Geophysical Journal International

Geophys. J. Int. (2023) **234,** 1681–1700
Advance Access publication 2023 March 21
GJI Seismology

https://doi.org/10.1093/gji/ggad123

Application of the EEPAS earthquake forecasting model to Italy

E. Biondini [1], D.A. Rhoades [2] and P. Gasperini [1,3]

[1] *Dipartimento di Fisica e Astronomia, Università di Bologna,* 40129 *Bologna BO, Italy. E-mail: emanuele.biondini2@unibo.it*
[2] *Earthquake Physics and Statistics, GNS Science P.O. Box 30-368, Lower Hutt* 5040*, New Zealand*
[3] *Sezione di Bologna, Istituto Nazionale di Geofisica e Vulcanologia, Sezione di Bologna,* 40127 *Bologna BO, Italy*

Accepted 2023 March 20. Received 2023 March 1; in original form 2022 September 6

SUMMARY

The Every Earthquake a Precursor According to Scale (EEPAS) forecasting model is a space–time point-process model based on the precursory scale increase (ψ) phenomenon and associated predictive scaling relations. It has been previously applied to New Zealand, California and Japan earthquakes with target magnitude thresholds varying from about 5–7. In all previous application, computations were done using the computer code implemented in Fortran language by the model authors. In this work, we applied it to Italy using a suite of computing codes completely rewritten in Matlab. We first compared the two software codes to ensure the convergence and adequate coincidence between the estimated model parameters for a simple region capable of being analysed by both software codes. Then, using the rewritten codes, we optimized the parameters for a different and more complex polygon of analysis using the Homogenized Instrumental Seismic Catalogue data from 1990 to 2011. We then perform a pseudo-prospective forecasting experiment of Italian earthquakes from 2012 to 2021 with $M_{\rm w} \geq 5.0$ and compare the forecasting skill of EEPAS with those obtained by other time independent (Spatially Uniform Poisson, Spatially Variable Poisson and PPE: Proximity to Past Earthquakes) and time dependent [Epidemic Type Aftershock Sequence (ETAS)] forecasting models using the information gain per active cell. The preference goes to the ETAS model for short time intervals (3 months) and to the EEPAS model for longer time intervals (6 months to 10 yr).

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

INTRODUCTION

EEPAS is an earthquake forecasting method based on the statistical analysis of seismicity (Rhoades & Evison 2004). Its basic assumption is that magnitudes and rates of minor seismicity increase before a strong shock. This phenomenon (called ψ-phenomenon) was described by Evison & Rhoades (2004) for some regions of the world in which high quality earthquakes catalogues are available. They analysed 47 earthquakes with magnitude ranging between 5.8 and 8.2 to derive three empirical scaling relations: for time, magnitude and area. These relate the magnitude of main shock (M_m) with the precursor magnitude (M_P) the precursor time (T_P) and the precursor area (A_p). Such empirical scaling relations show that in general the magnitude of precursor events is smaller than the magnitude of the main shock by at least one magnitude unit. The EEPAS model considers each earthquake as an individual precursor according to the scale indicated by its magnitude, rather than as a possible member of a ψ-phenomenon.

The details of the EEPAS method are described in a number of papers (e.g. Rhoades & Evison 2004; Evison & Rhoades 2005; Rhoades 2007, 2011; Rhoades *et al.* 2020), some of which contain typos that make the formulation not perfectly identical in all of them. For such reason in Appendix A we describe again the method as well as some assumptions made without explicit mentions in previous papers.

We implemented such formulations in a suite of Matlab codes that we first compared with the code EEPSOF (Rhoades 2021) used in all previous applications of EEPAS methods. The results of this comparison are described in Appendix B and indicate a tight agreement between parameters estimates from the two codes before the introduction of spatial parameters in the optimization procedure. After that, the differences become a little more pronounced for parameters a_T, σ_A and b_T, where the two codes differ by 12.5, 10.2 and 24.0 per cent, respectively. These differences can be explained by the different numerical algorithm adopted by the two codes for spatial integration. However, the final maximum log-likelihood scores and the numbers of forecasted earthquakes are very similar (Table A1).

For comparison purposes we also consider other forecasting models and in particular the Epidemic Type Aftershock Sequence

(ETAS) model (Ogata 1989, 1998) and two time-independent forecasting models: the Spatially Uniform Poisson (SUP) and the Spatially Variable Poisson (SVP) models (Console *et al.* 2006). The description of our implementations of such models is reported in Appendix C.

The fitting of the free parameters of various models is carried out by maximizing the log-likelihood function of an inhomogeneous Poisson point process, which is given by:

$$\ln L = \sum_{t_i \in (t_a, t_b); m_i \geq m_T; (x_i, y_i) \in R} \ln \lambda\,(t_i, m_i, x_i, y_i) - \int_{t_a}^{t_b} \int_{m_T}^{m_u} \iint_R \lambda\,(t, m, x, y)\ \mathrm{d}y\ \mathrm{d}x\ \mathrm{d}m\ \mathrm{d}t, \quad (1)$$

where $\lambda(t, m, x, y)$ is the rate density function for PPE (eq. A12), EEPAS (eq. A10), ETAS (eq. C6) and SUP or SVP (eq. C7) models; (t_a, t_b) is the time interval of the fitting period; (m_T, m_u) is the magnitude range of target earthquakes and R is the spatial region of analysis.

APPLICATION TO ITALY

We chose $m_T = 5.0$ as the lower magnitude limit for target shocks, because in Italy such earthquakes potentially cause damage to buildings and threaten the health and life of inhabitants. This choice is also consistent with most of the applications of EEPAS model to other regions of the World (Rhoades & Evison 2004; Evison & Rhoades 2005; Rhoades 2007, 2011).

We chose the learning time interval from 1990 to 2011 for fitting the EEPAS model, because the accuracy and completeness of the Italian catalogue has improved significantly since 1990 (Gasperini *et al.* 2013), We use the independent 10-yr interval from 2012 to 2021 for pseudo-prospective testing of the model.

As application region R, we consider a regular tessellation of the Italian territory made of square cells with side $L = 30\sqrt{2}$ km from 7°E to 19°E in longitude and from 36°N to 47°N in latitude. The choice of L is made for compatibility with the previous work by Gasperini *et al.* (2021), so that each square cell is (almost) perfectly inscribed in a circular cell with radius of 30 km. Because the completeness of the earthquake catalogue is poor in offshore areas, according to Gasperini *et al.* (2021), we consider only the cells within which at least one earthquake with $M \geq 4.0$ occurred inland from 1600 to 1959 according to the CPTI15 catalogue (Rovida *et al.* 2020) and from 1960 to 2021 according to the Homogenized instrRUmental Seismic (HORUS) catalogue (Lolli *et al.* 2020). We also excluded the cells that are not contiguous to the main analysis polygon (such as insulated cells on islands). In total, 177 square cells constitute the region of analysis R (Fig. 1).

For fitting of model parameters, an earthquake catalogue with a completeness magnitude (m_c) at least two units lower than the target magnitude (m_T) is desirable (Rhoades & Evison 2004). For Italy an earthquake catalogue with homogeneous magnitudes and high resolution is the HORUS catalogue (Lolli *et al.* 2020) reporting earthquakes from 1960 to the present. According to Lolli *et al.* (2020), HORUS can be considered complete within the Italian mainland for $m \geq 4.0$ since 1960, for $m \geq 3.0$ since 1981, for $m \geq 2.5$ since 1990, for $m \geq 2.1$ since 2003 and for $m \geq 1.8$ since 2005. In the data set for this work, we used only shallow earthquakes with depth $Z \leq 40$ km. To avoid edge effects in the fitting of model parameters, the contribution of earthquakes in the neighbourhood of the region R must also be considered (Rhoades & Evison 2004). We assume as neighbourhood region the area included in the CPTI15 polygon (Fig. 2) according to Rovida *et al.* (2020).

To account for the limited accuracy of magnitude data, we binned all magnitudes to the nearest tenth of a unit:

$$m_{\text{binned}} = \frac{\text{int}\,(m_{\text{raw}} \times 10 + 0.5)}{10}. \quad (2)$$

This also means that a magnitude lower threshold rounded to the nearest tenth of a unit (e.g. $m_T = 5.0$) implies an effective threshold 0.05 units smaller (e.g. $m_T = 4.95$).

The HORUS catalogue reports 27 target shocks with $M_w \geq 5.0$ from 1990 to 2011 and 27 from 2012 to 2021. Thus, the rate of target shocks in the testing period is about twice that in the learning period. Hence, the forecasting of the correct number of earthquakes in the testing period by any forecasting method will be difficult.

After the first target shock ('main shock') of a seismic sequence, the forecasting of successive target shock ('aftershocks') is easier, owing to the presence of small aftershocks. Hence, we also fit and test the models against a set of target earthquakes only including the first target shock of each sequence. According to Gasperini *et al.* (2021), we eliminate the target shocks, occurring within a spatial window of 50 km and a time window of 1 yr after any other target shock. In order to prevent various models from trying to forecast removed target earthquakes, we also remove all other (minor) shocks belonging to such spatial and time windows. The numbers of considered target shocks for the main shock only set then reduce to 12 and 9 for the learning and testing time intervals, respectively.

ESTIMATION OF PARAMETERS OF EEPAS AND OTHER FORECASTING MODELS USING THE LEARNING DATA SET

Considering the high number of free parameters to be determined for the EEPAS model (in principle about 20), the maximization of the log-likelihood function (eq. 1) would be very time consuming and subject to numerical instability. However, according to Rhoades & Evison (2004), simultaneous optimization of all parameters is not necessary because some of them, such as the b-value of the Gutenberg & Richter (1944) relation and the parameters of the aftershock epidemic decay model (p, k, c, ν), can be, in fact, separately fitted or even be simply assigned based on previous works in the same area.

The b-value of the Gutenberg & Richter (1944) relation is chosen to be representative of the behaviour of the frequency magnitude distribution of target events in the fitting time interval. For the main shocks + aftershocks and main shocks only target earthquake sets, the values $b = 1.084$ and $b = 1.176$, respectively, are computed so that the number of predicted target events by the spatial uniform Poisson model (SUP, see below and Appendix C) exactly match the number of observed events (27 and 12 respectively) in the learning set. This also makes the numbers of shocks predicted by the other models implemented for comparison consistent with observed ones. The parameters of the aftershock model are not particularly critical for the EEPAS model; however, they are necessary to determine the weight w (eq. A21) of the contribution of each earthquake ($M \geq m_c$), by defining the probability of an earthquake being an aftershock of a previous seismic event. The parameters $p = 1.2$ and $c = 0.03$ of eq. (A18) were chosen as typical parameters of Omori's law (Ogata 1983). The two parameters ν and k in eq.

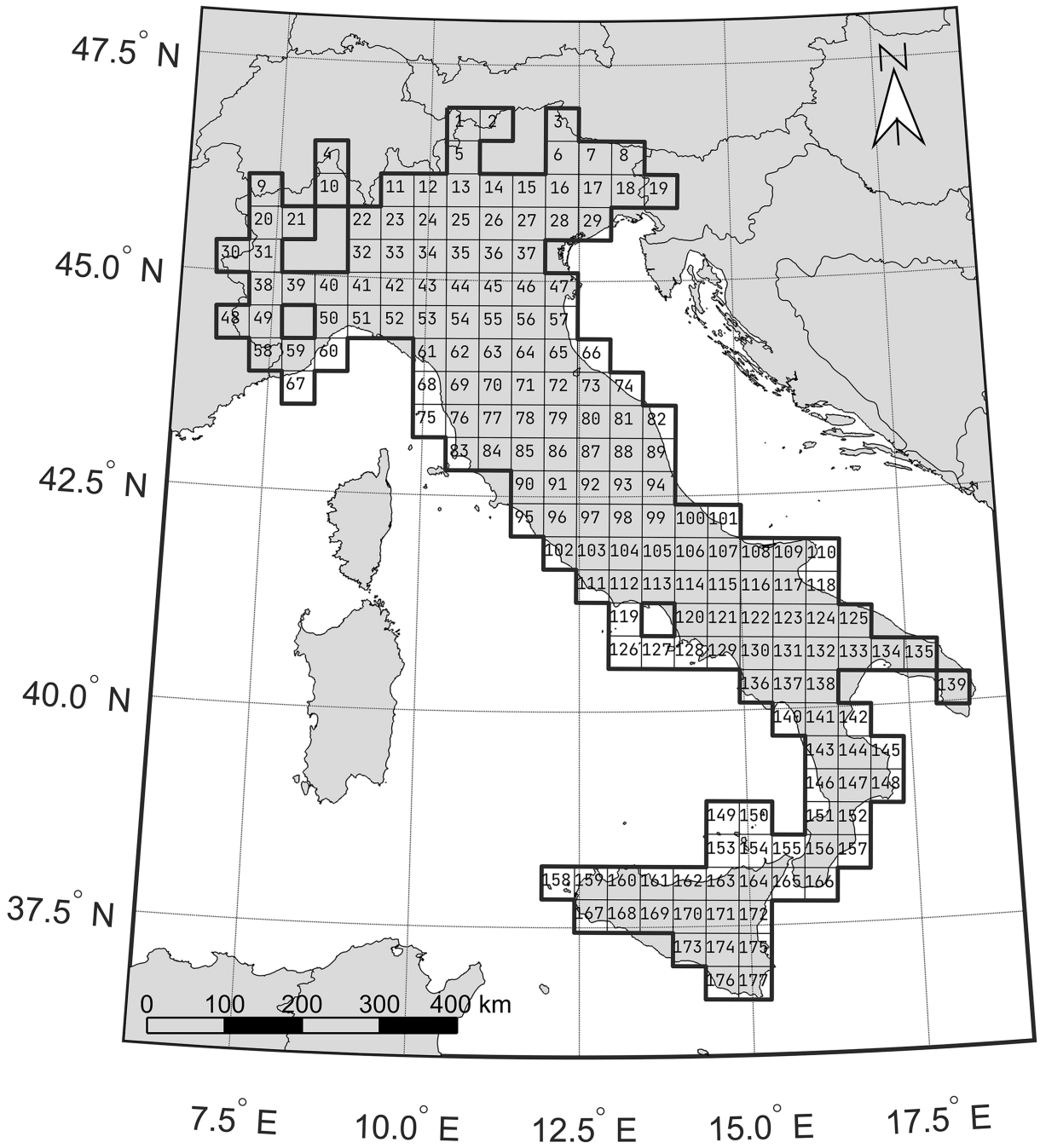

Figure 1. Tessellation of the Italian territory region used for the fitting of parameters and for the pseudo-prospective experiment. The thick black line delimits the analysis region R. The cells that R comprises are only those within which at least one earthquake with $M \geq 4.0$ from 1600 to 2021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

(A16) were fitted by maximizing the likelihood of earthquakes with $m \geq m_T$ occurring within R in the period 1990–2011.

