Geophysical Journal International

Geophys. J. Int. (2021) **225,** 1192–1206
Advance Access publication 2020 December 16
GJI Seismology

doi: 10.1093/gji/ggaa592

Retrospective short-term forecasting experiment in Italy based on the occurrence of strong (fore) shocks

P. Gasperini [1,2] E. Biondini [1] B. Lolli [2] A. Petruccelli [1,3] and G. Vannucci [2]

[1] *Dipartimento di Fisica e Astronomia, Università di Bologna, Bologna, I-40127, Italy. E-mail: paolo.gasperini@unibo.it*
[2] *Istituto Nazionale di Geofisica e Vulcanologia, Sezione di Bologna, Bologna, I-40128, Italy*
[3] *Swiss Seismological Service, ETH Zurich, Zurich, CH-8092, Switzerland*

Accepted 2020 December 8. Received 2020 November 4; in original form 2020 April 28

SUMMARY

In a recent work, we computed the relative frequencies with which strong shocks ($4.0 \leq M_w < 5.0$), widely felt by the population were followed in the same area by potentially destructive main shocks ($M_w \geq 5.0$) in Italy. Assuming the stationarity of the seismic release properties, such frequencies can be tentatively used to estimate the probabilities of potentially destructive shocks after the occurrence of future strong shocks. This allows us to set up an alarm-based forecasting hypothesis related to strong foreshocks occurrence. Such hypothesis is tested retrospectively on the data of a homogenized seismic catalogue of the Italian area against a purely random hypothesis that simply forecasts the target main shocks proportionally to the space–time fraction occupied by the alarms. We compute the latter fraction in two ways (i) as the ratio between the average time covered by the alarms in each area and the total duration of the forecasting experiment (60 yr) and (ii) as the same ratio but weighted by the past frequency of occurrence of earthquakes in each area. In both cases the overall retrospective performance of our forecasting algorithm is definitely better than the random case. Considering an alarm duration of three months, the algorithm retrospectively forecasts more than 70 per cent of all shocks with $M_w \geq 5.5$ occurred in Italy from 1960 to 2019 with a total space–time fraction covered by the alarms of the order of 2 per cent. Considering the same space–time coverage, the algorithm is also able to retrospectively forecasts more than 40 per cent of the first main shocks with $M_w \geq 5.5$ of the seismic sequences occurred in the same time interval. Given the good reliability of our results, the forecasting algorithm is set and ready to be tested also prospectively, in parallel to other ongoing procedures operating on the Italian territory.

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

INTRODUCTION

Even if the deterministic prediction of earthquakes is presently not feasible and perhaps it will never be (Geller *et al.* 1997), several methods of probabilistic operational forecasting have been proposed in the last decades (see Jordan & Jones 2010 and Jordan *et al.* 2011 for an overview). Many of such methods take advantage of the well-known property of earthquakes to cluster in space and time (Mulargia & Geller 2003; Kagan 2014) and in particular of the possibility that relatively small shocks, occurring in advance (foreshocks) of destructive main shocks, might be used as precursory signal.

Jones & Molnar (1976, 1979) first observed that the property of worldwide strong earthquakes of being preceded by a few days or weeks of smaller shocks could have been used to predict somehow their occurrence. Jones (1984, 1985) noted that in California the occurrence of a weak shock increased of several order of magnitude the probability of occurrence of a main shock in the following hours or days and Agnew & Jones (1991) and Jones (1994) computed the probability of a major earthquake along the San Andreas fault in California, given the occurrence of a potential foreshock nearby the fault. The occurrence of foreshocks was then adopted as one of possible precursor of large earthquakes by the Southern San Andreas Working Group (1991) and Reasenberg (1999a,b) estimated the prospective frequency of potential foreshock being followed by stronger earthquakes in California and worldwide.

In Italy, Caputo *et al.* (1977, 1983) analysed earthquakes' swarms as forerunners of strong earthquakes, Grandori *et al.* (1988) proposed an alarm system based on the occurrence of a pair of foreshocks, Console *et al.* (1993) and Console & Murru (1996) studied

© The Author(s) 2020. Published by Oxford University Press on behalf of The Royal Astronomical Society. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com

Downloaded from https://academic.oup.com/gji/article/225/2/1192/6039343 by guest on 11 March 2021

the foreshock statistics and their possible relationship to earthquake prediction and Di Luccio *et al.* (1997) and Console *et al.* (1999) set up a forecasting hypothesis for the occurrence of earthquakes conditioned by prior events.

More recently, Gasperini *et al.* (2016), by the retrospective analysis of a homogeneous seismic catalogue of the Italian region, computed the relative frequencies with which strong shocks (defined as $4.0 \leq M_w < 5.0$) were followed in the same area by potentially destructive main shocks (defined as $M_w \geq 5.0, 5.5, 6.0$). In particular, they found that just after strong shocks, the relative frequency of potentially destructive main shocks in the same area increases with respect to quiet periods by a factor up to about 100 000. Then, as time goes by without any main shock occurring, such factor decreases logarithmically down to less than 10 for time windows of months to years. Within one day after the occurrence of a strong shock, the frequencies of main shocks with $M_w \geq 5.0$ and ≥ 5.5 range from 5 per cent to 2 per cent while within one month they range from 14 per cent to 6 per cent. Frequencies remain quite stable for about one hour after the strong shock and then start to decrease logarithmically at a rate of about 1 per cent for a doubling of the time elapsed from the strong shock. The frequencies of large main shocks ($M_w \geq 6.0$) are generally lower than 1 per cent except from about one month after a strong shock with $4.5 \leq M_w < 5.0$ when they become of the order of 4 per cent, but they decrease well below 1 per cent about two or three months after the strong shock if the main shock did not actually occur in the meantime. About 30 per cent of main shocks have been preceded by strong shocks in the day before, about 50 per cent one in the month before and about 60 per cent in the year before.

All such evidences suggest us to formulate an alarm-based forecasting hypothesis related to the simple occurrence of strong shocks in a given area. In this work, we first set up such hypothesis and then optimize it by the retrospective analysis of the HOmogenized instRUmental Seismic catalogue (HORUS) of the Italian area from 1960 to 2019 (Lolli *et al.* 2020) which is an improved and updated version of the seismic catalogue used by Gasperini *et al.* (2016).

In our knowledge, this is the first alarm-based forecasting experiments applied to the Italian region after the one by Grandori *et al.* (1988) cited above and after Console *et al.* (2010) and Murru *et al.* (2009) who converted to an alarm-based approach previous probabilistic forecasting studies by Console & Murru (2001) and Console *et al.* (2003, 2006). In fact, the latter studies, as well as others forecasting efforts in Italy (see Schorlemmer *et al.* 2010 and Marzocchi *et al.* 2014 for an overview), mostly based on the Epidemic-Type Aftershock Sequence (ETAS) model (Kagan & Knopoff 1987; Ogata 1988), were developed to reproduce at best the general behaviour of future seismicity, not to issue a warning of a possibly impending damaging earthquake.

The present forecasting hypothesis will be possibly submitted for prospective testing and validation to the testing facilities of the Collaboratory Study of Earthquake Predictability (Jordan 2006; Zechar *et al.* 2010).

SETTING UP THE FORECASTING HYPOTHESIS

We issue an alarm of duration Δt within a circular area (CA) of radius R every time a strong shock with $M_{min} \leq M < M_{max}$ occurs inside the CA. As target events to be forecasted we consider all the shock, with magnitude above a threshold $M_m \geq M_{max}$.

We must note that after the actual occurrence of a target shock, the forecast of further target shocks in the same area and in the following weeks or months is somehow favoured by the strong aftershocks of the previous target event. Hence, we also verify the ability of our method to forecast only the first target shock of each sequence. We then consider also a declustered set of target shocks obtained by eliminating those target shocks occurred within a distance $D = 50$ km and a time window of a year after another target shock of the sequence, even if they are larger than the first target shock of the sequence. This kind of declustering is somehow different with respect to that adopted for example in seismic hazard assessment (e.g. Gardner & Knopoff 1974; Reasenberg 1985) in which each sequence is usually represented by the largest shock, even if it is not the first one in the sequence. We choose the declustering space and time windows based on our experience on past Italian seismic sequences but we also checked visually that none possible secondary main shock remains not declustered. Also note that the chosen declustering windows approximately correspond to those determined by the algorithm of Gardner & Knopoff (1974) for $M = 5.5$.

As source areas we consider a regular tessellation of the Italian territory made of partially overlapping CAs with fixed radius R. Starting from an initial CA, centred at latitude 47° and longitude 7°, we compute the centres of the neighbour CAs by moving with steps $D = R\sqrt{2}$ both in longitude (from 7° to 19°) and in latitude (from 47° to 36°) covering then the whole Italian area with partial overlapping (Fig. 1).

Based on the results of our previous analysis (Gasperini *et al.* 2016), we choose a radius $R = 30$ km, as a good compromise between the opposing demands of having short spatial resolution and a sufficiently high number of earthquakes within each CA, so obtaining a total of 695 partially overlapping CAs. However, as the completeness of the seismic catalogue is poor in offshore areas, we consider in our analysis only the CAs within which at least one earthquake with $M_w \geq 4.0$ occurred inland from 1600 to 1959 (so as to be independent of the seismicity from 1960 to 2019 that will be used for the retrospective testing and optimization of the forecasting method), according to the CPTI15 catalogue (Rovida *et al.* 2016, 2020).

According to Gasperini *et al.* (2016), we consider as target shocks the earthquakes with $M_w \geq 5.0$, ≥ 5.5 and ≥ 6.0, which, in Italy, usually cause moderate, heavy and very heavy damage to buildings and none, a few and many victims respectively. Larger thresholds cannot be investigated because only three shocks with $M_w \geq 6.5$ (1976 Friuli with $M_w = 6.5$, 1980 Irpinia with $M_w = 6.8$ and 2016 Norcia with $M_w = 6.6$) occurred during the time interval covered by our seismic catalogue.

We count a success if a target shock occurs during one or more alarm time windows Δt and within one or more CA. On the contrary we count a missed forecast if a target shock occurs outside any alarm window of any CA. According to Molchan (1990, 1991), we compute the miss rate 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.

