



OPEN

Improvements to seismicity forecasting based on a Bayesian spatio-temporal ETAS model

Hossein Ebrahimian¹✉, Fatemeh Jalayer^{1,2}, Behnam Maleki Asayesh^{3,4,5}, Sebastian Hainzl⁴ & Hamid Zafarani⁵

The epidemic-type aftershock sequence (ETAS) model provides an effective tool for predicting the spatio-temporal evolution of aftershock clustering in short-term. Based on this model, a fully probabilistic procedure was previously proposed by the first two authors for providing spatio-temporal predictions of aftershock occurrence in a prescribed forecasting time interval. This procedure exploited the versatility of the Bayesian inference to adaptively update the forecasts based on the incoming information provided by the ongoing seismic sequence. In this work, this Bayesian procedure is improved: (1) the likelihood function for the sequence has been modified to properly consider the piecewise stationary integration of the seismicity rate; (2) the spatial integral of seismicity rate over the whole aftershock zone is calculated analytically; (3) background seismicity is explicitly considered within the forecasting procedure; (4) an adaptive Markov Chain Monte Carlo simulation procedure is adopted; (5) leveraging the stochastic sequences generated by the procedure in the forecasting interval, the N-test and the S-test are adopted to verify the forecasts. This framework is demonstrated and verified through retrospective early forecasting of seismicity associated with the 2017–2019 Kermanshah seismic sequence activities in western Iran in two distinct phases following the main events with Mw7.3 and Mw6.3, respectively.

Within the first days elapsed after the occurrence of an earthquake and in the presence of an ongoing seismic sequence, emergency decision-making can benefit enormously from the short-term operational seismicity forecasts^{1–10}. The Epidemic Type Aftershock Sequence (ETAS) model is a widely used stochastic model to describe earthquake temporal¹¹ and spatio-temporal¹² occurrence and clustering of seismicity within a seismic sequence (see also^{13–15}). It is an epidemic-type stochastic point process¹⁶ in which every earthquake within the sequence is a potential triggering event for subsequent earthquakes by generating its own aftershocks based on a Modified Omori (MO¹⁷) decay. According to the study by¹⁸, the ETAS model is the best model for describing short-term seismicity. The ETAS model performed quite well in retrospectively forecasting the seismicity within various operational frameworks in California, Greece, Italy, Japan, Spain, New Zealand, and Iceland^{6,19–34}. It has been employed for time-dependent seismic hazard^{8,25,33,35,36}, and risk and loss forecasting^{8,35,37–40}.

Ogata¹¹ employed the maximum likelihood (ML) criterion parameter estimation (which serves currently as the most well-established and widespread method) for the temporal ETAS model. The method was extended also for the spatio-temporal ETAS model^{12–14}. Several attempts were made for developing improved algorithms to attain ML estimates of ETAS parameters^{41,42}. Algorithms based on numerical optimization methods for ML estimation may not be very efficient for seismicity forecasting within an ongoing seismic sequence for the following reasons: (a) they are computationally demanding and may encounter convergence problems (especially when the log-likelihood function is extremely flat or multimodal); (b) some algorithm parameters need to be tuned; something that sounds difficult to do automatically. Thus, the model parameters are usually calibrated a priori. New algorithms based on Simulated Annealing optimization technique that allows for a more automatic ML estimation of model parameters were developed^{43,44}; however, even this procedure is not totally autonomous and proper choice of tuning parameters is warranted. In case of providing early forecasts immediately after a large earthquake (e.g., using incomplete catalogues), ML estimation of the ETAS model parameters may include large estimation errors⁴⁵. This may cause bias in ETAS parameters that arises from missing data^{46,47}.

¹Department of Structures for Engineering and Architecture, University of Naples Federico II, Naples, Italy. ²Institute for Risk and Disaster Reduction (IRDR), University College London, London, UK. ³Institute of Geosciences, University of Potsdam, Potsdam, Germany. ⁴GFZ German Research Center for Geosciences, Potsdam, Germany. ⁵International Institute of Earthquake Engineering and Seismology (IIEES), Tehran, Iran. ✉email: ebrahimian.hossein@unina.it

Bayesian parameter estimation coupled with efficient simulation procedures, such as Markov Chain Monte Carlo Simulation, has the advantage of adaptively tuning into the ongoing sequence. This procedure can provide both joint probability distributions for the ETAS model parameters and the predicted spatio-temporal seismicity evolution. This is particularly important for early forecasts based on incomplete datasets. Bayesian parameter estimation has been employed in different aftershock models (MO model and spatio-temporal MO model both for L'Aquila 2009 seismic sequence in Italy^{24,37}; temporal ETAS model for L'Aquila 2009 seismic sequence²⁶; for 38 aftershock sequences in Japan from 1990–2014⁴⁵; for 2016 Kumamoto earthquake sequence in Japan⁴⁸; for 2019 Ridgecrest California seismic sequence⁴⁹; spatio-temporal ETAS model for 2010–2019 southern California dataset⁵⁰). It is to note that ETAS-based seismicity rate forecasts can be improved by incorporating additional main shock information including the rupture geometry and the coulomb stress changes⁵¹ or other stress scalars such as maximum shear and von-mises stress⁵², which might be available within minutes to hours after the events.

Ebrahimian and Jalayer²⁸ proposed a fully simulation-based framework for both Bayesian updating of spatio-temporal ETAS model parameters as well as a *robust* estimate of spatial distribution of events in a prescribed forecasting time interval after the main event. The procedure was applied for retrospective forecasting of early seismicity associated with the 2016 Amatrice seismic sequence in central Italy. The forecasting is robust (see^{53,54} for robust reliability assessment), since both the uncertainty in the ETAS model parameters and the uncertainty in the sequence of events that is going to occur during the forecasting interval is considered. This simulation-based framework consists of two parts: (a) Bayesian updating to learn the ETAS model parameters conditioned on the events that have already taken place (registered) in the ongoing seismic sequence. Markov Chain Monte Carlo (MCMC) simulation scheme⁵⁴ is used to sample directly from the conditional posterior probability distribution for ETAS model parameters (see also^{26,45,48} for the use of MCMC in ETAS parameter estimation). (b) Adaptively generate plausible sequences of events during the forecasting interval. For a given forecasting interval, this leads to spatial distribution of the forecasted events and probability distribution of the number of events. The outcomes of this robust seismicity forecasting are directly applicable in adaptive daily aftershock hazard (e.g.,^{24,33,55,56}) and risk assessment procedures (see e.g.,^{35,37,39}).

This work strives to improve different aspects of the fully probabilistic seismicity forecasting framework proposed in²⁸. These improvements are: (1) the likelihood function for Bayesian updating has been modified to adopt the piece-wise continuous formulation for a marked point process⁵⁷, whereas the previous work assumed magnitude and interarrival time to be jointly Poissonian. (2) The spatial integral of the conditional seismicity rate over the aftershock zone is calculated analytically (previously it was assumed to be equal to unity). (3) An adaptive MCMC simulation procedure for Bayesian updating of the ETAS model parameters have been employed by using the concept of multi-dimensional kernel sampling density function (which decreases the computational time; previously we have considered a simple component-wise MCMC procedure). (4) The background seismicity has been incorporated within the robust seismicity forecasting framework (previously, we have considered the background seismicity as an added rate to the forecasted rate of seismicity). (5) The stochastic process for generating sequences, i.e. part (b) described above, within the forecasting interval has been adjusted to the updated likelihood function and to the exact integral over the areal extent of the aftershock zone. (6) Leveraging the stochastic sequences generated by the procedure in the forecasting interval, the N-test and the S-test (see^{58,59}) are adopted to verify the forecasts. It is noted that this framework can automatically “tune-in” into the sequence of observed events and model updating, and forecasting are carried out without human interference and use of expert judgement. It is very efficient for early forecasts and in the presence of incomplete data. The refined framework is applied retrospectively to the Kermanshah seismic sequence 2017–2019 (see subsequent “[The Kermanshah 2017–2019 seismic sequence](#)” section) to provide early seismicity forecasting for two distinct phases of the sequence. The paper is organized as following: “[The Kermanshah 2017–2019 seismic sequence](#)” section provides an overview of the 2017–2019 Kermanshah seismic sequence. “[Results](#)” section shows the results of the seismicity forecasting for two important seismic sub-sequences including the Azgeleh mainshock (the main event with M_w 7.3 and its early triggered aftershocks, Phase 1), and the Sarpol-e Zahab event (the second main event with M_w 6.3 and its early triggered aftershocks, Phase 3). We have also investigated the sensitivity of the forecasts to some possible variations in the proposed workflow, which include: comparing two different spatial kernel density functions in the ETAS model; not considering the background seismicity; approximating the spatial integral of the conditional seismicity rate over the aftershock zone; and finally estimating the productivity coefficient of the ETAS model by two different approaches. “[Discussion and conclusions](#)” section is the general discussion and conclusions, and finally “[Methods](#)” section describes the proposed improved methodology for seismicity forecasting.

The Kermanshah 2017–2019 seismic sequence. Iranian plateau is a seismically active region whose active tectonic is dominated by the convergence of Arabian and Eurasian plates (~20 to 30 mm/year, see the inset map in Fig. 1). Approximately one third of this overall convergence rate is accommodated by the “Zagros” range as a central part of Alpine-Himalayan orogenic belt in southwestern Iran (see the inset map in Fig. 1)⁶⁰. The Zagros orogen is one of the youngest, and most tectonically active intracontinental belts in the world at the leading edge of the Arabian-Eurasian continental collision zone^{61,62}. On 12 November 2017 (18:18 UTC), an earthquake with M_w 7.3 struck the Zagros fold-and-thrust belt (ZFTB) in the Western Iran near the Iraq border in the area where large earthquakes had not been documented for several centuries. This severe earthquake occurred near the two small cities of Azgeleh and Sarpol-e Zahab (Az and SZ, see Fig. 1). This event which is called herein “Azgeleh” main shock (Azgeleh MS) inflicted around 630 casualties and caused immense buildings’ damages and economic losses⁶³. Based on the comprehensive study performed in⁶⁴, this event is an oblique-slip thrust with a considerable dextral component (Strike 354°, Dip 16°, and Rake 137°). They proposed a listric fault geometry for characterizing the Azgeleh MS (see the corresponding fault “AzF” Listric in Fig. 1), with a shallow

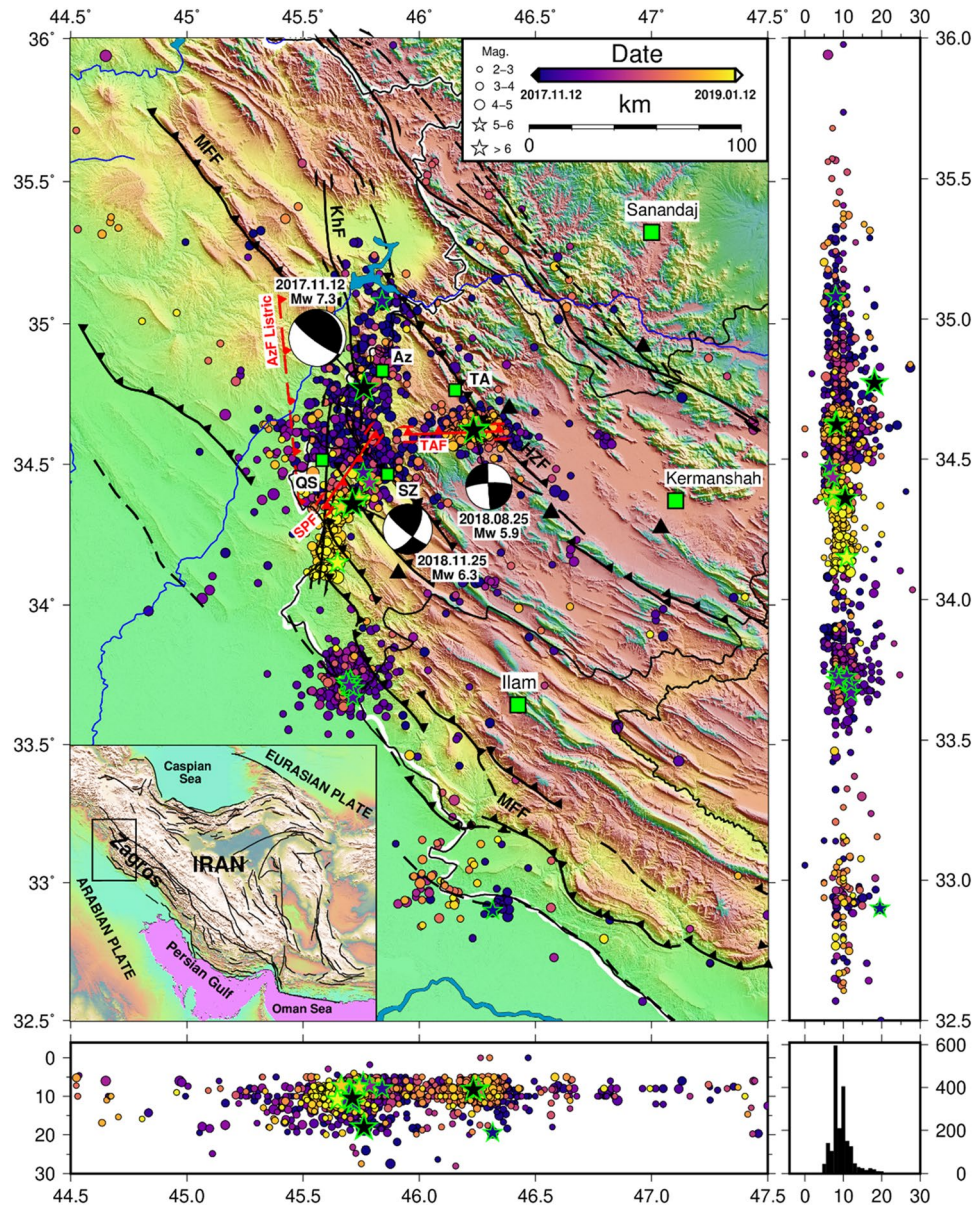


Figure 1. Main tectonic features of Iranian plateau; epicentral and depth distribution of the main events of Kermanshah seismic sequence 2017–2019 including Azgeleh, Tazehabad, and Sarpol-e Zahab earthquakes with black stars, and the distribution of seismic events (circles and green stars) with $M_w \geq 2.5$ registered by IRSC from 12/11/2017 up to 12/01/2019. The green stars show the events with $M_w \geq 5$. The color of stars and circles are based on the evolution of events. The black bitch balls show the focal mechanism of main events (based on the observations in⁶⁴). The big green squares show the main cities and small squares illustrate the small cities and towns nearby the epicentral area of the mains events (Az: Azgeleh, TA: Tazehabad, SZ: Sarpol-e Zahab, and QS: Qasr-e Shirin). The red solid and dashed lines show the surface projection of causative faults of the main events (AzF Listic: Azgeleh Listic slip model, TAF: Tazehabad fault, and SPF: Sarpol –e Zahab fault). The black solid lines in the map and the inset map show the main active faults in the area and the Iranian plateau (MFF: Main Front Fault, HZF: High Zagros Fault, and KhF: Khanaqin Fault); black triangles are the IRSC stations. The histogram in the right lower part illustrates the depth distribution of the seismic events.

angle plane (eastern plane) dipping $\sim 3^\circ$, a ramp dipping $\sim 16^\circ$ (central plane), and a steeper ramp dipping 25° (western plane). This MS triggered intense seismicity within the aftershock zone shown in Fig. 1 (the position of this area within the map of Iran is shown in the inset). From 12/11/2017 up to 18/04/2020 (i.e., in the time interval of around 2.5 years after the Azgeleh MS), about 9000 seismic events were recorded based on the catalog of Iranian Seismological Center (IRSC) issued by the Institute of Geophysics of the University of Tehran (IGUT seismic network) in the area shown in Fig. 1. From this pool of seismicity, 2318 events have $M_w \geq 2.5$. In addition to the Azgeleh MS, 19 events with $M_w \geq 5.0$ (one event, which will be discussed later, has $M_w \geq 6.0$), and more

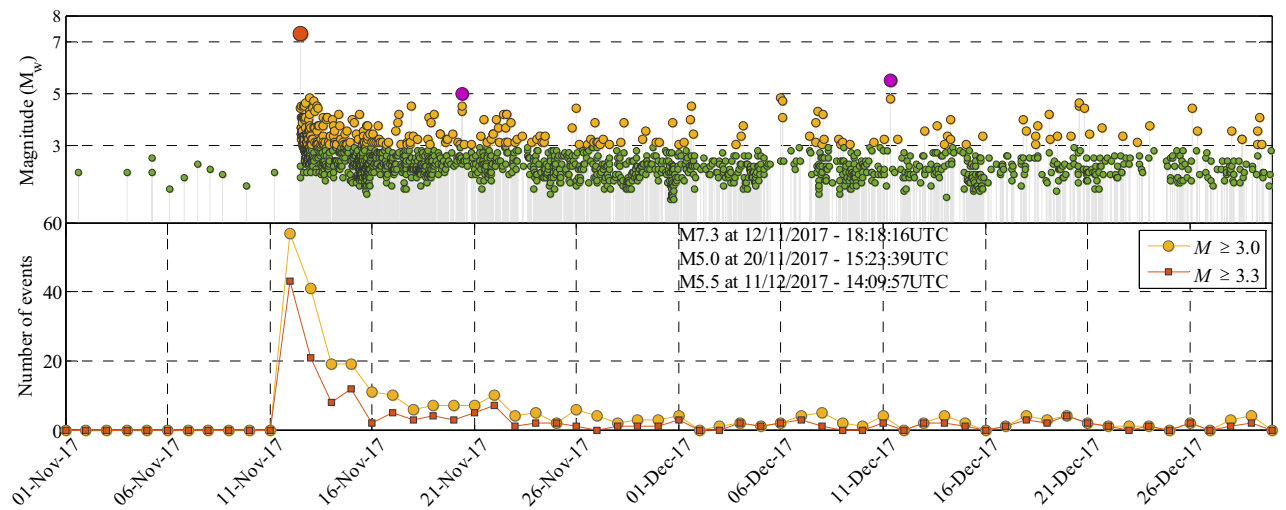


Figure 2. Magnitude vs. time (top) and the number of events with $M \geq 3.0$ and $M \geq 3.3$ (bottom) occurred within a 24-h interval starting from 6:00 UTC of each day from 01/11/2017 (T_0) until 30/12/2017 including the Azgeleh MS of $M_w 7.3$ within phase 1 of the complex Kermanshah 2017–2019 seismic sequence.

than 125 events with magnitude larger than 4 and less than 5 ($4 \leq M_w < 5$) were recorded within this sequence. Since this sequence has initiated in the Kermanshah province of Iran, it is called herein “*Kermanshah seismic sequence*”. Figure 1 illustrates Kermanshah seismic sequence 2017–2019 in the time span of 12/11/2017 up to 12/01/2019, where events with $M_w \geq 2.5$ are colored according to their occurrence time. The data is extracted from IRSC catalog issued by IGUT. The $M_w 7.3$ Azgeleh MS initiated “Phase 1” of Kermanshah seismic sequence. About 9 months after Azgeleh MS, on 25 August 2018 (22:13 UTC), an earthquake with $M_w 5.9$ occurred 45 km to the east of the epicenter of Azgeleh MS, close to the town Tazehabad (TA, see Fig. 1) with three casualties. At first glance, this new event was considered as an aftershock of the Azgeleh MS or a triggered event on the known faults of the study area (i.e., HZF or MFF faults shown in Fig. 1); however, reported focal mechanism from national and international seismological agencies and interferometric observation revealed that this event was associated to a previously unknown fault. The non-linear and linear inversion of the down-sampled data by⁶⁴ showed a triggered east–west trending fault (Strike 267°, Dip 78°, and Rake 2°) with dip to the north for this event. The event is called “Tazehabad” earthquake, and the corresponding left-lateral strike slip fault is called “TAF” in Fig. 1). The $M_w 5.9$ Tazehabad earthquake initiated “Phase 2” of this seismic sequence. About 3 months after the Tazehabad event and one year after the Azgeleh MS, on 25 November 2018 (16:37 UTC) another earthquake with $M_w 6.3$ struck near the Sarpol-e Zahab and Qasr-e Shirin cities (SZ and QS in Fig. 1). This event is called “Sarpol-e Zahab” earthquake, which started “Phase 3” of the Kermanshah seismic sequence. Fathian et al.⁶⁴ showed that the best-fitting inversion on the coseismic surface deformations of the Sarpol-e Zahab event agrees well with a single fault plane with a NE trend dipping to the southeast (Strike 34°, Dip 63°, and Rake 170°, see right-lateral strike slip fault “SPF” in Fig. 1). As a result of the characterized main events, Kermanshah seismic sequence 2017–2019 is composed of three main sub-sequences. These sub-sequences can be seen within the evolution of seismic activities in Fig. 1, where the events are colored according to their occurrence time.