Finally, the parameter $\sigma_U = 0.006$ of eq. (A20) is chosen to be consistent with the mean value of the cluster diffusion parameter for Italy (Musmeci & Vere-Jones 1992). The parameter $\delta = 0.7$ of eq. (A19) is taken from previous works for New Zealand, California and Japan (Rhoades & Evison 2004; Evison & Rhoades 2005; Rhoades 2007, 2011). The parameters of the PPE model (eq. A12) a, d and s are fitted simultaneously using the maximum likelihood method.

Regarding EEPAS parameters, the fit is made in three successive iterations. The parameter b_M is fixed to 1 for all three iterations; that means there is perfect scaling between precursor and target magnitudes (Rhoades & Evison 2004).

In the first iteration, the parameters b_T and b_A are fixed to 0.40 and 0.35, respectively, based on analyses conducted on scaling relationships obtained from the analysis of individual earthquakes. The parameters σ_M and σ_T are also fixed to 0.32 and 0.23, respectively. Such values correspond to the residual standard deviation for the

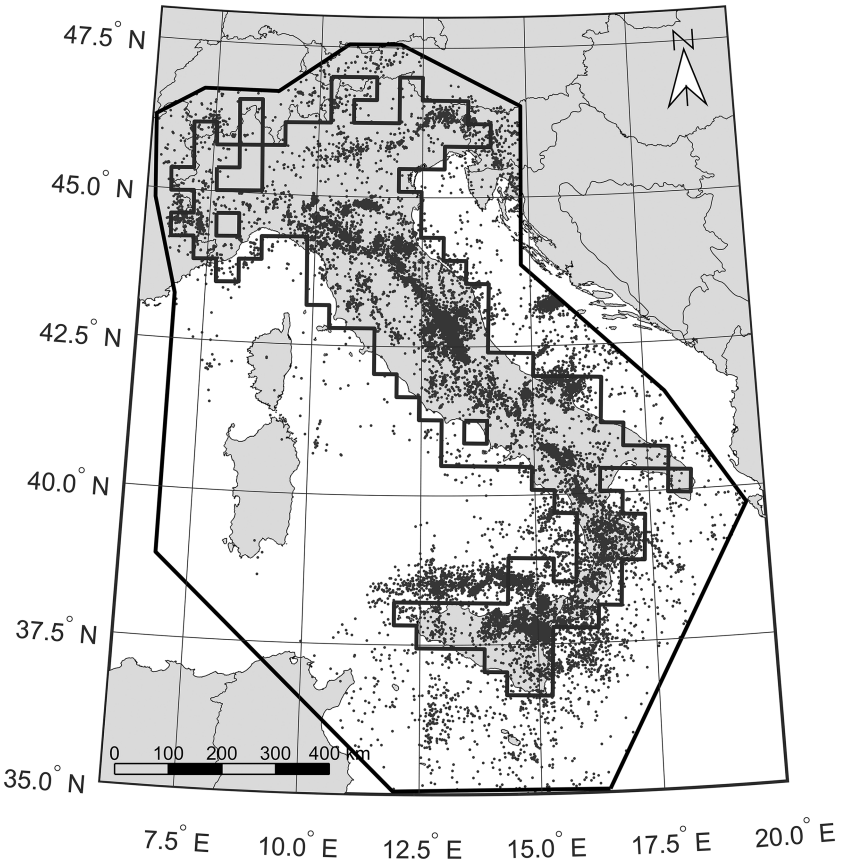

Figure 2. Epicentres of earthquakes with magnitude ≥2.5 that occurred within the CPTI15 polygon (outer thick polygon) between 1990 and 2021. The inner thick polygon represents the forecasting area R.

magnitude and time scaling relations (Rhoades & Evison 2004). Finally, parameters a_T, a_M, σ_A and μ are computed by maximum likelihood estimation.

In the second iteration, the previously fitted parameters a_T, a_M, σ_A are kept fixed at the obtained values and the parameters b_T, b_A, σ_M, σ_T and μ are computed instead by the maximum likelihood.

In the third and last iteration a final computation is made of all parameters (a_T, a_M, σ_A, b_T, b_A, σ_M, σ_T and μ) simultaneously, by providing the optimizer with starting values of the parameters as obtained in previous optimizations. The parameter μ, responsible for mixing the two models PPE and EEPAS, is the only parameter fitted in all three iterations of optimization.

Short descriptions of the parameters of the PPE and EEPAS models are listed in Table 2. The parameter values obtained by maximizing the likelihood are reported in Tables 3 and 4 for the main shocks + aftershocks and the main shocks only target sets respectively. In the same tables we also report the parameters of the other forecasting models (SUP, SVP, ETAS-SUP and ETAS-SVP) computed for comparison. A short description of all implemented models is reported in Table 1.

In Tables 5 and 6, we report the information gain per event (IGPE) and the Akaike information criterion (AIC, Akaike 1974) goodness of fit estimators of various models for the main shocks + aftershocks and the aftershocks only target sets, respectively. The IGPE is defined as

$$\text{IGPE} = \frac{L - \hat{L}}{N}, \tag{3}$$

where L and $\hat{L}$ are the likelihoods obtained by a model and reference model, respectively. The AIC defined as

$$\text{AIC} = -2\log(L) + 2n, \tag{4}$$

where L and n are the maximum likelihood and the number of fitted parameters for the model involved, respectively. The lower the AIC statistic, the better the fit to the data. We can note how both the ETAS models have better scores (higher loglikelihood, information gain per event and lower AIC) than EEPAS and other models. Both EEPAS models also have lower loglikelihoods than SVP and higher AIC (worse) than SUP for the main shock only set. Such scores are not particularly significant because they only represent the goodness of the fit of the models to learning data set and might include some degree of data overfitting.

PSEUDO-PROSPECTIVE COMPARISON OF FORECASTING MODELS ON THE INDEPENDENT TESTING DATA SET

We apply the suite of tests defined by the Collaboratory for the Study of Earthquake Predictability (CSEP, Jordan 2006; Zechar *et al.* 2010b) and particularly the new ones described by Bayona *et al.* (2022).

Such tests assess the consistency of observed earthquakes with a forecast model by (i) the conditional loglikelihood (cL-test), (ii) the observed number of earthquakes (N-test), (iii) their spatial distribution (S-test) and (iv) their magnitude distribution (M-test). However,

Table 1. Summary of applied forecasting models.

Model	Main features	Reference
SUP	Stationary uniform Poisson model based on the observed seismicity rate $M \geq m_T$.	Console & Murru (2001)
SVP	Space variable Poisson model based on the smoothed seismicity rate $M \geq m_T$ observed in each cell.	Console & Murru (2001)
PPE	Quasi time-dependent model based on the hypothesis that future earthquakes tend to occur near the place of the recent ones. The rate-density of future earthquakes is modelled as proportional to a smoothed version of past seismicity	Jackson & Kagan (1999), Rhoades & Evison (2004)
EEPAS	Space/time-dependent model based on the hypothesis that each earthquake ($M \geq m_c$) contributes to the transient increment of the future rate of $M \geq m_T$ in its vicinity according to ψpredictive relations.	Rhoades & Evison (2004)
EEPAS-W	As above but the contribution of aftershocks and triggered events are downweighted	Rhoades & Evison (2004)
ETAS-SUP	Epidemic-type aftershock model based on the hypothesis that each earthquake can perturb the rate of earthquakes and generate its own Omori-like decay sequence. The SUP is used as background model.	Ogata (1988, 1989), Ogata & Zhuang (2006)
ETAS-SVP	As above but with SVP as background model.	Ogata (1988, 1989), Ogata & Zhuang (2006)

Table 2. Summary of PPE and EEPAS parameters.

Model	Parameter	Description	Restriction
PPE	a	Normalizing constant.	≥ 0
	d	Smoothing kernel kilometric distance.	≥ 1
	s	Small value to account for earthquakes far from past epicentres.	≥ 0
EEPAS	a_M	Intercept of scaling relation between precursor magnitude and target magnitude (eq. A1).	$1.0 - 2.0$
	b_M	Slope of scaling relation between precursor magnitude and target magnitude (eq. A1).	1 (fixed)
	σ_M	Standard deviation of scaling relation between precursor magnitude and target magnitude (see Fig.1, Rhoades 2011)	$0.2 - 0.65$
	a_T	Intercept of scaling relation between precursor time and target magnitude (eq. A2)	$1 - 3$
	b_T	Slope of scaling relation between precursor time and target magnitude (eq. A2)	$0.3 - 0.65$
	σ_T	Standard deviation of the scaling relation between precursor magnitude and target magnitude (see Fig.1, Rhoades & Evison 2011).	$0.15 - 0.6$
	b_A	Slope of scaling relation between precursory area and target magnitude (eq. A3).	$0.2 - 0.6$
	σ_A	Related to Aa of scaling relation between precursory area and magnitude (see Fig.1, Rhoades & Evison 2011).	$1 - 30$
	μ	Proportion of target shocks that occurs without an appreciable sequence of precursory shocks.	$0 - 1$
Aftershocks (EEPAS)	c	c-parameter of Omori–Utsu law	$0 - 0.5$
	p	p-parameter of Omori–Utsu law	$1 - 1.6$
	k	Normalizing constant.	≥ 0
	ν	Proportion of earthquake that are not aftershocks	$0 - 1$
	δ	Average magnitude difference between the main shock and the largest aftershock	0.7 (Fixed)
	σ_U	Cluster diffusion parameter	0.006 (Fixed)

Table 3. Estimated parameters for various models (main shock + aftershocks).

SUP		SVP		PPE		EEPAS-NW		EEPAS-W		ETAS-SUP		ETAS-SVP	
b	1.08	b	1.08	b	1.08	b	1.08	b	1.08	b	1.08	b	1.08
		d_c	14.5	a	$0.62 \pm \|0.18$	a	0.62 ± 0.18	a	0.62 ± 0.18	K	0.029 ± 0.001	K	$0.0207 \pm 5 \times 10^{-4}$
				d	30 ± 12	d	30 ± 12	d	30 ± 12	c	0.004 ± 0.001	c	$0.0021 \pm 1 \times 10^{-4}$
				s	9.0×10^{-13}	s	9.0×10^{-13}	s	9.0×10^{-13}	p	1.042 ± 0.004	p	1.084 ± 0.003
					$\pm 1.5 \times 10^{-3}$		$\pm 1.5 \times 10^{-3}$		$\pm 1.5 \times 10^{-3}$				
						a_M	1.22 ± 0.17	a_M	1.23 ± 0.16	D	1.04 ± 0.03	D	0.895 ± 0.022
						b_M	1^a	b_M	1^a	γ	0.45 ± 0.04	γ	0.53 ± 0.42
						σ_M	0.25 ± 0.11	σ_M	0.24 ± 0.11	α	1.12 ± 0.03	α	1.22 ± 0.02
						a_T	2.55 ± 0.07	a_T	$2.7\,1 \pm 0.07$	q	1.5^a	q	1.5^a
						b_T	0.35 ± 0.02	b_T	0.32 ± 0.02	f_r	0.264 ± 0.006	f_r	0.416 ± 0.006
						σ_T	0.150 ± 0.004	σ_T	0.150 ± 0.004				
						b_A	0.52 ± 0.06	b_A	0.51 ± 0.05				
						σ_A	1.00 ± 0.03	σ_A	1.00 ± 0.03				
						μ	0.18 ± 0.13	μ	0.16 ± 0.12				
								c	0.03^a				
								p	1.2^a				
								k	0.13^a				
								ν	0.61^a				
								δ	0.7^a				
								σ_U	0.006^a				

[a]Fixed or estimated independently.