We also compute the total time duration d_c of alarms as the union of all alarm windows within each CA. This can also be computed by multiplying the window length Δt by the number n of issued alarms and then subtracting the sum of time intersections between

Downloaded from https://academic.oup.com/gji/article/225/2/1192/6039343 by guest on 11 March 2021

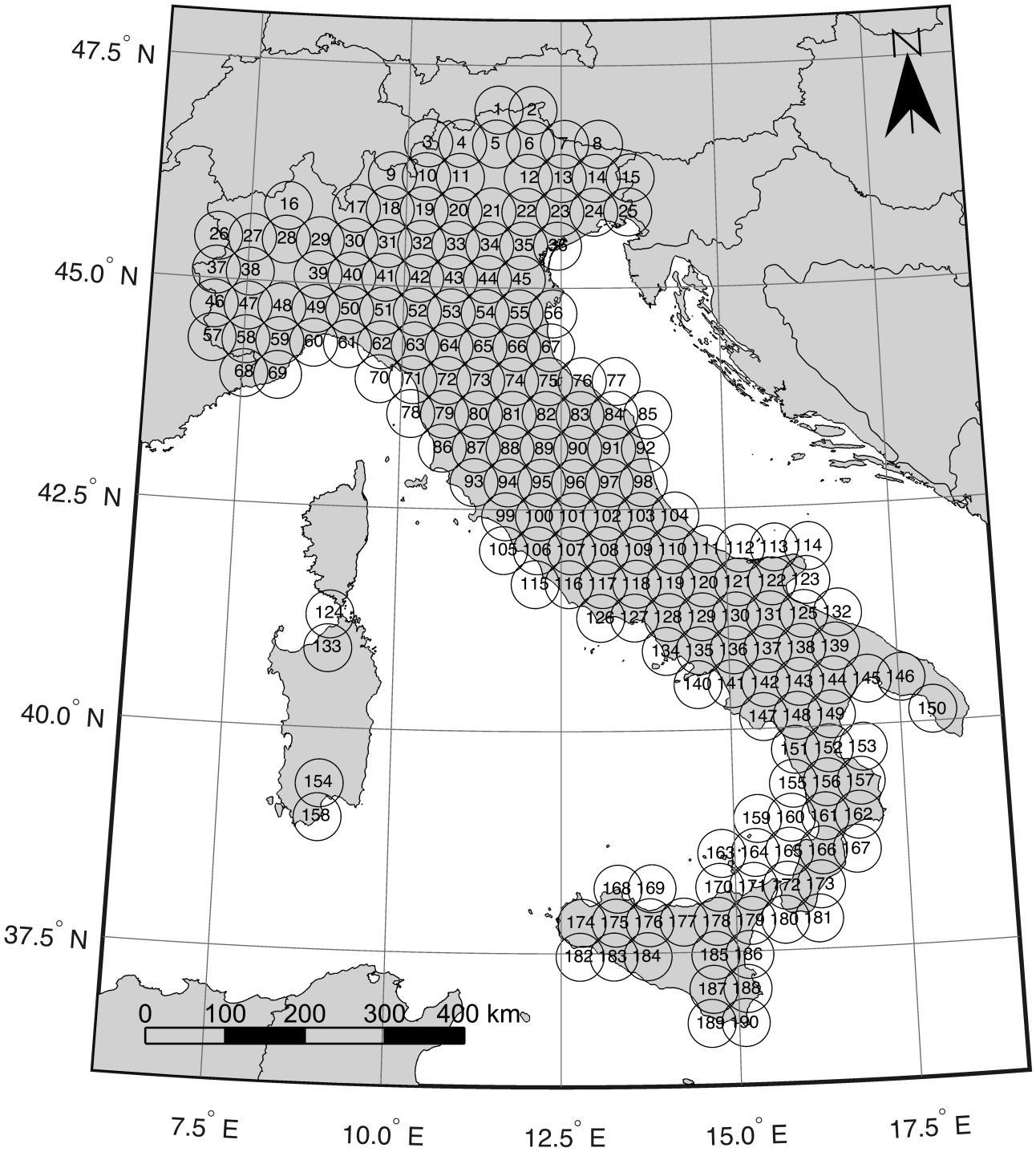

Figure 1. Tessellation of the Italian territory used for the retrospective forecast experiment. CA with $R = 30$ km within which at least one earthquake with $M_w \geq 4.0$ occurred inland from 1600 to 1959 according to the CPTI15 catalogue (Rovida *et al.* 2020).

alarm windows $\cap t_s$

$$d_c = \cup \Delta t = n\Delta t - \sum \cap t_s \tag{2}$$

The fraction of time occupied by alarms within each CA is then computed as

$$\tau_c = \frac{d_c}{T} \tag{3}$$

where T is the total duration of the forecasting experiment.

Finally, the overall fraction of space-time occupied by alarms is computed as the average of τ_c over all CAs

$$\tau_u = \frac{1}{M} \sum \tau_c \tag{4}$$

where M is the number of CAs. Note that such definition of fraction of space–time occupied by alarms is consistent with strong shocks occurring in the overlapping region of two adjoining CAs because in such case we sum the alarm fraction of time τ_c for both CAs.

Downloaded from https://academic.oup.com/gji/article/225/2/1192/6039343 by guest on 11 March 2021

Table 1. Magnitudes of completeness of the CPTI15 catalogue (Rovida *et al.* 2016, 2020).

Magnitude threshold Mc	Time interval of completeness	ΔT (yr)
$M_w \geq 4.5$	1880–1959	80
$M_w \geq 5.0$	1880–1959	80
$M_w \geq 5.5$	1780–1959	180
$M_w \geq 6.0$	1620–1959	340

Following Shebalin *et al.* (2011), we also compute the fraction of space–time occupied by alarms by weighting each alarm with the long-term rate of earthquakes within each CA. We compute such rate based on the data of the CPTI15 catalogue (Rovida *et al.* 2016, 2020) using different completeness thresholds Mc for different time intervals from 1620 to 1959 (Table 1). We count the numbers of earthquakes $N(Mc)$ above each magnitude threshold Mc occurred within each CA and within the corresponding time interval of completeness $\Delta T(Mc)$. Then we compute for each magnitude threshold the expected rate λ (event yr^{-1}) of earthquakes with $M_w \geq 4.0$, assuming the b-value of the frequency–magnitude distribution (Gutenberg & Richter 1944) equal to 1 (Rovida *et al.* 2020):

$$\lambda = \frac{N\,(Mc)}{\Delta T\,(Mc)}\,10^{M_c - 4.0} \tag{5}$$

In each CA, we then compute the average λ_{ave} of rates $\lambda > 0$ from different magnitude thresholds. For those CAs for which such average frequency cannot be computed because there are no earthquakes within the completeness time window of any magnitude threshold, we assign the minimum rate computed overall.

Finally, the weighted fraction of space–time occupied by alarms is computed from all CAs as

$$\tau_w = \frac{\sum \lambda_{ave}\,\tau_c}{\sum \lambda_{ave}} \tag{6}$$

See the details of such computations for each CA in Table S1 of the Supporting Information.

DATA SET USED FOR TESTING AND OPTIMIZATION

To test and optimize our algorithm, we apply it retrospectively to the HORUS catalogue of Italian instrumental seismicity from 1960 to 2019 (Lolli *et al.* 2020). For the time interval from 1960 to 1980, HORUS coincides with the data set prepared by Lolli *et al.* (2018) and that can be downloaded from the electronic supplement of such paper. For the period from 1981 to 2019, it is obtained by merging various data sources and homogenizing the magnitudes to M_w as described by Gasperini *et al.* (2012, 2013). The catalogue used here is updated up to the end of 2019 but we have implemented an automatic procedure able to continuously update such catalogue in near real-time (with daily to hourly updates) through the downloading of new data from online sources and the application of magnitude conversions (Lolli *et al.* 2020). We provide the final catalogue on the web (https://doi.org/10.13127/HORUS) for public dissemination and the possible prospective testing of the present and other forecasting methods.

The magnitude completeness threshold for the period 1960–1980 has been assessed by Lolli *et al.* (2018) to be about 4.0 whereas, according to Gasperini *et al.* (2013), it is definitely lower for the successive time periods. Such thresholds might be definitely larger in offshore areas owing to the large distances from the closest seismic stations, which are usually located on land (excepting for a few instruments deployed on the sea bottom). This is the reason why we only consider earthquakes with $M_w \geq 4.0$ occurred within the 190 CAs containing one inland earthquake at least. As our interest is to forecast earthquakes that potentially threaten lives and goods, we also limit the analysis to shocks shallower than 50 km. We show in Fig. S1 of Supporting Information the spatial distribution of inland earthquakes from the HORUS catalogue (Lolli *et al.* 2020) with $M_w \geq 4.0$ and depth < 50 km used for testing and optimization and in Fig. S2 in the Supporting Information the time distribution of magnitudes of all inland earthquakes with depth < 50 km.

The catalogue provides uncertainties for all magnitude estimates, ranging from less than 0.1 (for M_w estimated by moment tensor inversion) to about 0.5 (for M_w proxies from body wave magnitude mb observed by a few stations). In general, magnitude and location errors have the effect to increase the randomness of the catalogue and then to penalize skilled forecasting methods with respect to unskilled ones.

Owing to the Gutenberg Richter (1944) law, errors tend on average to overestimate all magnitudes because there are more earthquakes below a given threshold which can be overestimated than earthquakes above the same threshold which can be underestimated. The larger the error the larger the overestimation.

On the other hand, magnitude errors are generally larger for small earthquakes because the latter are observed by less stations and because accurate method of magnitude determination, like moment tensor inversion, cannot be applied to them. This means that in general small earthquakes are overestimated more than larger ones and then that foreshocks are overestimated more than target shocks.

One possible consequence in the present case is that errors in magnitude might improperly increase the number of alarms and then the space–time fraction occupied by alarms, particularly in earlier times when the coverage of seismic networks was coarser, so that to slightly underestimate the real skill of the method. Conversely the number of target shocks should not be affected much by magnitude errors because in HORUS catalogue the most (about 80 per cent) of $M_w \geq 5.0$ are accurately computed by moment tensor inversions.

