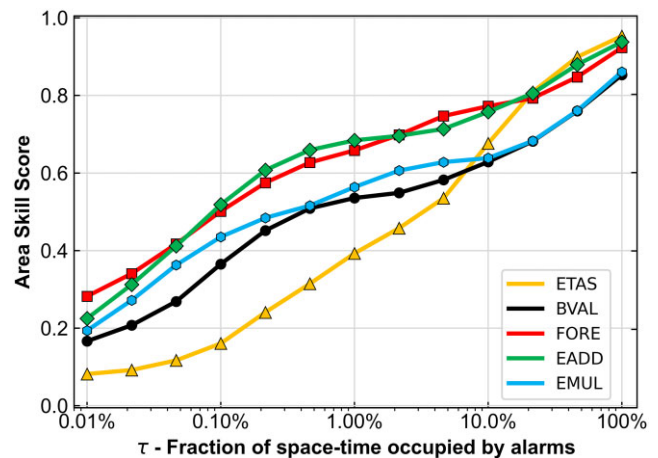


Figure 14. AS scores of various forecasting models as a function of the weighted fraction of space occupied by alarms τ_w for $M \geq 5.0$ target shocks in the testing period 2005–2022.

forecasting models because its power generally improves with increasing τ . Although the ETAS model achieves high overall AS scores, the frequency and duration of the alarms it generates may limit its overall effectiveness. A forecasting method that can achieve a good predictive efficiency with alarms concentrated in relatively short-time intervals (like BVAL and FORE) could be probably more effective and may be better received by the users.

In Fig. 14, we compare the performance of various methods by reporting the AS scores, for targets with $M_w \geq 5.0$, for τ_w varying from 0.01 per cent to 100 per cent. For $\tau_w < 20$ per cent, the higher scores are obtained by the FORE and EADD methods. Even though the ETAS model has the highest score for $\tau_w \geq 20$ per cent, for $\tau_w < 4$ per cent it results even worse than BVAL and EMUL. Note

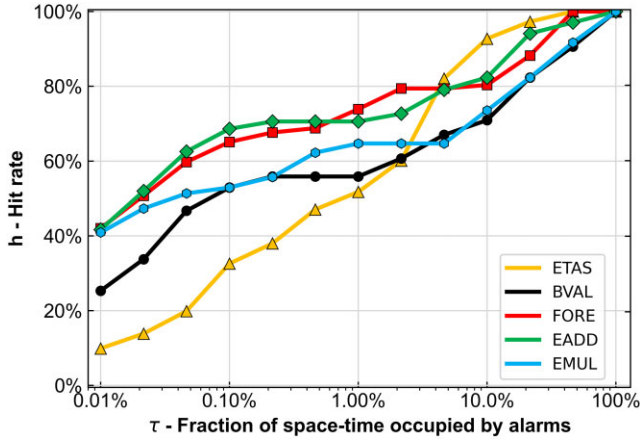


Figure 15. Hit rates of various forecasting models as a function of the weighted fraction of space occupied by alarms τ_w for $M \geq 5.0$ target shocks in the testing period 2005–2022.

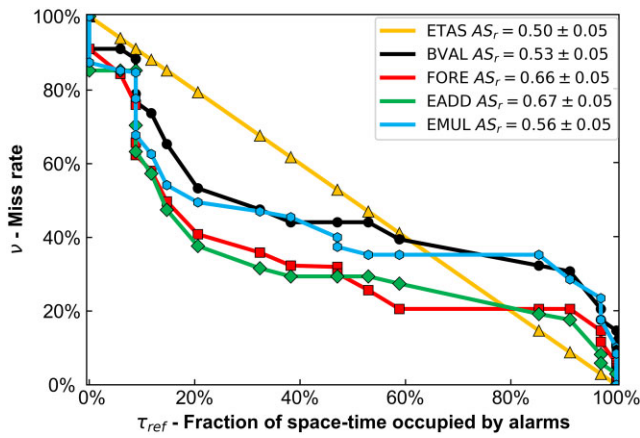


Figure 16. Molchan–Shebalin diagram and MSAS scores (see text) of all compared models for target earthquakes with $M_w \geq 5.0$, considering the ETAS model as reference and weighted fraction of space-time occupied by alarms (τ_w). The continuous diagonal line (orange) represents the miss rates of the ETAS model, distinguishing models with higher performance (below the line) from those with lower performance (above the line) compared to the reference.

that $\tau_w = 0.01$ per cent means that in each cell there is on average less than one hour of alarm each year, $\tau_w = 0.10$ per cent, less than 9 hr each year, $\tau_w = 1$ per cent, less than 4 d each year, and $\tau_w = 10$ per cent about 36 d.

Similar inferences can be drawn from Fig. 15 where the rate of successful predictions (hit rate, $h = 1 - \nu$) is plotted versus τ_w . In this case, ETAS is the worst model for $\tau_w < 2$ per cent and becomes the best one for $\tau_w > 5$ per cent, which mean about 8 and 20 d of alarm each year, respectively, on average in each cell. While ETAS excels in scenarios with extended alarm durations, methods like FORE and EADD may be more suitable for situations where the goal is to maintain high predictive accuracy with minimal disruption to daily life, especially where frequent alarms could lead to public desensitization.

MOLCHAN-SHEBALIN DIAGRAM

Following the methodology introduced by Shebalin *et al.* (2011), we can compare the miss rates (ν) of various analysed models to a

skilled reference model (for this study, the ETAS model) playing the role of the purely random prediction on the Molchan diagonal. The skilled reference model is characterized by its space-time fractions occupied by alarms (τ_{ref}) and miss rates (ν_{ref}). The miss rates (ν_{ref}) are plotted along the diagonal line of the Molchan diagram, where $\tau_{\text{ref}} = 1 - \nu_{\text{ref}}$. For the other models being compared, their expected miss rates (ν) must also be plotted against $\tau = 1 - \nu_{\text{ref}}$. Since these models may have been computed at different τ values compared to the τ_{ref} values of the ETAS model, linear interpolation is required to align the τ values appropriately. Specifically, for each τ_{ref} of the ETAS reference model, we compute the miss rates (ν_{int}) for the other models (BVAL, FORE, EADD and EMUL) using the following linear interpolation formula (Biondini & Gasperini 2023):