Table 4. Estimated parameters for various models (main shocks).

SUP		SVP		PPE		EEPAS-NW		EEPAS-W		ETAS-SUP		ETAS-SVP	
b	1.18	b	1.18	b	1.18	b	1.18	b	1.18	b	1.18	b	1.18
		d_c	14.5	a	0.39 ±|0.20	a	0.39 ±|0.20	a	0.39 ±|0.20	K	0.028 ± 0.001	K	0.018 ±5 × 10^{-4}
				d	3 3 ± 15	d	3 3 ± 15	d	3 3 ± 15	c	0.002 ± 2 × 10^{-4}	c	0.001 ±8 × 10^{-5}
				s	1.58 × 10^{-7} ± 0.002	s	1.58 × 10^{-7} ± 0.002	s	1.58 × 10^{-7} ± 0.002	p	1.002 ± 0.003	p	1.0 67±0.003
						a_M	1.34 ± 0.13	a_M	1.3 4 ± 0.13	D	1.28 ± 0.04	D	1.04 ±0.03
						b_M	1^a	b_M	1^a	γ	0.36 ± 0.07	γ	0.48 ± 0.08
						σ_M	0.200 ± 0.005	σ_M	0.200 ± 0.005	α	0.81 ± 0.03	α	0.86 ±0.03
						a_T	1.35 ± 0.11	a_T	1.36 ± 0.11	q	1.5^a	q	1.5^a
						b_T	0.60 ± 0.03	b_T	0.60 ± 0.03	f_r	0.39 ± 0.01	f_r	0.55 ± 0.01
						σ_T	0.15 ± 0.06	σ_T	0.15 ± 0.07				
						b_A	0.45 ± 0.07	b_A	0.48 ± 0.07				
						σ_A	1 .64± 0.58	σ_A	1.45 ± 0.44				
						μ	0.34 ± 0.30	μ	0.36 ± 0.29				
								c	0.03^a				
								p	1.2^a				
								k	0.13^a				
								ν	0.61^a				
								δ	0.7^a				
								σ_U	0.006^a				

[a]Fixed or estimated independently.

Table 5. Performance estimators of various models in the learning time interval (1990–2011) (main shock + aftershocks).

	SUP	SVP	PPE	EEPAS-NW	EEPAS-W	ETAS-SVP	ETAS-SUP
E	27	27.22	27	27.67	27.73	27.49	27.52
lnL	−524.63	−465.47	−514.11	−500.39	−496.06	−363.87	−363.58
IGPE	0.00	2.19	0.39	0.90	1.06	5.95	5.97
AIC	1051.3	934.9	1036.2	1026.8	1018.1	727.7	727.1
ΔAIC	0.00	2.15	0.28	0.45	0.61	5.99	6.00

E, number of predicted events; lnL, loglikelihood score; IGPE, information gain per event with respect to model SUP; AIC, Akaike Information Criteria; ΔAIC, Akaike Information Criteria difference with respect to model SUP per event.

Table 6. Performance estimators of various models in the learning time interval (1990–2011) (main shocks only).

	SUP	SVP	PPE	EEPAS-NW	EEPAS-W	ETAS-SVP	ETAS-SUP
E	12.00	12.19	11.99	14.26	14.75	12.01	11.97
lnL	−246.15	−237.68	−243.52	−239.92	−239.79	−215.57	−212.91
IGPE	0.00	0.75	0.22	0.52	0.54	3.03	3.16
AIC	494.30	479.37	495.04	505.84	505.40	435.49	432.57
ΔAIC	0.00	0.62	−0.03	−0.48	−0.46	1.88	2.10

E, number of predicted events; lnL, loglikelihood score; IGPE, information gain per event with respect to model SUP; AIC, Akaike Information Criteria and ΔAIC, Akaike Information Criteria difference with respect to model SUP.

we do not report the results for the latter, because all forecasting models assume a Gutenberg–Richter frequency–magnitude distribution and all pass the M-test.

Traditional CSEP tests are based on a likelihood function that regards earthquakes in individual cells or bins as independent and Poisson distributed (Schorlemmer *et al.* 2007, 2010; Zechar *et al.* 2010b). However, the Poisson distribution insufficiently captures the spatiotemporal variability of earthquakes, especially in the presence of clusters of seismicity (Werner & Sornette 2008; Lombardi & Marzocchi 2010; Nandan *et al.* 2019). The new CSEP tests are characterized by a lower sensitivity to clustering of target events rather than the traditional ones.

The new CSEP N-test compares the number of predicted earthquakes in all (time–space–magnitude) bins with the number of target earthquakes observed and is based on the negative binomial distribution (NBD)

$$p\ [(\omega|\tau,\nu)] = \frac{\Gamma(\tau+\omega)}{\Gamma(\tau)\,\omega!}\,\nu^{\omega}(1-\nu)^{\omega}, \tag{5}$$

where $\omega = 1, 2, \ldots$ is the number of events, $\tau > 0$ and $0 \leq \nu \leq 1$ are parameters and Γ is the Gamma function. The mean and the variance of NBD are given by

$$\mu = \tau\frac{1-\nu}{\nu};\ \ \sigma^2 = \frac{1-\nu}{\nu^2} \tag{6}$$

According to Werner *et al*. (2010) and Bayona *et al.* (2022) we used the number of expected earthquakes as the mean value μ of the NBD. The variances are determined considering the numbers of events with $M_w \geq 5.0$ within 10 yr non overlapping intervals from 1882 to 2011 (Fig. 3) from the Italian historical catalogue CPTI15 (Rovida *et al.* 2020). The computed variances are $\sigma^2_{ND} = 67.76$ and $\sigma^2_D = 35.94$ for the main shocks + aftershocks and the main shocks only data sets, respectively.

The numbers of earthquakes corresponding to 95 per cent ($p = 0.025 \div 0.975$) and 97.5 per cent ($p = 0.0125 \div 0.9875$) predictive limits, based on the NBD cumulative distribution function, are computed. If the observed number of earthquakes falls within such limits, the model satisfactorily describe the observed data.

The new binary cL-test compares the joint binary log-likelihood (JBLL) of the forecasted events by a model with the observed seismicity, with the distribution of joint binary log-likelihoods obtained by the simulation of random catalogues.

The rates forecasted by the model within each active bin (i.e. time–space–magnitude bins containing observed events) are first normalized to the total number of observed active bins (so that their sum is 1). The active bins are then sorted according to the increasing value of the normalized rate and a vector of cumulated normalized rates is computed (ranging from 0 and 1). The number $N_{\rm sim}$ of bins to simulate is fixed to the number of observed active bin. For each simulated bin, a uniformly distributed random number is extracted in the interval]0,1]. An earthquake is then placed in the first bin for which the cumulated normalized rate exceeds the random sampled value (Zechar *et al.* 2010).

For each simulated catalogue the JBLL is estimated and, after $N = 10\,000$ simulated catalogues, the simulated JBLL distribution is obtained. If the JBLL obtained by a model lies in the lower tail of the simulated JBLL distribution, the forecasting model does not reproduce well the real seismicity pattern and then the test fails.

The new binary log-likelihood is obtained by calculating the probability of an earthquake in a forecast bin. Assuming the Poissonian distribution, the probability of observing $\omega = 0$ events, given an expected rate λ, is $P_0 = \exp(-\lambda)$, while the probability of observing more than zero events is $P_1 = 1 - P_0$ (Bayona *et al.* 2022). The binary log-likelihood for each bin is thus given by

$$\mathrm{BLL} = X_i\ \ln\left(1-\exp\left(-\lambda\right)\right) + \left(1-X_i\right)\ln\left(\exp(-\lambda)\right), \tag{7}$$

where $X_i = 1$ if the ith bin contains at least one event and $X_i = 0$ otherwise. The observed binomial joint log-likelihood is given by the summation of the BLL over all space–magnitude–time bins:

$$\mathrm{JBLL} = \sum_{l=1}^{s}\sum_{j=1}^{m}\sum_{k=1}^{t} X(l,j,k)\ln\left[1-\exp\left(-\lambda\left(l,j,k\right)\right)\right] + \left[1-X(l,j,k)\ln(\exp\left(-\lambda\left(l,j,k\right)\right)\right]\ . \tag{8}$$

The S-test evaluates the consistency of the spatial occurrence of target earthquakes regardless of their magnitudes. For the new S-test the joint binary log-likelihood of the forecasted catalogue

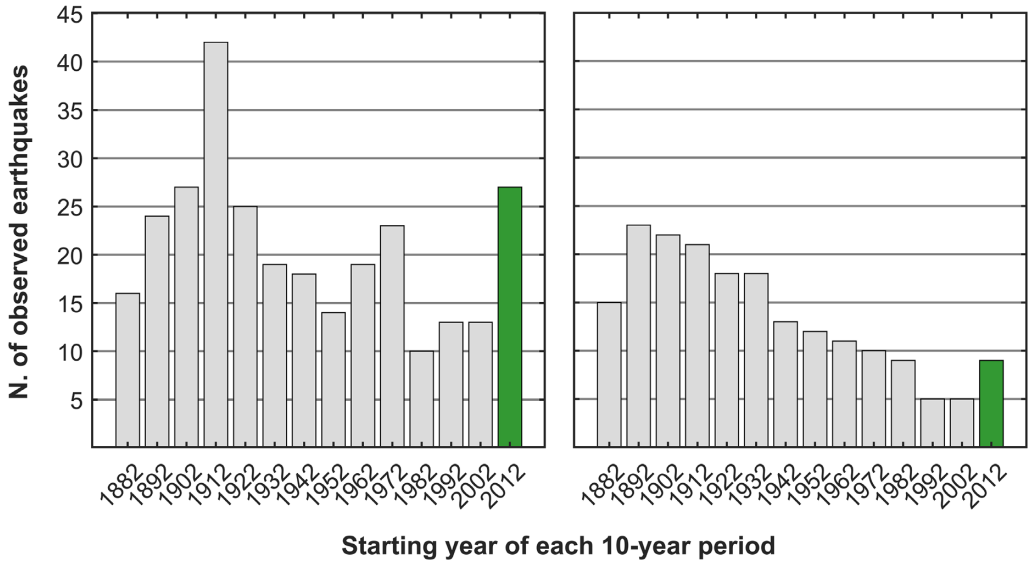

Figure 3. Number of target events ($M \geq 5.0$) reported in the CPTI15 earthquake catalogue (from 1880 to 1959) and in the HORUS seismic catalogue (from 1960 to 2021) in non-overlapped 10-yr intervals for main shocks + aftershocks (left-hand panel) and main shocks only (right-hand panel) used to estimate the variance of the catalogues. The green bars represent the number of target events occurred during the testing period (from 2012 to 2021) not used to estimate the variances.

is calculated considering only the spatial distribution of forecasted events. To isolate the spatial distribution, the forecasted events are summed over the magnitude and time bins. In addition, to assess the JBLL, the forecasted catalogue is normalized to the number of active spatial cells. The simulation procedure to obtain the spatial simulated JBLL is similar to that described above for the binary cL-test, but N_{sim} is fixed to the number of active spatial cells. For the cL, N and S test, the computed statistic is the quantile score: that is, the fraction of simulated likelihoods that are less than or equal to the likelihood observed by the model. A small value, lower than the usual significance level $\alpha = 0.05$ or than the Bonferroni-adjusted significance level $\alpha_{Bf} = 0.05/2 = 0.025$ means that the model inadequately describes the seismicity pattern.

To evaluate the relative skill of the forecasting models, we use the information gain per active bin (IGPA, Bayona *et al.* 2022), which is based on the likelihood difference with respect to a reference (baseline) forecasting model divided by the number of earthquakes or by the number of active bins (the bins in which the likelihood contribution is not zero) respectively. The IGPA is thus given by

$$IGPA = \frac{N_{\text{base}} - N_{\text{mod}}}{M} + \frac{1}{M}\sum_{m=1}^{M}\left[X_{\text{mod}}(m) - X_{\text{base}}(m)\right], \quad (9)$$

where N_{base} and N_{mod} are the total number of earthquakes expected by the baseline and the model respectively, M is the number of active bins, and $X_{\text{mod}}(m)$ and $X_{\text{base}}(m)$ are the joint log-likelihood score obtained in the bin with the $m-th$ target earthquake by the model and the reference baseline model respectively. According to Rhoades *et al.* (2011) the variance of $X_{\text{mod}}(m) - X_{\text{base}}(m)$ is given by

$$s^2 = \frac{1}{M-1}\sum_{m=1}^{M}\left(X_{\text{mod}}(m) - X_{\text{base}}(m)\right)^2 - \frac{1}{M^2 - M}\left[\sum_{m=1}^{M} X_{\text{mod}}(m) - X_{\text{base}}(m)\right]^2. \quad (10)$$

The IGPA error is estimated as $\pm ts\sqrt{M}$, where t is the 95th\ (or 97.5th) percentile of the Student's t distribution with $M-1$ degrees of freedom.

As baseline model we take the SUP, which is the simpler one. We do not need a correction for the number of free parameters, as proposed by Rhoades *et al.* (2014), because the fitting of models is independent of the testing set targets, being made using the learning set. In addition, we do not use the parimutuel gambling score (PGS) by Zhuang (2010) and Zechar & Zhuang (2014), because Serafini *et al.* (2022) recently demonstrated that PGS is improper when the number of forecasting methods being tested is greater than two.