TESTING AND OPTIMIZING THE FORECASTING HYPOTHESIS

We here follow the approach proposed by Zechar & Jordan (2008, 2010) based on the so-called 'Molchan error diagram' (Molchan 1990, 1991; Molchan & Kagan 1992). The latter consists of a plot (e.g. Fig. 2) of the miss rate ν (eq. 1) as a function of the fractions of space–time occupied by alarms τ (τ_u of eq. 4 or τ_w of eq. 6). For a paradoxical forecasting method not issuing any alarm, the space–time occupied by alarms is 0 and no target events can be forecasted (all target event are missed) then it is represented by the point $(\tau, \nu) = (0, 100$ per cent) at the upper left corner of the Molchan diagram. On the other hand, for a forecasting method issuing an alarm at any time and in any place, so occupying the entire space-time volume, no target events are missed and then the forecasting method is represented by the point $(\tau, \nu) = (100$ per cent, $0)$ at the lower right corner of the diagram. The points on the diagonal line connecting such two points (e.g. the black continuous line in Fig. 2), with equation

$$\nu = 1 - \tau \tag{7}$$

Downloaded from https://academic.oup.com/gji/article/225/2/1192/6039343 by guest on 11 March 2021

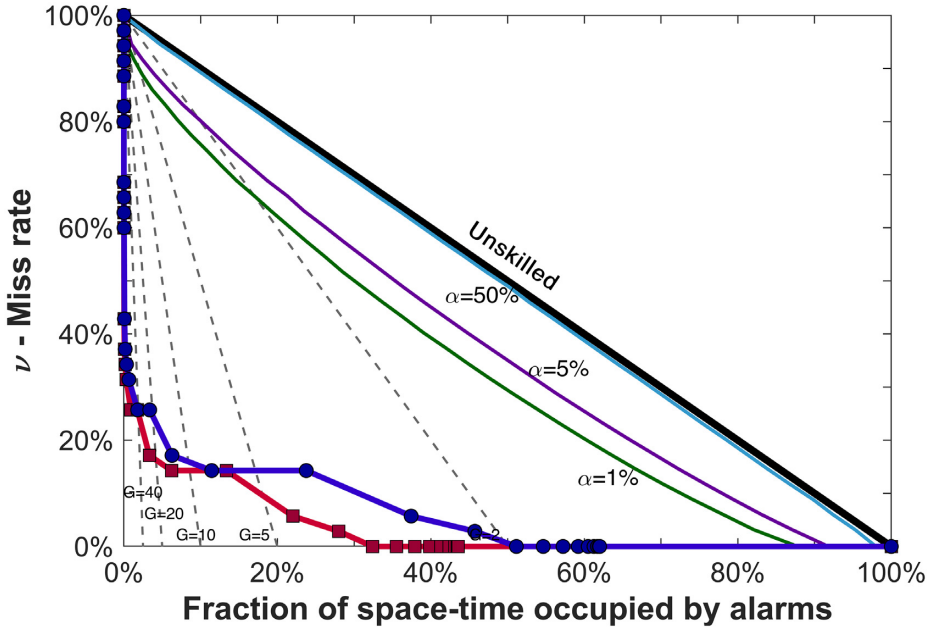

Figure 2. Molchan diagram for all target shocks with $M_w \geq 5.5$ (not-declustered). Red and dark blue lines indicate the forecasting performance of foreshocks with $4.4 \leq M_w < 4.8$ for unweighted (τ_u) and weighted (τ_w) fractions of space–time occupied by alarms respectively (see the text). The 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 $\alpha = 50$ per cent, 5 per cent and 1 per cent, respectively. The black dashed lines indicate probability gains $G = 2$, 5, 10, 20 and 40.

indicate the expected performance of a purely random forecasting method that simply forecasts target events proportionally to the space–time fraction occupied by the alarms.

On the diagonal line, the ratio between the success rate and the space-time fraction

$$G = \frac{1-\nu}{\tau} \tag{8}$$

is 1 for any τ, while for a skilled forecasting method, located below the line, $G > 1$ represents the 'probability gain' factor with respect to the random case.

Following Zechar & Jordan (2008), τ (τ_u or τ_w) can be assumed as the probability of forecasting a target events by chance and then can be used to measure the performance of a forecasting method under the reasonable assumption that the probability of having exactly h successful forecasts over N targets is given by the binomial probability function

$$B(h|N\tau) = \binom{N}{h} (\tau)^h (1-\tau)^{N-h} \tag{9}$$

Then, the cumulative probability of having by chance h or more successful forecasts is

$$\alpha = \sum_{n=h}^{N} B(n|N\tau) = 1 - \sum_{n=0}^{h-1} B(n|N\tau) \tag{10}$$

Such statistic allows to measure the skill of a forecasting methods, given the miss rate ν and the fraction of space–time occupied by alarms τ. In particular, the lower the statistic the higher the skill. Moreover, by inverting eq. (10), we can compute the expected miss rate ν at a given τ, for a hypothetical forecasting method with given probability α, and then to plot confidence limits on the Molchan diagram (e.g. the blue, violet and green lines in Fig. 2).

This statistic can be used to validate a forecasting method using a prospective data set (collected after the final fixing of the forecasting hypothesis) but even to optimize the forecasting hypothesis by searching the values of the parameters of the forecasting algorithm (if any) for which the statistic is minimum, by using a retrospective data set.

A given forecasting method with fixed parameter values is represented by a single point (τ, ν) on the Molchan diagram. However, one can even consider curves (Molchan trajectories) connecting different points referred to the same general forecasting approach but obtained by varying one of the free parameters of the forecasting algorithm. In our case, we can vary the alarm time window Δt from 0 to the total duration T of the experiment. In this way, we span the total space–time occupied by the alarms and correspondingly the number of successful forecasts, which increase with increasing Δt.

In the light of such definition, the diagonal line in the Molchan diagram can be seen as the Molchan trajectory of a purely random forecasting method. If a forecasting method performs better than the random one, its trajectory mainly lies in the lower left half of the Molchan diagram below the random line.

Zechar & Jordan (2008, 2010) proposed to use as a measure of the performance of an alarm-based forecasting method the integral of the success rate function $1 - \nu_f(\tau)$ normalized to the alarm space–time coverage τ

$$a_f(\tau) = \frac{1}{\tau} \int_0^{\tau} \left[1 - \nu_f(t)\right] dt \tag{11}$$

As the integral corresponds to the area above the Molchan random trajectory, the statistic was named area skill (AS) score. The AS score is normalized so that its value ranges between 0 and 1: the larger the statistic the better the performance.

The expected value of the AS score for a purely random method can be derived by substituting the eq. (7) of the random line $\nu_f(t) =$

Downloaded from https://academic.oup.com/gji/article/225/2/1192/6039343 by guest on 11 March 2021

$1 - t$ in eq. (11). This gives

$$\langle a_f(\tau)\rangle = \frac{1}{\tau}\int_0^{\tau}[1-(1-t)]\,dt = \frac{1}{\tau}\,\frac{\tau^2}{2} = \frac{\tau}{2} \qquad (12)$$

Such expectance function is represented in a plot as a function of τ by a straight line connecting the axes origin (0,0) with the point (100 per cent, 50 per cent) (e.g. the black line in Fig. 3). In such plot, the skilled forecasting methods lie above such random line.

Zechar & Jordan (2008, 2010) explored the AS score distribution and found that, for a continuous alarm function, the AS score at $\tau = 1$ is asymptotically Gaussian with a mean of 1/2 and a variance of 1/(12 N). They also found that the kurtosis excess is -6/(5 *N*) and hence, for *N* of the order of a dozen at least, the Gaussian approximation provides a good estimate of confidence bounds. Finally, they argued that even if the AS score can be computed for any τ, 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.

RESULTS OF RETROSPECTIVE TESTING

In Fig. 2, we show the Molchan trajectories for all target shocks (35) with $M_w \geq 5.5$ (not-declustered) preceded by strong shocks with $4.4 \leq M_w < 4.8$, by varying Δt from a width of a few seconds to the total duration $T = 60$ yr of the catalogue. 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 in the Supporting Information the numerical values of plotted curves).

The adopted foreshock M_w range ($M_w = 4.6 \pm 0.2$) was chosen after a comparative analysis of the relative performance of various ranges with lower and upper magnitude bounds varying from the completeness threshold of the catalogue ($M_w = 4.0$) to the minimum magnitude of target shocks ($M_w = 5.0$). Such analysis was aimed at maximizing the overall AS score and at the same time minimizing the total number of alarms (Fig. 4).

Both the red and dark blue lines in Fig. 1 lie well below the $\alpha = 1$ per cent confidence curve (green) for all explored Δt. All the target shocks are successfully forecasted ($\nu = 0$) for $\Delta t =$ 20 yr (corresponding to $\tau_u = 32$ per cent and $\tau_w = 51$ per cent) or larger. For $\Delta t = 1$ yr, about 83 per cent of target shocks (29) are successfully forecasted, with space–time coverages $\tau_u = 3.3$ per cent and $\tau_w = 6.3$ per cent. 40 per cent of target shocks (14) are forecasted with $\Delta t = 1$ d for which $\tau_u = 0.01$ per cent and $\tau_w = 0.03$ per cent. The AS diagram in Fig. 3 (see Table S2 in the Supporting Information for numerical values) confirms such good performance with the scores of the forecasting method (red and dark blue lines) well above the random expectation (black) and the 1 per cent confidence line (green) for any Δt. The overall AS scores $a_f\ (\tau_u = 1) = 0.96 \pm 0.05$ and $a_f\ (\tau_w = 1) = 0.94 \pm 0.05$, based on the Student's *t*-test, are significantly larger than the expectance of a random method (0.5) with significance level (s.l.) $\ll 0.01$.

As noted above the aftershocks produced by the first target shocks of seismic sequences may significantly contribute to forecast subsequent target shocks with $M_w \geq 5.5$ within the same sequence. We then proceed to analyse in the same way the declustered set of target shocks with $M_w \geq 5.5$ obtained by discarding all target shocks occurred within a spatial distance $R = 50$ km and a time window of a year after the first and all subsequent $M_w \geq 5.5$ shocks of the sequence. This reduces the number of considered target shocks with $M_w \geq 5.5$ from 35 to 14.

In Figs 5 and 6, we report the same plots as in Figs 2 and 3 but for the (declustered) set of only the first target shocks with $M_w \geq 5.5$ of each sequence (see Table S3 in the Supporting Information for numerical values). The performance is worse than for the not-declustered set but remains well below the random line and the $\alpha = 1$ per cent confidence curve in the Molchan diagram of Fig. 5 and also well above the $\alpha = 1$ per cent confidence line of AS diagram of Fig. 6. Even in this case all 14 target shocks are successfully forecasted with $\Delta t = 20$ yr or larger. For $\Delta t = 1$ yr, 64 per cent of target shocks (9) are forecasted and 29 per cent (4) for $\Delta t = 1$ d. The overall AS score $a_f\ (\tau_u = 1) = 0.93 \pm 0.08$ and $a_f\ (\tau_w = 1) = 0.87 \pm 0.08$ are lower than for the not-declustered set but anyhow they are significantly larger than the expectance (0.5) of a random method with s.l. $\ll 0.01$.