Results

Daily forecasts of seismicity for the first days after the event $M_w 7.3$ of 12/11/2017 (Phase 1).

In this section, we strive to perform robust forecasts for the spatio-temporal evolution of the events in specific time intervals within phase 1 of the complex seismic sequence started by the Azgeleh MS of $M_w 7.3$, which took place in November 12, 2017, at 18:18 (UTC). Figure 2 (top) shows the evolution of events within the catalog characterized by their magnitude within the aftershock (AS) zone in phase 1 starting from 01/11/2017 up to 30/12/2017. The AS zone (A, see “[The epidemic-type aftershock sequence \(ETAS\) model for space–time clustering of aftershocks](#)” section in “[Methods](#)”) is selected to be smaller than the whole zone shown in Fig. 1 in the latitudinal range of [32.50–35.50] and longitudinal range of [45–47]. Figure 2 (bottom) shows the evolution of the number of events within the AS zone in phase 1 with magnitude $M \geq 3.0$ (orange circles) and $M \geq 3.3$ (red squares) within a 24-h interval starting from 6:00 UTC each day (also reported at 6:00 UTC each day). The time origin T_0 (the reference time) is set to 6:00 UTC of 01/11/2017 (i.e., 11 days before the main event). The seismicity after the occurrence of the mainshock increases considerably up to 15 November (afterwards, less than 20 events with $M \geq 3.0$ within a day have occurred). The occurrence of two triggered events of $M_w 5.0$ at 20/11/2017 and $M_w 5.5$ at 11/12/2017 (with very close epicenters) did not significantly increase the seismicity of the site. We examine the proposed improved ETAS framework for the following time windows [T_{start} , T_{end}]: (a) [12/11/2017–21:00UTC, 13/11/2017–06:00UTC] (i.e., 9 h forecasting with T_{start} at 2 h and 42 min after the main event); (b) [13/11/2017–00:00UTC, 13/11/2017–06:00UTC] (i.e., 6 h forecasting with T_{start} at 5 h and 42 min after the main event); (c) [13/11/2017–06:00UTC, 14/11/2017–06:00UTC]; (d) [14/11/2017–06:00UTC, 15/11/2017–06:00UTC]; (e) [15/11/2017–06:00UTC, 16/11/2017–06:00UTC]. The latter three are one-day (24 h) forecasting

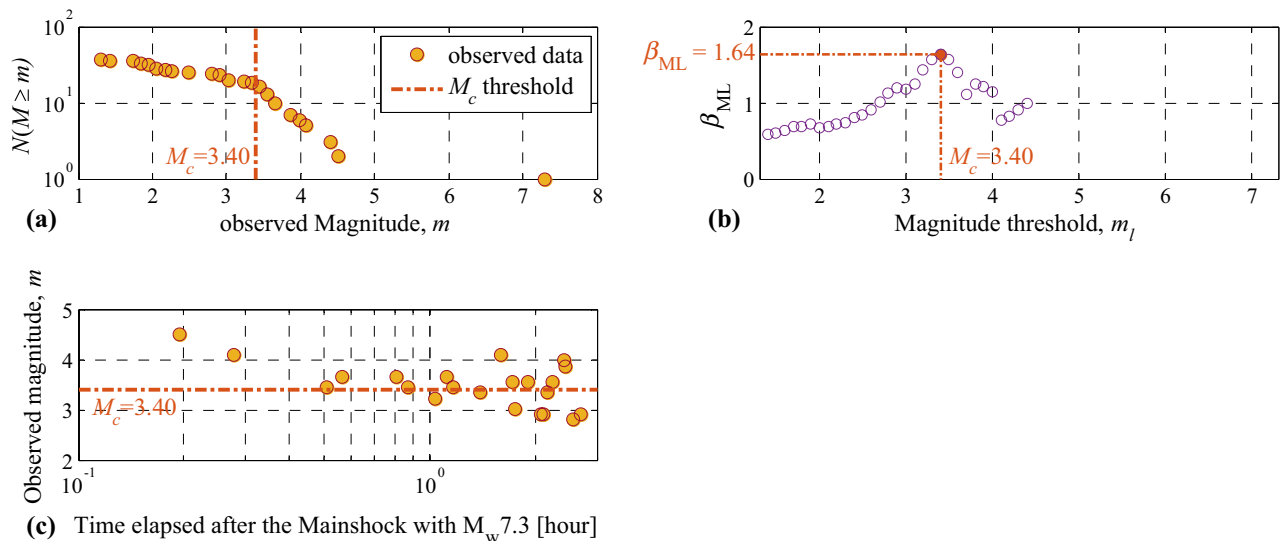


Figure 3. The three methods (Section SI-3-Results) for estimation/validation of M_c for the first forecasting interval [12/11/2017–21:00UTC, 13/11/2017–06:00UTC] in phase 1 of the seismic sequence.

intervals issued at 06:00 UTC each day. For each forecasting time window $[T_{start}, T_{end}]$, the observation history, **seq**, comprises all the events in the interval $[T_o, T_{start}]$ with magnitude greater than lower magnitude (M_l), set herein to be equal to the completeness magnitude M_c (i.e., $M_l = M_c$).

Estimating the magnitude of completeness M_c for each forecasting interval. M_c is estimated for each of the five forecasting intervals (see Supplementary Information Section SI-3-Results for description of three different strategies for estimation of M_c). Figure 3 illustrates the three M_c estimation procedures for the first forecasting interval (a) described above. Figure 3a shows a frequency-magnitude semi-logarithmic plot, which shows a value of $M_c = 3.40$ (the first method introduced in Section SI-3-Results). Figure 3b illustrates the mode (i.e. posterior maximum likelihood estimate) of the posterior probability distribution β_{ML} estimated through Bayesian inference as a function of magnitude thresholds m_l (the second method introduced in Section SI-3-Results). It shows that the maximum value $\beta_{ML} = 1.64$ leads to $M_c = 3.40$. Figure 3c (the third method discussed in SI-3-Results) is merely a visual check to see if the observation history **seq** is complete at the time of issuing the forecast by setting $M_c = 3.40$ (i.e., 2 h and 42 min elapsed after the main event). M_c estimations based on the three methods for time intervals (b) to (e) are illustrated in Section SI-3-Results. In summary, we set $M_c = 3.40$ for the first (early) forecasting interval (a), $M_c = 3.30$ for the two subsequent forecasting intervals (b, and c), and $M_c = 3.0$ for the ultimate two forecasting intervals (d, and e). Based on the background seismicity data provided in *Supplementary Information* (Section SI-4-Results, Figure SI-5b), the long-term background seismicity slope β for $M \geq 4.50$ is equal to 1.67.

Bayesian Inference for θ . The first step towards providing seismicity forecasts (Eq. 9) is sampling from the distribution of model parameters θ based on posterior (target) probability distribution $p(\theta|\mathbf{seq}, M_l)$ (“**Sampling θ from the distribution $p(\theta|\mathbf{seq}, M_l)$** ” section in “**Methods**”). The vector $\theta = [\beta, \alpha, c, p, d, q, \gamma]$, considering magnitude-dependent spatial kernel density (see “**The epidemic-type aftershock sequence (ETAS) model for space-time clustering of aftershocks**” section in “**Methods**”) is updated for each forecasting interval mentioned before based on the data provided (observed) within **seq** by applying the Bayesian updating routine illustrated in Eq. (18). It is considered that parameters K (Eq. 17), K_i (Eq. 3 as a function of c and p), K_r (Eq. 5 as a function of d , q and γ) are derived as functions of other parameters within vector θ . It is to mention that K is calculated as described in “**Calculating K** ” section using Eq. (17) and is not learned directly through the Bayesian updating. Samples for θ are generated using an adaptive MCMC procedure from $p(\theta|\mathbf{seq}, M_l)$, as noted in “**Sampling θ from the distribution $p(\theta|\mathbf{seq}, M_l)$** ” section (see also Section SI-1-Method) with 6 chains (simulation levels). p and q values equal to or smaller than one are rejected according to Eqs. (3) and (4) and the constraints are $p > 1$ (making sure that the temporal process is asymptotically described by a long-term Gutenberg Richter seismicity model) and $q > 1$ (making sure that the spatial kernel integral is asymptotically equal to unity).

The first simulation level, for which a component-wise updating approach is employed, contains 520 original seeds ($N_{seed} = 520$). To bypass the initial transient effect of the Markov chain, the first 20 samples are discarded and $N_{seed} = 500$ is used for the first simulation level. For the next simulation levels (i.e., chains 2 to 6) we have employed the adaptive kernel estimate (Section SI-1-Method) with $N_{seed} = 1000$, by performing MCMC updating in a block-wise manner. The $N_d \leq N_{seed}$ shows the number of distinct Markov chain samples generated within the last (the 6th) simulation level. The background seismicity is also considered within the likelihood estimation as addressed in “**Calculating the likelihood of the observed sequence $p(\mathbf{seq}|\theta, M_l)$** ” section (according to background seismicity data provided in Section SI-4-Results). Table 1 illustrates the histograms of marginal posterior PDFs for the model parameters $\theta = [\beta, \alpha, c, p, d, q, \gamma]$ (shown with bar plots) together with their prior

$[T_{start}, T_{end}]^*$ (dd/mm-hour)	β	α	c [day]	p	d [km]	q	γ	K
(a) [12/11-21:00, 13/11-06:00]; $N_d = 381$ $M_l = 3.40$	mean=1.76 CI=[1.09-2.63]	mean=1.05 CI=[0.45-1.96]	mean=0.05 CI=[0.02-0.09]	mean=1.41 CI=[1.02-2.03]	mean=1.82 CI=[0.54-3.84]	mean=1.48 CI=[1.01-2.54]	mean=0.31 CI=[0.10-0.53]	mean=4.60 CI=[0.03-22.24]
(b) [13/11-00:00, 13/11-06:00]; $N_d = 361$ $M_l = 3.30$	mean=1.64 CI=[1.06-2.28]	mean=1.12 CI=[0.55-1.79]	mean=0.05 CI=[0.01-0.10]	mean=1.35 CI=[1.01-1.96]	mean=1.88 CI=[0.66-4.22]	mean=1.88 CI=[1.11-3.33]	mean=0.36 CI=[0.16-0.53]	mean=3.13 CI=[0.06-9.78]
(c) [13/11-06:00, 14/11-06:00]; $N_d = 311$ $M_l = 3.30$	mean=1.77 CI=[1.22-2.24]	mean=1.03 CI=[0.62-1.40]	mean=0.05 CI=[0.02-0.11]	mean=1.27 CI=[1.01-1.78]	mean=1.55 CI=[0.61-3.25]	mean=1.73 CI=[1.22-2.70]	mean=0.38 CI=[0.20-0.57]	mean=2.71 CI=[0.28-12.40]
(d) [14/11-06:00, 15/11-06:00]; $N_d = 354$ $M_l = 3.00$	mean=1.68 CI=[1.37-2.02]	mean=1.09 CI=[0.86-1.30]	mean=0.06 CI=[0.02-0.12]	mean=1.27 CI=[1.02-1.69]	mean=1.28 CI=[0.56-2.42]	mean=1.75 CI=[1.31-2.39]	mean=0.39 CI=[0.24-0.53]	mean=0.96 CI=[0.32-4.56]
(e) [15/11-06:00, 16/11-06:00]; $N_d = 427$ $M_l = 3.00$	mean=1.77 CI=[1.45-2.11]	mean=1.09 CI=[0.87-1.29]	mean=0.05 CI=[0.02-0.09]	mean=1.27 CI=[1.02-1.68]	mean=1.17 CI=[0.50-2.10]	mean=1.68 CI=[1.31-2.24]	mean=0.39 CI=[0.24-0.52]	mean=1.30 CI=[0.33-4.25]
Statistics of prior marginal PDFs	median=2.30 COV=0.50	median=2.30 COV=0.50	median=0.03 COV=0.50	median=1.10 COV=0.50	median=1.00 COV=0.50	median=1.50 COV=0.50	median=0.20 COV=0.50	—

*The time window denotes the forecasting interval; note that “hour” is based on UTC.

Table 1. Distribution of the ETAS model parameters including marginal PDFs of posterior (shown as histograms) and lognormal PDFs of prior (shown with dashed orange lines), and the statistics of posterior (including mean and [2%-98%] confidence interval (CI) for providing early seismicity forecasts after $M_w 7.3$ at 12–November 2017.

PDFs (illustrated with dashed-orange lines); in addition, the histogram of parameter K derived based on Eq. (17) is shown in the last column. The statistics of the samples including mean and [2%-98%] confidence interval (CI) of the posterior of model parameters $\{\theta = [\beta, \alpha, c, p, d, q, \gamma], K\}$ are also shown on the corresponding marginal distribution in Table 1. The last row of Table 1 reports the statistics for the prior marginal PDFs of θ . To capture the outliers of K , we have set the upper limit of the horizontal axis to be around 98th percentile of K . Since all ETAS parameters are positive, we assigned a multivariate lognormal distribution as prior to the model parameters θ (see Eq. 19; zero correlations are considered at the prior level) with a coefficient of variation (COV) equal to 0.5 for all the parameters (high COV is assumed to avoid using an over-informative prior distribution). The prior median values for both β and α are set to $1.0 \times \ln 10 = 2.3026$. This selection was done deliberately to allow the Bayesian updating algorithm to infer these two parameters. It can be observed in Table 1 that $\alpha < \beta$ holds for all generated samples based on MCMC simulation. This is interesting investigation as we did not put any restriction on these two parameters while performing the Bayesian updating ($\alpha < \beta$ is a necessary condition for a stable “subcritical” process⁶⁵). Prior median value $c = 10^{-1.53} \approx 0.03$ (day) is set based on the value provided by the MO Italian generic model parameters⁶⁶, which is close to the California generic model (=0.05; see^{67,68}), and equal to the New Zealand generic model (=0.03; see⁶⁹). Prior median values assigned to $[p, d, q]$ are equal to [1.1, 1.0, 1.5], which were used in²⁸. The prior median for $\gamma = 0.20$ is selected based on⁷⁰. With reference to Table 1, the mean value of parameter β matches quite well the values obtained based on β_{ML} (see Fig. 3 for the first forecasting interval and also figures in Section SI-3-Results for the subsequent forecasting intervals). The mean values of Parameter $\alpha \cong 1.05 - 1.12$ are quite invariant as the sequence evolves. The mean value of parameter $c \cong 0.05$ (day) seems to be invariant in the first few days after the main event, which is reasonable as it measures the time offset. Parameters p and d gradually decrease as the observed data in **seq** accumulates, while q increases. Finally, the mean of the magnitude-dependent coefficient of the kernel density $\gamma \cong 0.31 - 0.39$. It is to note that as the sequence evolves and observed data in **seq** accumulates, the variation in the posterior of model parameters reduces. The temporal evolution of ETAS model parameters for both phases 1 and 3 of the seismic sequence is shown in Fig. 8 and will be discussed later in “Daily forecasts of seismicity after the event Mw6.3 of 25/11/2018 (Phase 3)” section.

Discussion on the correlations between pairs of model parameters θ in Phase 1. Figure 4 illustrates the contour lines showing the correlation structure between each of the sample marginal posteriors $[\beta, \alpha, c, p, d, q, \gamma]$

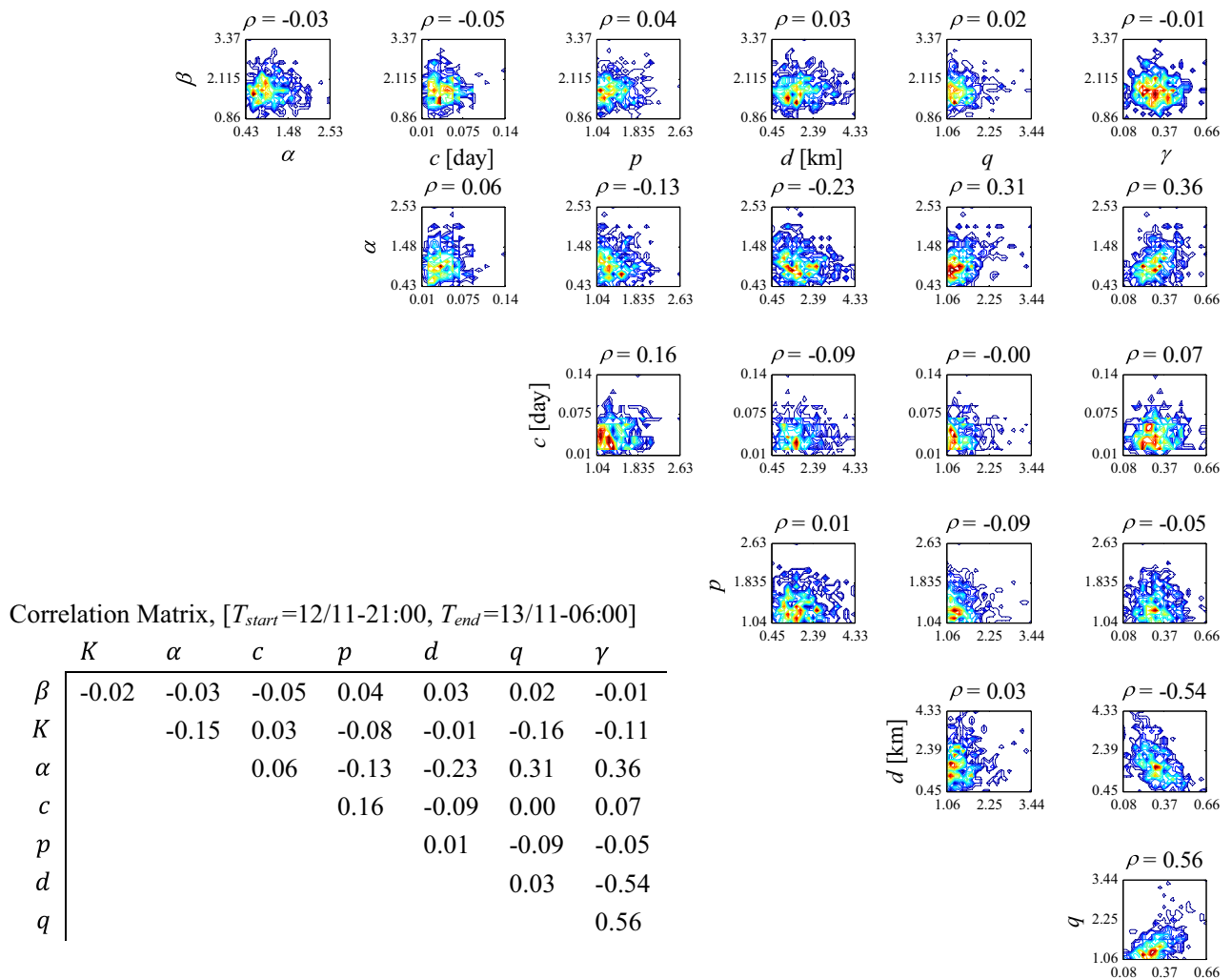


Figure 4. Contour lines representing the correlation between pairs of the posterior marginal samples of $\theta = [\beta, \alpha, c, p, d, q, \gamma]$ for the first forecasting interval (1st row in Table 1) and the corresponding correlation matrix.

shown in the first row of Table 1 for the first forecasting interval (a). The correlation coefficient ρ is reported above each figure. We have reported the complete correlation matrix of the ETAS parameters $\theta = [\beta, K, \alpha, c, p, d, q, \gamma]$ in the left side of the contour plots in Fig. 4. It is revealed that β has almost no correlation or small correlation (up to the order of around 10%) with other ETAS parameters, as expected from the functional form of the seismicity rate λ_{ETAS} in Eqs. (1) and (2). Parameter K reveals small correlations with all ETAS parameters. Indeed, K is related to other ETAS parameters through Eq. (17). Parameter α shows correlation with temporal parameters (c and p) and even more significant correlation with the parameters of the spatial kernel density (d, q, γ). There is correlation between the temporal parameters (c, p), significant negative correlation between the spatial parameters (d, γ), and significant positive correlation between (q, γ). However, parameters (d, q) show no dependence. The correlation structure for ETAS parameters for the subsequent forecasting intervals (b) to (e) are shown in *Supplementary Information* (Section SI-5-Results). As the sequence evolves, parameter K tends to show more correlation with p . The correlations between α and temporal/spatial parameters hold more-or-less at the same amount. On the same page, the absolute values of correlations between (c, p) and (d, γ) increase, while the correlation between (q, γ) decreases.