RESULTS

In Fig. 4 and Table 7, we report the numbers of main shocks + aftershocks targets predicted by various models using different time intervals (3 months, 6 months, 1 yr, 5 yr and 10 yr) of prediction. All models definitely underestimate the total number of target earthquakes (27) that actually occurred. The reason is that the average rate of targets in the testing set (about 2.7 per yr) is more than twice than that in learning set (about 1.2 per yr). However, according to Werner *et al.* (2010) , the negative binomial N-test is characterized by wider confidence intervals than the traditional Poissonian N-test and then the forecast models are found to be consistent as the observed number earthquakes is within both the 95 and 97.5 per cent confidence intervals (Fig. 5 and Table S1). The binary cL-tests show that all forecasted models adequately describe the observed seismicity as the quantile scores exceed the 0.025 and 0.05 significance levels (Fig. 6 and Table S2). The S-test (Fig. 7 and Table S3) confirm the spatial consistency between the forecasts and the observed data set. The results of the IGPA (T-test) for main shocks + aftershocks targets in Fig. 8 and Table S4 indicate that for the shortest prediction interval of 3 months the best performing models are the ETAS-SVP and ETAS-SUP. For longer prediction intervals, the best performing

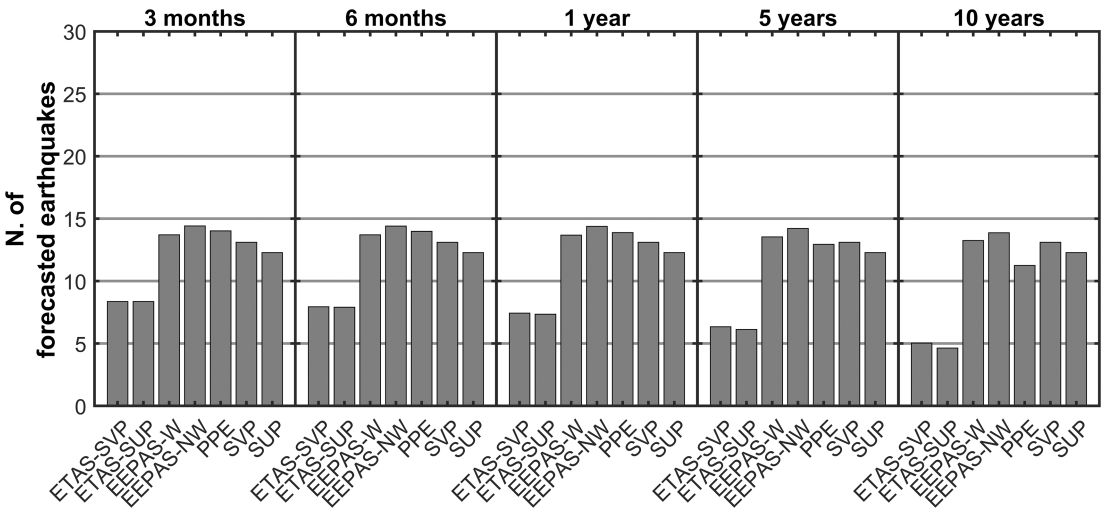

Figure 4. Numbers of targets (main shocks + aftershocks) in the testing set (2012–2021) predicted by various models using different prediction intervals. The effective total number of targets is 27.

Table 7. Numbers of earthquakes predicted by various models in the testing time interval (2012–2021) (main shocks + aftershocks).

Time interval	SUP	SVP	PPE	EEPAS-NW	EEPAS-W	ETAS-SUP	ETAS-SVP
3 months	12.27	13.11	14.03	14.41	13.70	8.36	8.00
6 months	12.27	13.11	13.98	14.40	13.69	7.43	7.73
1 yr	12.27	13.11	13.87	14.38	13.67	6.94	7.42
5 yr	12.27	13.11	12.94	14.21	13.53	5.90	6.79
10 yr	12.27	13.11	11.25	13.86	13.25	4.66	6.05

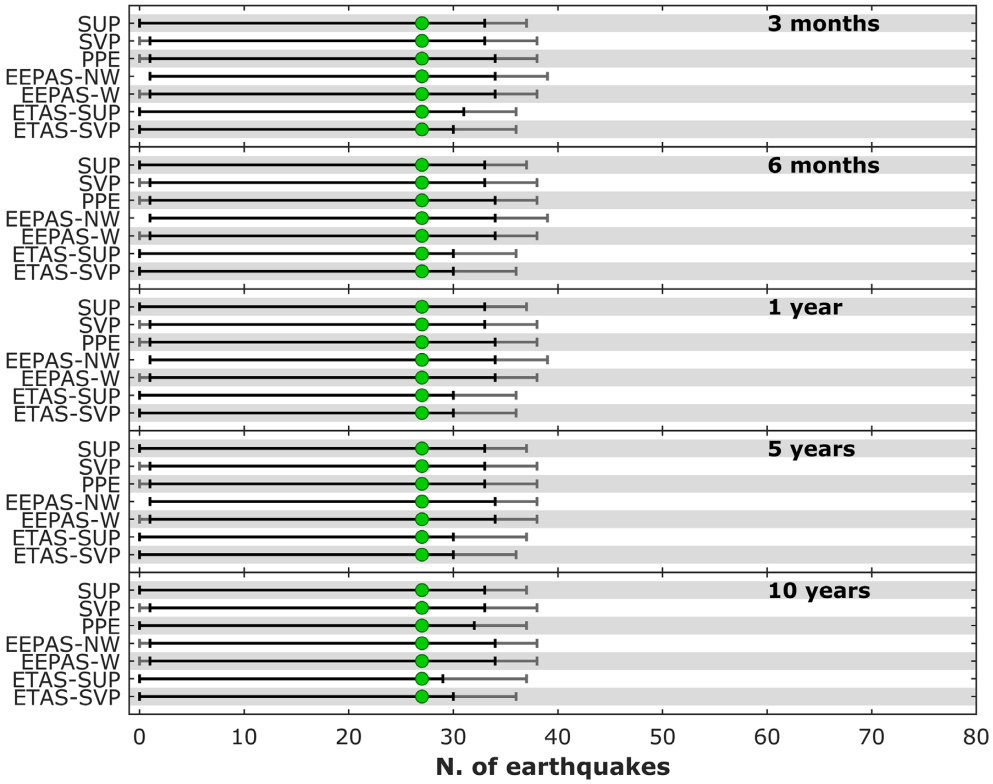

Figure 5. Results of number consistency test (N-test) in the testing set (2012–2021) for various models using different prediction intervals (main shocks + aftershocks). Coloured circles indicate the number of observed events in the testing set. Black and grey bars indicate the 95 and 97.5 per cent predictive intervals, respectively. Green coloured circles indicates that all models passed the test.

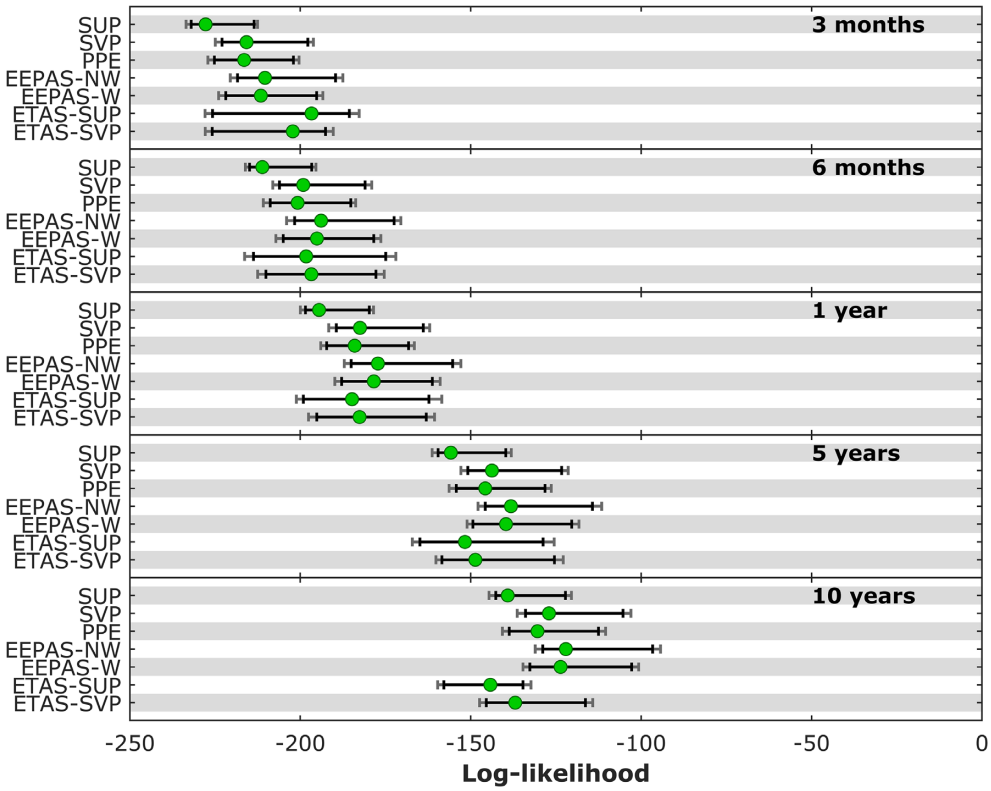

Figure 6. Results of conditional likelihood consistency test (cL-test) in the testing set (2012–2021) for various models using different prediction intervals (main shocks + aftershocks). Black and grey bars indicate the 95 and 97.5 per cent confidence limits, respectively. Green coloured circles indicates that all models passed the test.

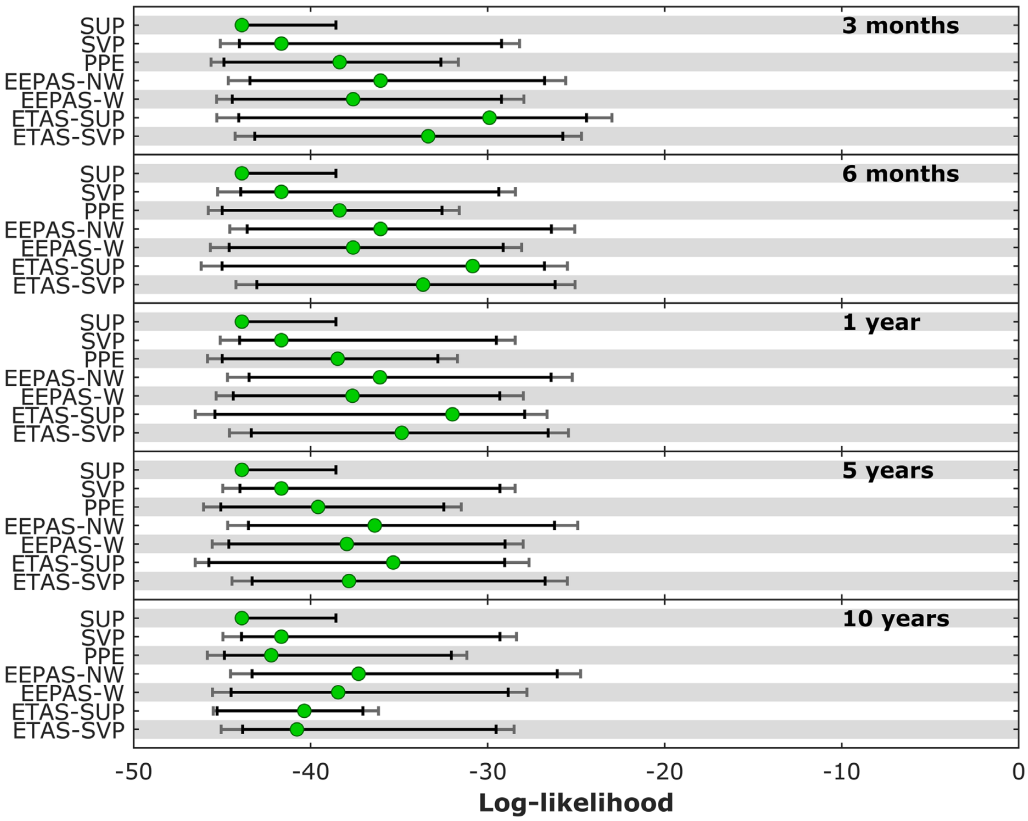

Figure 7. Results of spatial consistency test (S-test) in the testing set (2012–2021) for various models using different prediction intervals (main shocks + aftershocks). Black and grey bars indicate the 95 and 97.5 per cent confidence limits, respectively. Green coloured circles indicates that all models passed the test.