In Figs S3–S6 of Supporting Information, we report the same plots of Figs 2, 3, 5 and 6 for target shocks with $M_w \geq 5.0$ (numerical values in Tables S4 and S5, Supporting Information). The performance is definitely worse than for $M_w \geq 5.5$, but still better than the 1 per cent confidence limit. In particular, even for $\Delta t = 60$ yr, only 89 over 98 (91 per cent) target shocks for the not-declustered set and only 36 over 44 (82 per cent) for the declustered set are successfully forecasted. The reason is that even when Δt is equal to the total duration of the catalogue, in some CAs there remains a fraction of time (before the first strong shock) without any strong shock and then without any alarm. Actually, the maximum fraction of space–time occupied by alarms (τ_u) is only about 44 per cent of the total space–time and nine target shocks with $M_w \geq 5.0$ occurred in the remaining 56 per cent. Here, the last part of the Molchan trajectories, consisting of a linear decrease from the last point defined by the algorithm ($\tau_u = 44$ per cent and $\tau_w = 62$ per cent with $\nu = 9$ per cent for not-declustered and 18 per cent for declustered) to the lower left corner ($\tau = 100$ per cent, $\nu = 0$), can be interpreted as the application to the remaining earthquakes, not predicted by any foreshock, of a purely random forecasting method with success rate proportional to the fraction of the remaining space–time region not covered by our forecasting algorithm.

The overall AS scores are $a_f\ (\tau_u = 1) = 0.89 \pm 0.03$ and $a_f\ (\tau_w = 1) = 0.85 \pm 0.03$ for the not-declustered set and $a_f\ (\tau_u = 1) = 0.78 \pm 0.04$ and $a_f\ (\tau_w = 1) = 0.70 \pm 0.04$ for the declustered set. In all cases they are significantly larger than the expectance (0.5) of a random method with s.l. $\ll 0.01$.

In Figs S7–S10 of the Supporting Information, we also report the plots for targets with $M_w \geq 6.0$ (see numerical values in Tables S6 and S7, Supporting Information). The performance is similar to that for $M_w \geq 5.5$ but as the number of target events is smaller (10 not-declustered and 7 declustered), the power of the tests and the reliability of possible inferences are relatively poorer. This is actually reflected by the fact that the confidence limits in this case are relatively close to the Molchan and AS trajectories.

All not-declustered targets are successfully forecasted with $\Delta t =$ 20 yr, 80 per cent with $\Delta t = 1$ yr and 50 per cent with $\Delta t = 1$ d. For declustered targets, the corresponding forecasting rates are 100 per cent, 71 per cent and 43 per cent respectively. The overall AS scores are $a_f\ (\tau_u = 1) = 0.95 \pm 0.09$ and $a_f\ (\tau_w = 1) = 0.91 \pm 0.09$ for not-declustered and $a_f\ (\tau_u = 1) = 0.93 \pm 0.11$ and $a_f\ (\tau_w = 1) = 0.87 \pm 0.11$ for declustered. In all cases, they are significantly larger than the expectance (0.5) of a random method with s.l. $\ll 0.01$.

One question that may come to mind when looking at the results of such space–time analysis is how much of the observed forecasting performance is due to spatial clustering and how much to time clustering. In order to try to answer such question, we made some further computations in which the time clustering is eliminated by

Downloaded from https://academic.oup.com/gji/article/225/2/1192/6039343 by guest on 11 March 2021

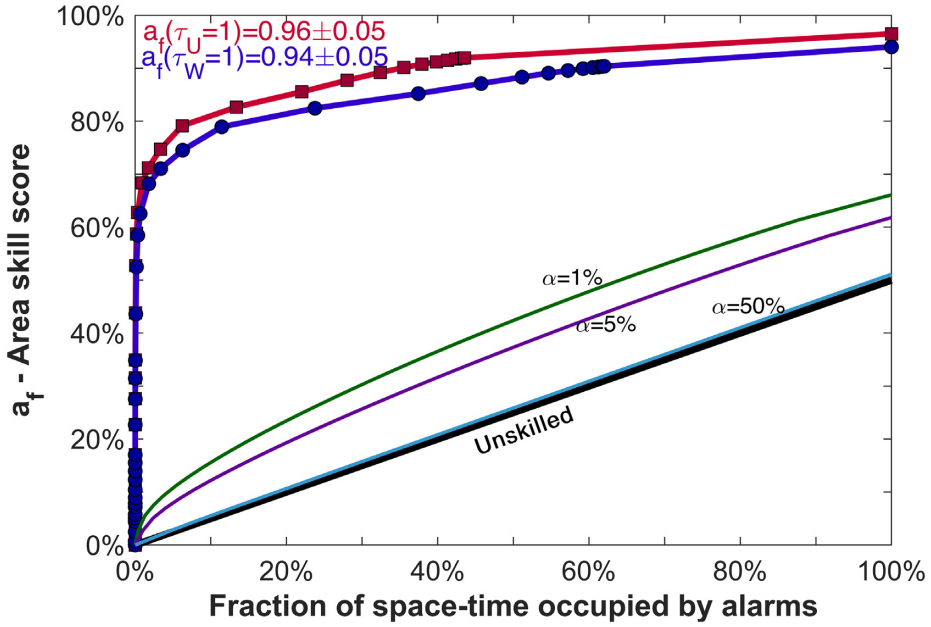

Figure 3. AS score diagram for all target shocks with $M_w \geq 5.5$ (not-declustered). Red and dark blue lines indicate the forecasting performance of foreshocks with $4.4 \leq M_w < 4.8$ for unweighted (τ_u) and weighted (τ_w) fractions of space–time occupied by alarms, respectively (see the text). The black continuous line indicates the performance of a purely random forecasting method that separates skilled (above the line) from unskilled (below) forecasting methods. The light blue, violet and green lines indicate the confidence limits for $\alpha = 50$ per cent, 5 per cent and 1 per cent, respectively.

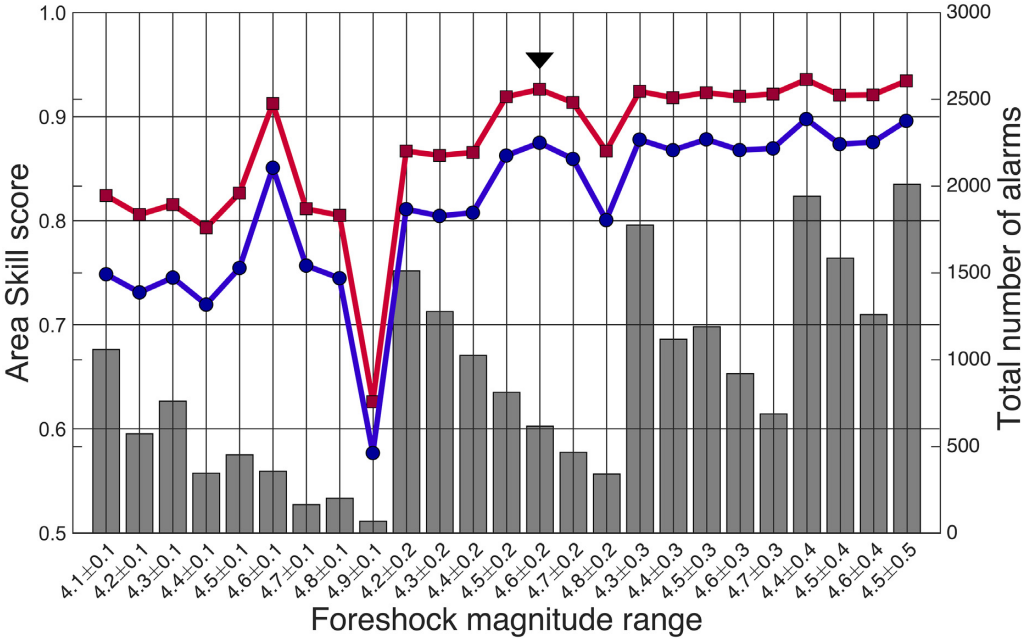

Figure 4. AS score computed for declustered targets with $M_w \geq 5.5$, using unweighted (red line) and weighted (blue) fractions of space–time occupied by alarms, and total number of alarms (grey bars) as a function of the foreshock magnitude range. The arrows indicate the range $M_w = 4.6 \pm 0.2$, chosen as best compromise between high AS score and low number of alarms.

assuming in each CA a permanent alarm for the entire duration of the forecasting experiment ($T = 60$ yr). We computed the time-independent Molchan and AS score trajectories by adding step by step one CA at a time, starting from the CA with highest weight (highest long-term seismic activity) and then going on, up to add all CAs. At each step, the unweighted and weighted fractions of space occupied by alarms are computed by simply taking $\tau_c = 1$ in eqs (4) and (6), respectively, for the included CAs and $\tau_c = 0$ for the not included CAs.

The results of such time-independent analysis for declustered (first) target shocks with $M_w \geq 5.5$ is shown in Figs 7 and 8. Even if they are not fully comparable with the time-dependent analysis of Figs 5 and 6 because the trajectories depend on the adopted ordering of the CAs, from the most to the least active, we can note

Downloaded from https://academic.oup.com/gji/article/225/2/1192/6039343 by guest on 11 March 2021

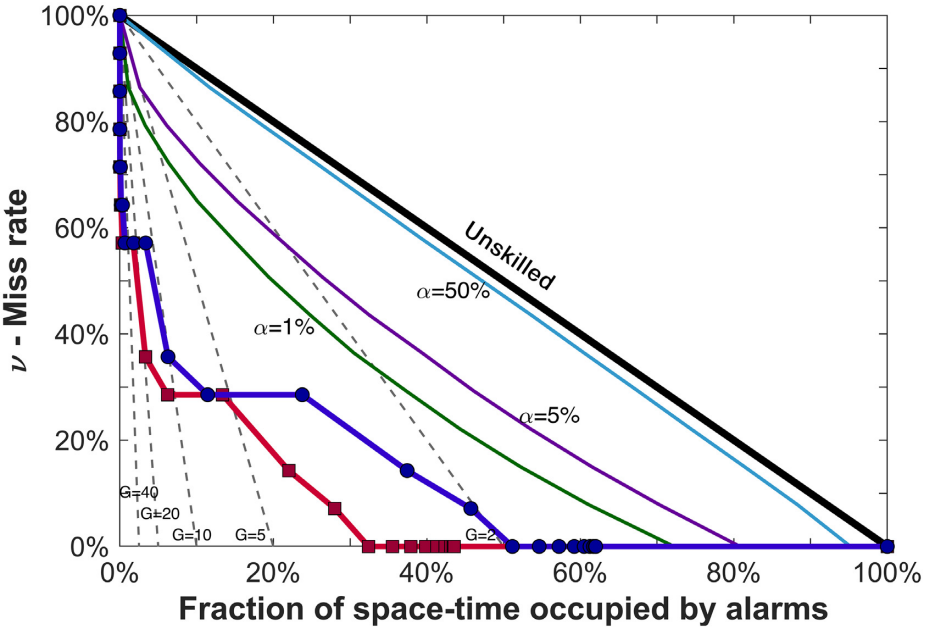

Figure 5. Same as Fig. 2 for declustered (first) target shocks with $M_w \geq 5.5$.