Representing the robust forecasting results. Given that θ and seq are known, sequences of events taking place during the forecasting interval, denoted as seqg , are generated according to “Generating sequences according to $p(\text{seqg}|\theta, \text{seq}, M_i)$ ” section (“Methods”). Based on the description provided in Section SI-4-Results, $M_{max} = 7.5$ is considered for generating seqg for all the forecasting intervals within Phase 1 (“Daily forecasts of seismicity for the first days after the event Mw7.3 of 12/11/2017 (Phase 1)” section) and Phase 3 (“Daily forecasts of seismicity after the event Mw6.3 of 25/11/2018 (Phase 3)” section). Based on samples generated for θ and seqg in Eq. (9), the robust estimate for the number of events with $M \geq m$ in a spatial cell units centered at (x, y) within the aftershock zone is obtained. This robust estimate is calculated as the expected number of events considering the uncertainties in the spatio-temporal distribution of the sequence of events.

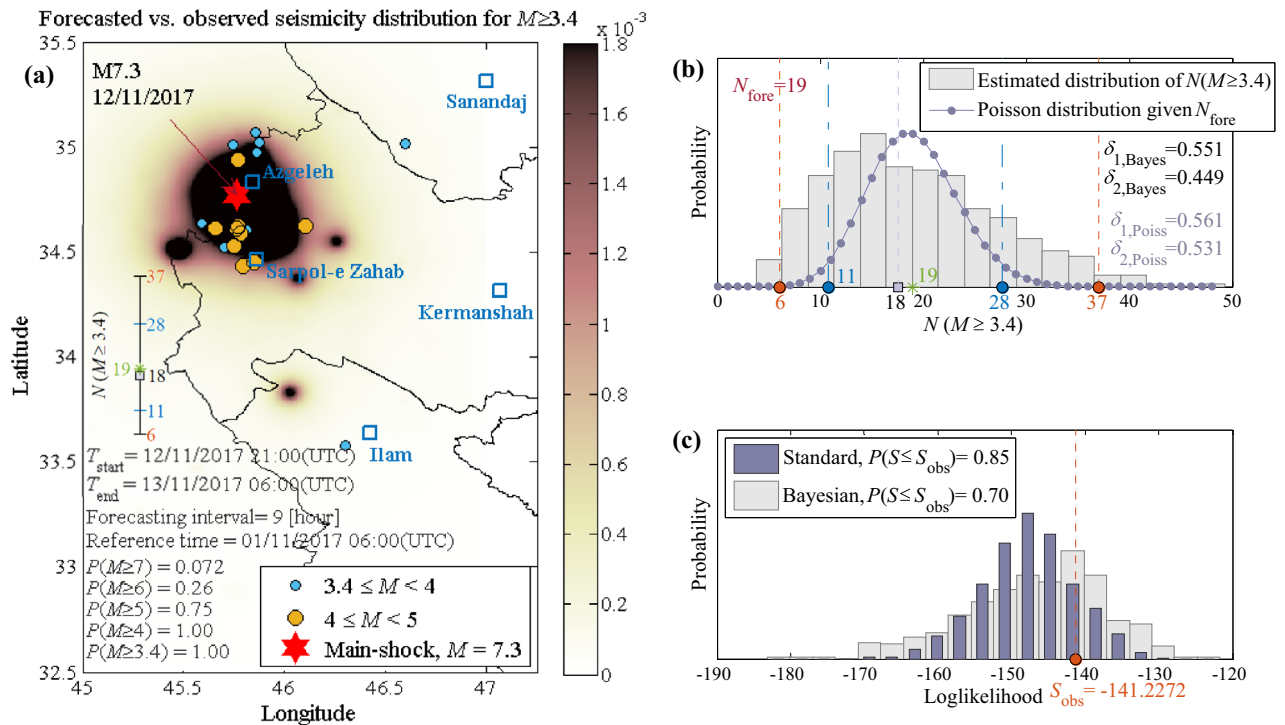


Figure 5. (a) The map showing forecasted vs. observed seismicity distribution in the aftershock zone for the 1st forecasting time window [$T_{start} = 12/11/2017-21:00\text{UTC}$, $T_{end} = 13/11/2017-06:00\text{UTC}$] including: the expected value for the number of events in each cell unit with $M \geq 3.40$; the reported $P(M \geq m)$; the earthquakes that occurred during the corresponding forecasting time window; the mainshock of $M_w 7.3$; and bar chart showing the observed vs. the percentiles of the forecasted number of events. (b) N-test based on the simulation-based Bayesian workflow and the Poisson distribution; consistent with the error-bar on the left-side of (a), the green star shows the observed number of events within the forecasting interval (= 19), the grey-filled square is the median value or the 50th percentile (= 18); blue circles are the 16th and 84th percentiles (= 11 & 28); red circles show the 2nd and 98th percentiles (= 6 & 37); $N_{fore} = 19$ is the expected number of events $N(M \geq 3.40)$. (c) S-test based on the simulation-based Bayesian framework and the standard method.

Figure 5a shows the forecasted seismicity maps for the number of events with $M \geq M_I = 3.40$ issued for the first forecasting time window (a) in Phase 1 (as shown in Table 1) within each spatial cell unit (latitude/longitude cells of a $0.01^\circ \times 0.01^\circ$ grid). The maps illustrate the expected number of events in each cell unit ($E[N(x, y, M \geq M_I | \text{seq}, M_I)]$ in Eq. (9), Section “Methods”). The earthquakes of interest occurred within the corresponding forecasting interval, [T_{start} , T_{end}], are illustrated as colored dots (distinguished by their magnitudes) with the main event of $M_w 7.3$ (red star) illustrated on all plots. We also report the forecasted probabilities of having earthquakes with magnitude equal to or larger than $m = [M_I, 4, 5, 6, 7]$, denoted as $P(M \geq m)$, in the whole aftershock zone and in the corresponding forecasting interval (see Eq. 12). Inside Fig. 5a, the observed (shown as a green star) vs. forecasted number of events (shown in an error-bar format) is illustrated for events with $M > M_I$ for the entire aftershock zone. The forecasted number of events on the error-bar chart features: the median value (the 50th percentile, equivalent to the exponential of the logarithmic mean) marked with a grey-filled square; the (logarithmic) mean \pm one (logarithmic) standard deviation indicating the interval between 16 and 84th percentiles (marked with blue horizontal lines and numbered in blue); the (logarithmic) mean \pm two (logarithmic) standard deviations indicating the interval between 2nd and 98th percentiles (marked with black horizontal lines and numbered in red). Obviously, we have rounded the forecast statistics to the nearest integers. This bar chart helps in locating the observed number of events (marked and numbered in green) within plus or minus certain number of standard deviations from the median estimate.

The N-test (see “N-test based on the simulation-based Bayesian framework” section in “Methods”) is used to validate the total number of events forecasted by the proposed method. Figure 5b shows the N-test associated with each forecasting interval. The robust estimate for the number of events within the forecasting interval using Eq. (9) will lead to the distribution of events with $M \geq M_I$, denoted as $N(M \geq M_I | \theta, \text{seq}, M_I)$ in Eq. (12), as shown with the histogram in these figures. This distribution is shown also with error-bar chart in Fig. 5a. The observed number of events within the forecasting interval, N_{obs} , is shown as a green star (as also shown on the bar chart). For the N-test, we use the same colors as the bar chart to represent the forecasted number of events: the median value (the 50th percentile) is marked with a grey-filled square; the 16th and 84th percentiles (marked with blue circles and numbered in blue); the 2nd and 98th percentiles (marked with red circles and numbered in red). The classical N-test based on a Poisson distribution with the mean value equal to the forecasted number

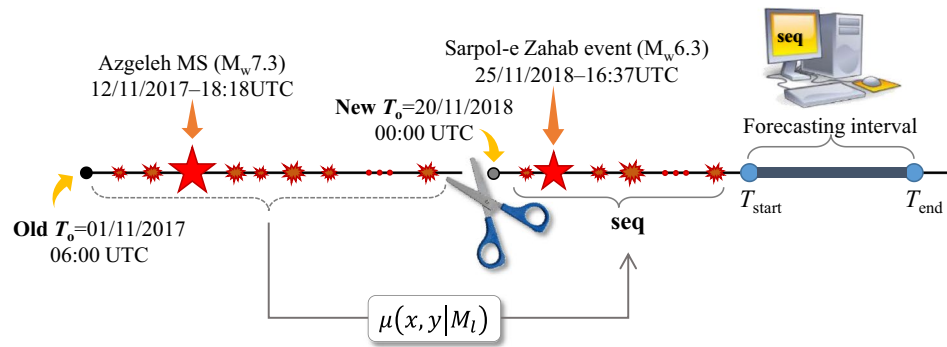


Figure 6. Sketch showing the shift in the time of origin T_o .

of events $N_{\text{fore}} = \mathbb{E}[N(M \geq M_l | \theta, \text{seq}, M_l)]$ (see “Methods” “N-test based on the simulation-based Bayesian framework” section and Eq. 12) is also shown on the same figure. It can be noticed that the distribution of $N(M \geq M_l | \theta, \text{seq}, M_l)$ according to the histogram shows a larger spread with respect to the equivalent Poisson distribution. The terms $\delta_{1,\text{Poiss}}$ and $\delta_{2,\text{Poiss}}$ (based on the Poisson distribution), and $\delta_{1,\text{Bayes}}$ and $\delta_{2,\text{Bayes}}$ (based on the Bayesian simulation-based workflow) represent the probabilities that the number of events $n \leq N_{\text{obs}}$ and $n \geq N_{\text{obs}}$ given N_{fore} as described in “N-test based on the simulation-based Bayesian framework” section (“Methods”). They are used as an indicator for validating the forecasts within different time intervals.

The S-test (see “S-test based on the simulation-based Bayesian framework” section in “Methods”) is used to validate the spatial distribution of the forecasts provided by the proposed method. Figure 5c illustrates the S-test results regarding each forecasting interval based on the two methods described, namely, the standard and the Bayesian methods. It shows the distribution of the vector of the simulated log-likelihood values S_{sim} (see Eq. 34) based on standard (in dark blue) and Bayesian (in light grey) histograms. The observed likelihood (S_{obs} ; see Eq. 33) is marked with red circle and numbered in red. The probability $P(S \leq S_{\text{obs}})$ based on both methods are reported on each figure. In the *Supplementary Information* (Section SI-6-Results), the forecasted seismicity maps, N-test and S-test results for $M \geq M_l$ are illustrated for the subsequent forecasting time windows (b-e) in Phase 1 (as shown in Table 1). Moreover, Fig. 9 (in “Daily forecasts of seismicity after the event Mw6.3 of 25/11/2018 (Phase 3)” section) shows the evolution of the N-test and S-test in the five forecasting intervals in Phase 1.

For this first and early forecasting interval (2 h and 42 min after the main event), **seq** includes 16 events with $M \geq 3.4$. It is noted that the observed number of events (= 19) is equal to the $N_{\text{fore}} = \mathbb{E}[N(M \geq M_l | \theta, \text{seq}, M_l)]$ reported in Fig. 5b. The probabilities δ_1 and δ_2 (based on Bayesian method or Poisson distribution) are around 50% which indicate that the forecast matches observed number of events. Both methods for S-test indicate that the spatial distribution of forecasts reasonably matches the observed one.

Daily forecasts of seismicity after the event Mw6.3 of 25/11/2018 (Phase 3). Phase 3 of the complex Kermanshah seismic sequence 2017–2019 (described in Section “Introduction”) initiates by the Sarpol-e Zahab event of $M_w 6.3$, which took place in November 25, 2018, at 16:37 (UTC), around one year after the Azgeleh MS of $M_w 7.3$ in Phase 1. Given the time elapsed from the occurrence of the Azgeleh MS, it is computationally unjustifiable and tedious to consider entire **seq** consisted of the events of interest starting from the time origin T_o (i.e., 01/11/2017–06:00 UTC used for Phase 1). To this end, we have shifted the time origin T_o from old $T_o = 01/11/2017-06:00$ UTC to new $T_o = 20/11/2018-00:00$ UTC (where new T_o is more than 5 days before the occurrence of $M_w 6.3$ earthquake), as shown in Fig. 6. With this shift in the time origin T_o , the spatial rate of background seismicity, $\mu(x, y | M_l)$ (see “Robust estimation for the number of aftershock events” section), is updated with respect to the long-term seismicity used for Phase 1 of the sequence. All the seismicity data from the old T_o up to the new T_o , in addition to the long-term background seismicity is used to calculate the updated $\mu(x, y | M_l)$ that is added as a constant term to the contribution of the triggering events (see “Robust estimation for the number of aftershock events” section). Herein, we use this background seismicity to conservatively approximate the triggering effect of the events that occurred in the first part of the sequence in the time interval [old T_o – new T_o] (see the following “Estimating the updated background seismicity” section for the calculation of the updated $\mu(x, y | M_l)$ in detail).

The reference-time shift has been shown²⁸ to work well in providing operational seismicity forecasts and avoids considering all the observed data within the **seq** (this claim will be supported by further studies in “6-h forecasting for the time interval [25/11/2018–18:00 UTC–26/11/2018–00:00 UTC]” section). We set $M_l = M_c = 2.70$ (see *Supplementary Information* Section SI-7-Results, Figure SI-14a and its discussion). Figure 7 (top) shows the evolution of the events within Phase 3, and Fig. 7 (bottom) shows the evolution of the number of events in phase 3 starting from new T_o in the AS zone and with $M \geq 2.7$ (orange circles) within a 24-h interval starting from 0:00 UTC each day (also reported at 0:00 UTC each day). The seismicity after the occurrence of $M_w 6.3$ increases considerably on 25 November and decreases very quickly the day after (around 20 events with $M \geq 2.70$ occurred). Note also the two triggered events with very close epicenters of $M_w 5.3$ and $M_w 5.0$ occurred on 25/11/2018 and 26/11/2018, respectively. In the following, we describe the forecasting for the period from 25 to 26th of November 2018.

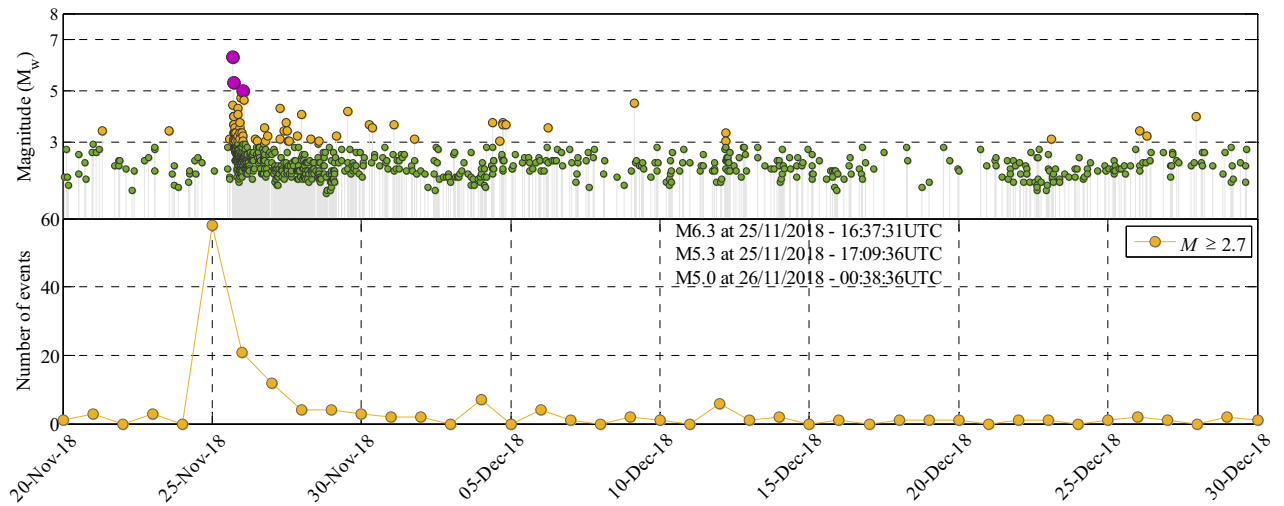


Figure 7. Magnitude vs. time (top) and the number of events with $M \geq 2.7$ (bottom) occurred within a 24-h interval starting from 0:00 UTC of each day from 20/11/2018 (new T_o) until 30/12/2017 including the Sarpol-e Zahab event of $M_w,6.3$ within phase 3 of the Kermanshah 2017–2019 seismic sequence.

Estimating the updated background seismicity. The rate of seismicity is estimated based on **seq** consisting of 1337 events with $M \geq M_l = 2.7$ (Supplementary Information Section SI-7-Results, Figure SI-14a) in the interval [old T_o – new T_o] (Fig. 6). The ETAS model parameter θ is estimated with the same priors described in “Bayesian Inference for θ ” section (the last row in Table 1). The background seismicity is drawn from the long-term seismicity of the site that is the same used in Phase 1. The histograms of marginal posterior PDFs for the model parameters θ together with their statistics are shown in first row of Table SI-2 (Supplementary Information Section SI-8-Results, labeled a). It is noted that parameter p is estimated almost close to 1.0 with a very small CI = [1.04, 1.12]. The median of these model parameters will be used as the new priors for the upcoming forecasting intervals. The correlations between ETAS parameters are also reported and discussed in Section SI-9-Results (first correlation matrix). The forecasted seismicity within the 24-h interval, [20/11/2018–00:00 UTC – 21/11/2018–00:00 UTC], is estimated in terms of the expected number of events in the spatial cell unit with $M \geq M_l = 2.7$ denoted as $\mathbb{E}[N(x, y, M \geq M_l | \mathbf{seq}, M_l)]$ (see Eq. 9); thus, the updated background seismicity rate is $\mu(x, y | M_l) = \mathbb{E}[N(x, y, M \geq M_l | \mathbf{seq}, M_l)]$ (“Robust estimation for the number of aftershock events” section in “Methods”), which includes the rate of background events due to the long-term seismicity plus the triggering effect of the events in **seq**. The updated $\mu(x, y | M_l)$ is used as a constant background rate in Eq. (7) to Eq. (9), and it conservatively does not represent the expected decay with time after the new T_o . $\mu(x, y | M_l)$ is plotted in Figure SI-18a (Section SI-10-Results). It is observed that the updated background seismicity has increased more than 300 times and that the probability of occurrence of $M \geq 6$ is more than 12 times higher the same probability estimated based on the long-term background seismicity, which reveals an alarming level (this was verified by Sarpol-e Zahab event of $M_w,6.3$).