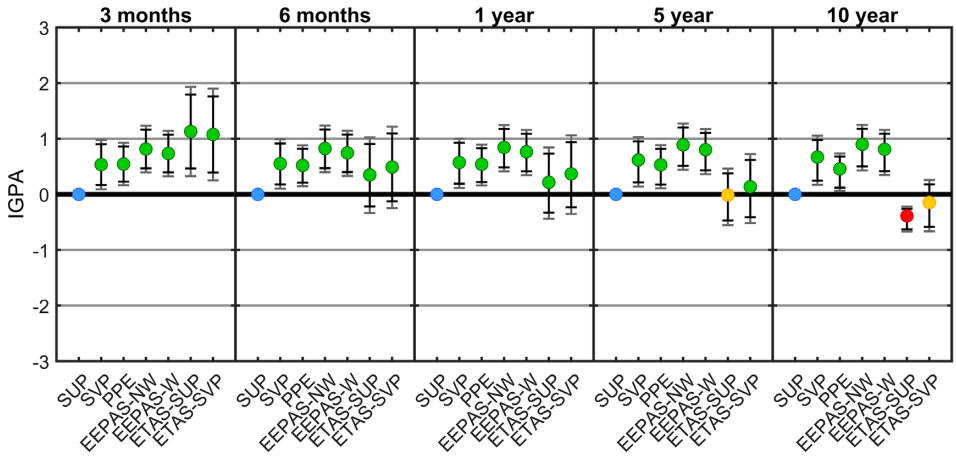

Figure 8. Comparison between various models in different time intervals (main shocks + aftershocks) in the testing set (2012–2021) by the IGPA (T-test). Black and grey bars indicate the 95 and 97.5 per cent confidence limits, respectively. Coloured circles are green if the IGPA of a model is larger than the reference value 0 corresponding to the SUP model, yellow if IGPA of a model is lower than the reference model, but not significantly and red if a model IGPA is significantly lower than the reference model.

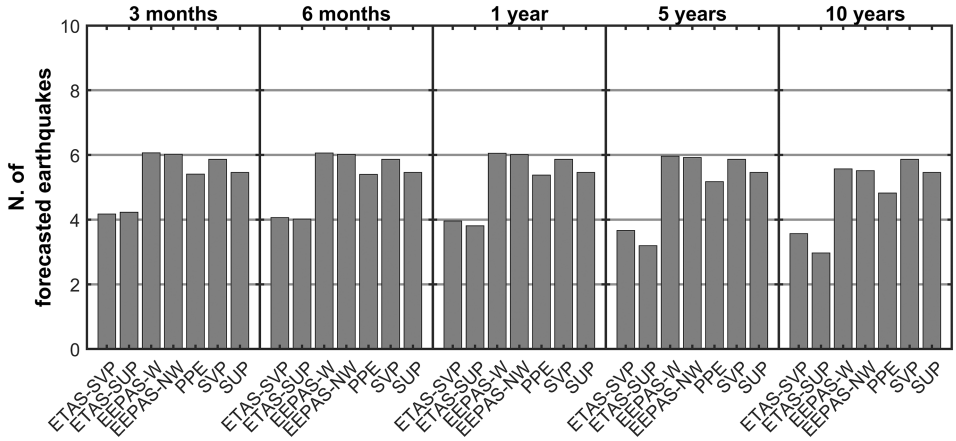

Figure 9. Numbers of targets (main shocks only) in the testing set (2012–2021) predicted by various models and prediction intervals. The effective total number of targets is 9.

Table 8. Numbers of earthquakes predicted by various models in the testing time interval (2012–2021) (main shocks only).

Time interval	SUP	SVP	PPE	EEPAS-NW	EEPAS-W	ETAS-SUP	ETAS-SVP
3 months	5.46	5.86	5.41	6.02	6.06	4.23	4.17
6 months	5.46	5.86	5.40	6.02	6.06	4.02	4.06
1 yr	5.46	5.86	5.38	6.01	6.05	3.82	3.96
5 yr	5.46	5.86	5.18	5.92	5.96	3.19	3.67
10 yr	5.46	5.86	4.82	5.51	5.57	2.97	2.57

models are the EEPAS-NW and EEPAS-W but such superior performance appears to be statistically significant only for time intervals of 5 and 10 yr.

In Fig. 9 and Table 8, we report the numbers of main shocks only targets predicted by various models using different time intervals of prediction. All models still underestimate the total number of targets (eq. 9) that actually occurred, as even in this case the average rate of targets in the testing set (0.9 per yr) is greater than in the learning set (0.5 per yr). All models pass the negative binomial N-test (Fig. 10 and Table S5), and the binary cL-test (Fig. 11 and Table S6) and S-test (Fig. 12 and Table S7) for all time intervals of prediction. The results of the IGPA (T-test) main shock only targets in Fig. 13 and Table S8 confirm that, based on the IGPA, the best performing models are ETAS-SVP and ETAS-SUP for the shortest prediction interval of 3 months and the EEPAS-NW and EEPAS-W for longer prediction intervals. However, such superior performance is not significant for any time interval.

CONCLUSIONS

We applied the EEPAS earthquake forecasting model to Italy, similarly to previous application in other seismic regions of the world (e.g. Rhoades & Evison 2004; Evison & Rhoades 2005; Rhoades 2007, 2011; Rhoades *et al.* 2020), using a suite of computing codes completely rewritten in Matlab and implementing both EEPAS formulations with the input earthquakes not weighted (EEPAS-NW) and weighted (EEPAS-W). We calibrated and fitted the model parameters using earthquakes of the HORUS seismic catalogue of

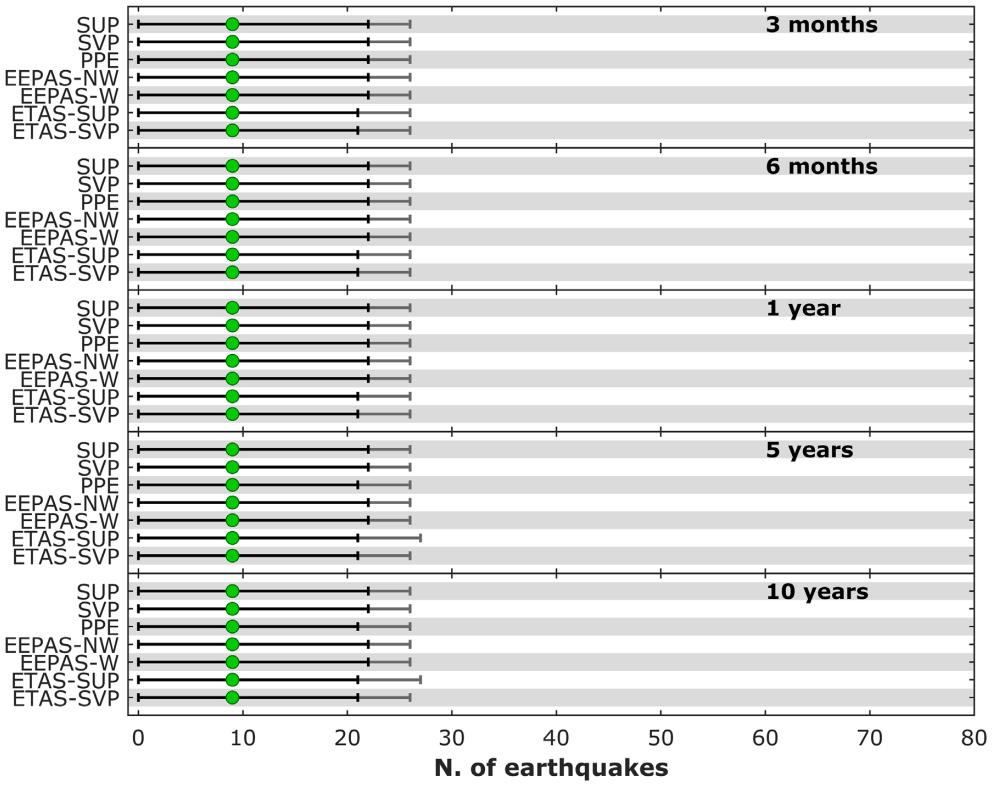

Figure 10. Results of number consistency test (N-test) in the testing set (2012–2021) for various models using different prediction intervals (main shocks only). Black and grey bars indicate the 95 and 97.5 per cent confidence limits, respectively. Coloured circles indicate the number of observed events in the testing set. Green coloured circles indicates that all model passed the test.

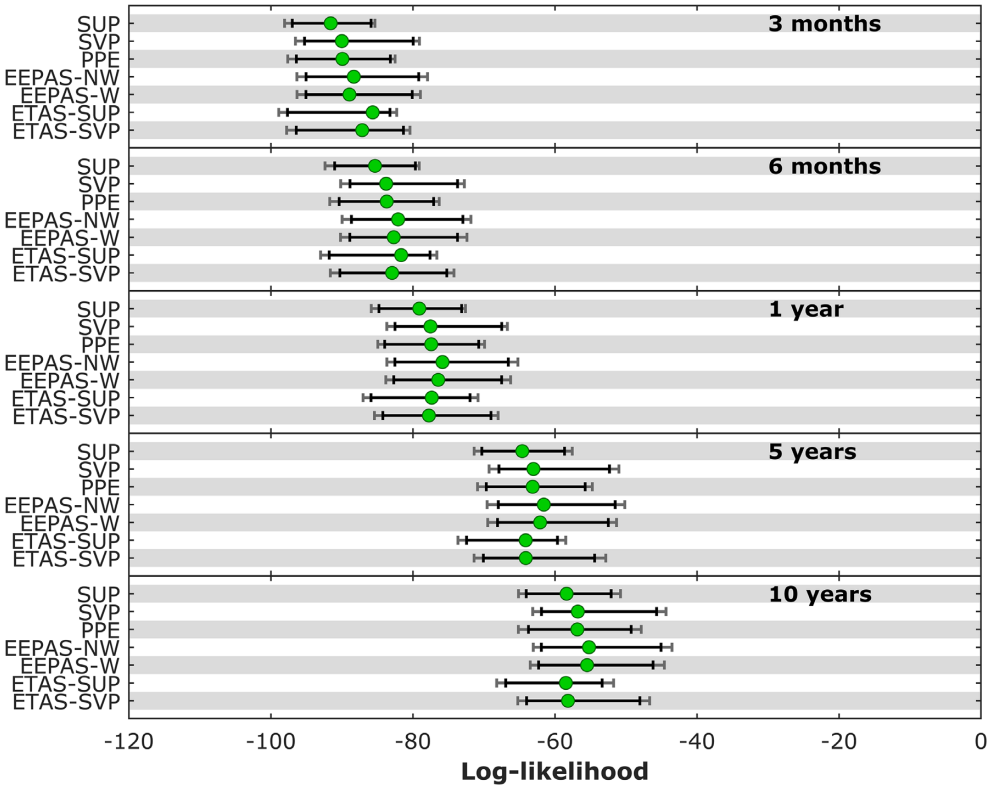

Figure 11. Results of conditional likelihood consistency test (cL-test) in the testing set (2012–2021) for various models using different time intervals (main shocks only). Black and grey bars indicate the 95 and 97.5 per cent confidence limits, respectively. Green coloured circles indicates that all model passed the test.

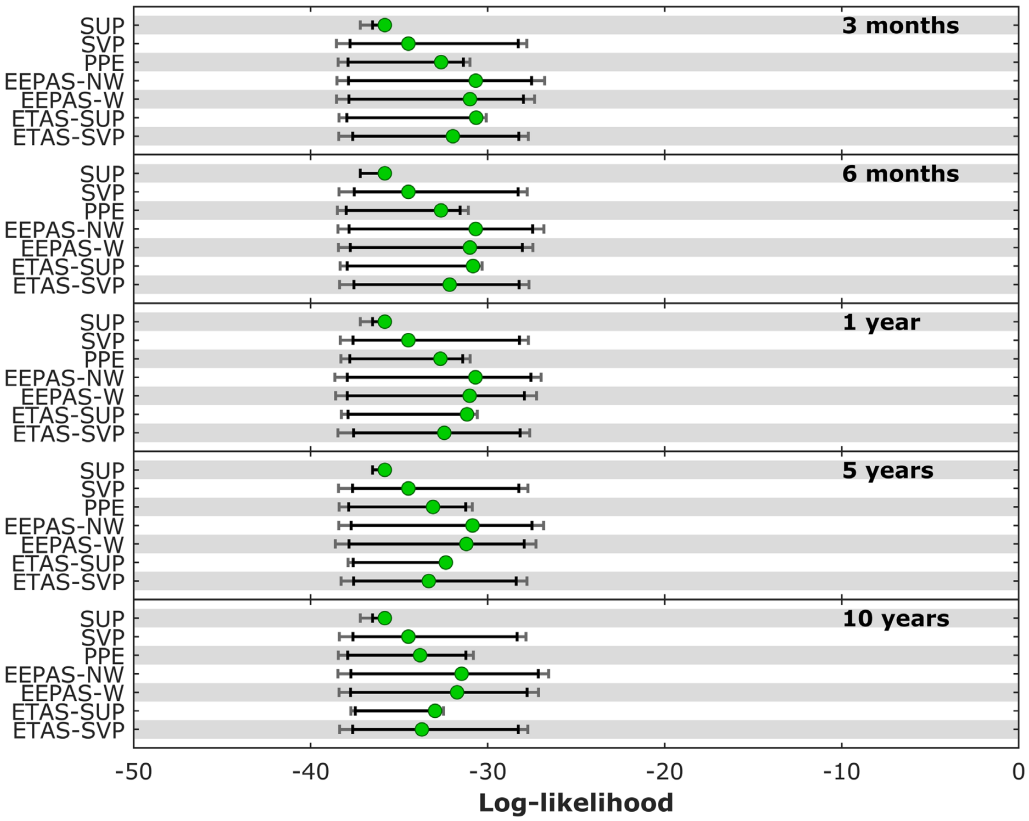

Figure 12. Results of spatial consistency test (S-test) in the testing set (2012–2021) for various models using different prediction intervals (main shocks only). Black and grey bars indicate the 95 and 97.5 per cent confidence limits, respectively. Green coloured circles indicates that all model passed the test.