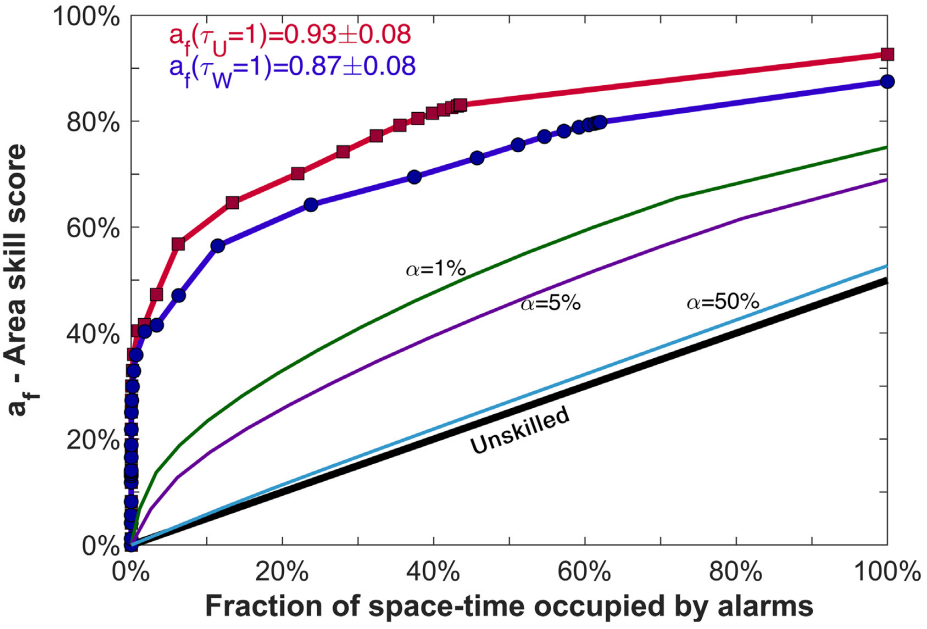

Figure 6. Same as Fig. 3 for declustered (first) target shocks with $M_w \geq 5.5$.

that the skill of time-independent analysis appears definitely lower, particularly at small τ and for the weighted trajectories (blue lines). This can be easily explained by the higher time clustering at short times (and then at small τ) and by the fact that the weights based on the long-term seismic activity penalize more the CAs where the target shocks actually occurred in the last 60 yr.

The results for declustered (first) target shocks with $M_w \geq 5.0$ and ≥ 6.0 are reported in Figs S11–S14 of Supporting Information. For $M_w \geq 5.0$, the comparison of Figs S11 and S12 in the Supporting Information with the time-dependent analysis of Figs S5 and S6 in the Supporting Information is similar to the case for $M_w \geq 5.5$ described before. For $M_w \geq 6.0$, the comparison of Figs S13 and S14 in the Supporting Information with the time-dependent analysis of Figs S9 and S10 in the Supporting Information, apart for small τ, apparently indicates an overall higher skill for the time-independent analysis with respect to the time-dependent one. This is due to the fact that for $M_w \geq 6.0$ all declustered target shocks occurred in CAs with very high long-term seismic activity and that, as noted above, time-independent and time-dependent statistics are not fully comparable between them.

Downloaded from https://academic.oup.com/gji/article/225/2/1192/6039343 by guest on 11 March 2021

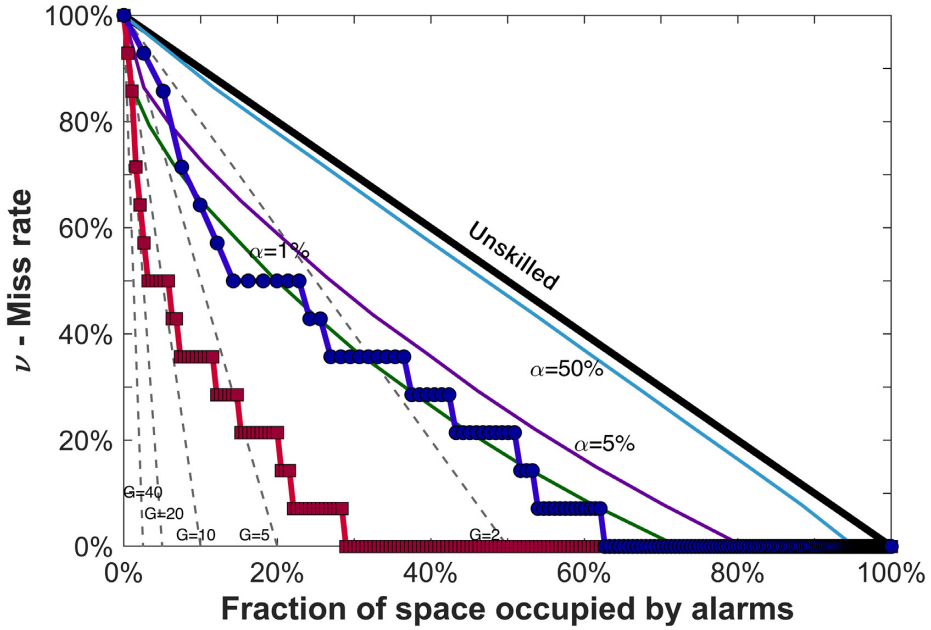

Figure 7. Same as Fig. 2 for time-independent analysis of declustered (first) target shocks with $M_w \geq 5.5$.

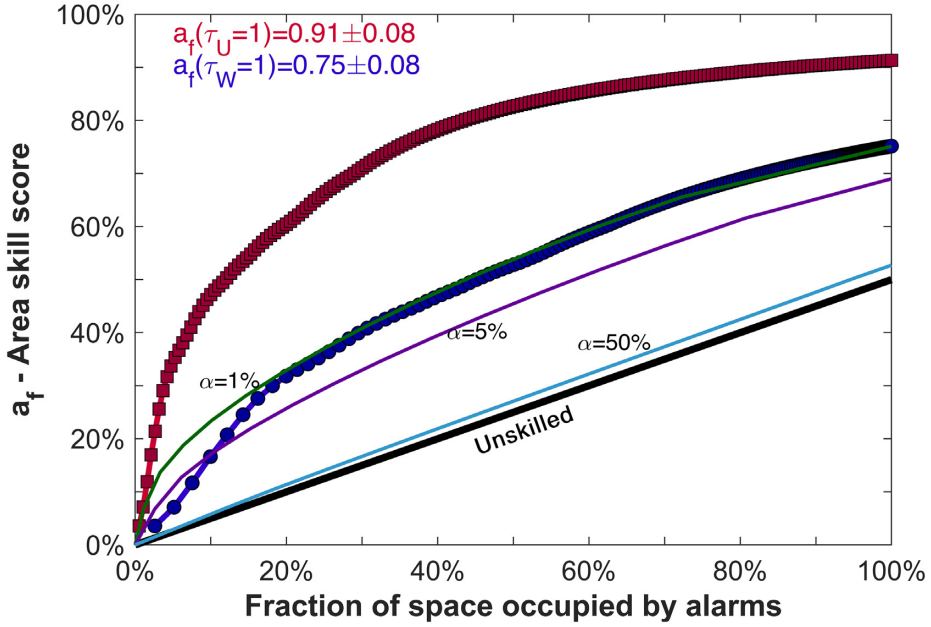

Figure 8. Same as Fig. 3 for time-independent analysis of declustered (first) target shocks with $M_w \geq 5.5$.

OPTIMIZATION OF THE FORECASTING ALGORITHM

For a practical application of the forecasting method, it might be useful to determine the values of the algorithm parameter Δt for which the forecasting method is more efficient and useful for risk mitigation. To accomplish this purpose, we analyse the behaviour of some statistics that depend on the alarm time window Δt.

In Fig. 9 we report, for declustered targets and weighted fraction of space–time occupied by alarms (τ_w), the binomial probability (eq. 9), that is the probability that the observed number of successful forecasts is obtained by chance, as a function of Δt. The lower the probability the higher the strength of the forecast. In general, probabilities are relatively low within a wide range going from one day to some years. For $M_w \geq 5.0$ (red line), very low probabilities are observed around $\Delta t = 2 \div 10$ d. For $M_w \geq 5.5$ (blue line) and $M_w \geq 6.0$ (green line) the minimum probabilities are larger than the ones for $M_w \geq 5.0$, and they remain relatively low from a few hours to a few months. Within such ranges, the forecasting ability of our method reaches its higher efficiency.

The behaviour of the probability gain G (eq. 8) as a function Δt (Fig. 10) shows, for all the three magnitude thresholds, monotonically descending trends from more than 100 000 at very short Δt (less than a minute) to slightly more than 1 at very long Δt (tens

Downloaded from https://academic.oup.com/gji/article/225/2/1192/6039343 by guest on 11 March 2021

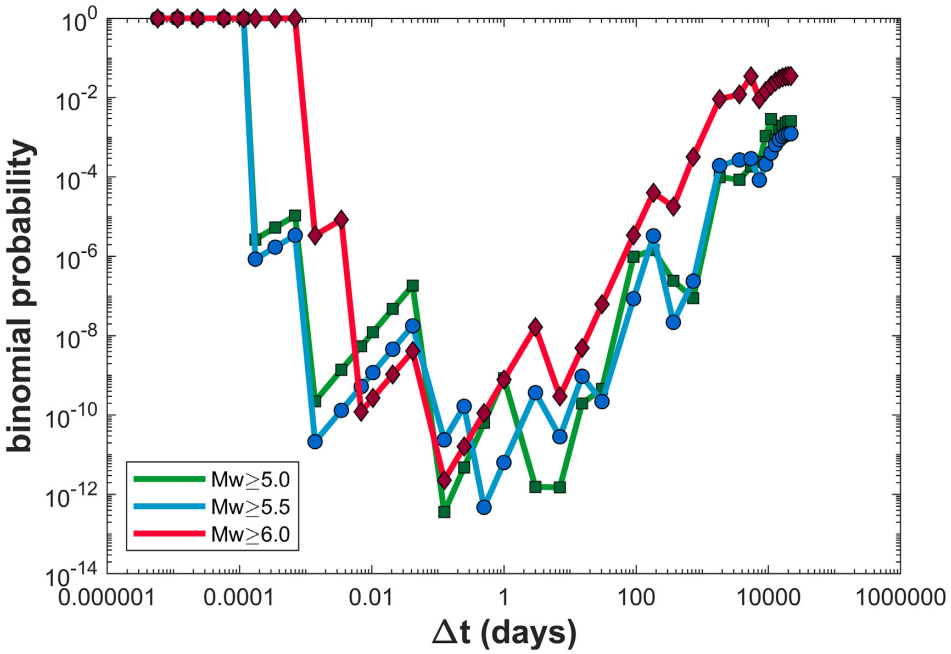

Figure 9. Binomial probability density for declustered (first) target shocks and weighted fraction of space–time occupied by alarms for different magnitude thresholds (see inset) as a function of the alarm duration Δt.