6-h forecasting for the time interval [25/11/2018–18:00 UTC–26/11/2018–00:00 UTC]. After the occurrence of Sarpol-e Sahab event with $M_w,6.3$ at 16:37 UTC of 25/11/2018, we provide a 6-h prediction of seismicity for the forecasting interval starting from T_{start} set to 18:00 UTC (i.e., 1 h and 23 min after the occurrence of $M_w,6.3$ event). Considering new T_o (Fig. 6), the data within the time interval [new $T_o = 20/11/2018–00:00$ UTC, $T_{start} = 25/11/2018–18:00$ UTC] forms the **seq**, which includes 27 events with $M \geq M_l = 2.7$ ($M_l = M_c$; see Section SI-7-Results, Figure SI-14b and its discussion). The histograms of marginal posterior PDFs for the model parameters θ together with their statistics are shown in second row of Table SI-2 (Section SI-8-Results, labeled b). The spatial distribution of the forecasted seismicity with $M \geq 2.7$ is shown in Figure SI-19a (Section SI-10-Results). Observed events with $M \geq 2.7$ are also shown on the map. The Sarpol-e Zahab $M_w,6.3$ is shown with red-colored hexagram to show the epicenter of the main event – although this event is not within the observed events in the forecasting time interval. According to the error-bar plot in the same figure, the total number of registered events within the 6-h forecasting interval (green star; = 38) is higher than the forecasted values (median; = 18). This can be attributed (1) to the small number of observed input data in **seq**; (2) the presence of the triggered event with $M_w,5.3$ at 17:09. The N-test result in Figure SI-19b shows the histogram for the distribution of the number of events based on the Bayesian workflow with $N_{fore} = \mathbb{E}[N(M \geq 2.7 | \theta, \mathbf{seq}, M_l = 2.7)] = 18$ and other related statistics (similar to the error-bar chart in Figure SI-19a). The probabilities δ_1 and δ_2 (based on Bayesian method or Poisson distribution) show that the real number of events falls in the upper tail of the distributions (i.e., underestimation in the forecasted seismicity). However, the distribution of the forecasted seismicity is validated through the standard/Bayesian S-test method (Figure SI-19c) that shows a 30% value for the probability $P(S \leq S_{obs})$. Although less successful in predicting the number of events, Figure SI-19a reports exceedance probabilities $P(M \geq 5)$ more than 30 times and $P(M \geq 6)$ around 90 times compared to their corresponding initial estimates based on the updated background seismicity in Figure SI-18a. In order to validate the reference-time shift

consideration, we perform the same forecast done here considering the whole events with $M \geq M_l = 2.7$ in the sequence (= 1364) starting from [old $T_o = 01/11/2017-06:00$ UTC, $T_{start} = 25/11/2018-18:00$ UTC], which form the **seq**. The marginal distributions of the ETAS model parameters have smaller confidence interval compared to those reported in Table SI-2 (row b) as the Bayesian updating of model parameters is done based on higher data points in the observation history. Figure SI-19d shows the spatial distribution of the forecasted seismicity, where the Azgeleh $M_w 7.3$ mainshock is indicated by a red-colored hexagram (while the Sarpol-e Zahab $M_w 6.3$ is illustrated with magenta-colored triangle). The error-bar plot in the same figure shows the distribution of the forecasted number of events with median = 15 (event) and a smaller confidence interval (as noted previously) with respect to the similar error-bar in Figure SI-19a. The N-test result in Figure SI-19e shows the histogram for the distribution of the number of events with smaller dispersion (compared to Figure SI-19b) with $N_{fore} = 15$ being close to the previous estimate in Figure SI-19b (= 18) that is obtained based on the shift in the time origin. With reference to the S-test (Figure SI-19f.), the distribution of the forecasted seismicity is even better predicted with $P(S \leq S_{obs})$ more than 70%. As a result, the shift in the time of origin (herein, more than one year) of the sequence, by introducing a non-uniformly distributed background seismicity, does not affect the results of our forecast overall while it relieves the computational cost of summing up the triggering properties of all the events that took place in an ongoing sequence in the time interval of more than one year. Neglecting the time-decay in the triggering of the events before “new T_o ” can even furnish more conservative estimates.

5-h forecasting for the time interval [25/11/2018–19:00 UTC–26/11/2018–00:00 UTC] and two subsequent forecasts. To get more reliable early forecasts, we set $T_{start} = 25/11/2018-19:00$ UTC (i.e., 2 h and 23 min after the occurrence of $M_w 6.3$ event) and keep the same T_{end} . For this forecasting interval, $M_c = 2.70$ (Section SI-7-Results). Thus, **seq** includes 43 events occurred from new T_o up to T_{start} with $M \geq M_l = 2.7$ (where $M_l = M_c$). The marginal distribution of the updated ETAS model parameters θ is shown in Table SI-2, (Section SI-8-Results, labeled c). The spatial distribution of the forecasted seismicity (the expected value for the number of events with $M \geq 2.7$), the N-test, and S-test results are shown in Figure SI-20 (Section SI-10-Results) for this 5-h forecasting interval. The total number of registered events (green star; = 22) is within + 1 standard deviation confidence interval (also verified by the probabilities δ_1 and δ_2). The S-test also shows that the forecasted spatial distribution of seismicity matches quite well the observed spatial pattern.

Seismicity forecasts are also provided for two subsequent back-to-back intervals; namely, a 6-h time interval with $T_{start} = 26/11/2018-00:00$ UTC and a 18-h forecast with $T_{start} = 6:00$ UTC of the same day with $M_c = 2.70$ (Section SI-7-Results). The last two rows of Table SI-2 (d, e) illustrate the statistics of updated model parameters θ . The evolution of ETAS model parameters for Phase 3 of the seismic sequence is shown in Fig. 8. It can be seen that all parameters start to stabilize after the early forecasts. As expected, parameter β stabilizes to a slightly larger value in Phase 3 (1.95); in both phases it is estimated to be larger than background β (1.67). In both phases, α is smaller than β (sub-critical process). As expected, α for Phase 3 is smaller than that of Phase 1, which is also expected given the lower productivity of this phase. Note that the value of K shows discontinuity between the two phases, this is due to the shift in the time of origin. The larger p value in Phase 3 indicates a steeper time decay. The same is through for parameter q which indicates a stronger spatial attenuation in Phase 3. Parameter γ indicates a stronger magnitude dependence of the spatial kernel in Phase 1 compared with Phase 3.

The predicted seismicity distribution maps and the corresponding error-bar plots, together with the N-test and S-test results are shown in Figures SI-21 and SI-22 (Section SI-10-Results). For the 6-h forecasting interval (Figure SI-21a), the $M_w 5.0$ event occurred 38 min after T_{start} (see also Fig. 7). The forecasts for these two latter time intervals manage to properly predict the observed seismicity in terms of the number of events with $M \geq 2.7$, as well as its spatial content through the seismicity tests. The correlations between ETAS parameters for the four forecasting intervals in Phase 3 are reported and discussed in detail in Section SI-9-Results. Figure 9 summarizes the evolution of the N-test and S-test in the four forecasting intervals in Phase 3.

Sensitivity analyses. In this section, the sensitivity of the resulting robust forecasts to some possible variations to the proposed method are investigated.

Using the simple spatial kernel density compared to the magnitude-dependent kernel density. An alternative kernel density, denoted herein as the simple spatial kernel density with two spatial term parameters (d, q), can be used instead of the magnitude-dependent kernel with the three related parameters (d, q, γ). Table SI-3 (Section SI-11-Results) shows the histograms of marginal posterior PDFs for the six ETAS model parameters $\theta = [\beta, \alpha, c, p, d, q]$ where the simple spatial kernel density term is used. The distribution of parameter β and its statistics matches very well those reported in Table 1. The mean values of Parameter $\alpha \cong 0.70 - 0.80$ are around 2/3 of those reported in Table 1. The distribution of temporal parameters (c, p) are more-or-less invariant. The posterior distributions of spatial parameters (d, q) are different (see Table 1) in the sense that parameter d has a higher mean and dispersion, while parameter q behaves in an opposite manner. Parameter K has a higher estimate in case of simple spatial kernel density (to be expected since it has to fulfill Eq. (13) for each realization of θ). The correlation matrices of 7 model parameters are illustrated in Section SI-11-Results for different forecasting intervals. Comparing the correlation structure with those obtained based on the magnitude-dependent spatial kernel density (Fig. 4 and Section SI-5), it is revealed that β has no correlation and K reveals a sort of small correlations with other ETAS parameters (similar behavior for both kernels). Parameter α has smaller correlation with temporal parameters (c and p), and spatial parameters (d, q) in case of simple kernel density. This is the reason for having different estimates for this parameter between the two spatial models. In the magnitude-dependent kernel density model, α reveals more correlation with the magnitude-related parameter γ compared to the other parameters (see Fig. 4 and Section SI-5). The correlations between temporal parameters (c, p) and spatial param-

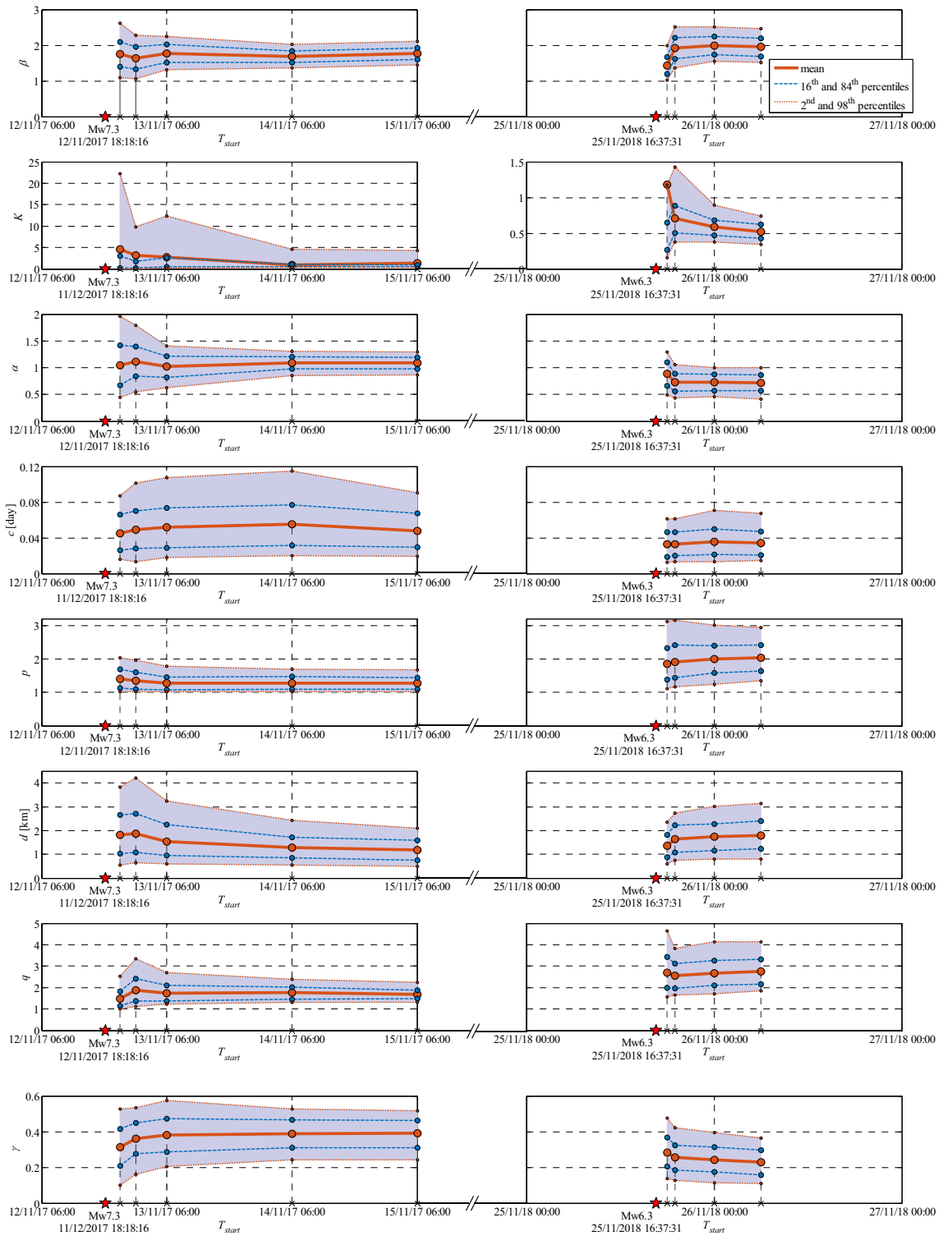


Figure 8. Summary and timeline of the evolution of ETAS model parameters for Phase 1 (left column) and Phase 3 (right column).

eters (d, q) increases as the sequence evolves. It is to note that in the magnitude-dependent spatial kernel density model, there is no dependence between the spatial parameters (d, q).

Table 2 compares the results of seismicity forecasting test, N-test and S-test, based on alternative spatial kernel density terms, for the five forecasting intervals indicated in Table 1. Both models for spatial kernel density are similarly evaluated through the N-test and S-test. It is interesting to highlight that S-test results in terms of $P(S \leq S_{obs})$ becomes higher as the forecast evolves showing that the simple kernel density provides a better

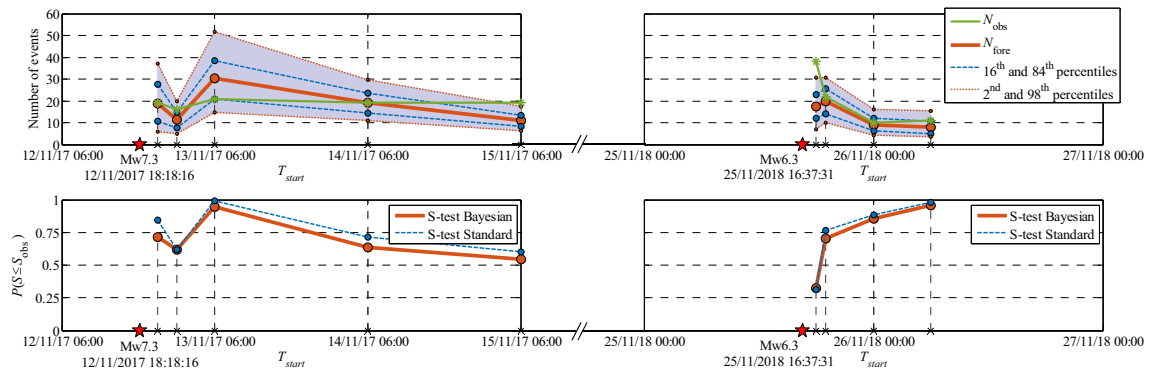


Figure 9. Summary and timeline of the evolution of N-Test and S-test results for Phase 1 (left column) and Phase 3 (right column).

spatial forecast of seismicity compared to the more complex magnitude-dependent model (although this might not be a general conclusion).

Considering background seismicity. Since the background seismicity is low compared to the big jump in seismicity due to the $M_w 7.3$ mainshock for Phase 1 (i.e., the first term N_b compared to the second term in Eq. (7), “Methods”), it is expected that the effect of not considering the background seismicity (BGS) spatial data $\mu(x, y|M_I)$ within the short term operational forecasting may not be significant. Table 2 compares the results of observations vs. forecasts in terms of N-test and S-test for the cases with BGS (*Proposed method*) with the case without BGS ($BGS=0$) for all forecasting intervals indicated in Table 1. It can be observed from Table 2 that assuming $BGS=0$ leads to no apparent difference with respect to considering BGS based on both N-test and S-test results. That is, neglecting the effect of long-term BGS within the proposed robust seismicity forecasting framework will not significantly affect the forecasted number of events. This appears to be useful while operational forecasting is undertaken within an ongoing seismic sequence with limited time and input data.

Effect of some simplifications within the robust seismicity forecasting framework. The proposed framework can be simplified/approximated by considering the following assumptions (see “Robust estimation for the number of aftershock events” section in “Methods”):

1. Characterizing parameter K : this parameter can be calculated through the closed-form expression in Eq. (17) (“Calculating K ” section, *Calculate K*), or can be learnt through the Bayesian updating framework (“Calculating K ” section, *Learn K*). It is to note that the parameter K learned through the Bayesian inference does not necessarily satisfy Eq. (13) for each sample θ .
2. The integral over the whole aftershock zone \underline{A} : the term I_r (see Section “Methods”, Eq. (16)), can be solved analytically. However, by assuming an infinite spatial domain, I_r can be approximated with \tilde{I}_r over the infinite space.

Figure 10 shows a general overview of schemes to provide seismicity forecasts based on different levels of approximation that encompass: (a) *Proposed method* is the suggested method; (b) *Semi-Fast method* approximates the integral over the aftershock zone with \tilde{I}_r ; (c) *Fast method* learns K through Bayesian inference (i.e., $\theta = [\beta, K, \alpha, c, p, d, q, \gamma]$) and approximates the integral over the aftershock zone with \tilde{I}_r .

Discussion on the effect of considering the integration over the whole aftershock zone (using *Semi-Fast method*). Table SI-4 (Section SI-12-Results) shows the histograms of marginal posterior PDFs for the model parameters θ together with their prior PDFs using *Semi-Fast method* as opposed to results shown in Table 1 for the *Proposed method*. Comparing these two Tables, apart from the spatial kernel density parameters of distance d and q , the statistics (mean and CI) of the other parameters are very close (as the parameter estimation evolves, they become even closer). Parameters d and q have higher estimates (consequently higher mean, 2%, and 98% statistics) in case of *Semi-Fast method* compared to the *Proposed method*. In terms of the correlation structure among the parameters, we have reported the correlation matrix between the ETAS model parameters $\theta = [\beta, K, \alpha, c, p, d, q, \gamma]$ in Section SI-12-Results for *Semi-Fast method* which shows the same trends as those observed for the *Proposed method*.

Table 2 summarizes the N-test and S-test results of the *Semi-Fast method* to be contrasted with the first column of Table 2 associated with the *Proposed method*. There are some differences between the two: (1) The first forecasting interval for *Semi-Fast* has lower expected value (N_{fore}) with respect to *Proposed method*; (2) the *Proposed method* outperforms the *Semi-Fast* based on S-test results. In terms of N-test, as the sequence evolves, the differences between the test results becomes negligible. This observation has been also made in⁷¹, where the assumption of an infinite spatial domain was shown to have negligible effect on likelihood. (3) Overall, the computational time for the *Semi-Fast method* is observed to be around 75% of that of the *Proposed method*.

Measure of forecast		Proposed method*	Using simple spatial kernel density	No background seismicity, BGS=0	Semi-Fast method	Fast method	
(a) [12/11-21:00, 13/11-06:00]; $M_I = 3.40$	Distribution of N	N_{fore}, N_{obs}	19, 19	21, 19	19, 19	17, 19	25, 19
		2 nd percentile	6	6	5	6	8
		16 th percentile	11	11	11	10	14
		Median	18	19	18	16	24
		84 th percentile	28	30	28	26	35
		98 th percentile	37	41	38	36	50
	N-test	$\delta_{1, Poiss}, \delta_{2, Poiss}$	0.56, 0.53	0.38, 0.70	0.56, 0.53	0.74, 0.34	0.13, 0.91
		$\delta_{1, Bayes}, \delta_{2, Bayes}$	0.55, 0.45	0.47, 0.53	0.54, 0.46	0.63, 0.37	0.28, 0.72
	S-test	Standard $P(S \leq S_{obs})$	0.85	0.62	0.83	0.49	0.64
Bayesian $P(S \leq S_{obs})$		0.70	0.56	0.75	0.44	0.54	
(b) [13/11-00:00, 13/11-06:00]; $M_I = 3.30$	Distribution of N	N_{fore}, N_{obs}	11, 16	12, 16	11, 16	11, 16	13, 16
		2 nd percentile	5	5	5	5	7
		16 th percentile	8	7	8	8	9
		Median	11	11	11	11	12
		84 th percentile	15	16	15	14	17
		98 th percentile	20	20	19	19	22
	N-test	$\delta_{1, Poiss}, \delta_{2, Poiss}$	0.94, 0.09	0.90, 0.16	0.94, 0.09	0.94, 0.09	0.83, 0.24
		$\delta_{1, Bayes}, \delta_{2, Bayes}$	0.90, 0.10	0.86, 0.14	0.89, 0.11	0.90, 0.10	0.77, 0.23
	S-test	Standard $P(S \leq S_{obs})$	0.59	0.77	0.74	0.33	0.42
Bayesian $P(S \leq S_{obs})$		0.62	0.76	0.62	0.35	0.40	
(c) [13/11-06:00, 14/11-06:00]; $M_I = 3.30$	Distribution of N	N_{fore}, N_{obs}	30, 21	31, 21	30, 21	29, 21	34, 21
		2 nd percentile	15	14	15	14	19
		16 th percentile	21	21	21	20	24
		Median	30	30	29	29	34
		84 th percentile	39	41	39	38	43
		98 th percentile	52	52	50	50	53
	N-test	$\delta_{1, Poiss}, \delta_{2, Poiss}$	0.05, 0.96	0.04, 0.98	0.05, 0.96	0.08, 0.95	0.01, 0.99
		$\delta_{1, Bayes}, \delta_{2, Bayes}$	0.16, 0.84	0.17, 0.83	0.14, 0.86	0.22, 0.78	0.05, 0.95
	S-test	Standard $P(S \leq S_{obs})$	0.99	1.00	0.99	0.95	0.93
Bayesian $P(S \leq S_{obs})$		0.94	0.99	0.96	0.96	0.89	
(d) [14/11-06:00, 15/11-06:00]; $M_I = 3.00$	Distribution of N	N_{fore}, N_{obs}	19, 19	19, 19	19, 19	19, 19	23, 19
		2 nd percentile	11	10	12	12	15
		16 th percentile	14	14	15	15	19
		Median	19	18	19	19	23
		84 th percentile	23	24	23	23	27
		98 th percentile	30	27	29	29	33
	N-test	$\delta_{1, Poiss}, \delta_{2, Poiss}$	0.56, 0.53	0.56, 0.53	0.56, 0.53	0.56, 0.53	0.24, 0.82
		$\delta_{1, Bayes}, \delta_{2, Bayes}$	0.52, 0.48	0.53, 0.47	0.49, 0.51	0.50, 0.50	0.19, 0.81
	S-test	Standard $P(S \leq S_{obs})$	0.72	0.91	0.72	0.51	0.53
Bayesian $P(S \leq S_{obs})$		0.66	0.88	0.65	0.45	0.48	
(e) [15/11-06:00, 16/11-06:00]; $M_I = 3.00$	Distribution of N	N_{fore}, N_{obs}	11, 19	11, 19	11, 19	11, 19	13, 19
		2 nd percentile	6	6	6	6	9
		16 th percentile	8	8	8	9	11
		Median	11	10	10	11	13
		84 th percentile	13	13	13	14	16
		98 th percentile	18	16	18	17	18
	N-test	$\delta_{1, Poiss}, \delta_{2, Poiss}$	0.99, 0.02	0.99, 0.02	0.99, 0.02	0.99, 0.02	0.96, 0.07
		$\delta_{1, Bayes}, \delta_{2, Bayes}$	0.99, 0.01	1.00, 0.00	0.99, 0.01	0.99, 0.01	0.99, 0.01
	S-test	Standard $P(S \leq S_{obs})$	0.56	0.82	0.58	0.33	0.32
Bayesian $P(S \leq S_{obs})$		0.51	0.76	0.52	0.37	0.33	

*The "Proposed Method" employs magnitude-dependent spatial kernel and considers the background seismicity.