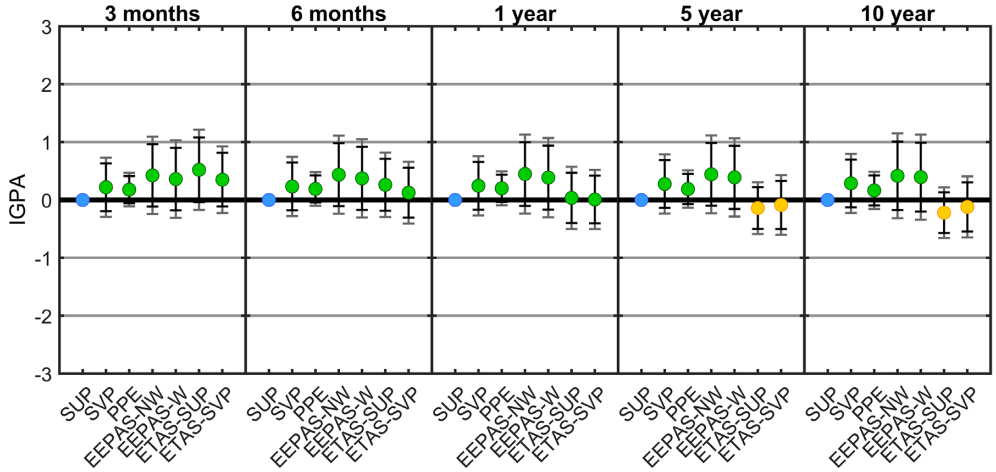

Figure 13. Comparison between various models in different time intervals (main shocks only) in the testing set (2012–2021) by the IGPA (T-test). Black and grey bars indicate the 95 and 97.5 per cent confidence limits, respectively. Coloured circles are green if the IGPA of a model is larger than the reference value 0 corresponding to the SUP model, yellow if IGPA of a model is lower than the reference model, but not significantly and red if a model IGPA is significantly lower than the reference model.

Italy (Lolli *et al.* 2020) for the learning period 1990–2011. The EEPAS model was then applied to forecast all earthquakes (main shocks + aftershocks) of the same seismic catalogue with $M \geq 5.0$ and only the main shocks that occurred within the polygon of analysis for the test period 2012–2021. We compared the forecasting skill of EEPAS with the ones obtained by other time dependent (ETAS-SUP and ETAS-SVP) and time independent (SUP, SVP and PPE) models implemented on the same data set. We used a set of new CSEP consistency tests based on a binary likelihood function as described in Bayona *et al.* (2022). This latter likelihood function reduces the sensitivity of spatial log-likelihood scores to the occurrence of seismic events (Bayona *et al.* 2022) compared to previous versions of the tests based on a Poisson distribution assumption. The number of expected target earthquakes forecasted by each model tends to decrease as the forecasting interval increases. The highest expected number of earthquakes is for a window of 3 months. However, all models tend to underestimate the total numbers of expected

events that actually occurred, 27 and 9 for the main shocks + aftershocks and main shocks only data sets, respectively. This is due to the different average rate of target events in the learning and testing period. This difference does not affect the performance of consistency tests, which are passed by all models. In particular, the cL and S-test showed an adequate consistency between the forecasted and the observed events distribution as the quantile scores exceed the significance values of 0.025 and 0.05.

The difference in the seismicity rate between the learning and the testing period is less pronounced for the main shocks only data set and the consistency tests show similar results. We also assess the relative forecasting skill of various models using the IGPA (Rhoades *et al.* 2011; Bayona *et al.* 2022) considering as baseline reference model the SUP. For both main shocks + aftershocks and main shocks only data sets, the best performing model is the ETAS-SUP and ETAS-SVP for the shortest prediction interval of 3 months and the EEPAS-NW and EEPAS-W for the longer prediction intervals. These results confirm the different characteristics of the models ETAS and EEPAS. The latter model seems more appropriate than the ETAS model for making forecasts of the long-term seismicity, even if the small number of target shocks suggest some caution. The reason of such different behaviours can be that ETAS model is focused on short time clustering while EEPAS model to long time clustering. In particular, the weighted version of EPAS is designed so that to weigh less earthquakes that are likely to be aftershocks of a previous one (see Appendix A). These results also suggests that EEPAS could be a valid candidate for hybrid forecasting models in combination to EPAS.

ACKNOWLEDGMENTS

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). One author (DAR) was supported by the New Zealand Ministry of Business, Innovation and Employment (MBIE) through the Hazards and Risks Management programme (Strategic Science Investment Fund, contract C05 × 1702).

DATA AVAILABILITY

We provide the Matlab source codes used for fitting the EEPAS model in supplementary information. The data used are available from various public providers.

CONFLICT OF INTEREST

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

REFERENCES

Akaike, H., 1974. A new look at the statistical model identification, *IEEE Trans. Autom. Contl.,* **19**(6), 716–723.

Båth, M., 1965. Lateral inhomogeneities of the upper mantle, *Tectonophysics,* **2**(6), 483–514.

Bayona, J. A., Savran, W. H., Rhoades, D. A. & Werner, M. J., 2022. Prospective evaluation of multiplicative hybrid earthquake forecasting models in California, *Geophs. J. Int.,* **229**(3), 1736–1753.

Console, R. & Murru, M., 2001. A simple and testable model for earthquake clustering, *J. geophys. Res.,* **106**(B5), 8699–8711.

Console, R., Murru, M. & Catalli, F., 2006. Physical and stochastic models of earthquake clustering, *Tectonophysics,* **417**(1–2), 141–153.

Console, R., Murru, M. & Lombardi, A. M., 2003. Refining earthquake clustering models, *J. geophys. Res.,* **108**(B10), 2468.

Evison, F. & Rhoades, D., 2004. Long-term seismogenesis and self-organized criticality, *Earth, Planets Space,* **56**(8), 749–760.

Evison, F. & Rhoades, D., 2005. Multiple - mainshock events and long - term seismogenesis in Italy and New Zealand, *N. Zeal. J. Geol. Geophys.,* **48,** 523–536.

Frankel, A., 1995. Mapping seismic hazard in the central and eastern United States, *Seismol. Res. Lett.,* **66**(4), 8–21.

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

Gasperini, P., Lolli, B. & Vannucci, G., 2013. Empirical calibration of local magnitude data sets versus moment magnitude in Italy, *Bull. seism. Soc. Am.,* **103**(4), 2227–2246.

Gutenberg, R. & Richter, C., 1944. Frequency of earthquakes in California, *Bull. seism. Soc. Am.,* **34**(4), 185–188.

Jackson, D. D. & Kagan, Y. Y., 1999. Testable earthquake forecasts for 1999, *Seismol. Res. Lett.,* **70**(4), 393–403.

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

Lombardi, A. M. & Marzocchi, W., 2010. The ETAS model for daily forecasting of Italian seismicity in the CSEP experiment, *Ann. Geophys.,* **53**(3), 155–164.

Musmeci, F. & Vere-Jones, D., 1992. A space-time clustering model for historical earthquakes, *Ann. Inst. Stat. Math.,* **44**(1), 1–11.

Nandan, S., Ouillon, G. & Sornette, D., 2019. Magnitude of earthquakes controls the size distribution of their triggered events, *J. geophys. Res.,* **124**(3), 2762–2780.

Nelder, J. A. & Mead, R., 1965. A simplex method for function minimization, *Comput. J.,* **7**(4), 308–313.

Ogata, Y., 1983. Estimation of the parameters in the modified Omori formula for aftershock frequencies by the maximum likelihood procedure, *J. Phys. Earth,* **31**(2), 115–124.

Ogata, Y., 1988. Statistical models for earthquake occurrences and residual analysis for point processes, *J. Am. Stat. Assoc.,* **83**(401), 9–27.

Ogata, Y., 1989. Statistical model for standard seismicity and detection of anomalies by residual analysis, *Tectonophysics,* **169**(1–3), 159–174.

Ogata, Y., 1998. Space-time point-process models for earthquake occurrences, *Ann. Inst. Stat. Math.,* **50**(2), 379–402.

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

Rhoades, D. A., 2007. Application of the EEPAS model to forecasting earthquakes of moderate magnitude in Southern California, *Seismol. Res. Lett.,* **78**(1), 110–115.

Rhoades, D. A., 2011. Application of a long-range forecasting model to earthquakes in the Japan mainland testing region, *Earth, Planets Space,* **63**(3), 197–206.

Rhoades, D. A. & Evison, F. F., 2004. Long-range earthquake forecasting with every earthquake a precursor according to scale, *Pure appl. Geophys.,* **161**(1), 47–72.

Rhoades, D. A., Gerstenberger, M. C., Christophersen, A., Zechar, J. D., Schorlemmer, D., Werner, M. J. & Jordan, T. H., 2014. Regional earthquake likelihood models II: information gains of multiplicative hybrids, *Bull. seism. Soc. Am.,* **104**(6), 3072–3083.

Rhoades, D. A., Rastin, S. J. & Christophersen, A., 2020. The effect of catalogue lead time on medium-term earthquake forecasting with application to New Zealand data, *Entrophy,* **22**(11), 1–15.

Rhoades, D. A., Schorlemmer, D., Gerstenberger, M. C., Christophersen, A., Zechar, J. D. & Imoto, M., 2011. Efficient testing of earthquake forecasting models, *Acta Geophys.,* **59**(4), 728–747.

Rhoades, D.A., 2021. Manual for EEPAS software (EEPSOF) Version 2.3 w, GNS Science Internal Report, 49pp.

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

Schorlemmer, D., Gerstenberger, M. C., Wiemer, S., Jackson, D. D. & Rhoades, D. A., 2007. Earthquake likelihood model testing, *Seismol. Res. Lett.,* **78**(1), 17–29.

Schorlemmer, D., Zechar, J. D., Werner, M. J., Field, E. H., Jackson, D. D. & Jordan, T. H., 2010. First results of the regional earthquake likelihood models experiment, *Pure appl. Geophys.,* **167**(8–9), 859–876.

Serafini, F., Naylor, M., Lindgren, F., Werner, M. J. & Main, I., 2022. Ranking earthquake forecasts using proper scoring rules: binary events in a low probability environment, *Geophys. J. Int.,* **230**(2), 1419–1440.

Utsu, T., 1961. A statistical study on the occurrence of aftershocks, *Geophys. Mag.,* **30,** 521–605.

Werner, M. J. & Sornette, D., 2008. Magnitude uncertainties impact seismic rate estimates, forecasts, and predictability experiments, *J. geophys. Res.,* **113**(8), 1–20.

Zechar, J. D., Gerstenberger, M. C. & Rhoades, D. A., 2010. Likelihood-based tests for evaluating space-rate-magnitude earthquake forecasts, *Bull. seism. Soc. Am.,* **100**(3), 1184–1195.

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

Zechar, J. D. & Zhuang, J., 2014. A parimutuel gambling perspective to compare probabilistic seismicity forecasts, *Geophys. J. Int.,* **199**(1), 60–68.

Zhuang, J., 2010. Gambling scores for earthquake predictions and forecasts, *Geophys. J. Int.,* **181**(1), 382–390.

Zhuang, J., Ogata, Y. & Vere-Jones, D., 2004. Analyzing earthquake clustering features by using stochastic reconstraction, *J. geophys. Res.,* **109**(B5), 369–380.

Zhuang, J., Werner, M.J., Hainzl, S., Harte, D. & Zhou, S., 2011. Basic models of seismicity: spatiotemporal models, Community Online Resource for Statistical Seismicity Analysis, doi:10.5078/corssa-07487583, Available at: http://www.corssa.org.

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**(6), 3208–3222.

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

SUPPORTING INFORMATION

Supplementary data are available at *GJI* online.

Table S1. Binary N test in the testing time interval (2012–2021) (main shocks + aftershocks).
Table S2. Binary cL test in the testing time interval (2012–2021) (main shocks + aftershocks).
Table S3. Binary S test in the testing time interval (2012–2021) (main shocks + aftershocks).
Table S4. Information Gain per active bin in the testing time interval (2012–2021) (main shocks + aftershocks).
Table S5. Binary N test in the testing time interval (2012–2021) (main shocks).
Table S6. Binary cL test in the testing time interval (2012–2021) (main shocks).
Table S7. Binary S test in the testing time interval (2012–2021) (main shocks).
Table S8. Information Gain per active bin in the testing time interval (2012–2021) (main shocks).

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.