$$\nu_{\text{int}} = \frac{\nu_a - \nu_b}{\tau_a - \tau_b} (\tau_{\text{ref}} - \tau_b) + \nu_b, \quad (8)$$

where τ_a and τ_b are the fractions of space-time occupied by alarms for the compared models, immediately larger and smaller than τ_{ref} , respectively, and ν_a and ν_b the corresponding miss rates, respectively. We make such interpolation using as τ_{ref} both τ_u and τ_l of the reference ETAS model. The rationale for this interpolation is to provide a consistent basis for comparing the miss rates of different models at the same values of τ . Similarly to the Molchan diagrams showed above (Figs 5, and 9–12), if the miss rates of a compared model (ν_{int}) for the corresponding $\tau = 1 - \nu_{\text{ref}}$, are lower than ν_{ref} , such model has a better predictive ability than the reference model. This method provides a rigorous framework for evaluating the predictive capabilities of seismic forecasting models, allowing for a fair comparison based on equivalent conditions. For each model, we also compute a sort of a Molchan–Shebalin area skill (now on MSAS) score, according to eq. (7) where ν is substituted by ν_{int} and τ by τ_{ref} . For the reference ETAS model $\nu_{\text{int}} = \nu_{\text{ref}} = 1 - \tau_{\text{ref}}$, then its AS score is exactly 0.5.

Using the weighted τ_w of the ETAS model for $M_w \geq 5.0$ as τ_{ref} , in Fig. 16 (with values in Table 2) we show that all other models lie below the diagonal representing the ETAS model for $\tau_{\text{ref}} < 60$ per cent and the FORE and EADD models for $\tau_{\text{ref}} < 80$ per cent. All models have an overall MSAS score higher than the reference model ETAS. For FORE and EADD models, the score is significantly higher than for the ETAS according to the Student's t -test. The same plot for the $M_w \geq 5.5$ is reported in Fig. S6 of the Supporting Information. The results are like those of Fig. 16 with a slightly better performance of BVAL, FORE, EADD and EMUL with respect to ETAS.

DISCUSSION AND CONCLUSIONS

We developed the BVAL method to forecast potentially damaging target earthquakes ($M_w \geq 5.0$) in Italy, based on the temporal variations of the slope (b -value) of the frequency–magnitude distribution (Gutenberg & Richter 1944), and we compared it with other forecasting methods based on the statistical properties of seismicity: the FORE method using the occurrence of strong foreshocks (Gasperini *et al.* 2021), the ETAS method using the epidemic principle that any shock occurred acts as a trigger of the following ones (Ogata 1988), and two ensemble methods obtained by the union (EADD) and the intersection (EMUL) of space time alarm windows of BVAL and FORE.

We were inspired by the hypothesis formulated by Scholz (1968, 2015) and Amitrano (2003) that b -value decreases with increasing differential stress and with the increasing probability of occurrence

Table 2. Values of variables in Molchan–Shebalin plot of Fig. 12 for target shocks with $M \geq 5.0$ and unweighted fractions of space-time occupied by alarms (τ_u).

τ_{re}	ν_{ref}	$a_f(\tau_{re})$	ν_{BVAL}	$a_{BVAL}(\tau_{re})$	ν_{FORE}	$a_{FORE}(\tau_{re})$	ν_{EADD}	$a_{EADD}(\tau_{re})$	ν_{EMUL}	$a_{EMUL}(\tau_{re})$
0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000
0.000	1.000	0.000	0.912	0.000	0.893	0.000	0.853	0.000	0.853	0.000
0.059	0.941	0.029	0.907	0.091	0.805	0.151	0.853	0.147	0.850	0.149
0.088	0.912	0.044	0.865	0.098	0.708	0.182	0.853	0.147	0.819	0.154
0.088	0.912	0.044	0.786	0.098	0.643	0.182	0.690	0.147	0.695	0.154
0.088	0.912	0.044	0.759	0.098	0.600	0.182	0.608	0.147	0.618	0.154
0.118	0.882	0.059	0.725	0.138	0.543	0.243	0.555	0.215	0.556	0.219
0.147	0.853	0.074	0.638	0.174	0.475	0.293	0.460	0.270	0.515	0.268
0.206	0.794	0.103	0.519	0.245	0.399	0.370	0.366	0.361	0.484	0.335
0.324	0.676	0.162	0.474	0.339	0.350	0.463	0.315	0.469	0.471	0.403
0.382	0.618	0.191	0.441	0.370	0.324	0.494	0.294	0.504	0.441	0.425
0.471	0.529	0.235	0.441	0.406	0.317	0.529	0.294	0.542	0.381	0.455
0.471	0.529	0.235	0.441	0.406	0.299	0.529	0.294	0.542	0.368	0.455
0.529	0.471	0.265	0.441	0.423	0.254	0.550	0.294	0.560	0.353	0.476
0.588	0.412	0.294	0.400	0.438	0.206	0.572	0.277	0.576	0.353	0.493
0.853	0.147	0.426	0.325	0.500	0.206	0.641	0.201	0.633	0.353	0.541
0.912	0.088	0.456	0.324	0.511	0.206	0.651	0.176	0.645	0.279	0.550
0.971	0.029	0.485	0.206	0.525	0.154	0.661	0.105	0.658	0.235	0.562
0.971	0.029	0.485	0.176	0.525	0.118	0.661	0.059	0.658	0.176	0.562
1.000	0.000	0.500	0.147	0.534	0.059	0.669	0.039	0.666	0.108	0.570
1.000	0.000	0.500	0.147	0.534	0.059	0.669	0.029	0.666	0.105	0.570
1.000	0.000	0.500	0.133	0.534	0.059	0.669	0.029	0.666	0.105	0.570
1.000	0.000	0.500	0.120	0.534	0.059	0.669	0.029	0.666	0.105	0.570
1.000	0.000	0.500	0.118	0.534	0.059	0.669	0.029	0.666	0.104	0.570
1.000	0.000	0.500	0.117	0.534	0.059	0.669	0.029	0.666	0.104	0.570
1.000	0.000	0.500	0.116	0.534	0.059	0.669	0.029	0.666	0.103	0.570
1.000	0.000	0.500	0.115	0.534	0.059	0.669	0.029	0.666	0.102	0.570
1.000	0.000	0.500	0.114	0.534	0.059	0.669	0.029	0.666	0.101	0.570
1.000	0.000	0.500	0.114	0.534	0.059	0.669	0.029	0.666	0.101	0.570
1.000	0.000	0.500	0.113	0.534	0.059	0.669	0.029	0.666	0.100	0.570
1.000	0.000	0.500	0.112	0.534	0.059	0.669	0.029	0.666	0.100	0.570
1.000	0.000	0.500	0.112	0.534	0.059	0.669	0.029	0.666	0.099	0.570
1.000	0.000	0.500	0.112	0.534	0.059	0.669	0.029	0.666	0.099	0.570
1.000	0.000	0.500	0.111	0.534	0.059	0.669	0.029	0.666	0.099	0.570
1.000	0.000	0.500	0.111	0.534	0.059	0.669	0.029	0.666	0.099	0.570
1.000	0.000	0.500	0.111	0.534	0.059	0.669	0.029	0.666	0.098	0.570
1.000	0.000	0.500	0.111	0.534	0.059	0.669	0.029	0.666	0.098	0.570
1.000	0.000	0.500	0.111	0.534	0.059	0.669	0.029	0.666	0.098	0.570
1.000	0.000	0.500	0.111	0.534	0.059	0.669	0.029	0.666	0.098	0.570
1.000	0.000	0.500	0.111	0.534	0.059	0.669	0.029	0.666	0.098	0.570
1.000	0.000	0.500	0.111	0.534	0.059	0.669	0.029	0.666	0.098	0.570
1.000	0.000	0.500	0.111	0.534	0.059	0.669	0.029	0.666	0.098	0.570
1.000	0.000	0.500	0.111	0.534	0.059	0.669	0.029	0.666	0.098	0.570
1.000	0.000	0.500	0.111	0.534	0.059	0.669	0.029	0.666	0.098	0.570
1.000	0.000	0.500	0.000	0.534	0.000	0.669	0.000	0.666	0.000	0.570
1.000	0.000	0.500	0.000	0.534	0.000	0.669	0.000	0.666	0.000	0.570