Table 2. N-test and S-test results in different forecasting intervals considering different hypotheses.

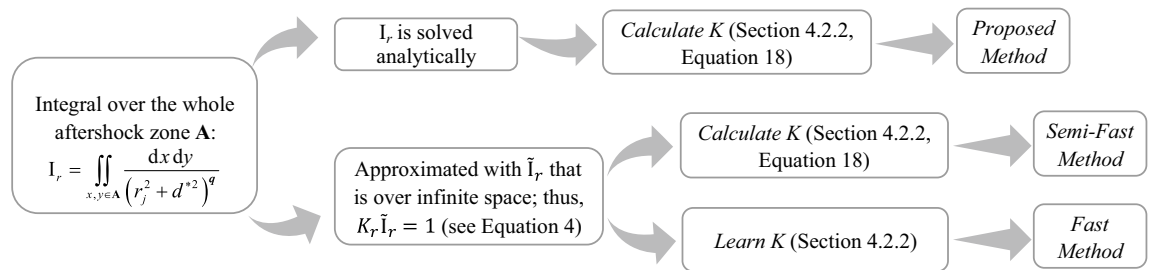


Figure 10. A schematic view of different approximations in the procedure to provide seismicity forecasts.

Discussion on the effect of calculating K . Table SI-5 in Section SI-13-Results illustrates the histograms of marginal posterior PDFs for the model parameters θ using *Fast method*. It can be compared with Table SI-4 based on *Semi-Fast method*. The parameter K is learnt through the MCMC procedure assuming that the model parameter θ has a multivariate lognormal prior (see “*Bayesian Inference for θ* ” section), for which we set a high COV = 1.0 for parameter K compared to other model parameters with COV = 0.5 (“*Bayesian Inference for θ* ” section). In the ballpark, the mean estimates for parameter K are quite the same between *Fast* and *Semi-Fast* methods with the CI for *Semi-Fast method* being wider. Although low number of observed data is available for learning K through the Bayesian framework at the very beginning of Phase 1, K does not show sensitivity to the choice of the prior in the *Fast method* (see Table SI-5). Parameter α has lower estimates based on the *Fast method*. However, as the sequence evolves, it gets closer to *Semi-Fast* estimates. The most critical parameter in the *Fast method* seems to be the temporal decay random variable p , which has a mean estimate very close to its lower threshold (= 1.0) and a very low variability. The temporal parameter c as well as the spatial parameters (d , q , γ) are quite close to those from the *Semi-Fast method*.

We have reported the correlation matrix between the ETAS model parameters in *Supplementary Information* (Section SI-13-Results) for *Fast method*. Compared to previous discussions on the correlation of the ETAS parameters by employing the *Proposed* and *Semi-Fast* methods, parameter β shows higher correlations with other model parameters while using *Fast method* (although the absolute value of the correlation is small). Parameter K shows a very high correlation with the temporal parameter p , and as the sequence evolves, this negative correlation increases (changing from around -55% to -80%). The negative correlations also appear in the *Proposed/Semi-Fast* methods but with lower intensities. This high correlation between (K , p) affects the correlation between (c , p) so that the amount of correlation between the latter two parameters are smaller than those appear in the *Proposed/Semi-Fast*. Parameter α shows the same order of correlation with temporal/spatial parameters in the *Fast method* compared to the other two methods. Higher correlations appear between p and the parameters of the spatial kernel density (compared to the *Semi-Fast method*). Like the previous two methods, there is a high negative correlation between (d , γ), and as the sequence evolves, the positive correlation between (q , γ) decreases.

Table 2 summarizes the N-test and S-test results of the *Fast method*. Comparing *Fast/Semi-Fast* (see the last two columns of Table 2): (1) For the first forecasting interval, *Fast method* slightly overestimated the mean number of forecasted events ($N_{\text{fore}} = 25$ vs 19 registered events with $M \geq 3.4$ took place) and the distribution of the forecasted number of events shows higher dispersion with respect to *Semi-Fast method*; (2) As the sequence evolves, the forecasts issued by the *Fast method* gets closer to those of *Semi-Fast method* (and consequently the *Proposed method*). This shows that *Fast method* is more reliable as the sequence evolves. However, like the *Semi-Fast method*, S-test shows that this method cannot predict the spatial pattern of seismicity as perfect as the *Proposed method* (it stills passes the test). (3) Overall, the time of conducting the *Fast method* is less than 50% of that for *Semi-Fast method* and 37% of that for *Proposed method*, making the procedure appealing for operational earthquake forecasting.

Based on the above discussions, it is recommended to do the *Proposed method* at least for very early forecasts. As the sequence evolves, the computationally less time-consuming *Fast* (or even the *Semi-Fast*) methods can be used.

Discussion and conclusions

We have improved and tested a Bayesian and fully simulation-based workflow for spatio-temporal early seismicity forecasting based on ETAS model. This workflow encompasses new/modified features of the Bayesian and fully probabilistic seismicity forecasting framework proposed previously by the first two authors²⁸. Regarding the Bayesian updating of the ETAS model parameters, the improvements concern (a) modification of the likelihood function and its calculation; (b) adoption of an adaptive MCMC simulation technique by using multi-dimensional kernel sampling density functions; (c) derivation of a close-form expression for productivity parameter K as a function of other ETAS parameters; (d) incorporation of seismicity tests such as N-test and S-test in the proposed workflow to see how well the forecasted number of events and the forecasted spatial distribution of seismicity match the observed ones. Additionally, the improved workflow is capable of (a) analytically solving the integral of the spatial kernel density over the whole aftershock zone instead of assuming it to be equal to unity; (b) incorporate the background seismicity within the robust seismicity forecasting framework; (c) consider magnitude-dependent spatial kernel density in the ETAS model.

As a demonstration of this procedure, we have applied the fully simulation-based workflow for providing retrospective early forecasts of seismicity in two distinct phases of the recent Kermanshah 2017–2019 seismic sequence in western Iran starting with the events of $M_w7.3$ and $M_w6.3$, respectively. Here are some observations:

- After an initial transition time in the order of few hours (2 h and 42 min after the main event of $M_w7.3$ in Phase 1, and 2 h and 23 min after the occurrence of $M_w6.3$ in Phase 3), the model tunes into the sequence and provides very reliable forecasts.
- The workflow traces the time evolution of the ETAS model parameters as the sequence goes on. Overall, the parameters tend to stabilize after the first early forecast intervals. Evolution of seismicity parameter β matches the expectations (a slight increase with respect to background seismicity). Parameter α is always smaller than β ensuring sub-critical behavior in the point process. Comparing Phase 1 and Phase 3, the ETAS parameters in Phase 3 reveal a seismic sequence with weaker productivity and stronger spatio-temporal decay, as matched by the observations.
- The productivity coefficient K is derived based on the constraint that the observed number of events previous to the time of starting the forecast is equal to the forecast provided based on each realization of the vector of model parameters (*Proposed method*). Alternatively, one can choose to learn K together with other ETAS model parameters (called herein *Fast method*) through the Bayesian inference. This is going to reduce the computational time by around 37% at the cost of losing accuracy in the early forecasting intervals.
- Neglecting the long-term background seismicity does not significantly affect the forecasted number of events in this study (Phase 1 forecasts).
- For providing early forecast of seismicity for $M_w6.3$ event in Phase 3, we performed a shift in the time of origin (more than one year) of the sequence by introducing a non-uniformly distributed background seismicity (which is calculated by exploiting the proposed method). This shift relieves the computational burden of summing up the triggering properties of all the events that took place in an ongoing sequence at the cost of neglecting the (negligible) time-decay (beyond the first day of Phase 3) in their triggering contribution.
- Adoption of a magnitude-dependent spatial kernel (compared to a simpler functional form) shows no significant differences in the seismicity forecasting estimates. However, S-test results reveal that the simple kernel performs better in forecasting the spatial distribution of seismicity.
- The integral of the spatial kernel density over the whole aftershock zone can be solved analytically (*Proposed method*) or can be approximated by assuming that the integral is over an infinite spatial domain (*Semi-Fast method*). In the early stage of forecasting, the *Proposed method* outperforms the *Semi-Fast method* based on both N-test and S-test results. As the sequence evolves, the differences between the test results becomes negligible. However, the computational time of conducting the *Semi-Fast method* is around 75% of the required time for performing *Proposed method*. Therefore, the use of *Semi-Fast method* is only recommended for providing forecasts as the sequence evolves.

Methods

The epidemic-type aftershock sequence (ETAS) model for space–time clustering of aftershocks. The ETAS model^{11,12} is a marked stochastic point process model widely used to describe the temporal and spatial clustering of seismicity (see also^{13–16,57}). In this model, every earthquake (being spontaneous or triggered) is a potential triggering event for its own subsequent earthquakes (aftershocks). Let $\lambda_{\text{ETAS}}(t, x, y, m|\theta, \text{seq}_t, M_l)$ denote the conditional rate of occurrence of events with magnitude M greater than or equal to m (the seismicity rate) based on the ETAS model at time t in the cell unit centered at the Cartesian coordinate $(x, y) \in \mathbf{A}$ where \mathbf{A} is the aftershock zone. The Cartesian area can be discretized into mutually exclusive and collectively exhaustive, MECE, subsets or spatial cell units centered at x and y . The rate λ_{ETAS} is conditioned on: the vector of ETAS model parameters θ (defined subsequently); observation history up to the time t denoted as $\text{seq}_t = \{(t_j, x_j, y_j, m_j), t_j < t, m_j \geq M_l\}$ where t_j is the occurrence time of the j th event (with $t_j < t$) with magnitude m_j and location $(x_j, y_j) \in \mathbf{A}$; the lower magnitude M_l (should be greater than or equal to the completeness magnitude M_c of the catalog of events up to the time t ; i.e. $M_l \geq M_c$; see the note on the calculation of M_c at the end of this section). $\lambda_{\text{ETAS}}(\cdot)$ can be shown as follows:

$$\lambda_{\text{ETAS}}(t, x, y, m|\theta, \text{seq}_t, M_l) = P(M \geq m|\theta, M_l) \cdot \lambda_{\text{ETAS}}(t, x, y|\theta, \text{seq}_t, M_l) \quad (1)$$

where $P(M \geq m|\theta, M_l)$ or the “marker” is the conditional probability that $M \geq m$ given θ and M_l and $\lambda_{\text{ETAS}}(t, x, y|\theta, \text{seq}_t, M_l)$ is the conditional seismicity rate based on ETAS model given θ , seq_t and M_l . The spatio-temporal triggering effect of a given sequence on the seismicity rate can be seen through $\lambda_{\text{ETAS}}(t, x, y|\theta, \text{seq}_t, M_l)$. Using a truncated exponential (Gutenberg-Richter) model, it is easy to show that $P(M \geq m|\theta, M_l) = \exp(-\beta(m - M_l))$, which is the CCDF (complementary cumulative density function) of an exponential distribution, and β is related to Gutenberg-Richter relation. Equation (1) can be written according to the general ETAS model as follows:

$$\lambda_{\text{ETAS}}(t, x, y, m|\theta, \text{seq}_t, M_l) = \underbrace{\frac{e^{-\beta(m - M_l)}}{P(M \geq m|\theta, M_l)} \cdot \sum_{t_j < t} K e^{\alpha(m_j - M_l)} \cdot \frac{K_t}{(t - t_j + c)^p} \cdot \frac{K_r}{(r_j^2 + d^{*2})^q}}_{\lambda_{\text{ETAS}}(t, x, y|\theta, \text{seq}_t, M_l)} \quad (2)$$

The conditional (triggering) rate $\lambda_{\text{ETAS}}(t, x, y|\theta, \text{seq}_t, M_l)$ in Eq. (2) is a summation taken over every j th event occurred before time t . It is comprised of three terms as follows:

- (a) The productivity function $Ke^{\alpha(m_j - M_i)}$ represents the number of events with magnitudes equal to or greater than M_i (productivity of event m_j ; see¹¹), where α (magnitude⁻¹) determines the exponential increase due to events with magnitudes larger than M_i and K (shocks or events) is the productivity coefficient and measures the intensity of the exponential function in generating triggered aftershocks.
- (b) The typical normalized aftershock time decay law with the functional form $K_t / (t - t_j + c)^p$ is a temporal distribution based on modified Omori (MO) model^{17,72}. Utsu¹⁷ showed that the occurrence rate of aftershocks has a functional form being inversely proportional to the time to the power of p , for which the p value affects the decay rate of aftershocks in time. The parameter c (time, e.g. day) eliminates the singularity issue at $t = t_j$, and partially reflects the effect of incomplete detection of small aftershocks taken place shortly after the mainshock⁷². The parameter K_t is a normalizing coefficient that satisfies the achievement of asymptotic compatibility between ETAS prediction and the long-term seismicity, i.e., integrating the time-dependent term over infinite time will in limit be equal to unity (see also^{11,72,73}):

$$K_t \int_{t_j}^{+\infty} \frac{dt}{(t - t_j + c)^p} = 1 \Rightarrow K_t \left(\frac{c^{1-p}}{p-1} - \frac{(t - t_j + c)^{1-p}}{p-1} \Big|_{t \rightarrow +\infty} \right) = 1 \xrightarrow{p > 1} K_t = (p-1) c^{p-1} \quad (3)$$

It is to note that when $p > 1$, the integral converges at $t \rightarrow +\infty$, while for $p < 1$, it goes to infinity as $t \rightarrow +\infty$. Thus, the condition $p > 1$ is an important constraint in the ETAS parameter estimation. Detailed discussions on other decay models have been furnished in⁷⁴.

- (c) The term $K_r / (r_j^2 + d^{*2})^q$ is the joint (two-dimensional) probability density function of coordinates $(x, y) \in \mathbf{A}$ around the epicenter of the j th event (x_j, y_j) , with r_j the distance between the location (x, y) and the epicenter (x_j, y_j) ; i.e., $r_j^2 = (x - x_j)^2 + (y - y_j)^2$. Herein, the following two functions are used as d^* :

- $d^* = d$: the so-called *simple spatial kernel density* herein is characterized by two parameters d (distance, e.g. km) and q defining the spatial distribution (decay) of the direct offspring triggered by the j th event with epicenter at (x_j, y_j) . This spatial decay footprint model is a function of only $(x - x_j, y - y_j)$ with a circular symmetry around (x_j, y_j) . It has been used by many researchers before (see e.g.,^{12,21,75}). The normalizing coefficient K_r is defined such that integrating the spatial term over infinite space will in limit be equal to unity (see also⁷⁵):

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \frac{K_r dx dy}{((x - x_j)^2 + (y - y_j)^2 + d^2)^q} = K_r \int_0^{+\infty} \frac{2\pi r dr}{[r^2 + d^2]^q} = 1 \xrightarrow{q > 1} K_r = \frac{(q-1)}{\pi} d^{2(q-1)} \quad (4)$$

$$\Rightarrow \frac{K_r}{(r_j^2 + d^{*2})^q} = \frac{(q-1)}{\pi} d^{2(q-1)} \cdot \frac{1}{(r_j^2 + d^2)^q}$$

when $q > 1$, the integral converges at $r \rightarrow +\infty$. Thus, the condition $q > 1$ is another key constraint the ETAS model parameter estimation. It is shown in different studies (see e.g.,^{12,75}) that this spatial model fits the data better than the normal distribution model (being also proposed as a spatial model) expressed by $\frac{1}{2\pi d^2} e^{-r_j^2/2d^2}$. The latter is characterized by one parameter while the model used herein has two parameters.

- $d^* = de^{\gamma m_j}$: the so-called *magnitude-dependent spatial kernel density* herein is a widely used model^{15,76}, has three parameters (d, q , and γ), and is characterized by introducing a decay model that takes into account also the magnitude of the triggered events m_j . The kernel density, that accounts for the observed exponential increase of the rupture dimensions with earthquake magnitude⁷⁷, is expressed as follows:

$$\frac{K_r}{(r_j^2 + d^{*2})^q} = \underbrace{\frac{(q-1)}{\pi} (de^{\gamma m_j})^{2(q-1)}}_{=K_r} \cdot \frac{1}{(r_j^2 + (de^{\gamma m_j})^2)^q} = \frac{(q-1)}{\pi d^2 e^{2\gamma m_j}} \cdot \left(1 + \frac{r_j^2}{d^2 e^{2\gamma m_j}} \right)^{-q} \quad (5)$$

These two spatial kernels are both isotropic. Recent study⁷⁸ shows the superiority of anisotropic spatial kernel especially in the presence of earthquake doublets. With reference to Eq. (2), the ETAS model parameters θ can have seven parameters $\theta = [\beta, K, \alpha, c, p, d, q]$ using the simple spatial kernel density with $d^* = d$, or eight parameters $\theta = [\beta, K, \alpha, c, p, d, q, \gamma]$ employing the magnitude-dependent spatial kernel density with $d^* = de^{\gamma m_j}$. Strictly speaking, the derived parameters K_t (see Eq. (3), function of c and p), and $K_r = \frac{q-1}{\pi} d^{*2(q-1)}$ (see Eq. (4) as function of d and q , or Eq. (5) as a function of d, q and γ) satisfy the achievement of asymptotic compatibility between ETAS predictions and the long-term seismicity. Parameter K , which has a direct relation with the number of events taken place before time t , will be discussed later (see “Calculating K ” section). The rate of events with magnitude (exactly) equal to m , denoted as $\lambda_{\text{ETAS}}(t, x, y, M = m | \theta, \text{seq}_t, M_i)$ herein, is calculated by taking the derivative of Eq. (1) with respect to magnitude m :

$$\lambda_{\text{ETAS}}(t, x, y, M = m | \theta, \text{seq}_t, M_i) = \left| \frac{\partial}{\partial m} \lambda_{\text{ETAS}}(t, x, y, m | \theta, \text{seq}_t, M_i) \right| = \underbrace{\beta e^{-\beta(m - M_i)}}_{p(M=m|\theta, M_i)} \cdot \lambda_{\text{ETAS}}(t, x, y | \theta, \text{seq}_t, M_i) \quad (6)$$

where $p(M = m | \theta, M_l)$ is the probability density function, PDF, of magnitude at m .