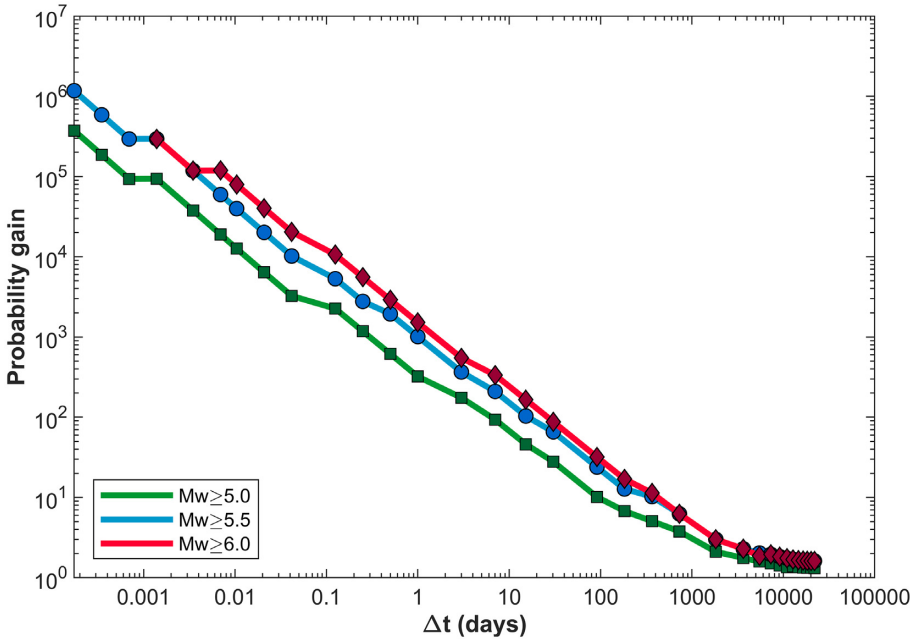

Figure 10. Probability gain for declustered (first) target shocks and weighted fraction of space–time occupied by alarms for different magnitude thresholds (see inset) as a function of the alarm duration Δt.

of years). Such curves also show relatively milder slopes in correspondence of steep decreases of binomial probabilities in Fig. 9 (i.e. around 0.001 d and a few days).

In Fig. 11, we show the miss rate ν as a function of Δt. In general, it decreases with increasing Δt. The (negative) trends—with respect to $\log_{10}\Delta t$—are in between the −5 per cent and −10 per cent per decade, for Δt ranging from a few seconds to about 1 yr. Then they start to decrease more rapidly (about −20 per cent per decade) reaching 0 for $M_w \geq 5.5$ and ≥ 6.0 and 19 per cent for $M_w \geq 5.0$ at very large Δt.

The behaviour of the same statistic for the full set of target events (not-declustered) is reported in Figs S15–S17 of the Supporting

Downloaded from https://academic.oup.com/gji/article/225/2/1192/6039343 by guest on 11 March 2021

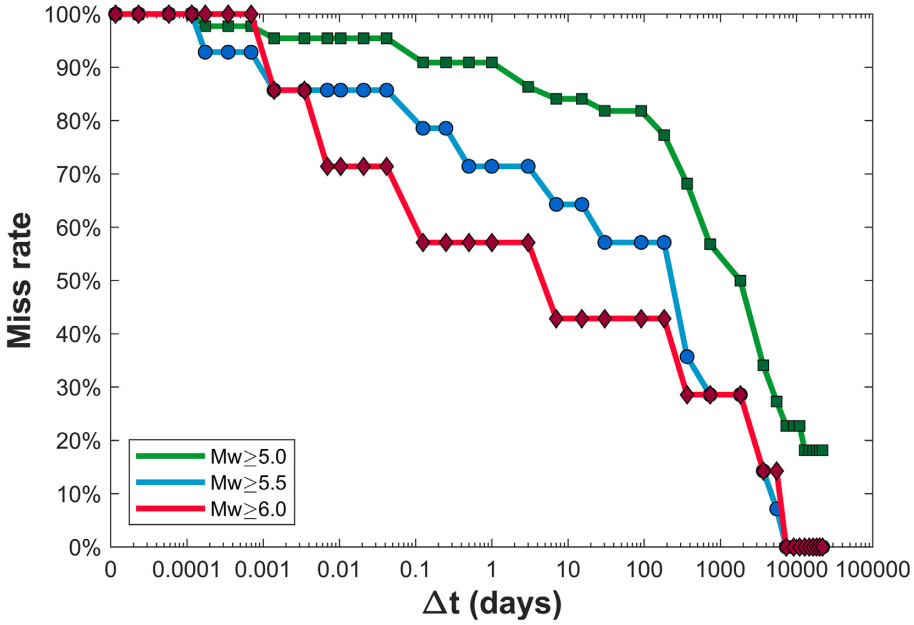

Figure 11. Miss rate for declustered (first) target shocks and different magnitude thresholds (see inset) as a function of the alarm duration Δt.

Information. It is similar to those of the declustered set but the binomial probabilities are lower, the probability gains are higher and the miss rates decrease more rapidly with Δt.

Another aspect to be considered for the practical application of the forecasting method is the dependence on Δt of the fractions of space–time occupied by alarms τ_u and τ_w (Fig. 12). A long alarm interval Δt (with a corresponding long fraction of space–time occupied by alarms τ) allows to forecast more target earthquakes but at the same time it has relatively lower probabilities of occurrence than a shorter Δt. Furthermore, a longer duration of alarms would impact more with life activities of the population in the involved area. Even if any decision on the possible practical application in real situations would eventually require a careful evaluation by decision makers even considering a cost-benefits analysis (e.g. van Stiphout *et al.* 2010; Hermann *et al.* 2016), we examine here as an example the choice of $\Delta t = 3$ months (0.25 yr). This choice, in most cases, results in a fairly trade-off between a good efficiency and a narrow space–time fraction covered by alarms $\tau \approx 2$.

We can see in Table 2 that in this case the method is able to retrospectively forecast more than 50 per cent of not-declustered target shocks with $M_w \geq 5.0$ and more than 70 per cent of those with $M_w \geq 5.5$ and ≥ 6.0. We also report in Table 2 the statistic of the numbers of successful alarms with respect to the total number of alarms indicating higher rates for target with $M_w \geq 5.0$. About one-fifth of alarms actually forecast an earthquake, while the fraction of successful alarms definitely decreases for larger targets and further decreases for declustered sets down to about 1 per cent. Note that several alarm time windows are actually overlapped and then the total duration of alarms is shorter than the simple sum of alarm windows (eq. 2).

The performance of the method is definitely worse for the first target shocks (declustered set) but it improves by increasing the magnitude of target shocks. Actually, 4 over 7 first target shocks with $M_w \geq 6.0$ over the last 60 yr in Italy are retrospectively forecasted in this way.

We tested the stability with time of the forecasting performance by subdividing the seismic catalog in two equal parts of 30 yr: before and after 1990 january 01. The same computations of Table 2 for $\Delta t = 3$ months for intervals 1960–1989 and 1990–2019 are reported in Tables 3 and 4 respectively. The rates of successfully forecasted target shocks (declustered or not) are similar in the two periods whereas the space–time fraction occupied by alarms is definitely lower in the most recent period, consistently with the higher ratios between successful and total alarms. We could argue that smaller magnitude errors in most recent times, owing to the continuous improvement of the Italian seismic network, reduce the amount of false alarms and then increase the observed skill of the forecasting method with respect to the previous period.

In Tables 5 and 6, we report the lists of retrospective forecast of the first (declustered) target shocks with $M_w \geq 5.5$ and ≥ 6.0, respectively, occurred in Italy from 1960 to 2019 (also see the results for the declustered first shocks with $M_w \geq 5.0$ in Table S8 in the Supporting Information and the results for not-declustered targets with $M_w \geq 5.0$, 5.5 and 6.0 in Tables S9–S11, respectively, of the Supporting Information).

We can note that for two target shocks (1976 Friuli and 1990 Potentino) the forecast could have hardly been used by civil protection services to adopt safety countermeasures because the forecasting strong shocks occurred too shortly before the main shock (67 and 13 s, respectively). In other cases, the time delay between the forecasting shock and the main shock (going from a couple of hours to a few weeks) would have been sufficient to take some countermeasures.

We could note that a foreshock did actually occur a couple of days before the first main shock of 2012 May 20 ($M_w = 6.1$) in the area of Pianura Emiliana but its magnitude ($M_w = 4.2$) was only slightly below the lower threshold of $M_w = 4.4$ we adopted. The retrospective ability to predict $M_w \geq 6.0$ earthquakes might have been improved then by slightly reducing such lower threshold but at a cost of a general reduction of the performance of the algorithm, because of the increment of the number of alarms and of the fraction of space–time covered by alarms.

Downloaded from https://academic.oup.com/gji/article/225/2/1192/6039343 by guest on 11 March 2021

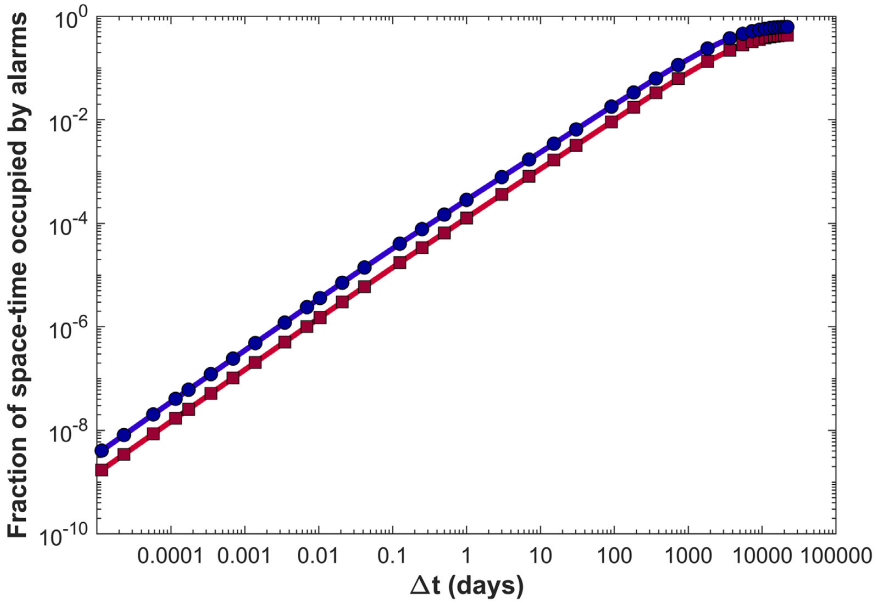

Figure 12. Unweighted (red) and weighted (dark blue) fraction of space–time occupied by alarms as a function of the alarm duration Δt.