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

of strong earthquakes. The same hypothesis brought to the formulation of the Foreshock Traffic Light System (Gulia & Wiemer 2010, 2019, 2021; Gulia *et al.* 2018) predicting whenever a shock with $M_w \geq 6.0$ is the strongest in a seismic sequence (green light, high b -value) or an even stronger shock is likely to occur close in space in the near future (red light, low b -value).

Our algorithm declares an alarm of duration Δt , which varies from a fraction of second to the entire duration of the experiment, in a square cell with sides of about 42.4 km, every time the b -value in that cell decreases below the threshold $b_{th} = 0.9$ from a higher value. Such threshold was determined by optimizing the ability of the algorithm to forecast earthquakes with $M_w \geq 5.0$ of the HORUS catalogue (Lolli *et al.* 2020) from 1990 to 2004.

If a target earthquake occurs in a cell where an alarm window is active, we count a hit, if a target earthquake occurs outside any alarm window we count a miss. To not miss the prediction of target earthquakes located only slightly outside an alarmed cell, we used a double tessellation with two grids staggered by half the grid side.

We pseudo-prospectively evaluated the predictive efficiency of the forecasting method using the AS score (Zechar & Jordan 2008, 2010) measuring the area above the Molchan trajectory (Molchan 1990, 1991), based on the data of the HORUS catalogue from 2005 to 2022. We obtained the Molchan trajectory by plotting the miss rate ν as a function of the fraction of space-time occupied by alarms τ , varying the alarm duration Δt from a fraction of second to the

total duration of the testing period (18 yr). The lower the miss rates of the Molchan trajectory the larger the AS score and then the more accurate the forecasting method.

We also penalized the forecasting performance of models in highly seismic areas, where random forecasts are easier, by weighting the fraction of space-time occupied by alarms proportionally to the long-term earthquake occurrence rate deduced from the catalogue CPTI15 V4.0 (Rovida *et al.* 2020, 2022) from 1600 to 1959.

Based on the overall AS score (computed for $\tau = 100$ per cent) the best performing model would be the ETAS, which achieves a near-perfect AS score (97–99 per cent). While this might indicate ETAS as the most effective model, it might also evidence a critical limitation of the Molchan diagram for testing models, like ETAS, issuing extremely frequent and dense alarms which results in a rapid reduction of the miss rate to zero even for notably short alarms of a few seconds or minutes. This creates an illusion of perfect predictive capability, as all target earthquakes are technically predicted, but at the cost of an overwhelming and nearly constant state of alarm in some cells.

This characteristic of ETAS, while beneficial in a purely statistical sense, undermines its broader applicability. The constant stream of alarms not only risks desensitising the population and creating ‘alarm fatigue’, effective management of emergency responses to natural disasters. In highly alarmed cells, the ETAS model alarms are so continuous that it becomes difficult to distinguish between periods of genuine increased risk and the model default state of alert. This diminishes its overall effectiveness for the operational utility where the goal is to balance predictive accuracy and minimizing unnecessary alerts.

More in general, we can note that high fractions (e.g. >10 per cent) of space-time occupied by alarms may lead to a diminished trust in the model, as the constant presence of alarms could make it harder to discern meaningful signals from the background noise. Hence, we also computed the AS scores of BVAL and other forecasting models even at smaller τ down to 0.01 per cent.

If compared with other methods, the BVAL is worse than FORE at any τ but better than ETAS for $\tau < 2 - 4$ per cent. The ensemble model EADD obtained by the sum of alarms issued by both BVAL and FORE is the best performing for space time occupations between 0.04 per cent to 2 per cent, which is the range of time-space occupation for which the forecasting algorithms might be more useful for civil protection purposes.