Note on the estimation of the completeness magnitude M_c : It is known that aftershocks catalogs compiled immediately following a strong, shallow mainshock are not completely registered^{72,79–82}. Therefore, for providing early forecasts, proper estimation of M_c is a significant component for defining the observed aftershock sequence (observation history) to estimate the rate λ_{ETAS} (see Eq. (2)). It is noted that having a high estimate for M_c would deprive the data of potentially valuable information that are crucial for Bayesian updating of ETAS model parameters (see “Sampling θ from the distribution $p(\theta | \text{seq}, M_l)$ ” section). Herein, we have employed three different strategies for estimation of M_c (see *Supplementary Information* Section SI-3-Results for a detailed discussion): (1) Plotting the frequency-magnitude distribution of events in the observation history (available data at the time of forecasting) in a semi-logarithmic scale and visually detecting M_c as the point where the curve becomes approximately linear after the flat lower magnitude ranges. (2) Employing Bayesian inference for plotting the maximum likelihood of the posterior probability distribution of β versus various magnitude thresholds; M_c can be detected as the magnitude threshold, where the plot of maximum likelihood estimates of β becomes quasi-invariant or reaches its peak value. (3) The semi-logarithmic plot, showing the observed earthquake magnitudes as a function of the time elapsed after the mainshock, is a visual check for the value of M_c estimated by the previous two approaches. This is to ensure that the observed catalog of data at the time of issuing a forecast or the time of interest is complete for magnitudes $\geq M_c$ and does not include small magnitude ranges missed in an early time interval right after the occurrence of the main event.

In principle, M_c can be defined as a function of the mainshock magnitude and the elapsed time since the mainshock (see⁷⁶). Parameter estimation of this time-dependent detection probability function could be incorporated directly into our Bayesian approach. Methods to fit the corresponding parameters of the detection function along with the ETAS parameters are proposed in^{5,83}. However, this would be computationally too expensive in our Bayesian approach at this time and is deferred for future work.

Robust estimation for the number of aftershock events. The conditional number of events, denoted as $N(x, y, m | \theta, \text{seq}, M_l)$, in the spatial cell unit centered at (x, y) with magnitude greater than or equal to m in the forecasting interval $[T_{start}, T_{end}]$ can be calculated as:

$$\begin{aligned}
 N(x, y, m | \theta, \text{seq}, M_l) &= \int_{T_{start}}^{T_{end}} \left(\underbrace{\mu(x, y, m | \theta, M_l) + \lambda_{ETAS}(t, x, y, m | \theta, \text{seq}, M_l)}_{\lambda(t, x, y, m | \theta, \text{seq}, M_l)} \right) dt = \int_{T_{start}}^{T_{end}} \lambda(t, x, y, m | \theta, \text{seq}, M_l) dt \\
 &= N_b(x, y, m | \theta, M_l) + \int_{T_{start}}^{T_{end}} \lambda_{ETAS}(t, x, y, m | \theta, \text{seq}, M_l) dt
 \end{aligned}
 \tag{7}$$

where $\text{seq} = \{(t_i, x_i, y_i, m_i), T_0 \leq t_i < T_{start}, m_i \geq M_l, i = 1:N_0\}$; t_i is the arrival time for the i th event with magnitude m_i and location $(x_i, y_i) \in \mathbf{A}$; $\mu(x, y, m | \theta, M_l)$ is the time-invariant spatial rate representing the background seismicity of the area showing events with magnitude greater than or equal to m ($M \geq m$) in the cell unit centered at $(x, y) \in \mathbf{A}$. Moreover, $\mu(x, y, m | \theta, M_l)$ can be expressed as the product of the “marker” $P(M \geq m | \theta, M_l)$ and the spatial (inhomogeneous) seismicity term $\mu(x, y | M_l)$: $\mu(x, y, m | \theta, M_l) = e^{-\beta(m-M_l)} \cdot \mu(x, y | M_l)$. The time scale of $\mu(x, y | M_l)$ is the same as λ_{ETAS} (e.g., daily or weekly). N_b represents the average number of events occurred due to the background seismicity with $M \geq m$ in the forecasting interval $[T_{start}, T_{end}]$, and estimated as $N_b(x, y, m | \theta, M_l) = e^{-\beta(m-M_l)} \cdot \mu(x, y | M_l) \cdot (T_{end} - T_{start})$. The term $\lambda(\cdot) = \mu(\cdot) + \lambda_{ETAS}(\cdot)$ defines the total spatio-temporal conditional intensity function as the sum of time-invariant (spatial) background seismicity plus the ETAS seismicity rate. A *robust* estimate^{26,28,53,54,84} of the average number of events in the spatial cell unit centered at (x, y) with magnitude greater than or equal to m in the forecasting interval $[T_{start}, T_{end}]$, denoted as $\mathbb{E}[N(x, y, m | \theta, \text{seq}, M_l)]$, can be calculated over the domain of the model parameters Ω_θ as follows:

$$\mathbb{E}[N(x, y, m | \theta, \text{seq}, M_l)] = \int_{\Omega_\theta} \left(N_b(x, y, m | \theta, M_l) + \int_{T_{start}}^{T_{end}} \lambda_{ETAS}(t, x, y, m | \theta, \text{seq}, M_l) dt \right) \cdot p(\theta | \text{seq}, M_l) d\theta
 \tag{8}$$

where $\mathbb{E}[\cdot]$ denotes the expectation, and $p(\theta | \text{seq}, M_l)$ is the posterior conditional PDF for θ given the seq and M_l . As mentioned above, seq denotes the sequence of events taking place before the beginning of the forecasting interval, i.e. $[T_0, T_{start}]$. Nonetheless, the triggering effect of the events taking place during the forecasting interval $[T_{start}, T_{end}]$ is expected to play a major role²⁸. We denote this sequence as seqq , which is unknown at the time of forecasts and is simulated/generated herein. Let us assume that a plausible seqq is defined as the events within the forecasting interval so that $\text{seqq} = \{(IAT_i, x_i, y_i, m_i), T_{start} \leq t_i \leq T_{end}, m_i \geq M_l\}$, where $IAT_i = t_i - t_{i-1}$ stands for the inter-arrival time. The *robust* estimate for the average number of aftershock events in Eq. (8) should also consider all the plausible sequences of events seqq (i.e., the domain Ω_{seqq}) that can happen during the forecasting time interval:

$$\mathbb{E}[N(x, y, m | \theta, \text{seq}, M_l)] = \int_{\Omega_\theta} \left[N_b(x, y, m | \theta, M_l) + \int_{\Omega_{\text{seqq}}} \left(\int_{T_{start}}^{T_{end}} \lambda_{ETAS}(t, x, y, m | \text{seqq}, \theta, \text{seq}, M_l) dt \right) \cdot p(\text{seqq} | \theta, \text{seq}, M_l) d\text{seqq} \right] \cdot p(\theta | \text{seq}, M_l) d\theta
 \tag{9}$$

where $\lambda_{ETAS}(t, x, y, m|\mathbf{seqg}, \boldsymbol{\theta}, \mathbf{seq}, M_i)$ is the space–time clustering ETAS model considering also the sequence of events taking place within the forecasting interval denoted as \mathbf{seqg} ; $p(\mathbf{seqg}|\boldsymbol{\theta}, \mathbf{seq}, M_i)$ is the PDF for the generated sequence \mathbf{seqg} given that $\boldsymbol{\theta}$ and \mathbf{seq} are known. The integral with respect to time in Eq. (9) is calculated piece-wise by summing over the $[t_{i-1}, t_i]$ sub-intervals to span the entire interval $[T_{start}, T_{end}]$:

$$\begin{aligned} \int_{t_{i-1}}^{t_i} \lambda_{ETAS}(t, x, y, m|\mathbf{seqg}_{i-1}, \boldsymbol{\theta}, \mathbf{seq}, M_i) dt &= e^{-\beta(m-M_i)} \sum_{j: t_j < t_i} K e^{\alpha(m_j-M_i)} \cdot K_t \int_{t_{i-1}}^{t_i} \frac{dt}{(t-t_j+c)^p} \cdot \frac{K_r}{(r_j^2+d^{*2})^q} \\ &= e^{-\beta(m-M_i)} \sum_{j: t_j < t_i} K e^{\alpha(m_j-M_i)} \cdot K_t I_t(t_i, t_{i-1}, t_j) \cdot \frac{K_r}{(r_j^2+d^{*2})^q} \end{aligned} \tag{10}$$

where \mathbf{seqg}_{i-1} is the generated sequence up to the $(i-1)$ th event, and the sequence of events that precede the i th generated event is $\{\mathbf{seq}, \mathbf{seqg}_{i-1}\}$. The integral over time I_t is calculated as:

$$I_t(t_i, t_{i-1}, t_j) = \int_{t_{i-1}}^{t_i} \frac{dt}{(t-t_j+c)^p} = \frac{1}{1-p} [(t_i-t_j+c)^{1-p} - (t_{i-1}-t_j+c)^{1-p}]; \quad p > 1 \tag{11}$$

The term “robust” herein implies that the set of all possible model parameters is used to estimate the conditional number of events $N(x, y, m|\mathbf{seq}, M_i)$ rather than a single nominal model parameter. Equation (9) can be solved via a fully simulation-based framework. First, vector of model parameters $\boldsymbol{\theta}$ are sampled from $p(\boldsymbol{\theta}|\mathbf{seq}, M_i)$; then those samples are used to generate plausible sequences \mathbf{seqg} taking place within the forecasting interval $[T_{start}, T_{end}]$. The sequence of events that precede T_{end} is $\{\mathbf{seq}, \mathbf{seqg}\}$, where \mathbf{seq} remains unchanged (observed data) among plausible samples. In the next two sections, we are going to discuss how samples are drawn from the probability distributions $p(\boldsymbol{\theta}|\mathbf{seq}, M_i)$ (“Sampling $\boldsymbol{\theta}$ from the distribution $p(\boldsymbol{\theta}|\mathbf{seq}, M_i)$ ” section) and $p(\mathbf{seqg}|\boldsymbol{\theta}, \mathbf{seq}, M_i)$ (“Generating sequences according to $p(\mathbf{seqg}|\boldsymbol{\theta}, \mathbf{seq}, M_i)$ ” section) through a simulation-based workflow.

The probability that an event exceeding a given magnitude level m occurs within the aftershock zone in the forecasting interval can be calculated as follows (see also^{28,48}):

$$P(M \geq m) = 1 - \exp(-\mathbb{E}[N(M \geq m|\boldsymbol{\theta}, \mathbf{seq}, M_i)]) = 1 - \exp\left(-\int_{x,y \in \mathbf{A}} \mathbb{E}[N(x, y, m|\boldsymbol{\theta}, \mathbf{seq}, M_i)] dx dy\right) \tag{12}$$

where $\mathbb{E}[N(x, y, m|\boldsymbol{\theta}, \mathbf{seq}, M_i)]$ is calculated from Eq. (9).

Calculating K . The coefficient K in calculating λ_{ETAS} (see Eq. 2) is calibrated so that the number of events with magnitude greater than or equal to M_i taking place in time interval $[T_o, T_{start})$ over the whole aftershock zone \mathbf{A} is equal to observed number N_o based on \mathbf{seq} (see also²⁸):

$$\int_{T_o}^{T_{start}} \iint_{(x,y) \in \mathbf{A}} \lambda(t, x, y|\boldsymbol{\theta}, \mathbf{seq}, M_i) dx dy dt = N_o \tag{13}$$

where the term $\lambda(t, x, y|\boldsymbol{\theta}, \mathbf{seq}, M_i)$ is the conditional rate, and according to Eq. (7), we have:

$$\begin{aligned} \int_{T_o}^{T_{start}} \iint_{x,y \in \mathbf{A}} \lambda(t, x, y|\boldsymbol{\theta}, \mathbf{seq}) dx dy dt &= \int_{T_o}^{T_{start}} \iint_{x,y \in \mathbf{A}} [\mu(x, y|M_i) + \lambda_{ETAS}(t, x, y|\boldsymbol{\theta}, \mathbf{seq}, M_i)] dx dy dt \\ &= N_{bo} + \int_{T_o}^{T_{start}} \iint_{x,y \in \mathbf{A}} \lambda_{ETAS}(t, x, y|\boldsymbol{\theta}, \mathbf{seq}, M_i) dx dy dt \end{aligned} \tag{14}$$

where N_{bo} is the number of events due to the background rate $\mu(x, y|M_i)$ with magnitude greater than or equal to M_i taking place in time interval $[T_o, T_{start})$ over the whole aftershock zone \mathbf{A} . The second (triple) integral term in Eq. (14) is calculated piece-wise by summing over the sub-intervals $[t_{i-1}, t_i]$ (where $i=2:N_o$) and the last interval $[t_{N_o}, T_{start}]$ (where t_{N_o} is the arrival time of the event N_o) as follows (see also Eq. (10) and Eq. (11); $\lambda_{ETAS}=0$ in the interval $[T_o, t_1]$):

$$\begin{aligned} \int_{T_o}^{T_{start}} \iint_{x,y \in \mathbf{A}} \lambda_{ETAS}(t, x, y|\mathbf{seq}, M_i) dx dy dt &= \sum_{i=2}^{N_o} \int_{t_{i-1}}^{t_i} \iint_{x,y \in \mathbf{A}} \lambda_{ETAS}(t, x, y|\boldsymbol{\theta}, \mathbf{seq}, M_i) dt + \int_{t_{N_o}}^{T_{start}} \iint_{x,y \in \mathbf{A}} \lambda_{ETAS}(t, x, y|\boldsymbol{\theta}, \mathbf{seq}, M_i) dt \\ &= \sum_{i=2}^{N_o} \sum_{j: t_j < t_i} K e^{\alpha(m_j-M_i)} \cdot K_t I_t(t_i, t_{i-1}, t_j) \cdot K_r I_r(x_j, y_j) + \sum_{j: t_j < T_{start}} K e^{\alpha(m_j-M_i)} \cdot K_t I_t(T_{start}, t_{N_o}, t_j) \cdot K_r I_r(x_j, y_j) \end{aligned} \tag{15}$$

where,

$$I_r(x_j, y_j) = \iint_{x,y \in A} \frac{dx dy}{(r_j^2 + d^{*2})^q} \tag{16}$$

Therefore, using Eq. (14) and Eq. (15), K can be derived as follows:

$$K = \frac{1}{K_t K_r} \cdot \frac{N_o - N_{bo}}{\sum_{i=2}^{N_o} \sum_{j: t_j < t_i} e^{\alpha(m_j - M_i)} I_t(t_i, t_{i-1}, t_j) I_r(x_j, y_j) + \sum_{j: t_j < T_{start}} e^{\alpha(m_j - M_i)} I_t(T_{start}, t_{N_o}, t_j) I_r(x_j, y_j)} \tag{17}$$

The integral over the whole aftershock zone A , denoted as I_r in Eq. (16), should be solved numerically. However, I_r can be approximated with integration over infinite space (denoted as \tilde{I}_r); thus, it can be shown (see Eq. 4) that $K_r \tilde{I}_r = 1$. Schoenberg⁷¹ has shown that the assumption of an infinite spatial domain has negligible effect and so this approximation can be used to reduce the computational cost of evaluating the triple integral (double integral on spatial domain and another integral on time domain). This formal assumption was employed in many research efforts (e.g.,^{28,50}). However, this approximation may cause significant errors when the considered aftershock zone A is not large enough to satisfy Eq. (4). Parameter K can be inferred based on two different approaches:

- **Calculate K :** The parameter K has a derived distribution based on Eq. (17) as a function of other ETAS model parameters. That is, the vector of model parameters $\theta = [\beta, \alpha, c, p, d, q]$ has six parameters considering simple spatial kernel density, and the vector $\theta = [\beta, \alpha, c, p, d, q, \gamma]$ has seven parameters employing the magnitude-dependent spatial kernel density.
- **Learn K :** Parameter K is learned within the Bayesian updating framework. In this approach, $\theta = [\beta, K, \alpha, c, p, d, q]$ using the simple spatial kernel density, $\theta = [\beta, K, \alpha, c, p, d, q, \gamma]$ employing the magnitude-dependent spatial kernel density.

Both methods for characterizing K are examined and compared herein for the case-study (see Sect. [Sensitivity analyses, Results](#)).

Sampling θ from the distribution $p(\theta|\text{seq}, M_i)$. The probability distribution $p(\theta|\text{seq}, M_i)$ is calculated using Bayesian parameter estimation (for more detail see²⁸):

$$\underbrace{p(\theta|\text{seq}, M_i)}_{\text{posterior}} = \frac{p(\text{seq}|\theta, M_i)p(\theta|M_i)}{\int_{\Omega_\theta} p(\text{seq}|\theta, M_i)p(\theta|M_i)d\theta} = C^{-1} \underbrace{p(\text{seq}|\theta, M_i)}_{\text{likelihood}} \underbrace{p(\theta|M_i)}_{\text{prior}} \tag{18}$$

where C^{-1} is a normalizing constant; $p(\text{seq}|\theta, M_i)$ is the likelihood of the observed sequence given the vector of model parameters θ and lower cut-off magnitude M_i , $p(\theta|M_i)$ is the prior distribution for the vector θ . The prior joint distribution $p(\theta|M_i)$ can be estimated as the product of marginal lognormal PDFs for each model parameter (i.e., a multivariate lognormal distribution with zero correlation between the pairs of model parameters θ) whose central statistics (median) can be assigned based on the regional models; thus,

$$p(\theta|M_i) = \frac{1}{\left(\prod_{k=1}^n \theta_k\right) \sqrt{(2\pi)^n |\mathbf{S}|}} \exp\left(-\frac{1}{2} (\ln \theta - \mu_{\ln \theta})^T \mathbf{S}^{-1} (\ln \theta - \mu_{\ln \theta})\right) \tag{19}$$

where n is the number of uncertain parameters in the vector $\theta = \{\theta_k, k = 1:n\}$; $\mu_{\ln \theta}$ is the vector of the mean value of $\ln \theta$ (= logarithm of the vector of median values) associated with the prior distribution; \mathbf{S} is the covariance matrix. In order to sample from the posterior distribution $p(\theta|\text{seq}, M_i)$ in Eq. (18), Markov Chain Monte Carlo (MCMC) simulation routine is employed. MCMC is particularly useful for drawing samples from the target posterior PDF $p(\theta|\text{seq}, M_i)$, while it is known up to a scaling constant C^{-1} (see⁵⁴). Although Eq. (18) looks daunting, we only need un-normalized PDFs in MCMC, as discussed in *Supplementary Information* (Section SI-1-Method). The MCMC routine here employs the Metropolis–Hastings (MH) algorithm^{85,86} in order to generate samples from the target probability distribution $p(\theta|\text{seq}, M_i)$, and later to estimate the robust estimation based on Eq. (9). The MH algorithm generates a Markov chain that produces a sequence of samples $[\theta_1 \rightarrow \theta_2 \rightarrow \dots \rightarrow \theta_i \rightarrow \dots]$, where θ_i represents the state of Markov chain at i th iteration (the first few samples are often discarded to reduce the initial transient effect). It can be shown that the samples from the chain after the initial transient ones reflect samples from the target posterior distribution $p(\theta|\text{seq}, M_i)$. To improve the rate of convergence of the simulation process, we used an adaptive MH algorithm (as proposed in⁵⁴) herein that introduces a sequence of intermediate candidate evolutionary PDFs that resemble more and more the target PDF. The Metropolis–Hastings routine and its adaptive version are described in detail in Section SI-1-Method.