APPENDIX A: FORMULATION OF PPE AND EEPAS FORECASTING MODELS

The EEPAS model is based on the increase in rate and in magnitude of the minor seismicity observed before the occurrence of major earthquakes (ψ-phenomenon, Rhoades & Evison 2004). Evison & Rhoades (2004) analysed the ψ-phenomenon that evolved before 47 major earthquakes in well catalogued regions (such as California, New Zealand, northern Mexico, Japan and Greece) developing three empirical scaling relations that relate the precursor magnitude (M_p) with the main shock magnitude (M_m), precursor time (T_p) and precursory area (A_p). Where M_p is assumed as the mean of the three largest precursor shocks, T_p is the time interval between the onset of the ψ-phenomenon and main shock and A_p (expressed in km) is the smallest rectangular spatial box containing the precursory events, main shock and aftershocks (for details see Evison & Rhoades 2004). The predictive scale relations are thus defined as

$$M_m = a_m + b_m M_p, \tag{A1}$$

$$\log\left(T_p\right) = a_t + b_t M_p, \tag{A2}$$

and

$$\log(A_p) = a_a + b_a M_p. \tag{A3}$$

In the EEPAS model each ith earthquake, occurring at time t_i with magnitude m_i and located at (x_i, y_i), is assumed to contribute to the transient increment of the rate density $\lambda_i(t, m, x, y)$ of future seismicity (defined as the derivative of the expected number of earthquakes with respect to time, magnitude and location coordinates) by the term

$$\lambda_i\ (t, m, x, y) = w_i\ f_{1i}\,(t)\, g_{1i}\,(m)\, h_{1i}\,(x, y)\,, \tag{A4}$$

where w_i is a weighting factor which depends on other earthquakes in its proximity (see below). $f_{1i}(t)$, $g_{1i}(m)$ and $h_{1i}(x, y)$ are the probability density functions of time, magnitude and location, respectively. The assumed forms for these distributions are defined consistently with the ψ scaling relations by Rhoades & Evison (2004) . The time distribution is assumed to be Lognormal with the form

$$f_{1i}\ (t) = \frac{H\,(t - t_i)}{(t - t_i)\ln(10)\,\sigma_T\sqrt{2\pi}}\ \exp\left[-\frac{1}{2}\left(\frac{\log\,(t - t_i) - a_T - b_T m_i}{\sigma_T}\right)^2\right], \tag{A5}$$

where $H(t - t_i)$ is the Heaviside step function, which takes the value 1 if $t - t_i > 0$, and 0 otherwise. This means that at the time t, the rate density function is contributed only by earthquakes occurring before t. Here a_T, b_T and σ_T are free parameters to be determined.

The magnitude distribution $g_{1i}(m)$ is assumed to be normal with the form:

$$g_{1i}\ (m) = \frac{1}{\sigma_M\sqrt{2\pi}}\ \exp\left[-\frac{1}{2}\left(\frac{m - a_M - b_M m_i}{\sigma_M}\right)^2\right] \tag{A6}$$

where a_M, b_M and σ_M are free parameters.

The space distribution is assumed to be bivariate Normal with circular symmetry with the form

$$h_{1i}\ (x, y) = \frac{1}{2\pi\sigma_A^2 10^{b_A m_i}}\ \exp\left[-\frac{(x - x_i)^2 + (y - y_i)^2}{2\sigma_A^2 10^{b_A m_i}}\right], \tag{A7}$$

where σ_A and b_A are free parameters.

An adjustment is necessary because of the missing contribution of earthquakes below the minimum completeness magnitude m_c

which causes the rate density at magnitude m to be underestimated on average by a fraction $\Delta(m)$ of its real value given by

$$\Delta\ (m) = \phi\left(\frac{m - a_M - b_M m_c - \sigma_M^2 \beta}{\sigma_M}\right), \tag{A8}$$

where ϕ is the Normal distribution integral. Then $\Delta(m)$ can also be written as

$$\Delta\ (m) = \frac{1}{2}\ \mathrm{erf}\left[\left(\frac{m - a_M - b_M m_c - \sigma_M^2 \beta}{\sigma_M \sqrt{2}}\right) + 1\right], \tag{A9}$$

where erf is the Error function.

Hence, to compensate for the lack of earthquakes with magnitude lower than the completeness magnitude m_c, $\lambda_i(t, m, x, y)$ is inflated by a factor $\frac{1}{\Delta(m)}$.

The total rate density is obtained by summing the contribution of all past earthquakes and also adding a background term that allows for the possibility that an earthquake can occur without an appreciable scale increase of precursory shocks:

$$\lambda\ (t, m, x, y) = \mu\lambda_0\,(t, m, x, y) \quad + \sum_{t_i \geq t_0; m_i \geq m_c}^{t-delay} \eta\,(m_i)\,\lambda_i\,(t, m, x, y), \tag{A10}$$

where $\lambda_0(t, m, x, y)$ is the background rate density, t_0 is the time of the beginning of the catalogue, μ is the mixing parameter and can be interpreted as the proportion of earthquakes that occur without precursory shocks. The normalizing function η is defined by

$$\eta\ (m_i) = \frac{b_M\,(1-\mu)}{E\,(w)}\ \exp\left[-\beta\left(a_M + (b_M - 1)\,m_i + \frac{\sigma_M^2 \beta}{2}\right)\right], \tag{A11}$$

where $E(w)$ is the mean weight of earthquakes in the catalogue; a_M, b_M and σ_M are free parameters; and $\beta = b \ln 10$, with b being the slope of the frequency–magnitude distribution of Gutenberg & Richter (1944). The normalizing function $\eta(m_i)$ ensures that the number of earthquakes expected by the model approximatively matches the actual number of target earthquakes. The *delay* term in eq. (A10) is to prevent the fit of the parameters being influenced by the short-term clustering of earthquakes (such as aftershocks and swarms). The EEPAS model is focused on the long-term clustering detected by the precursory scale increase phenomenon and its associated scaling relations. For this reason, the delay (usually assumed to be 50 d) after the time of occurrence of each earthquake is applied and no earthquake from the input catalogue is considered before such time interval elapsed after its occurrence.

The background rate density $\lambda_0(t, m, x, y)$ depends on the proximity of the location (x, y) with respect to previous seismicity. It is described by a quasi-time-invariant smoothed seismicity model, described by Rhoades & Evison (2004), which is similar to the forecasting model proposed by Jackson & Kagan (1999) and is called proximity to past earthquakes (PPE). It takes the form

$$\lambda_{0i}\ (t, m, x, y) = f_{0i}\ (t)\,g_{0i}\,(m)\,h_{0i}\,(x, y)\,, \tag{A12}$$

where $f_{0i}(t)$ is the time density function, $g_{0i}(m)$ is the magnitude density function and $h_{0i}(x, y)$ is the spatial density function. The time density function takes the form

$$f_{0i}\ (t) = \frac{1}{t_i - t_0}. \tag{A13}$$

This ensures that at any time the estimated rate of earthquakes with $m \geq m_T$ within the region R is similar to the past rate.

The magnitude density function is that implied by the frequency magnitude law of Gutenberg & Richter (1944):

$$g_{0i}\ (m) = \beta \exp\left[-\beta\,(m_i - m_c)\right]. \tag{A14}$$

Finally, $h_{0i}(x, y)$ is the sum over all earthquakes with $m_i \geq m_T$ from time t_0 up to, but not including time t of smoothing kernels with the form

$$h_{0i}\ (x_i, y_i) = \sum_{t_i > t_0; m_i > m_T}^{t\text{-delay}} a\,(m_i - m_T)\,\frac{1}{\pi}\left(\frac{1}{d^2 + r_i^2}\right) + s, \tag{A15}$$

where r_i is the distance in km between (x, y) and the epicentre (x_i, y_i); a is a normalizing parameter, d is a smoothing distance and s is a small term that includes the contribution from earthquakes that occur far from past epicentres.

The rate density $\lambda_0(t, m, x, y)$ of the PPE model decreases gradually with time elapsed after an earthquake occurrence and increases when a new earthquake occurs. The function $h_{0i}(x, y)$ considers the earthquake location and the function $f_{0i}(t)$ the passage of time.

The purpose of the weighting factor w_i in eq. (A10) is to give more weight to earthquakes that are more likely to be part of a long-term clustering, thus giving less weight to events that are aftershocks of previous earthquakes. Two different weighting strategies were applied in the past application of EEPAS. The simplest one is giving the same weight $w_i = 1$ to each earthquake in the catalogue. With this strategy aftershocks triggered by previous earthquakes have the same weight of any other shock. The other strategy is to assign a lower weight to any earthquake which is likely to be an aftershock of a previous earthquake. Therefore, the total rate density is mostly given by earthquakes that are part of long-term clustering.

This latter strategy requires estimating the rate density λ' for aftershock occurrence, incorporating epidemic-type aftershock behaviour (Ogata 1988, 1998; Console & Murru 2001). The aftershock model adopted for EEPAS takes the form:

$$\lambda'\ (t, m, x, y) = \nu\lambda_0\,(t, m, x, u) + k \sum_{t_i \geq t_0} \lambda_i^{'}\,(t, m, x, y)\,, \tag{A16}$$

where λ_0 is the rate density given by PPE model, ν is the proportion of earthquake that are not aftershocks, k is a normalization constant and $\lambda'(t, m, x, y)$ describes the aftershocks occurrence with the form:

$$\lambda_i^{'}\ (t, m, x, y) = f_{2i}\ (t)\,g_{2i}\,(m)\,h_{2i}\,(x, y)\,, \tag{A17}$$

where $f_{2i}(t)$, $g_{2i}(m)$ and $h_{2i}(x, y)$ are, respectively, the density functions for time, magnitude, and locations of the aftershocks of the ith earthquake. The time distribution is given by the modified Omori law (Utsu 1961; Ogata 1983):

$$f_{2i}\ (t) = H\,(t - t_i)\,\frac{p-1}{(t - t_i + c)^p}, \tag{A18}$$

where t_i is the time of the ith earthquake, and c and p are the Omori law parameters.

The magnitude distribution follows the Gutenberg & Richter (1944) law, and it is assumed that the magnitude of an aftershock is smaller than its main shock by at least δ units

$$g_{2i}\ (m) = H\,(m_i - \delta - m)\,\beta \exp\left[-\beta\,(m - m_i)\right]. \tag{A19}$$

The addition of the parameter δ is based on the so-called Bath's law (Båth 1965), according to which the largest aftershock typically has a magnitude about 1.2 units smaller than the main shock. Finally, the spatial distribution is assumed to be bivariate Normal with circular symmetry:

$$h_{2i}\ (x, y) = \frac{1}{2\pi\sigma_U^2 10^{m_i}}\ \exp\left[-\frac{(x - x_i)^2 + (y - y_i)^2}{2\sigma_U^2 10^{m_i}}\right], \tag{A20}$$

where σ_U is a free parameter. The weighting factor is then computed as

$$w_i = \frac{\nu\lambda_0(t_i, m_i, x_i, y_i)}{\lambda'(t_i, m_i, x_i, y_i)}, \tag{A21}$$

In this way, if an earthquake has the characteristics of an aftershock, it will have a weight close to 0; on the contrary, if an earthquake that in no way resembles an aftershock it will have a weight close to 1.

APPENDIX B: IMPLEMENTATION OF MATLAB AND PYTHON CODES AND COMPARISON WITH EEPSOF VERSION 2.3 W

We developed a suite of codes in Matlab language reproducing the formulation described in the Appendix A. We tested them against the EEPSOF code (Version 2.3 w) developed by D. A. Rhoades (2021) and provided as a binary Linux executable file compiled by Fortran77. To make the comparison, we adopted a simplified spatial geometry as EEPSOF hardly manages the complex shape made by a tessellation of the Italian area as described in the main text. The purpose of the comparison is to ensure that the optimized parameter values and the relative maximum log-likelihoods are satisfactorily similar.

One difference between the Matlab implementation and EEPSOF is the treatment of spatial data. While the EEPSOF code itself computes the kilometric distances directly from geographical coordinates, for the Matlab implementation we have chosen to first convert all coordinates from the WGS84 geographic reference to kilometric coordinates in the RDN2008 Italy Zone (E-N) EPSG: 7794 by the QGIS software.

We applied both codes to the data set of target earthquakes with magnitude $M \geq 5.0$ that occurred from 1990 to 2020 within the analysis polygon. The latter is a rectangle with sides of 576 km eastward and of 745 km northward (Fig. A1). The vertices of the polygon for EEPSOF were converted from kilometric coordinates in the RDN2008 Italy Zone (E–N) system to the WGS84 coordinate reference system. For fitting the EEPAS model, we used the earthquakes from the HORUS catalogue with $M \geq 2.5$ and $Z \leq 50$ km occurring inside the polygon from 1960 to 2020. To avoid edge effects in the fitting of model parameters, the contribution of earthquakes in the neighbourhood up to 200 km from the polygon were also considered (Fig. A1). The used data set contains 38 086 events, of which 24 816 are within the analysis polygon. For both software codes, the log-likelihood optimization is carried out using the downhill simplex method (Nelder & Mead 1965) as described in Rhoades & Evison (2004).

For the comparison, the fit of the EEPAS parameters is made in five iterations, one for the parameters a_T, a_M and ϑ_A and the others adding one at a time the parameters (ϑ_T, ϑ_M, b_A, b_T) to note the onset of possible deviations. In the first iteration the parameters ϑ_T, ϑ_M, b_A, b_T were set to 0.23, 0.32, 0.35 and 0.40, respectively, based on analyses conducted on scaling relationships obtained from the analysis of individual earthquakes (Rhoades & Evison 2004). With these parameters set, a_T, a_M and σ_A were fit by the maximum likelihood estimation using as starting values 1.5, 1.4 and 3.3, respectively. The fit procedure continued by adding one parameter at a time and considering the previously obtained values as initial values. The parameters, log-likelihoods values and the expected numbers of earthquakes are reported in Table A1.