Table 2. Retrospective forecasting performance of the algorithm for $\Delta t = 3$ months.

Target magnitude	≥5.0		≥5.5		≥6.0		τ_u (per cent)	τ_w (per cent)
				Not-declustered				
Forecasted/total shocks	55/98	56 per cent	26/35	74 per cent	7/10	70 per cent	0.9	1.9
Successful/total alarms	115/617	18.6 per cent	72/617	11.7 per cent	30/617	4.9 per cent	0.9	1.9
				Declustered				
Forecasted/total shocks	8/44	18 per cent	6/14	43 per cent	4/7	57 per cent	0.9	1.9
Successful/total alarms	13/617	2.1 per cent	9/617	1.5 per cent	8/617	1.3 per cent	0.9	1.9

Table 3. Same as Table 2 for the time interval 1960–1989.

Target magnitude	≥5.0		≥5.5		≥6.0		τ_u (per cent)	τ_w (per cent)
				Not-declustered				
Forecasted/total shocks	21/45	47 per cent	11/15	73 per cent	3/4	75 per cent	1.0	2.1
Successful/total alarms	45/336	12.9 per cent	22/336	6.6 per cent	9/336	2.7 per cent	1.0	2.1
				Declustered				
Forecasted/total shocks	3/25	12 per cent	3/7	43 per cent	2/3	67 per cent	1.0	2.1
Successful/total alarms	5/336	1.5 per cent	5/336	1.5 per cent	3/336	0.89 per cent	1.0	2.1

CONCLUSIONS

We analysed a simple algorithm to forecast shallow (depth < 50 km) main shocks ($M_w \geq 5.0$, 5.5 and 6.0) that threaten the life and the goods of the population living on the Italian mainland territory, based on the previous occurrence within CA of 30 km of radius of widely felt strong shocks ($4.4 \leq M_w < 4.8$) not particularly harmful in themselves. Based on a retrospective analysis of the HORUS seismic catalogue of Italy from 1960 to 2019 (Lolli *et al.* 2020) this method retrospectively forecast the majority of damaging earthquakes occurred in Italy in the past 60 yr by issuing alarms covering only a small fraction of the space–time coverage.

We estimated such fraction even considering the different levels of seismic activity in different areas of Italy by weighting more the alarm times in CA where the average seismicity rate, computed from the CPTI15 seismic catalogue (Rovida *et al.* 2016, 2020) from 1600 to 1959, is higher.

The retrospective testing using the Molchan diagram (Molchan 1990, 1991; Molchan & Kagan 1992) and the AS score (Zechar &

Downloaded from https://academic.oup.com/gji/article/225/2/1192/6039343 by guest on 11 March 2021

Table 4. Same as Table 2 for the time interval 1990–2019.

Target magnitude	≥5.0		≥5.5		≥6.0		τ_u (per cent)	τ_w (per cent)
			Not-declustered					
Forecasted/total shocks	34/53	64 per cent	15/20	75 per cent	4/6	67 per cent	0.4	0.7
Successful/total alarms	70/281	24.9 per cent	50/281	17.8 per cent	21/281	7.5 per cent	0.4	0.7
			Declustered					
Forecasted/total shocks	5/19	26 per cent	3/7	43 per cent	2/4	50 per cent	0.4	0.7
Successful/total alarms	8/281	3.5 per cent	4/281	1.4 per cent	5/281	1.8 per cent	0.4	0.7

Table 5. Results of retrospective forecast of first main shocks (declustered targets) with $M_w \geq 5.5$ in Italy from 1960 to 2019, using $\Delta t = 3$ months (0.25 yr).

Year	Month	Day	Lat	Lon	M_w	t_a (d)		Epicentral area
1962	8	21	41.233	14.933	5.7	0.093	2.22 h	Irpinia
1968	1	15	37.700	13.100	5.7	0.425	10.2 h	Valle del Belice
1976	5	6	46.250	13.250	6.5	7.8×10^{-4}	67 s	Friuli
1979	*9*	*19*	*42.717*	*12.950*	*5.8*	*Missed*		*Valnerina*
1980	*11*	*23*	*40.800*	*15.367*	*6.8*	*Missed*		*Irpinia-Basilicata*
1984	*4*	*29*	*43.204*	*12.585*	*5.6*	*Missed*		*Umbria settentrionale*
1984	*5*	*7*	*41.666*	*13.820*	*5.9*	*Missed*		*Monti della Meta*
1990	5	5	40.650	15.882	5.8	1.5×10^{-4}	13 s	Potentino
1997	9	26	43.023	12.891	5.7	22.1		Appennino umbro-marchigiano
1998	*9*	*9*	*40.060*	*15.949*	*5.5*	*Missed*		*Appennino lucano*
2002	*10*	*31*	*41.717*	*14.893*	*5.7*	*Missed*		*Molise*
2009	4	6	42.342	13.380	6.3	6.5		Aquilano
2012	*5*	*20*	*44.896*	*11.264*	*6.1*	*Missed*		*Pianura Emiliana*
2016	*8*	*24*	*42.698*	*13.234*	*6.2*	*Missed*		*Monti della Laga*

Notes: t_a is the maximum time advance of the foreshock with respect to the main shock. '*Missed*' indicates that the target shock was not forecasted (in such cases all entries are in italics). Epicentral area identifiers are taken from the CPTI15 catalogue (Rovida *et al.* 2016, 2020).

Table 6. Same as Table 2 for first main shocks with $M_w \geq 6.0$.

Year	Month	Day	Lat	Lon	M_w	t_a (d)		Epicentral area
1962	8	21	41.233	14.933	6.2	0.100	2.40 h	Irpinia
1976	5	6	46.250	13.250	6.5	7.8×10^{-4}	67 s	Friuli
1980	*11*	*23*	*40.800*	*15.367*	*6.8*	*Missed*		*Irpinia-Basilicata*
1997	9	26	43.015	12.854	6.0	22.5		Appennino umbro-marchigiano
2009	4	6	42.342	13.380	6.3	6.5		Aquilano
2012	*5*	*20*	*44.896*	*11.264*	*6.1*	*Missed*		*Pianura Emiliana*
2016	*8*	*24*	*42.698*	*13.234*	*6.2*	*Missed*		*Monti della Laga*

Jordan 2008) methods indicates that such approach clearly overperforms a purely random method with high or very high confidence, depending on the target shock magnitude threshold.

As the secondary main shocks during seismic sequences are definitely easier to be forecasted by this method because the aftershocks of the first main shock usually generate alarms at weakly (if not daily) rate, we also tested the ability of our approach to predict only the first main shock of each sequence. We found that the forecasting ability remains high even if being lower than that considering all main shocks.

Even if the true verification of the efficiency of the method will only be made on a prospective data set, we believe that such simple forecasting algorithm could be useful, like other operational forecasting approaches presently considered by the Italian Civil Protection Department, for planning preparation measures in the field (e.g. Marzocchi *et al.* 2014).

The latter approaches are mainly based on the ETAS model (Kagan & Knopoff 1987; Ogata 1988) and, as well as that of this work, showed to retrospectively forecast the evolution of Italian seismicity better than an inhomogeneous random process with spatial rates corresponding to past seismicity. On the other hand, Marzocchi & Zhuang (2011) showed that ETAS models is able to describe quite well even the observed foreshock activity. However, a comparison of the relative efficiency of our approach with ETAS models and even with other forecasting approaches (like e.g. the EEPAS method (Rhoades & Evison 2004) would require that the probabilistic formulation of the latter methods is adapted to the alarm-based one (e.g. by selecting a particular probability thresholds above which to declare an alarm). However, such adaptation is not trivial and hence, the question on which of the different approaches is better in predicting future damaging earthquakes remains not answered presently and has to be deferred to future papers comparing all

Downloaded from https://academic.oup.com/gji/article/225/2/1192/6039343 by guest on 11 March 2021

methods in an alarm-based context by using, for example, the approach proposed by Shebalin *et al.* (2014).

One advantage of the present forecasting approach is that it is easy to implement and communicate because it does not require any other scientific analysis than the correct determination of the location and of the magnitude of the precursory shock. In principle every person could be informed very quickly by a notification sent by one of the already available mobile Apps which provide near real-time access to the INGV online earthquake list (http://terremoti.ingv.it/en#).

ACKNOWLEDGEMENTS

We thank the editor Margarita Segou, Jancang Zhuang, Rodolfo Console and two anonymous reviewers 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), which in particular fully financed the PhD grant of one of the authors (EB).

REFERENCES

Agnew, D.C. & Jones, L.M., 1991. Prediction probabilities from foreshocks, *J. geophys. Res.,* **96,** 11959–11971.

Caputo, M., Gasperini, P., Keilis-Borok, V., Marcelli, L. & Rotwain, I., 1977. Earthquakes swarms as forerunners of strong earthquakes in Italy, *Ann. Geofis.,* **30,** 269–283.

Caputo, M., Console, R., Gabrielov, A.M., Keilis-Borok, V.I. & Sidorenko, T.V., 1983. Long-term premonitory seismicity patterns in Italy, *Geophys. J. R. astr. Soc.,* **75,** 71–75.

Console, R., Murru, M. & Alessandrini, B., 1993. Foreshock statistics and their possible relationship to earthquake prediction in the Italian region, *Bull. seism. Soc. Am.,* **83**(4), 1248–1263.

Console, R. & Murru, M., 1996. Probability gain due to foreshock following quiescence tested by synthetic catalogs, *Bull. seism. Soc. Am.,* **86**(3), 911–913.

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

Console, R., Di Luccio, F., Murru, M., Imoto, M. & Stavrakakis, G., 1999. Short term and short range seismicity patterns in different seismic areas of the world, *Nat. Hazard,* **19,** 107–121.

Console, R, Murru, M & Lombardi, A.M., 2003. Refining earthquake clustering models, *J. geophys. Res.,* **108,** 2468, doi:10.1029/2002JB002130.