Calculating the likelihood of the observed sequence $p(\text{seq}|\theta, M_i)$. Figure 11 shows the observed sequence, $\text{seq} = \{(t_i, x_i, y_i, m_i), T_o \leq t_i < T_{start}, m_i \geq M_b, i = 1:N_o\}$, with N_o events where $i = 1$ is assigned to the first event (the schematic plot in Fig. 11 shows only the time scale). The likelihood of observing the sequence can be calculated as follows:

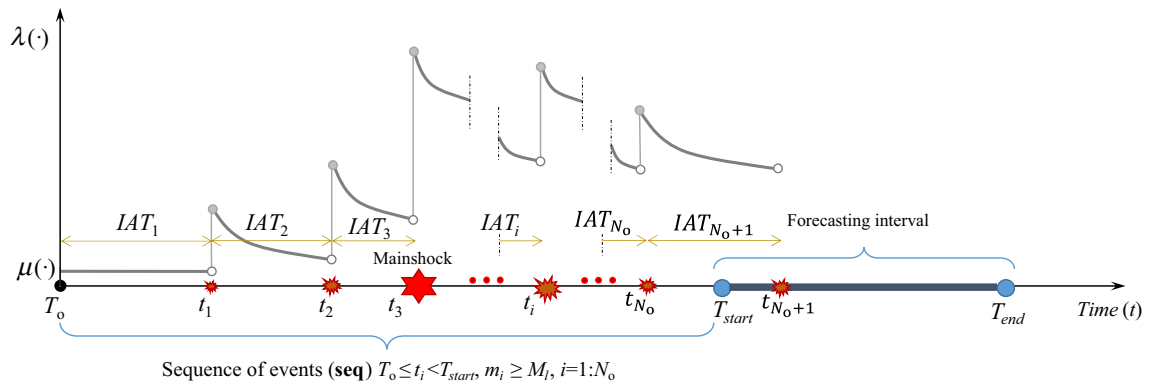


Figure 11. The main parameters required for constructing the likelihood function; the seismicity rate as a function of time showing how each event provokes a jump and is followed by a decay; and the illustration of **seq**.

$$\begin{aligned}
 p(\mathbf{seq}|\boldsymbol{\theta}, M_I) &= \left(\prod_{i=1}^{N_0} p(IAT_i, x_i, y_i, m_i | \boldsymbol{\theta}, \mathbf{seq}, M_I) \right) \cdot P(IAT_{N_0+1} \geq T_{start} - t_{N_0} | \boldsymbol{\theta}, \mathbf{seq}, M_I) \\
 &= \left(\prod_{i=1}^{N_0} p(m_i | \boldsymbol{\theta}, M_I) \right) \cdot \left(\prod_{i=1}^{N_0} p(IAT_i, x_i, y_i | \boldsymbol{\theta}, \mathbf{seq}, M_I) \right) \cdot P(IAT_{N_0+1} \geq T_{start} - t_{N_0} | \boldsymbol{\theta}, \mathbf{seq}, M_I)
 \end{aligned}
 \tag{20}$$

where $p(IAT_i, x_i, y_i, m_i | \boldsymbol{\theta}, \mathbf{seq}, M_I)$ is the conditional probability of observing event i within the **seq** with inter-arrival time $IAT_i = t_i - t_{i-1}$, magnitude equal to m_i and epicenter location $(x_i, y_i) \in \mathbf{A}$. The CCDF $P(IAT_{N_0+1} \geq T_{start} - t_{N_0} | \boldsymbol{\theta}, \mathbf{seq}, M_I)$ is the conditional probability that the $(N_0 + 1)$ th event takes place out of the time window $[t_{N_0}, T_{start}]$, where t_{N_0} is the arrival time of the N_0 th event (see Fig. 11). As shown in Eq. (6), the probability of observing an event with magnitude m_i , denoted as $p(m_i | \boldsymbol{\theta}, M_I)$, is history-independent. Thus, we can de-couple the probability of observing the magnitude m_i , $p(m_i | \boldsymbol{\theta}, M_I)$ and the conditional probability $p(IAT_i, x_i, y_i | \boldsymbol{\theta}, \mathbf{seq}, M_I)$ as shown in Eq. (20). In Fig. 11, the total conditional time- and history-dependent intensity function $\lambda(\cdot)$, as derived in Eq. (6), is shown with gray colored curves. Visibly, the value of $\lambda(\cdot)$ shows a spike right after the occurrence of an event and decays within the inter-arrival time to the next event. In the time interval $IAT_1 (= t_1 - T_0)$, the seismicity rate is constant and equal to the time-invariant spatial background seismicity $\mu(x, y | M_I)$ as shown in Fig. 11 (before T_0 , we consider only background seismicity; see also “Robust estimation for the number of aftershock events” section). We assume an exponential distribution for probability distribution of IAT_i consistent with a non-homogeneous Poisson point process. Thus, Eq. (20) can be written as:

$$\begin{aligned}
 p(\mathbf{seq}|\boldsymbol{\theta}, M_I) &= \left(\prod_{i=1}^{N_0} \beta e^{-\beta(m_i - M_I)} \right) \cdot \left(\mu(x_1, y_1 | M_I) \cdot e^{-\int_{T_0}^{t_1} \int_{x,y \in \mathbf{A}} \mu(x,y | M_I) dx dy dt} \right) \\
 &\cdot \left(\prod_{i=2}^{N_0} \lambda(t_i, x_i, y_i | \boldsymbol{\theta}, \mathbf{seq}, M_I) \cdot e^{-\int_{t_{i-1}}^{t_i} \int_{x,y \in \mathbf{A}} \lambda(t,x,y | \boldsymbol{\theta}, \mathbf{seq}) dx dy dt} \right) \cdot e^{-\int_{t_{N_0}}^{T_{start}} \int_{x,y \in \mathbf{A}} \lambda(t,x,y | \boldsymbol{\theta}, \mathbf{seq}) dx dy dt}
 \end{aligned}
 \tag{21}$$

Equation (21) is derived considering that $\lambda(t_i, x_i, y_i | \boldsymbol{\theta}, \mathbf{seq}, M_I) = \mu(x_i, y_i | M_I)$ in the time interval $[T_0, t_1]$, and that $P(IAT_{N_0+1} \geq T_{start} - t_{N_0} | \boldsymbol{\theta}, \mathbf{seq}, M_I)$ is the CCDF of the exponential distribution of IAT_{N_0+1} . The *likelihood function* can be expressed as follows by combining the exponential terms in Eq. (21):

$$p(\mathbf{seq}|\boldsymbol{\theta}, M_I) = \mu(x_1, y_1 | M_I) \cdot \left(\prod_{i=1}^{N_0} \beta e^{-\beta(m_i - M_I)} \right) \cdot \left(\prod_{i=2}^{N_0} \lambda(t_i, x_i, y_i | \boldsymbol{\theta}, \mathbf{seq}, M_I) \right) \cdot e^{-\int_{T_0}^{T_{start}} \int_{x,y \in \mathbf{A}} \lambda(t,x,y | \boldsymbol{\theta}, \mathbf{seq}, M_I) dx dy dt}
 \tag{22}$$

where,

$$\lambda(t_i, x_i, y_i | \boldsymbol{\theta}, \mathbf{seq}, M_I) = \mu(x_i, y_i | M_I) + \sum_{j: t_j < t_i} K e^{\alpha(m_j - M_I)} \cdot \frac{K_t}{(t_i - t_j + c)^p} \cdot \frac{K_r}{(r_{ji}^2 + d^2)^q}
 \tag{23}$$

where r_{ji} indicates the distance of the i th event with respect to the previously occurred j th event ($t_j < t_i$). With reference to the Bayes formula in Eq. (18), the term $\mu(x_i, y_i | M_I)$ is independent of any realization of the vector of model parameters $\boldsymbol{\theta}$; hence, this value can be eliminated from both nominator and denominator. Moreover, the MH algorithm employs the ratio of the likelihoods (see Section SI-1-Method), and constant terms can easily be disregarded. Thus, an estimator for the likelihood can be defined as follows (within the context of Bayesian inference):

$$p(\mathbf{seq}|\boldsymbol{\theta}, M_I) \approx \left(\prod_{i=1}^{N_o} \beta e^{-\beta(m_i - M_I)} \right) \cdot \left(\prod_{i=2}^{N_o} \lambda(t_i, x_i, y_i|\boldsymbol{\theta}, \mathbf{seq}, M_I) \right) \cdot e^{-\int_{T_o}^{T_{start}} \int_{x,y \in \mathbf{A}} \lambda(t, x, y|\boldsymbol{\theta}, \mathbf{seq}, M_I) dx dy dt} \tag{24}$$

With reference to “Calculating K ” section, we have two approaches to follow:

- In case that we want to calculate parameter K , as noted in “Calculating K ” section (*Calculate K*), the power of the exponential term in Eq. (24) becomes N_o (see Eq. 13); thus, the constant term e^{-N_o} can also be eliminated while employing the Bayesian updating framework in Eq. (18) (see also Section SI-1-Method). In this way, the likelihood estimator in Eq. (24) can be simplified as the product of the probabilities of observing m_i , for $i = 1:N_o$, and those of the total rate $\lambda(t_i, x_i, y_i|\boldsymbol{\theta}, \mathbf{seq}, M_I)$ for $i = 2:N_o$.
- If we learn parameter K within the MH algorithm (see “Calculating K ” section, *Learn K*), the triple integral (the power of the exponential term) in Eq. (24) can be solved based on Eq. (14) and Eq. (15). Even in this case, we can exclude the constant term e^{-N_o} from the likelihood estimator (as it is independent of any realization of $\boldsymbol{\theta}$). Thus, Eq. (15) is employed to estimate the triple integral of λ_{ETAS} over the time interval $[T_o, T_{start}]$. As shown in Eq. (15), it is solved piece-wise by summing over the sub-intervals $[t_{i-1}, t_i]$ (where $i = 2:N_o$), and $[t_{N_o}, T_{start}]$.

Generating sequences according to $p(\mathbf{seq}|\boldsymbol{\theta}, \mathbf{seq}, M_I)$. This section describes how the catalogue of stochastic events \mathbf{seq} , which will occur during the forecasting interval, is simulated. The event i in the generated catalogue is identified by the arrival time $t_i = t_{i-1} + IAT_i$, the Cartesian coordinates (x_i, y_i) of the epicenter, and the magnitude m_i . The sequence of events in the generated catalogue up to event i is denoted by \mathbf{seq}_{i-1} . Thus, $\mathbf{seq}_i = \{\mathbf{seq}_{i-1}, (IAT_i, x_i, y_i, m_i)\}$. Catalogue simulation continues until the arrival time of the simulated events do not fall outside the forecasting interval (i.e., $t_i \leq T_{end}$). The likelihood $p(\mathbf{seq}|\boldsymbol{\theta}, \mathbf{seq}, M_I)$ for the generated catalogue \mathbf{seq} can be calculated using the rule of product in probability; that is conditioning the occurrence of i th event on the previous $i-1$ events:

$$p(\mathbf{seq}|\boldsymbol{\theta}, \mathbf{seq}, M_I) = \prod_i p(IAT_i, x_i, y_i, m_i|\mathbf{seq}_{i-1}, \boldsymbol{\theta}, \mathbf{seq}, M_I) \tag{25}$$

where $p(IAT_i, x_i, y_i, m_i|\mathbf{seq}_{i-1}, \boldsymbol{\theta}, \mathbf{seq}, M_I)$ is the conditional probability of observing event i . It is noted that the sequence of events that precede the i th generated event is $\{\mathbf{seq}, \mathbf{seq}_{i-1}\}$. For event $i = 1$ in Eq. (25), $\mathbf{seq}_{i-1} = \mathbf{seq}_0 = \emptyset$. The probability distribution $p(IAT_i, x_i, y_i, m_i|\mathbf{seq}_{i-1}, \boldsymbol{\theta}, \mathbf{seq}, M_I)$ can be further expanded (again using the probability product rule) as follows:

$$p(IAT_i, x_i, y_i, m_i|\mathbf{seq}_{i-1}, \boldsymbol{\theta}, \mathbf{seq}, M_I) = p(m_i|\boldsymbol{\theta}, M_I) \cdot p(IAT_i|m_i, \mathbf{seq}_{i-1}, \boldsymbol{\theta}, \mathbf{seq}, M_I) \cdot p(x_i, y_i|IAT_i, m_i, \mathbf{seq}_{i-1}, \boldsymbol{\theta}, \mathbf{seq}, M_I) \tag{26}$$

where $p(m_i|\boldsymbol{\theta}, M_I)$ is the conditional marginal PDF for the magnitude m_i given the model parameters $\boldsymbol{\theta}$, and M_I ; $p(IAT_i|m_i, \mathbf{seq}_{i-1}, \boldsymbol{\theta}, \mathbf{seq}, M_I)$ is the conditional marginal PDF for inter-arrival time IAT_i given the history of past events $\{\mathbf{seq}, \mathbf{seq}_{i-1}\}$ and given that the value of magnitude is equal to m_i ; finally, $p(x_i, y_i|IAT_i, m_i, \mathbf{seq}_{i-1}, \boldsymbol{\theta}, \mathbf{seq}, M_I)$ is the conditional joint PDF for the spatial position $(x_i, y_i) \in \mathbf{A}$ given that IAT_i and m_i are known. The breakdown into the product of several conditional PDFs is necessary as we need to sample m_i , IAT_i , and (x_i, y_i) directly from the three consecutive probability terms, respectively. To generate a plausible sequence of events during the forecasting interval $[T_{start}, T_{end}]$, the procedure, illustrated originally in²⁸, is adopted with some modifications herein and described in the following sections.

Sampling magnitude, m . The magnitude m_i of the i th event within \mathbf{seq} is simulated according to a truncated Exponential PDF with rate β , denoted as $p(M = m_i|\boldsymbol{\theta}, M_I)$ in Eq. (6). The Exponential cumulative density function (CDF) used to generate m has the upper bound magnitude threshold M_{max} :

$$P(M \leq m|M_I \leq M \leq M_{max}) = \frac{1 - e^{-\beta(m - M_I)}}{1 - e^{-\beta(M_{max} - M_I)}} \tag{27}$$

To sample from this CDF, generate a uniform random number $u \sim \text{Uniform}(0, 1)$:

$$m_i = \frac{-1}{\beta} \cdot \ln \left[1 - u \cdot \left(1 - e^{-\beta(M_{max} - M_I)} \right) \right] + M_I \tag{28}$$

If the upper bound threshold M_{max} is not considered ($M_{max} \rightarrow +\infty$), Eq. (28) can be modified by eliminating the term $1 - \exp(-\beta(M_{max} - M_I))$ in the parenthesis. It is to note that we have considered the upper-bound threshold M_{max} just for generating the sequences within \mathbf{seq} . The ETAS rate (see Eqs. 2 and 6) has no magnitude upper threshold.

Sampling the inter-arrival time, IAT . To generate the inter-arrival time of the i th event, IAT_i , within \mathbf{seq} (given that its magnitude m_i is already known), the Thinning algorithm⁸⁷ (see also⁵⁷) is employed herein. The thinning algorithm is based on simulating inter-arrival times from a homogeneous Poisson process with sufficiently large intensity, denoted here as λ_{max} , and then thinning out (filtering) the points according to the conditional intensity function denoted as λ_{gen} . The Thinning algorithm is as follows:

Step (1): Calculate the temporal Poisson rate λ_{\max} over the entire aftershock zone at time t_{i-1} having integrated out the spatial term, as follows (see Eq. 7):

$$\begin{aligned} \lambda_{\max} &= \iint_{x,y \in \mathbf{A}} \lambda(t_{i-1}, x, y | \text{seqg}_{i-1}, \boldsymbol{\theta}, \text{seq}, M_i) dx dy \\ &= \iint_{x,y \in \mathbf{A}} (\mu(x, y | M_i) + \lambda_{\text{ETAS}}(t_{i-1}, x, y | \text{seqg}_{i-1}, \boldsymbol{\theta}, \text{seq}, M_i)) dx dy \\ &= \mu_b + \sum_{j: t_j < t_{i-1}} K e^{\alpha(m_j - M_i)} \cdot \frac{K_t}{(t_{i-1} - t_j + c)^p} \cdot K_r \underbrace{\left(\iint_{x,y \in \mathbf{A}} \frac{1}{(r_j^2 + d^2)^q} dx dy \right)}_{I_r(x_j, y_j)} \\ &= \mu_b + \sum_{j: t_j < t_{i-1}} \frac{K K_t K_r \cdot e^{\alpha(m_j - M_i)} \cdot I_r(x_j, y_j)}{(t_{i-1} - t_j + c)^p} \end{aligned} \tag{29}$$

where μ_b is the background rate over the whole area \mathbf{A} . It is noted that the ETAS rate is estimated by summing up the rate of events taken place before t_{i-1} . For generating the first event, t_{i-1} is set to T_{start} .

Step (2): Generate the inter-arrival time IAT_{gen} from a homogeneous Exponential CDF with the rate λ_{\max} expressed as $F(IAT) = 1 - \exp(-\lambda_{\max} \cdot IAT)$. Since the inverse function $F(IAT)^{-1}$ has a closed-form expression, we use the inverse transform sampling technique. To sample from this CDF, first generate a uniform random number $u \sim \text{Uniform}(0, 1)$; the generated IAT_{gen} can analytically be drawn as $IAT_{\text{gen}} = -\ln(1-u) / \lambda_{\max}$.

Step (3): Calculate the Poisson rate at time $t_{\text{gen}} = t_{i-1} + IAT_{\text{gen}}$, denoted by λ_{gen} , from Eq. (29) by substituting t_{i-1} with t_{gen} . It is noted that the generated magnitude m_i in the previous section has no effect on the rate λ_{gen} , as this rate is calculated based on events taking place before t_{gen} . The rate λ_{gen} should always be equal to or smaller than λ_{\max} ; i.e., $\lambda_{\text{gen}} \leq \lambda_{\max}$. If this condition does not hold, go to Step (5).

Step (4): Accept t_{gen} with the probability $pr = \lambda_{\text{gen}} / \lambda_{\max}$ by first generating a uniform random number $u \sim \text{Uniform}(0, 1)$; if $u \leq pr$, then the time t_{gen} is accepted, and thus $t_i = t_{\text{gen}}$. The procedure continues by generating the next inter-arrival time until $t_i > T_{\text{end}}$. λ_{\max} is going to change while generating each new t_i .

Step (5): Reject t_{gen} if $u > pr$ or $\lambda_{\text{gen}} > \lambda_{\max}$ (i.e., $pr = \lambda_{\text{gen}} / \lambda_{\max} > 1$). Let us denote the rejected t_{gen} as $t_{\text{gen}}^{(-)}$ (in order to keep track of it for the next simulation). In the case of rejection, the procedure continues by sampling a new inter-arrival time IAT_{gen} from the homogeneous Exponential PDF with rate λ_{gen} . The new generated arrival time is calculated as $t_{\text{gen}} = t_{\text{gen}}^{(-)} + IAT_{\text{gen}}$, $t_{\text{gen}} \leq T_{\text{end}}$. The quantities λ_{gen} and pr are calculated again to test whether the newly generated inter-arrival time is accepted or rejected.

Based on the above discussion, it is noted that $p(IAT_i | m_i, \text{seqg}_{i-1}, \boldsymbol{\theta}, \text{seq}, M_i) = p(IAT_i | \text{seqg}_{i-1}, \boldsymbol{\theta}, \text{seq}, M_i)$; i.e., m_i is independent of IAT_i .