This preference for other methods with respect to ETAS is also confirmed by the Molchan–Shebalin diagram (Shebalin *et al.* 2011) in which the role of the random forecast along the Molchan diagonal is played by a reference skilled method (in our case the ETAS method itself). In fact, all other methods describe Molchan trajectories below the reference ETAS diagonal for $\tau < 60$ per cent and have the area above the trajectory, larger than 0.5 that corresponds to the ETAS reference model score. In summary, the BVAL forecasting method works rather well so confirming the validity of the basic hypothesis that low b -values might indicate the future occurrence of potentially damaging earthquakes. It is anyhow slightly worse than the FORE method based on the occurrence of strong foreshocks (Gasparini *et al.* 2021). However, BVAL in additive combination with the FORE improves the efficiency of the latter, particularly at relatively low fractions (<5 per cent) of space-time occupied by alarm that are those that are most likely to be useful for seismic hazard mitigations.

ACKNOWLEDGMENTS

We thank Rodolfo Console and Aron Mirwald for thoughtful comments and suggestions that helped to improve the paper. This paper benefitted from funding provided by the European Union within the ambit of the H2020 project RISE (no. 821115) and the ETHZ, which in particular fully financed the Postdoc grant of one of the authors (BE). We thank the Istituto Nazionale di Geofisica e Vulcanologia, Italy, grant ‘Progetto INGV Pianeta Dinamico (NEMESIS)’—code CUP D53J19000170001—funded by Italian Ministry MIUR.

SUPPORTING INFORMATION

Supplementary data are available at [GJIRAS](https://doi.org/10.1002/gjras) online.

Figure S1. Molchan diagram and AS score of the BVAL model for $b_{th} = 0.9$ applied to target earthquake with $M_w \geq 5.5$ occurred from 2005 to 2022. Red and blue lines indicate the Molchan trajectories using unweighted (red) and weighted (blue) space-time occupied by alarms (τ_{nw} and τ_w), respectively (see the main text). The solid black diagonal line represents the Molchan trajectory of a purely random method, which forecasts earthquakes in proportion to the space-time occupied by alarms. This line divides the diagram into two distinct parts: the area above the diagonal, indicating unskilled performance, and the area below the diagonal, indicating skilled performance. The light blue, violet and green lines represent the confidence limits for $\alpha = 50, 5$ and 1 per cent, respectively. The grey dotted lines indicate the probability gains $G = 2, 5, 10, 20$ and 40.

Figure S2. Same of Fig. S1 but for the FORE model applied for target earthquake with $M_w \geq 5.5$.

Figure S3. Same of Fig. S1 but for the ETAS model applied for target earthquake with $M_w \geq 5.5$.

Figure S4. Same of Fig. S1 but for the EADD model applied for target earthquake with $M_w \geq 5.5$.

Figure S5. Same of Fig. S1 but for the EMUL model applied for target earthquake with $M_w \geq 5.5$.

Figure S6. Molchan–Shebalin diagram and MSAS scores of all compared models (BVAL, FORE, ETAS, EADD and EMUL) for target earthquakes with $M_w \geq 5.5$, considering the ETAS model as reference and not-weighted fraction of space-time occupied by alarms (τ_w). The continuous diagonal line (orange) represents the miss rates of the ETAS model, distinguishing models with higher performance (below the line) from those with lower performance (above the line) compared to the reference.

Table S1. Computation of cell weights.

Table S2. Area skill scores of Fig. 6 for the BVAL method computed for targets earthquakes with magnitude $M_w \geq 5.0$.

Table S3. Area skill scores of Fig. 7 for the FORE method computed for targets earthquakes with magnitude $M_w \geq 5.0$.

Table S4. Area skill scores of Fig. 8 for the ETAS method computed for targets earthquakes with magnitude $M_w \geq 5.0$.

Table S5. Values of variables in Molchan diagram of Fig. 5 for the BVAL model applied to predict target earthquake with $M_w \geq 5.0$.

Table S6. Values of variables in Molchan diagram of Fig. S1 for the BVAL model applied to predict target earthquake with $M_w \geq 5.5$.

Table S7. Values of variables in Molchan diagram of Fig. 9 for the FORE model applied to predict target earthquake with $M_w \geq 5.0$.

Table S8. Values of variables in Molchan diagram of Fig. S2 for the FORE model applied to predict target earthquake with $M_w \geq 5.5$.

Table S9. Values of variables in Molchan diagram of Fig. 10 for the ETAS model applied to predict target earthquake with $M_w \geq 5.0$.

Table S10. Values of variables in Molchan diagram of Fig. S3 for the ETAS model applied to predict target earthquake with $M_w \geq 5.5$.

Table S11. Values of variables in Molchan diagram of Fig. 11 for the EADD model applied to predict target earthquake with $M_w \geq 5.0$.

Table S12. Values of variables in Molchan diagram of Fig. S4 for the EADD model applied to predict target earthquake with $M_w \geq 5.5$.

Table S13. Values of variables in Molchan diagram of Fig. 12 for the EMUL model applied to predict target earthquake with $M_w \geq 5.0$.

Table S14. Values of variables in Molchan diagram of Fig. S5 for the EMUL model applied to predict target earthquake with $M_w \geq 5.5$.

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.

CONFLICT OF INTEREST

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

DATA AVAILABILITY

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.

The QGIS geographical software was downloaded from www.qgis.org.