The optimized parameter values for the first iteration are unequivocally similar for the two codes, because the differences are less than 0.9 per cent. In the second and third step, the estimates of parameters a_M and ϑ_T begin to slightly deviate, with maximum percentage differences up to about 3.0 per cent. With the introduction in the fit of the spatial parameters, the differences of the other parameters also increase. In the fifth and final iteration, the differences are more pronounced, particularly for the parameters a_T, σ_A and b_T, where the codes differ by 12.5, 10.2 and 24.0 per cent, respectively. However, the differences in log-likelihoods and expected numbers of earthquakes remain small. Such differences mainly concern spatial parameters and may be related to the different way in which distances are handled by the two software codes and to the different method used to integrate over space.

APPENDIX C: IMPLEMENTATION OF THE ETAS, SUP AND SVP MODELS

In the literature we can find several implementations of the Epidemic Type Aftershock Sequence (ETAS) model to earthquake forecasting in Italy (e.g. Console *et al.* 2006; Lombardi & Marzocchi 2010). In all of them the time dependence is formulated as a sum of Omori decays starting at the times of occurrence of each earthquake

$$f(t) = \sum_{i=1}^{n} \frac{H(t - t_i)\,K}{(t - t_i + c)^p}, \tag{C1}$$

where K, p and c are free parameters and $H(t - t_i)$ is the Heaviside step function which is 1 if $t - t_i > 0$ and is 0 otherwise.

The productivity of each earthquake of magnitude M_i is described by

$$r = e^{\alpha(M_i - M_c)}, \tag{C2}$$

where α is a free parameter and M_c is the minimum magnitude of completeness.

The decay of the productivity with the distance from the epicentre (x_i, y_i) of each earthquake can be described by many probability density functions (e.g. Ogata 1998; Console *et al.* 2003; Zhuang *et al.* 2004; Lombardi & Marzocchi 2010; Zhuang *et al.* 2011). In this work, we considered the spatial PDF as described in Ogata & Zhuang (2006) where the smoothing term is an exponential function of the magnitude:

$$g(x, y) = \frac{(q-1)\left[D^2 e^{\gamma(m_i - m_c)}\right]^{q-1}}{\pi\left[(x - x_i)^2 + (y - y_i)^2 + D^2 e^{\gamma(m_i - m_c)}\right]^q}, \tag{C3}$$

where q, D and γ are free parameters.

Finally, the frequency magnitude distribution of shocks is given by the Gutenberg & Richter (1944) law

$$h(m) = \beta e^{-\beta(m - M_c)}, \tag{C4}$$

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

Combining all the previous terms together, and adding a time invariant background seismicity term $\lambda_0(x, y, m)$, the rate density of ETAS models is given by

$$\lambda(t, x, y, m) = \nu\lambda_0(x, y, m) + [f(t)\, r\, g(x, y)\, h(m)]. \tag{C5}$$

That is

$$\lambda(t, x, y, m) = \nu\lambda_0(x, y, m) + \left\{ \sum_{i=1}^{n} \frac{H(t - t_i)\,K}{(t - t_i + c)^p} e^{\alpha(M_i - M_c)} \frac{(q-1)\left[D^2 e^{\gamma(m_i - m_c)}\right]^{q-1}}{\pi\left[(x - x_i)^2 + (y - y_i)^2 + D^2 e^{\gamma(m_i - m_c)}\right]^q} \beta e^{-\beta(m - M_c)} \right\}. \tag{C6}$$

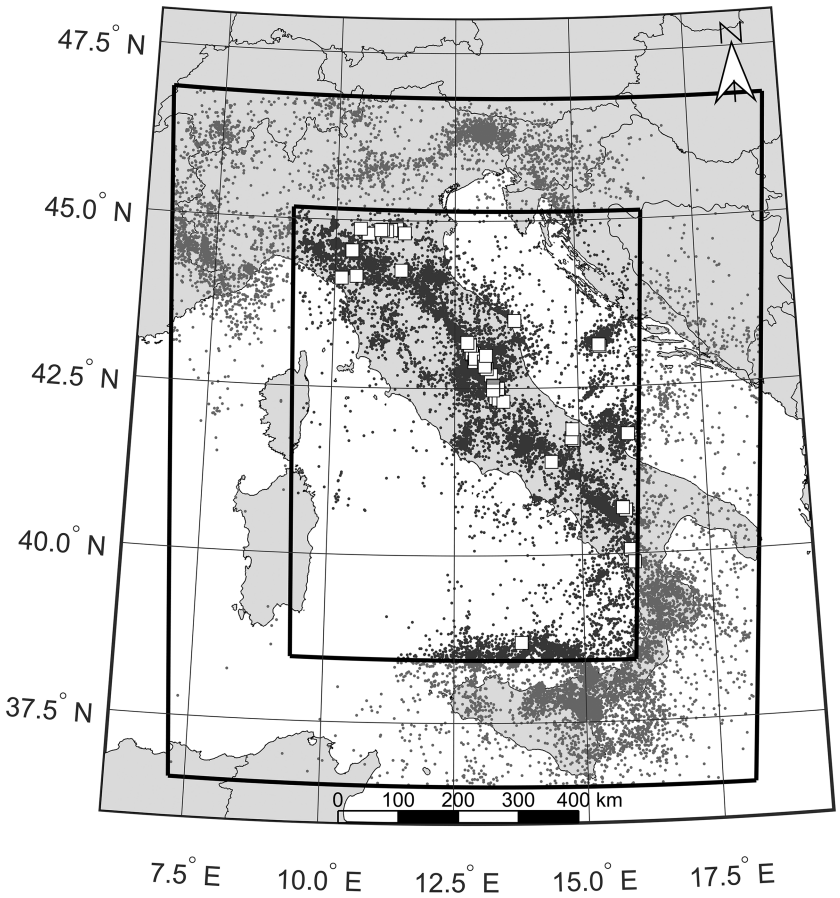

Figure A1. Map of epicentres of earthquakes with $M \geq 2.5$ and $Z \leq 50$ km that occurred from 1990 to 2020 within the region adopted for the software codes comparison. The interior rectangular area represents the analysis polygon for which the EEPAS model is applied. The black point represents the epicentres of earthquake occurred within the analysis polygon. The external rectangle represents the influence area for which earthquake indicated by the grey points are also considered for the parameters estimation to avoid edge effects. The white squares represent target earthquakes with $M \geq 5.0$ that occurred within the analysis polygon in the period 1990–2020.

The parameter ν 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.

The time invariant models of seismicity consist of stationary Poisson processes, in which the average earthquake occurrence rate may be spatially uniform (Spatially Uniform Poisson, SUP) or variable (Spatially Variable Poisson, SVP). SUP and SVP can also be seen as independent models of seismicity occurrence to compare with other forecasting models (Console *et al.* 2006).

Their rate density is given by:

$$\lambda_0\ (x, y, m) = \mu_0\ (x, y)\, \beta \exp\left[-\beta\,(m - M_c)\right] \tag{C7}$$

where $\mu_0(x, y)$ is the spatial rate density of earthquakes with magnitudes equal or larger than M_c. In the SUP model the space-density is assumed to be uniform and independent of the location (x, y). μ_0 is obtained by dividing the number of earthquakes with magnitude above M_c over the whole analysis region R by the total surface area considered.

In the SVP model, the spatial density $\mu_0(x, y)$ is considered as a continuous smooth function of the geographical location (x, y). To estimate it as a spatially varying function, it is necessary to divide the polygon into squared cells of suitable size. The number of earthquakes N_k with magnitude equal to or larger than M_c in each cell is estimated. Each N_k value, representative of a single cell is then smoothed by a Gaussian filter with correlation distance d_c and normalized so as to preserve the total number of events as described in Frankel (1995). For each cell, the smoothed N_k is given by

$$\tilde{N}_k = \frac{\Sigma_l N_k \exp\left(-\Delta_{kl}^2/d_c^2\right)}{\Sigma_l \exp\left(-\Delta_{kl}^2/d_c^2\right)}, \tag{C8}$$

where Δ_{kl} is the distance between the centre of the kth 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 value of $\mu_0(x, y)$ in each point of the space is computed by the weighted mean of the four nearest cells that surround the point. To determine d_c we follow the procedure suggested by Console & Murru (2001): the catalogue is divided into two subcatalogues of about the same temporal length and d_c is chosen as the value that maximizes the log-likelihood of a subcatalogue using the smoothed seismicity obtained from the other subcatalogue (Fig. A2). The analysis for the optimal d_c is conducted for both subcatalogues and the obtained values for d_c are, respectively, $\mathrm{d}_{c1} = 16.0$ and $\mathrm{d}_{c2} = 13.0$. The optimal correlation distance $\mathrm{d}_c = 14.5$ is given by the mean of these two estimates. Once the value d_c is optimized, the spatial density of earthquakes $\mu_0(x, y)$ of the SVP background model can be assessed for each cell and for each point in space (Fig. 1). The *b*-value of the Gutenberg & Richter (1944) distribution is taken to be the same as that computed for the EEPAS model, as described in the main text.

Table A1. Estimated parameters, expected number of target earthquake and log-likelihood values for each iteration step.

EEPSOF code									
1st step		2nd step (σ_T)		3rd step (σ_M)		4th step (b_A)		5th step (b_T)	
a_T	2.2264	a_T	2.2109	a_T	2.2204	a_T	2.2187	a_T	2.2173
a_M	1.2295	a_M	1.2289	a_M	1.0165	a_M	1.0339	a_M	1.0141
σ_A	2.3893	σ_A	2.3739	σ_A	2.3360	σ_A	1.2387	σ_A	1.1125
		σ_T	0.2757	σ_T	0.2696	σ_T	0.2686	σ_T	0.2683
				σ_M	0.4845	σ_M	0.4820	σ_M	0.4918
						b_a	0.4899	b_a	0.5181
								b_T	0.4021
L	−961.64	L	−961.491	L	−960.876	L	−960.632	L	−960.594
$\bar{E}$	40.8569	$\bar{E}$	40.0899	$\bar{E}$	39.7087	$\bar{E}$	39.8445	$\bar{E}$	39.5826
MATLAB code									
1st step		2nd step (σ_T)		3rd step (σ_M)		4th step (b_A)		5th step (b_T)	
a_T	2.2190	a_T	2.2074	a_T	2.2118	a_T	2.2075	a_T	2.4950
a_M	1.2188	a_M	1.2208	a_M	1.0477	a_M	1.0049	a_M	1.0025
σ_A	2.3929	σ_A	2.3769	σ_A	2.3456	σ_A	1.0015	σ_A	1.0097
		σ_T	0.2670	σ_T	0.2603	σ_T	0.2607	σ_T	0.2652
				σ_M	0.4493	σ_M	0.4771	σ_M	0.4642
						b_a	0.5413	b_a	0.5399
								b_T	0.3236
L	−961.791	L	−961.719	L	−961.264	L	−960.945	L	−960.755
$\bar{E}$	41.222	$\bar{E}$	40.6459	$\bar{E}$	40.5468	$\bar{E}$	40.6171	$\bar{E}$	40.9646

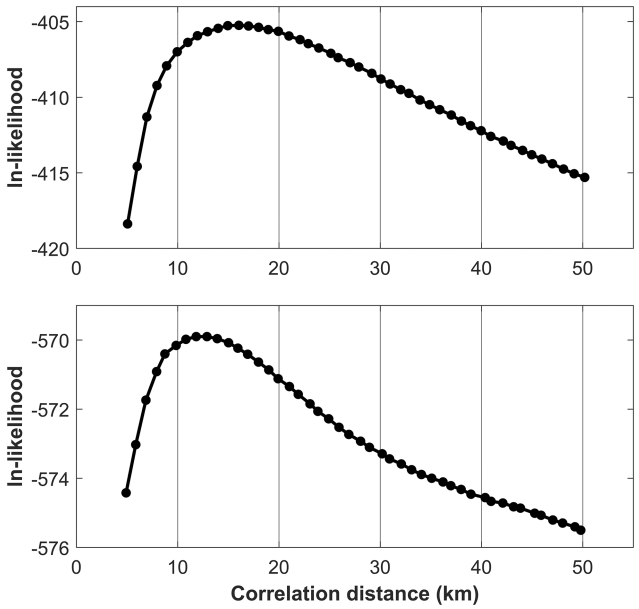

Figure A2. Upper frame: log-likelihood of the subcatalogue of earthquakes that occurred in the period 1990 to April 2009 under the time-independent SVP model obtained by the seismicity from April 2009 to 2021. Lower frame: log-likelihood of the subcatalogue of earthquakes that occurred in the period April 2009 to 2021 under the time-independent SVP model obtained by the seismicity from 1990 to April 2009.

The parameter q is set to 1.5, according to physical investigation showing that the static stress changes decrease with epicentral distance as r^{-3} (Lombardi & Marzocchi 2010). The other parameters (k, p, c, α, d, ν) are fitted by the maximization of the likelihood function (eq. 1) of main text) using the interior point method.