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.

Di Luccio, F., Console, R., Imoto, M. & Murru, M., 1997. Analysis of short time-space range seismicity patterns in Italy, *Ann. Geofis.,* **40/4,** 783–798.

Gardner, J.K. & Knopoff, L., 1974. Is the sequence of earthquakes in Southern California, with aftershocks removed, Poissonian? *Bull. seism. Soc. Am.,* **64,** 1363–1367.

Gasperini, P., Lolli, B., Vannucci, G. & Boschi, E., 2012. A comparison of moment magnitude estimates for the European–Mediterranean and Italian region, *Geophys. J. Int.,* **190,** 1733–1745.

Gasperini, P., Lolli, B. & Vannucci, G., 2013. Empirical calibration of local magnitude data sets versus moment magnitude in Italy, *Bull. seism. Soc. Am.,* **103,** 2227–2246.

Gasperini, P., Lolli, B. & Vannucci, G., 2016. Relative frequencies of seismic main shocks after strong shocks in Italy, *Geophys. J. Int.,* **207,** 150–159.

Geller, R., Jackson, D.D., Kagan, Y.Y. & Mulargia, F., 1997. Earthquakes cannot be predicted, *Science,* **275,** 1661, doi:10.1126/science.275.5306.1616.

Gutenberg, B. & Richter, C.F., 1944. Frequency of earthquakes in California, *Bull. seism. Soc. Am.,* **34,** 185–188.

Grandori, G., Guagenti, E. & Perotti, F., 1988. Alarm systems based on a pair of short-term earthquake precursors, *Bull. seism. Soc. Am.,* **78**(4), 1538–1549.

Hermann, M., Zechar, J.D. & Wiemer, S., 2016. Communicating time-varying seismic risk during an earthquake sequence, *Seismol. Res. Lett.,* **87**(2A), 30–312.

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,** 1669–1680.

Jones, L.M., 1994. Foreshocks, aftershocks, and earthquake probabilities accounting for the Landers earthquake, *Bull. seism. Soc. Am.,* **84**(1), 892–899.

Jones, L. & Molnar, P., 1976. Frequency of fore- shocks, *Nature,* **262,** 677–679, 1976.

Jones, L. & Molnar, P., 1979. Some characteristics of foreshocks and their possible relationship to earthquake prediction and premonitory slip on faults, *J. geophys. Res.,* **84**(B7), 3596–3608.

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

Jordan, T.H. *et al.*, 2011. Operational earthquake forecasting: state of knowledge and guidelines for implementation, final report of the international commission on earthquake forecasting for civil protection, *Ann. Geophys.,* **54**(4), 315–391.

Jordan, T.H. & Jones, L.M., 2010. Operational earthquake forecasting: some thoughts in why and how, *Seismol. Res. Lett.,* **81/4,** 571–574.

Kagan, Y.Y., 2014. *Earthquakes,* AGU-Wiley, 283 pp.

Kagan, Y. & Knopoff, L., 1987. Statistical short-term earthquake prediction, *Science,* **236,** 1563–1567.

Lolli, B., Gasperini, P. & Rebez, A., 2018. Homogenization in terms of Mw of local magnitudes of Italian earthquakes occurred before 1981, *Bull. seism. Soc. Am.,* **108**(1), 481–492.

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

Marzocchi, W., Lombardi, A.M. & Casarotti, E., 2014. The Establishment of an Operational Earthquake Forecasting System in Italy, *Seismol. Res. Lett.,* **85/5,** 961–969.

Marzocchi, W. & Zhuang, J., 2011. Statistics between mainshocks and foreshocks in Italy and Southern California, *Geophys. Res. Lett.,* **38,** L09310.

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 earth- quake prediction, *Tectonophysics*. **193,** 267–276.

Molchan, G.M. & Kagan, Y.Y., 1992, Earthquake predictionand its optimization, *J. geophys. Res.,* **97,** 4823–4838.

Mulargia, F. & Geller, R.J., 2003. *Earthquake Science and Seismic Risk Reduction,* Kluwer, 338pp.

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.

Reasenberg, P.A., 1985. Second-order moment of central California seismicity, 1969–1982, *J. geophys. Res.,* **90,** 5479–5495.

Reasenberg, P.A., 1999a. Foreshock occurrence before large earthquakes, *J. geophys. Res.,* **104**(B3), 4755–4768.

Reasenberg, P.A., 1999b. Foreshock occurrence rates before large earthquakes worldwide, *Pure appl., Geophys,* **155,** 355–379.

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., 2016. CPTI15. The 2015 version of the Parametric Catalogue of Italian Earthquakes, *Istituto Nazionale di Geofisica e Vulcanologia*.

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

Downloaded from https://academic.oup.com/gji/article/225/2/1192/6039343 by guest on 11 March 2021

Schorlemmer, D., Christophersen, A., Rovida, A., Mele, F., Stucchi, M. & Marzocchi, W., 2010. Setting up an earthquake forecast experiment in Italy, *Ann. Geophys.,* **53/3,** 1–9.

Southern San Andreas Working Group, 1991. Short-Term Earthquake Hazard Assessment for the San Andreas Fault in Southern California, U.S. Geol. Surv., Open-file Report 91-31.

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

Shebalin, P., , Narteau, C., Zechar, J. D. & Holschneider, M., 2014. Combining earthquake forecasts using differential probability gains, *Earth Planets Space,* **66,** 37.

van Stiphout, T., Wiemer, S. & Marzocchi, W., 2010. Are short-term evacuations warranted? Case of the 2009 L'Aquila earthquake, *Geophys. Res. Lett.,* **37**(6), 1–5.

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., 2010. The Collaboratory for the Study of Earthquake Predictability perspective on computational earthquake science, *Concurr. Comput.,* **22,** 1836–1847.

SUPPORTING INFORMATION

Supplementary data are available at *GJI* online.

Figure S1. Spatial distribution of inland earthquakes from the HORUS catalogue (Lolli *et al.* 2020) with $M_w \geq 4.0$ and depth < 50 km used for testing and optimization. Black dots indicate $4.0 \leq M_w < 5.0$, green dots $5.0 \leq M_w < 5.5$, blue dots $5.5 \leq M_w < 6.0$ and red dots $M_w \geq 6.0$.

Figure S2. Time distribution of magnitudes of inland earthquakes km from the HORUS catalogue (Lolli *et al.* 2020) with depth < 50 km used for testing and optimization. Black dots indicate $M_w < 5.0$, green dots $5.0 \leq M_w < 5.5$, blue dots $5.5 \leq M_w < 6.0$ and red dots $M_w \geq 6.0$.

Figure S3. Molchan diagram for all target shocks with $M_w \geq 5.0$ (not-declustered). Red and dark blue lines indicate the forecasting performance of foreshocks with $4.4 \leq M_w < 4.8$ for unweighted (τ_u) and weighted (τ_w) fractions of space–time occupied by alarms respectively (see the main text). The black continuous line indicates the performance of 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 α =50 per cent, 5 per cent and 1 per cent, respectively. The black dashed lines indicate probability gains $G = 2$, 5, 10, 20 and 40.

Figure S4. AS score diagram for all target shocks with $M_w \geq 5.0$ (not-declustered). Red and dark blue lines indicate the forecasting performance of foreshocks with $4.4 \leq M_w < 4.8$ for unweighted (τ_u) and weighted (τ_w) fractions of space–time occupied by alarms respectively (see the main text). The black continuous line indicates the performance of a purely random forecasting method that separates skilled (above the line) from unskilled (below) forecasting methods. The light blue, violet and green lines indicate the confidence limits for α =50 per cent, 5 per cent and 1 per cent, respectively.

Figure S5. Same as Fig. S2 for declustered (first) target shocks with $M_w \geq 5.0$ (see the text).

Figure S6. Same as Fig. S3 for declustered (first) target shocks with $M_w \geq 5.0$ (see the text).

Figure S7. Same as Fig. S2 for all target shocks with $M_w \geq 6.0$ (not-declustered).

Figure S8. Same as Fig. S3 for all target shocks with $M_w \geq 6.0$ (not-declustered).

Figure S9. Same as Fig. S2 for declustered (first) target shocks with $M_w \geq 6.0$.

Figure S10. Same as Fig. S3 for declustered (first) target shocks with $M_w \geq 6.0$.

Figure S11. Same as Fig. S2 for time-independent analysis of declustered (first) target shocks with $M_w \geq 5.0$.

Figure S12. Same as Fig. S3 for time-independent analysis of declustered (first) target shocks with $M_w \geq 5.0$.

Figure S13. Same as Fig. S2 for time-independent analysis of declustered (first) target shocks with $M_w \geq 6.0$.

Figure S14. Same as Fig. S3 for time-independent analysis of declustered (first) target shocks with $M_w \geq 6.0$.

Figure S15. Binomial probability density for all target shocks (not-declustered) and weighted fraction of space–time occupied by alarms for different magnitude thresholds (see inset) as a function of the alarm duration Δt.

Figure S16. Probability gain for all target shocks (not-declustered) and weighted fraction of space–time occupied by alarms for different magnitude thresholds (see inset) as a function of the alarm duration Δt.

Figure S17. Miss rate for all target shocks (not-declustered) for different magnitude thresholds (see inset) as a function of the alarm duration Δt.

Table S1. List of centre coordinates of CA with radius of 30 km.

Table S2. Values of variables in Molchan and AS score plots of Figs 2 and 3 for $M_w \geq 5.5$ not-declustered targets.

Table S3. Same as Table S2 for $M_w \geq 5.5$ declustered targets (Figs 4 and 5).

Table S4. Same as Table S2 for $M_w \geq 5.0$ not-declustered targets (Figs S1 and S2).

Table S5. Same as Table S2 for $M_w \geq 5.0$ declustered targets (Figs S3 and S4).

Table S6. Same as Table S2 for $M_w \geq 6.0$ not-declustered targets (Figs S5 and S6).

Table S7. Same as Table S2 for $M_w \geq 6.0$ declustered targets (Figs S7 and S8).

Table S8. Results of retrospective forecast of first main shocks (declustered targets) with $M_w \geq 5.0$ in Italy from 1960 to 2019, using Δt= 3 months (0.25 yr).

Table S9. Results of retrospective forecast of not-declustered targets with $M_w \geq 5.0$ in Italy from 1960 to 2019, using Δt= 3 months (0.25 yr).

Table S10. Same as Table S9 for not-declustered targets with $M_w \geq 5.5$.

Table S11. Same as Table S9 for not-declustered targets with $M_w \geq 6.0$.

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.

Downloaded from https://academic.oup.com/gji/article/225/2/1192/6039343 by guest on 11 March 2021