REFERENCES

- Aki, K., 1965. Maximum likelihood estimate of b in the formula $\log N = a - bM$ and its confidence limits, *Bull. Earthq. Res. Inst.*, **43**, 237–239.
- Amitrano, D., 2003. Brittle-ductile transition and associated seismicity: experimental and numerical studies and relationship with the b value, *J. geophys. Res.*, **108**(B1). doi:10.1029/2001JB000680
- 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. & Gasperini, P., 2023. Comparison between alarm-based and probability-based earthquake forecasting methods, *Geophys. J. Int.*, **235**(2), 1541–1551.
- Biondini, E., Rhoades, D.A. & Gasperini, P., 2023. Application of the EEPAS earthquake forecasting model to Italy, *Geophys. J. Int.*, **234**(3), 1681–1700.
- Bliss, J.P., Gilson, R.D. & Deaton, J.E., 1995. Human probability matching behaviour in response to alarms of varying reliability, *Ergonomics*, **38**, 2300–2312.
- Chan, C., Wu, Y., Tseng, T., Lin, T. & Chen, C., 2012. Spatial and temporal evolution of b -values before large earthquakes in Taiwan, *Tectonophysics*, **532–535**, 215–222.
- Chen, S.K., Chen, P.-Y., Wu, Y.-M., Chen, C.-C. & Chan, C.-H. (2023). Temporal variations of earthquake magnitude–frequency relation in the source area of $M \geq 6.0$ earthquakes: a systematic survey in Taiwan, *Earth Space Sci.*, **10**, e2023EA002927. doi:10.1029/2023EA002927
- Console, R. & Murru, M., 2001. A simple and testable model for earthquake clustering, *J. geophys. Res.*, **106**, 8699–8711.
- Console, R., Murru, M. & Lombardi, A.M., 2003. Refining earthquake clustering models, *J. geophys. Res.*, **108**(B10). doi:10.1029/2002JB002130
- Console, R., Rhoades, D.A., Murru, M., Evison, F.F., Papadimitriou, E.E. & Karakostas, V.G., 2006. Comparative performance of time-invariant, long-range and short-range forecasting models on the earthquake catalogue of Greece, *J. geophys. Res.*, **111**(B9), B09304. doi:10.1029/2005JB004113
- DeSalvio, N.D. & Rudolph, M.L., 2021. A retrospective analysis of b -value changes preceding strong earthquakes, *Seismol. Res. Lett.*, **93**, 364–375.
- El-Isa, Z.H. & Eaton, D.W., 2014. Spatiotemporal variations in the b -value of earthquake magnitude–frequency distributions: classification and causes, *Tectonophysics*, **615–616**, 1–11.
- Falcone, G., Spassiani, I., Ashkenazy, Y., Shapira, A., Hofstetter, R., Havlin, S. & Marzocchi, W., 2021. An operational earthquake forecasting experiment for Israel: preliminary results, *Front. Earth Sci.*, **9**, 1–10.
- Frankel, A., 1995. Mapping seismic hazard in the central and eastern United States, *Seismol. Res. Lett.*, **66**(4), 1–21.
- García-Hernández, R., D’Auria, L., Barrancos, J., Padilla, G.D. & Pérez, N.M., 2021. Multiscale temporal and spatial estimation of the b -value, *Seismol. Res. Lett.*, **92**, 3712–3724.
- Gasperini, P., Biondini, E., Lolli, B., Petruccioli, 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.
- Graham, K.C. & Cvach, M., 2010. Monitor alarm fatigue: standardizing use of physiological monitors and decreasing nuisance alarms, *Am. J. Crit. Care*, **19**, 28–34.
- Gulia, L. & Wiemer, S., 2010. The influence of tectonic regimes on the earthquake size distribution: a case study for Italy, *Geophys. Res. Lett.*, **37**(10), 1–6.
- Gulia, L. & Wiemer, S., 2019. Real-time discrimination of earthquake foreshocks and aftershocks, *Nature*, **574**, 193–199.
- Gulia, L. & Wiemer, S., 2021. Comment on “Two Foreshock Sequences Post Gulia and Wiemer (2019)” by Kelian Dascher-Cousineau, Thorne Lay, and Emily E. Brodsky, *Seismol. Res. Lett.*, **92**, 3251–3258.
- Gulia, L., Rinaldi, A.P., Tormann, T., Vannucci, G., Enescu, B. & Wiemer, S., 2018. The effect of a mainshock on the size distribution of the aftershocks, *Geophys. Res. Lett.*, **45**, 13277–87.
- Gulia, L., Wiemer, S., Biondini, E., Enescu, B. & Vannucci, G. (2024). Improving the foreshock traffic light systems for real-time discrimination between foreshocks and aftershocks, *Seismol. Res. Lett.*, **95**(6), 3579–3592.
- Guo, Z. & Ogata, Y., 1997. Statistical relations between the parameters of aftershocks in time, space, and magnitude, *J. Geophys. Res. Solid Earth*, **102**, 2857–2873.
- Gutenberg, B. & Richter, C.F., 1944. Frequency of earthquakes in California, *Bull. seism. Soc. Am.*, **34**, 185–118.
- Guttorp, P. & Hopkins, D., 1986. On estimating varying b values, *Bull. seism. Soc. Am.*, **76**, 889–895.
- Kagan, Y.Y., 2009. Testing long-term earthquake forecasts: likelihood methods and error diagrams, *Geophys. J. Int.*, **177**, 532–542.
- Lolli, B., Boschi, E. & Gasperini, P., 2009. A comparative analysis of different models of aftershock rate decay by maximum likelihood estimation of simulated sequences, *J. Geophys. Res. Solid Earth*, **114**(B01305), 1–25.
- 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. A double-branching model applied to long-term forecasting of Italian seismicity ($M_L \geq 5.0$) within the CSEP project, *Ann. Geophys.*, **53**(3), 31–39.
- Lu, S., 2017. Long-term b value variations of shallow earthquakes in New Zealand: a HMM-based analysis, *Pure appl. Geophys.*, **174**, 1629–1641.

- 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. & Keilis-Borok, V., 2008. Earthquake prediction: probabilistic aspect, *Geophys. J. Int.*, **173**, 1012–1017.
- Molchan, G.M. & Kagan, Y.Y., 1992. Earthquake prediction and its optimization, *J. geophys. Res.*, **97**, 4823. doi:10.1029/91JB03095
- 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., 1997. Earthquake prediction as a decision-making problem, *Pure appl. Geophys.*, **149**, 233–247.
- Montuori, C., Murru, M. & Falcone, G., 2016. Spatial variation of the b -value observed for the periods preceding and following the 24 August 2016, Amatrice earthquake (ML6.0) (Central Italy), *Ann. Geophys.*, **59**. doi:10.4401/ag-7273
- Murru, M., Console, R. & Falcone, G., 2009. Real time earthquake forecasting in Italy, *Tectonophysics*, **470**, 214–223.
- Murru, M., Console, R. & Lisi, A., 2004. Seismicity and mean magnitude variations correlated to the strongest earthquakes of the 1997 Umbria–Marche sequence (central Italy), *J. geophys. Res.*, **109**, B01304. doi:10.1029/2002JB002276
- Nuannin, P., Kulhanek, O. & Persson, L., 2005. 2004 Spatial and temporal b value anomalies preceding the devastating off coast of NW Sumatra earthquake of December 26, *Geophys. Res. Lett.*, **32**, L11307. doi:10.1029/2005GL022679
- Ogata, Y. & Zhuang, J., 2006. Space–time ETAS models and an improved extension, *Tectonophysics*, **413**, 13–23.
- Ogata, Y., 1988. Statistical models for earthquake occurrences and residual analysis for point processes, *J. Am. Stat. Assoc.*, **83**, 9–27.
- Ogata, Y., 1989. Statistical model for standard seismicity and detection of anomalies by residual analysis, *Tectonophysics*, **169**, 159–174.
- Ogata, Y., Jones, L.M. & Toda, S., 2003. When and where the aftershock activity was depressed: contrasting decay patterns of the proximate large earthquakes in southern California, *J. Geophys. Res. Solid Earth*, **108**(B6), 1–12.
- Öztürk, S., 2020. A study on the variations of recent seismicity in and around the Central Anatolian region of Turkey, *Phys. Earth Planet. Int.*, **301**, 1–11.
- Parsons, T., 2007. Forecast experiment: do temporal and spatial b value variations along the Calaveras fault portend $M \geq 4.0$ earthquakes?, *J. geophys. Res.*, **112**, 1–14.
- Pastorella, A.E., Murru, M., Taroni, M., Console, R., Montuori, C., Falcone, G. & Di Stefano, R., 2023. Temporal variations of seismicity rates and Gutenberg–Richter b -values for a stochastic declustered catalog: an example in Central Italy, *Seismol. Res. Lett.*, **94**, 1566–1578.
- 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., 2022. Parametric Catalogue of Italian Earthquakes (CPTI15), version 4.0, Istituto Nazionale di Geofisica e Vulcanologia (INGV), doi:10.13127/cpti/cpti15.4
- Scholz, C.H., 1968. The frequency-magnitude relation of microfracturing in rock and its relation to earthquakes, *Bull. seism. Soc. Am.*, **58**, 399–415.
- Scholz, C.H., 2015. On the stress dependence of the earthquake b value, *Geophys. Res. Lett.*, **42**, 1399–1402.
- Sharma, S., Hainzl, S. & Zöller, G., 2023. Seismicity parameters dependence on main shock-induced co-seismic stress, *Geophys. J. Int.*, **235**, 509–517.
- Shebalin, P., Narteau, C., Holschneider, M. & Schorlemmer, D., 2011. Short-term earthquake forecasting using early aftershock statistics, *Bull. seism. Soc. Am.*, **101**, 297–312.
- Tinti, S. & Gasperini, P., 2024. The estimation of b -value of the frequency-magnitude distribution and of its one-sigma intervals from binned magnitude data, *Geophys. J. Int.*, **238**, 433–458.
- Tsukakoshi, Y. & Shimazaki, K., 2008. Decreased b -value prior to the M 6.2 Northern Miyagi, Japan, earthquake of 26 July 2003, *Earth Planets Space*, **60**, 915–924.
- Utsu, T., 1965. A method for determining the value of b in a formula $\log n = a - bM$ showing the magnitude-frequency relation for earthquakes, *Geophys. Bull. Hokkaido Univ.*, **13**, 99–103.
- Utsu, T., 1966. A statistical significance test of the difference in b -value between two earthquake groups, *J. Phys. Earth*, **14**, 37–40.
- van der Elst, N., 2021. B-positive: a robust estimator of aftershock magnitude distribution in transiently incomplete catalogs, *J. Geophys. Res. Solid Earth*, **126**, 1–19.
- Varotsos, P.A., Sarlis, N.V. & Skordas, E.S., 2012. Order parameter fluctuations in natural time and b -value variation before large earthquakes, *Nat. Hazards Earth Syst. Sci.*, **12**, 3473–3481.
- Wiemer, S. & Wyss, M., 2000. Minimum magnitude of completeness in earthquake catalogs: examples from Alaska, the western United States, and Japan, *Bull. seism. Soc. Am.*, **90**, 859–869.
- Woessner, J. & Wiemer, S., 2005. Assessing the quality of earthquake catalogues: estimating the magnitude of completeness and its uncertainty, *Bull. seism. Soc. Am.*, **95**, 684–698.
- 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.
- Zhuang, J., Ogata, Y. & Vere-Jones, D., 2004. Analyzing earthquake clustering features by using stochastic reconstruction, *J. geophys. Res.*, **109**, 1–17.
- Zhuang, J., Werner, M.J., Zhou, S., Harte, D. & Hainzl, S., 2011. Basic models of seismicity: spatiotemporal models, Community Online Resource for Statistical Seismicity Analysis. doi:10.5078/corssa-07487583

APPENDIX A: ETAS MODEL FORMULATION AND PARAMETERS FITTING

The ETAS formulation and the background model SVP

The ETAS model, introduced by Ogata (1988), is a stochastic point process used to model aftershock sequences and provide probabilistic estimates of earthquake occurrences. The model addresses the limitation of using a single Omori–Utsu law to describe the temporal decay of aftershock sequences. Guo & Ogata (1997) and Ogata *et al.* (2003) demonstrated that significant aftershocks often generate their own secondary sequences of aftershocks. As a result, the ETAS model treats each aftershock as a potential trigger for additional aftershocks. The occurrence rate at a given time t is derived from the superposition of time-shifted Omori–Utsu decay functions, each weighted by the magnitude of the parent event (Ogata & Zhuang 2006).

The ETAS model is widely used for earthquake forecasting, including various applications in Italy (e.g. Console *et al.* 2006; Lombardi & Marzocchi 2010; Falcone *et al.* 2021; Biondini *et al.* 2023). The temporal dependence in these models is typically represented by a sum of

Omori's law decays, initiated at the times of each earthquake occurrence:

$$f(t) = \sum_{i=1}^n \frac{H(t-t_i) K e^{\alpha(m_i-m_c)}}{(t-t_i+c)^p}, \quad (\text{A1})$$

where K , c , p and α are free parameters, and $H(t-t_i)$ is the Heaviside step function, which equals 1 if $t-t_i > 0$ and 0 otherwise. n in the sum represents the total number of earthquakes considered, with each i -th term in the sum corresponding to an individual earthquake in the sequence. The earthquake productivity is assumed to be an exponential function of the magnitude of the i -th parent earthquake m_i and of the completeness magnitude m_c .

The spatial decay of the productivity can be described using various probability density function that represent the nucleation kernel (Ogata 1988; Console *et al.* 2003; Zhuang *et al.* 2004, 2011; Lombardi & Marzocchi 2010). In this work, we used the spatial probability density function proposed by Ogata & Zhuang (2006), where the smoothing term is an exponential function of the magnitude:

$$g(x, y) = \frac{(q-1) [D^2 e^{\gamma(m_i-m_c)}]^{q-1}}{\pi [(x-x_i)^2 + (y-y_i)^2 + D^2 e^{\gamma(m_i-m_c)}]^q}, \quad (\text{A2})$$

where q , D and γ are free parameters.

The frequency magnitude distribution of earthquakes adheres to the Gutenberg–Richter law (1944):

$$h(m) = \beta e^{-\beta(m-m_c)}, \quad (\text{A3})$$

where $\beta = b \ln 10$ and b is a free parameter.

By incorporating a time-invariant background seismicity term $\lambda_0(x, y, m)$, the rate density function for the ETAS model is given by:

$$\lambda(t, x, y, m) = \mu \lambda_0(x, y, m) + [f(t) g(x, y) h(m)], \quad (\text{A4})$$

which explicitly becomes:

$$\lambda(t, x, y, m) = \mu \lambda_0(x, y, m) + \left\{ \sum_{i=1}^n \frac{H(t-t_i) K e^{\alpha(m_i-m_c)}}{(t-t_i+c)^p} \frac{(q-1) [D^2 e^{\gamma(m_i-m_c)}]^{q-1}}{\pi [(x-x_i)^2 + (y-y_i)^2 + D^2 e^{\gamma(m_i-m_c)}]^q} \beta e^{-\beta(m-m_c)} \right\}. \quad (\text{A5})$$

The parameter μ represents the ratio between the expected number of independent events of the background seismicity $\lambda_0(x, y, m)$ and the total number of events.

To describe the background seismicity, we employ the spatially variable poisson (SVP) model. The SVP rate density is given by:

$$\lambda_0(x, y, m) = \mu_0(x, y) \beta e^{-\beta(m-m_c)}, \quad (\text{A6})$$

where $\mu_0(x, y)$ is the spatial rate density of earthquake with $m \geq m_c$, which is a continuous smoothed function of the geographic location. To assess the spatial rate density, we divide each cell of the tessellation into 16 subcells of side $l = \frac{30\sqrt{2}}{4}$ km. The number of earthquakes N_k with magnitude $m \geq m_c$ is calculated for each subcell. Each N_k value, representative of a single subcell, is then smoothed by a Gaussian filter with correlation distance d_c and normalized to preserve the total number of events as described in Frankel (1995). For each cell, the smoothed N_k is given by

$$\tilde{N}_k = \frac{\sum_l N_k \exp(-\Delta_{kl}^2/d_c^2)}{\sum_l \exp(-\Delta_{kl}^2/d_c^2)}, \quad (\text{A7})$$

where Δ_{kl} is the distance between the centre of the k -th and the l -th cells. To obtain \tilde{N}_k in terms of number of events per unit of time and area, it must be divided by the total duration of the earthquake catalogue and by the area of the cell. The spatial density $\mu_0(x, y)$ at any given point is calculated taking the weighted average of the four nearest surrounding subcells, where the weights are inversely proportional to the distance from each subcell to the point. To determine the optimal d_c of the SVP background model, we used the method proposed by Console & Murru (2001). This involves dividing the earthquake catalogue into two subcatalogues of approximately equal temporal length. The value of d_c is then chosen by maximizing the log-likelihood of one subcatalogue using the smoothed seismicity derived from the other subcatalogue. This analysis is performed separately for both subcatalogues, yielding $d_{c1} = 16.0$ and $d_{c2} = 13.0$. The optimal correlation distance d_c is given by the average of these two values, resulting in $d_c = 14.5$. Given the parameter d_c , the spatial density of earthquakes $\mu_0(x, y)$ for the SVP background model can be evaluated for each cell and point in space.

ETAS parameters fitting

The best fit for the free parameters of the ETAS model is achieved by maximizing the log-likelihood of an inhomogeneous Poisson point process, using the data set of earthquakes occurred in the learning period (1990–2004). The log-likelihood function is given by:

$$\ln L = \sum_{t_i \in (t_a, t_b); m_i \geq m_T; (x_i, y_i) \in R} \ln \lambda(t_i, m_i, x_i, y_i) - \int_{t_a}^{t_b} \int_{m_T}^{m_u} \iint_R \lambda(t, m, x, y) dy dx dm dt, \quad (\text{A8})$$

where $\lambda(t_i, m_i, x_i, y_i)$ is the rate density function of the ETAS model (eq. A5); (t_a, t_b) represent the fitting period; m_T and m_u are the lower and upper bounds of the magnitude range, set to 2.5 and 7.5, respectively; and R is the spatial region used for parameter fitting.

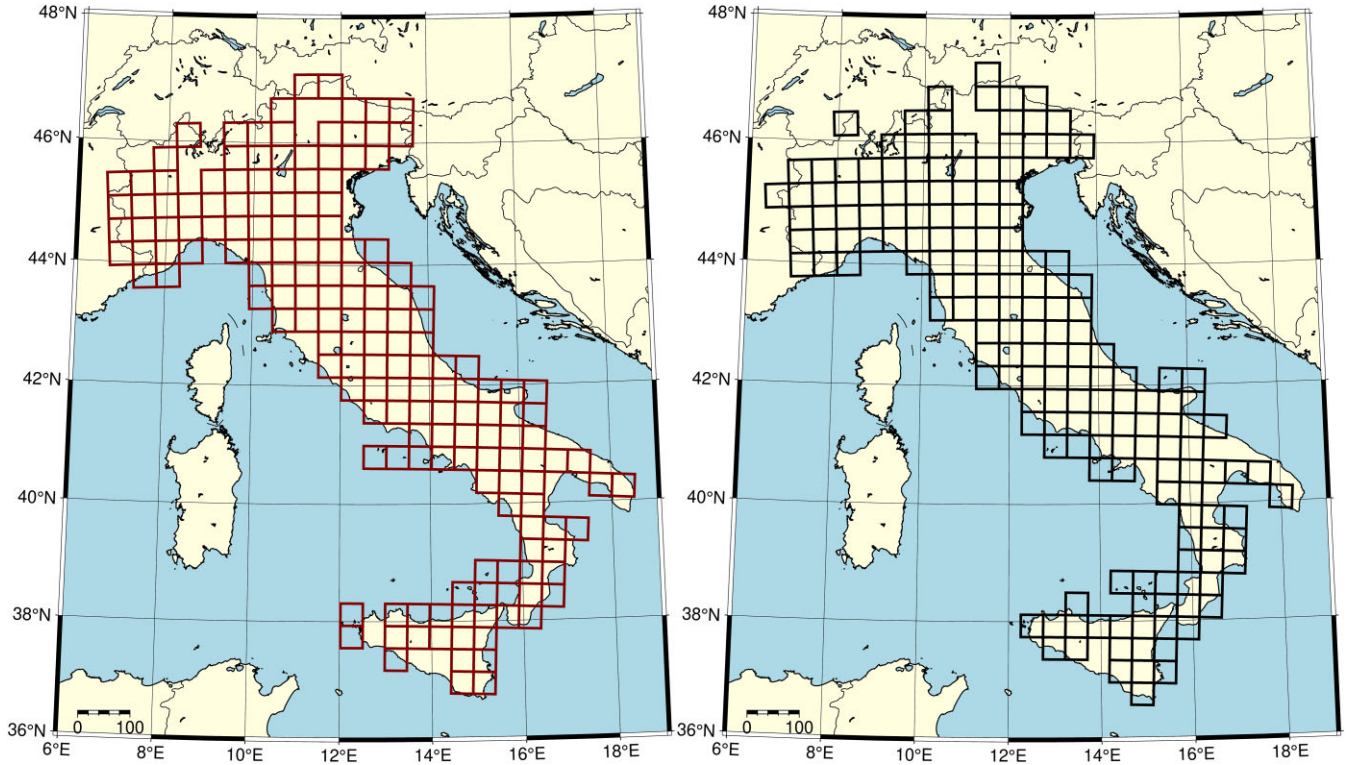


Figure A1. Tessellations R_1 (left) and R_2 (right) used to fit the optimal parameters of the ETAS model. Each cell has a side of $30\sqrt{2}$ km. See the main text for details.

Table A1. List of estimated best fit parameters for the ETAS model.

Parameter	Tessellation	
	R_1	R_2
b	1.035	1.033
k	0.0216 ± 0.0006	0.0228 ± 0.0006
c	0.0034 ± 0.0001	0.0034 ± 0.0001
p	1.0914 ± 0.0042	1.0884 ± 0.0039
d	1.0927 ± 0.0340	1.0735 ± 0.0320
γ	0.4499 ± 0.0565	0.4925 ± 0.0525
α	1.1544 ± 0.0314	1.1714 ± 0.0299
q	1.5*	1.5*
f_r	0.4560 ± 0.0076	0.4410 ± 0.0074
E	19.9814	21.2766
$\ln L$	326.2935	327.3842

*Parameter fitted separately (see text of Appendix A for more details).

The analysis region R consists of two different tessellations, which form a grid of partially overlapping cells. Consequently, the fitting of the free parameters was performed separately for each tessellation, R_1 and R_2 , (see Fig. A1). The b -value is estimated independently for each tessellation using the b^+ estimator introduced by van der Elst (2021) and recently explored by Tinti & Gasperini (2024). To assess the b -value, we use the seismicity data from the learning period (1990–2004) within each tessellation, resulting in b -values of 1.035 for R_1 and 1.033 for R_2 . The parameter q is set to 1.5 according to Lombardi & Marzocchi (2010) showing that the static stress change decreases with epicentral distance as r^{-3} . Similarly, the parameters k , p , c , α , d and ν have been fitted by maximizing the log-likelihood of earthquakes with magnitude $m \geq 2.5$ occurred within each individual tessellation, R_1 and R_2 , respectively, during the period 1990–2004 using the interior-point method (Wright 2004). The standard error (SE) for each parameter is computed as the square root of the corresponding diagonal element of the variance–covariance matrix computed as the inverse of the Hessian matrix for log-likelihood function at the maximum (Ogata 1988; Lolli *et al.* 2009). The optimal fitted parameters with their standard deviations are reported in Table A1. Additionally, the expected number of events (E) and the log-likelihood ($\ln L$), shown in Table A1 are calculated based on the fit to earthquakes with magnitudes $m \geq 5.0$. This approach ensures a more robust fit by initially using a broader range of magnitudes ($m \geq 2.5$), as described above, to obtain a more reliable

parameter estimation with a larger data set. Subsequently, the log-likelihood and the expected number of events for earthquakes with $m \geq 5.0$ are recalculated with the previously fitted parameters. This two step-process leverages the extensive data set for parameter estimation and then focuses on higher-magnitude events for assessing model performance.

Using the fitted parameters, the daily rate of target earthquakes ($m \geq 5.0$) is calculated for each cell within both the tessellations R_1 and R_2 for both the learning period (1990–2004) and testing period (2005–2023). The daily rate for each cell, which refers to the expected number of such events per day within the cell, is obtained by integrating the rate density function $\lambda(t, m, x, y)$ over the magnitude range $[m_T = 5.0, m_u = 7.5]$ and the spatial cell region C , while excluding the integration over time. The daily rate $\lambda_{\text{daily}}(m, x, y)$ is thus given by:

$$\lambda_{\text{daily}}(m, x, y) = \int_{m_T}^{m_u} \iint_C \lambda(t, m, x, y) \, dx \, dy \, dm. \quad (\text{A9})$$