Generating the epicentral coordinate, (x, y) . The spatial coordinates are simulated from a spatial kernel $p(x, y | IAT_i, m_i, \text{seqg}_{i-1}, \boldsymbol{\theta}, \text{seq}, M_i)$ that is obtained based on the epicenters of the previous events (i.e., $\{\text{seq}, \text{seqg}_{i-1}\}$), and conditioned on the generated m_i and IAT_i . This spatial kernel is a joint PDF of the epicentral (Cartesian) coordinates, from which (x, y) for the i th event can be generated given that the magnitude m_i , the time of occurrence t_i , and the previous $(i-1)$ events within the generated sequence seqg_{i-1} are known (see also²⁸):

$$p(x, y | IAT_i, m_i, \text{seqg}_{i-1}, \boldsymbol{\theta}, \text{seq}, M_i) = \frac{p(IAT_i, x, y, m_i | \text{seqg}_{i-1}, \boldsymbol{\theta}, \text{seq}, M_i)}{\iint_{x,y \in \mathbf{A}} p(IAT_i, x, y, m_i | \text{seqg}_{i-1}, \boldsymbol{\theta}, \text{seq}, M_i) dx dy} \tag{30}$$

In other words, the joint PDF $p(x, y | \cdot)$ in Eq. (30) is estimated for all Cartesian coordinates $(x, y) \in \mathbf{A}$ and is normalized to obtain the kernel. The likelihood in the nominator $p(IAT_i, x, y, m_i | \text{seqg}_{i-1}, \boldsymbol{\theta}, \text{seq}, M_i)$ of Eq. (30) can be calculated as (see also Eq. 20):

$$p(IAT_i, x, y, m_i | \text{seqg}_{i-1}, \boldsymbol{\theta}, \text{seq}, M_i) = \underbrace{\beta e^{-\beta(m_i - M_i)}}_{p(m_i | \boldsymbol{\theta}, M_i)} \cdot \underbrace{\lambda(t_i, x, y | \text{seqg}_{i-1}, \boldsymbol{\theta}, \text{seq}, M_i) \cdot e^{-\int_{t_{i-1}}^{t_i} \lambda(t, x, y | \text{seqg}_{i-1}, \boldsymbol{\theta}, \text{seq}, M_i) dt}}_{p(IAT_i, x, y | \text{seqg}_{i-1}, \boldsymbol{\theta}, \text{seq}, M_i)} \tag{31}$$

where $\lambda(t_i, x, y | \text{seqg}_{i-1}, \boldsymbol{\theta}, \text{seq}, M_i)$ is obtained for each Cartesian coordinate $(x, y) \in \mathbf{A}$ and for all $t < t_i$ (see Eq. 7 and Eq. 29) and the exponential power is derived as (see also Eq. (10) and Eq. (11)):

$$\int_{t_{i-1}}^{t_i} \lambda(t, x, y | \text{seqg}_{i-1}, \boldsymbol{\theta}, \text{seq}, M_i) dt = \mu(x, y | M_i) \cdot IAT_i + \sum_{j: t_j < t_i} K e^{\alpha(m_j - M_i)} \cdot K_t I_t(t_i, t_{i-1}, t_j) \cdot \frac{K_r}{(r_j^2 + d^2)^q} \tag{32}$$

The denominator in Eq. (30) is the sum over all the Cartesian coordinate $(x, y) \in \mathbf{A}$. Moreover, the PDF of observing m_i , i.e. $\beta \exp(-\beta(m_i - M_i))$, can be dropped from both nominator and denominator of Eq. (30) since it is not location-dependent. Thus, it is interesting to mention that $p(x, y | \cdot)$ in Eq. (30) is independent of m_i .

N-test based on the simulation-based Bayesian framework. The N-test is intended to measure (in a probabilistic manner) how well the forecasted number of earthquakes matches the observed number of events⁵⁹. According to this test, we fit a Poisson distribution to the forecasted number of events N_{fore} with magnitude greater than or equal to M_l . N_{fore} is the expected number of events in the forecasting interval integrated over the aftershock zone **A**, as in Eq. (12), $N_{\text{fore}} = \mathbb{E}[N(M \geq M_l | \boldsymbol{\theta}, \text{seq}, M_l)]$. The objective of the N-test is to verify if the observed number of events N_{obs} (with $M \geq M_l$ taking place within the forecasting interval) are not outliers. To this end, one needs to estimate the two probabilities that the number of events $n \leq N_{\text{obs}}$ and $n \geq N_{\text{obs}}$ given N_{fore} , which are denoted as $\delta_{1,\text{Poiss}} = P(n \leq N_{\text{obs}} | N_{\text{fore}})$ and $\delta_{2,\text{Poiss}} = P(n \geq N_{\text{obs}} | N_{\text{fore}})$, respectively. These two probabilities should be greater than a threshold (say 0.025 to reflect 95% confidence interval) to guarantee that N_{obs} will not lie within the tails of our forecast.

The proposed simulation-based workflow provides as output the posterior probability distribution for the forecasted number of events $N(M \geq M_l | \boldsymbol{\theta}, \text{seq}, M_l)$ according to Eq. (9) that is integrated over the aftershock zone **A**. Therefore, it is quite straightforward to compute the percentiles of this distribution (i.e., 50th, 16th, 84th, 2nd and 98th percentiles) to realize how well the forecasted number of earthquakes matches N_{obs} . It is possible to estimate empirically the probabilities $\delta_{1,\text{Bayes}} = P(n \leq N_{\text{obs}} | N_{\text{fore}})$ and $\delta_{2,\text{Bayes}} = P(n \geq N_{\text{obs}} | N_{\text{fore}})$ as the number of data that are ($\leq N_{\text{obs}}$) and ($\geq N_{\text{obs}}$) divided by the number of samples simulated for $\boldsymbol{\theta}$ (the number of samples $N(M \geq M_l | \boldsymbol{\theta}, \text{seq}, M_l)$ is equal to the number of generated samples of $\boldsymbol{\theta}$).

S-test based on the simulation-based Bayesian framework. The spatial distribution of forecast and observation can be compared based on the S-test^{58,59}, which provides an overall measure of how well the forecast in single cell units (associated with the mesh grid of the space **A** of the aftershock zone centered at x and y) matches the observations in single cell units. According to this test, we calculate the log-likelihood of observations $N_{\text{obs},n}$ given the forecasts $N_{\text{fore},n}$ in each grid, where $n = 1:N_{\text{grid}}$, assuming that the total number of cell units centered at the Cartesian coordinate $(x, y) \in \mathbf{A}$ is equal to N_{grid} . Consider that $N_{\text{fore}} = \sum_{n=1}^{N_{\text{grid}}} N_{\text{fore},n}$ and $N_{\text{obs}} = \sum_{n=1}^{N_{\text{grid}}} N_{\text{obs},n}$, where $N_{\text{fore},n} = \mathbb{E}[N(x, y, M_l | \boldsymbol{\theta}, \text{seq}, M_l)]$ (see Eq. 9) is the (expected) number of forecasted events with $M \geq M_l$ at the Cartesian coordinate $(x, y) \in \mathbf{A}$, and $N_{\text{obs},n}$ is the number of observed events (with $M \geq M_l$) in the same cell unit within the forecasting interval $[T_{\text{start}}, T_{\text{end}}]$. We isolate the spatial forecasts so that sum of the number of events over the whole grid cells matches the observation; thus, we set $N_{\text{fore},n} = N_{\text{fore},n} \cdot \frac{N_{\text{obs}}}{N_{\text{fore}}}$. We estimate the likelihood S_{obs} as:

$$S_{\text{obs}} = \ln \left(\prod_{n=1}^{N_{\text{grid}}} \frac{(N_{\text{fore},n})^{N_{\text{obs},n}} e^{-N_{\text{fore},n}}}{N_{\text{obs},n}!} \right) = \sum_{n=1}^{N_{\text{grid}}} [-N_{\text{fore},n} + N_{\text{obs},n} \ln(N_{\text{fore},n}) - \ln(N_{\text{obs},n}!)] \quad (33)$$

The S-test is done herein based on the *standard* S-test⁵⁹ and the *Bayesian* S-test. The two methods are described as follows:

1. The standard S-test accounts for the forecast uncertainty by simulating catalogues that are consistent with the forecast. For simulating each catalogue: (a) the number of events in each simulated catalog is normalized to match N_{obs} ; (b) For each event in the previous step, generate a uniform random number $\sim \text{Uniform}(0, 1)$ in order to locate the event in the corresponding spatial cell according to the proportion of $N_{\text{fore},n}$ in the cell units (i.e., using the inverse cumulative distribution function). Stage (b) will lead to construction of $N_{\text{sim},n}$ to be the number of simulated events (with $M \geq M_l$) in the cell unit $(x, y) \in \mathbf{A}$, where $n = 1:N_{\text{grid}}$ considering that $N_{\text{obs}} = \sum_{n=1}^{N_{\text{grid}}} N_{\text{sim},n}$. Repeating stage (b) many times (say e.g. 500 to 1000 iterations) leads to the construction of forecast-consistent simulated catalogues of events. Finally, for each simulated catalogue, the log-likelihood S_{sim} of $N_{\text{sim},n}$ simulations given the forecasts $N_{\text{fore},n}$ in all cell units can be calculated as follows:

$$S_{\text{sim}} = \ln \left(\prod_{n=1}^{N_{\text{grid}}} \frac{(N_{\text{fore},n})^{N_{\text{sim},n}} e^{-N_{\text{fore},n}}}{N_{\text{sim},n}!} \right) = \sum_{n=1}^{N_{\text{grid}}} [-N_{\text{fore},n} + N_{\text{sim},n} \ln(N_{\text{fore},n}) - \ln(N_{\text{sim},n}!)] \quad (34)$$

Repeating the calculation of Eq. (34) for each simulated catalog, we obtain the vector of log-likelihood \mathbf{S}_{sim} .

2. The presented simulation-based workflow furnishes, as a side-product, the Bayesian S-test and forecast-consistent catalogues. Stage (b) in method (1) described above can be accomplished by directly using the normalized spatial seismicity forecasts $[N(x, y, M \geq M_l | \boldsymbol{\theta}, \text{seq}, M_l) \cdot \frac{N_{\text{obs}}}{N_{\text{fore}}}]$ for the cell units based on different realizations of the vector of uncertain parameters $\boldsymbol{\theta}$. Thus, the number of simulated catalogues is equal to the number of generated samples of $\boldsymbol{\theta}$. The log-likelihood S_{sim} can be estimated according to Eq. (34).

This probability $P(S \leq S_{\text{obs}})$ can be estimated as follows:

$$P(S \leq S_{\text{obs}}) = \frac{\mathbb{N}(\mathbf{S}_{\text{sim}} \leq S_{\text{obs}})}{\mathbb{N}(\mathbf{S}_{\text{sim}})} \quad (35)$$

where $\mathbb{N}(\mathbf{S}_{\text{sim}} \leq S_{\text{obs}})$ indicates the number of the components in the vector \mathbf{S}_{sim} that are $\leq S_{\text{obs}}$, and the denominator $\mathbb{N}(\mathbf{S}_{\text{sim}})$ is the total number of elements in \mathbf{S}_{sim} which is equal to the number of simulated catalogues. If S_{obs}

falls in the lower tail of the distribution $P(S \leq S_{\text{obs}})$, this indicates that the observation is not consistent with the forecast in each cell; thus, the forecast is not accurate.

Data availability

The earthquake catalog of Kermanshah seismic sequence used in this study and the MATLAB code developed for this workflow is available on a Git repository hosting service GitHub https://github.com/HossEbi/Bayesian_spatiotemporal_ETAS_model_ver2. The data is also available upon request from the corresponding author.

Received: 1 August 2022; Accepted: 9 November 2022

Published online: 05 December 2022

References

- Gerstenberger, M. C., Wiemer, S., Jones, L. M. & Reasenberg, P. A. Real-time forecasts of tomorrow's earthquakes in California. *Nature* **435**, 328–331 (2005).
- Jordan, T. H. & Jones, L. M. Operational earthquake forecasting: Some thoughts on why and how. *Seism. Res. Lett.* **81**(4), 571–574 (2010).
- Jordan, T. H. *et al.* Operational earthquake forecasting: State of knowledge and guidelines for implementation. *Ann. Geophys.* **54**(4), 315–391 (2011).
- Jordan, T. H., Marzocchi, W., Michael, A. J. & Gerstenberger, M. C. Operational earthquake forecasting can enhance earthquake preparedness. *Seism. Res. Lett.* **85**(5), 955–959 (2014).
- Omi, T., Ogata, Y., Hirata, Y. & Aihara, K. Forecasting large aftershocks within one day after the main shock. *Sci. Rep.* **3**, 2218 (2013).
- Marzocchi, W., Lombardi, A. M. & Casarotti, E. The establishment of an operational earthquake forecasting system in Italy. *Seism. Res. Lett.* **85**(5), 961–969 (2014).
- Zechar, J. D., Marzocchi, W. & Wiemer, S. Operational earthquake forecasting in Europe: Progress, despite challenges. *Bull. Earthq. Eng.* **14**, 2459–2469 (2016).
- Zhang, L., Werner, M. J. & Goda, K. Spatiotemporal seismic hazard and risk assessment of aftershocks of M 9 megathrust earthquakes. *Bull. Seism. Soc. Am.* **108**(6), 3313–3335 (2018).
- Schorlemmer, D. *et al.* The collaboratory for the study of earthquake predictability: Achievements and priorities. *Seism. Res. Lett.* **89**(4), 1305–1313 (2018).
- Lippiello, E. *et al.* Forecasting of the first hour aftershocks by means of the perceived magnitude. *Nat. Commun.* **10**, 2953 (2019).
- Ogata, Y. Statistical models for earthquake occurrences and residual analysis for point processes. *J. Am. Stat. Assoc.* **83**, 9–27 (1988).
- Ogata, Y. Space-time point-process models for earthquake occurrences. *Ann. Inst. Statist. Math.* **50**(2), 379–402 (1998).
- Zhuang, J., Ogata, Y. & Vere-Jones, D. Stochastic declustering of space-time earthquake occurrences. *J. Am. Stat. Assoc.* **97**(458), 369–380 (2002).
- Zhuang, J., Ogata, Y. & Vere-Jones, D. Analyzing earthquake clustering features by using stochastic reconstruction. *J. Geophys. Res.* **109**, B05301 (2004).
- Ogata, Y. & Zhuang, J. Space-time ETAS models and an improved extension. *Tectonophysics* **413**(1), 13–23 (2006).
- Ogata, Y. Seismicity analysis through point-process modeling: A review. *Pure Appl. Geophys.* **155**(317), 471–507 (1999).
- Utsu, T. A statistical study of the occurrence of aftershocks. *Geophys. Mag.* **30**, 521–605 (1961).
- Console, R., Murru, M., Catalli, F. & Falcone, G. Real time forecasts through an earthquake clustering model constrained by the rate-and-state constitutive law: comparison with a purely stochastic ETAS model. *Seism. Res. Lett.* **78**(1), 49–56 (2007).
- Zhuang, J. Next-day earthquake forecasts for the Japan region generated by the ETAS model. *Earth planets space* **63**(3), 207–216 (2011).
- Marzocchi, W. & Lombardi, A. M. Real-time forecasting following a damaging earthquake. *Geophys. Res. Lett.* **36**, L21302 (2009).
- Lombardi, A. M. & Marzocchi, W. The ETAS model for daily forecasting of Italian seismicity in the CSEP experiment. *Ann. Geophys.* **53**, 155–164 (2010).
- Werner, M. J., Helmstetter, A., Jackson, D. D. & Kagan, Y. Y. High-resolution long-term and short-term earthquake forecasts for California. *Bull. Seism. Soc. Am.* **101**, 1630–1648 (2011).
- Marzocchi, W. & Murru, M. Daily earthquake forecasts during the May–June 2012 Emilia earthquake sequence (northern Italy). *Ann. Geophys.* **55**(4), 561–567 (2012).
- Ebrahimian, H. *et al.* Adaptive daily forecasting of seismic aftershock hazard. *Bull. Seism. Soc. Am.* **104**(1), 145–161 (2014).
- Gerstenberger, M., McVerry, G., Rhoades, D. & Stirling, M. Seismic hazard modeling for the recovery of Christchurch. *Earthq. Spectra* **30**, 17–29 (2014).
- Jalayer, F. & Ebrahimian, H. MCMC-based updating of an epidemiological temporal aftershock forecasting model. *Vulnerab. Uncertain. Risk*, 2093–2103; <https://doi.org/10.1061/9780784413609.210> (2014).
- Rhoades, D. A., Liukis, M., Christophersen, A. & Gerstenberger, M. C. Retrospective tests of hybrid operational earthquake forecasting models for Canterbury. *Geophys. J. Int.* **204**(1), 440–456 (2015).
- Ebrahimian, H. & Jalayer, F. Robust seismicity forecasting based on Bayesian parameter estimation for epidemiological spatio-temporal aftershock clustering models. *Sci. Rep.* **7**, 9803 (2017).
- Field, E. *et al.* A spatiotemporal clustering model for the third Uniform California Earthquake Rupture Forecast (UCERF3-ETAS): Toward an operational earthquake forecast. *Bull. Seism. Soc. Am.* **107**, 1049–1081 (2017).
- Yazdi, P., Hainzl, S. & Gaspar-Escribano, J. M. Statistical analysis of the 2012–2013 Torreyerogil-Sabiote seismic series Spain. *J. Seism.* **21**(4), 705–717 (2017).
- Cattania, C. *et al.* The forecasting skill of physics-based seismicity models during the 2010–2012 Canterbury, New Zealand, earthquake sequence. *Seism. Res. Lett.* **89**(4), 1238–1250 (2018).
- Kourouklas, C., Mangira, O., Iliopoulos, A., Chorozoglou, D. & Papadimitriou, E. A study of short-term spatiotemporal clustering features of Greek seismicity. *J. Seism.* **24**, 459–477 (2020).
- Azarbakht, A., Ebrahimian, H., Jalayer, F. & Douglas, J. Variations in hazard during earthquake sequences between 1995 and 2018 in western Greece as evaluated by a Bayesian ETAS model. *Geophys. J. Int.* **231**(1), 27–46 (2022).
- Darzi, A. *et al.* Calibration of a Bayesian spatio-temporal ETAS model to the June 2000 South Iceland seismic sequence. *Geophys. J. Int.* **232**(2), 1236–1258 (2023).
- Ebrahimian, H. *et al.* A performance-based framework for adaptive seismic aftershock risk assessment. *Earthq. Eng. Struct. Dyn.* **43**(14), 2179–2197 (2014).
- Yaghmaei-Sabegh, S., Shoaefar, P. & Shoaefar, N. Probabilistic seismic-hazard analysis including earthquake clusters. *Bull. Seism. Soc. Am.* **107**, 2367–2379 (2017).
- Jalayer, F., Asprone, D., Prota, A. & Manfredi, G. A decision support system for post-earthquake reliability assessment of structures subjected to aftershocks: an application to L'Aquila earthquake, 2009. *Bull. Earthq. Eng.* **9**(4), 997–1014 (2011).

38. Iervolino, I. *et al.* Operational (short-term) earthquake loss forecasting in Italy. *Bull. Seism. Soc. Am.* **105**, 2286–2298 (2015).
39. Jalayer, F. & Ebrahimiyan, H. Seismic risk assessment considering cumulative damage due to aftershocks. *Earthq. Eng. Struct. Dyn.* **46**(3), 369–389 (2016).
40. Field, E., Porter, K. & Milner, K. A prototype operational earthquake loss model for California based on UCERF3-ETAS—A first look at valuation. *Earthq. Spectra* **33**, 1279–1299 (2017).
41. Veen, A. & Schoenberg, F. P. Estimation of space-time branching process models in seismology using an EM-type algorithm. *J. Am. Stat. Assoc.* **103**, 614–624 (2008).
42. Lippiello, E., Giacco, F., de Arcangelis, L., Marzocchi, W. & Godano, C. Parameter estimation in the ETAS model: Approximations and novel methods. *Bull. Seism. Soc. Am.* **104**(2), 985–994 (2014).
43. Bottiglieri, M., Lippiello, E., Godano, C. & de Arcangelis, L. Comparison of branching models for seismicity and likelihood maximization through simulated annealing. *J. Geophys. Res. Solid Earth* **116**, B02303 (2011).
44. Lombardi, A. M. Estimation of the parameters of ETAS models by simulated annealing. *Sci. Rep.* **5**, 8417 (2015).
45. Omi, T., Ogata, Y., Hirata, Y. & Aihara, K. Intermediate-term forecasting of aftershocks from an early aftershock sequence: Bayesian and ensemble forecasting approaches. *J. Geophysical Res: Solid Earth* **120**(4), 2561–2578 (2015).
46. Seif, S., Mignan, A., Zechar, J. D., Werner, M. J., Wiemer, S. Estimating ETAS: The effects of truncation, missing data, and model assumptions. *J. Geophys. Res.: Solid Earth* **122**(1), 449–469 (2017).
47. Hainzl, S. ETAS-approach accounting for short-term incompleteness of earthquake catalogs. *Bull. Seism. Soc. Am.* **112**(1), 494–507 (2022).
48. Shcherbakov, R., Zhuang, J., Zöller, G. & Ogata, Y. Forecasting the magnitude of the largest expected earthquake. *Nat. Commun.* **10**, 4051 (2019).
49. Shcherbakov, R. Statistics and forecasting of aftershocks during the 2019 Ridgecrest, California, earthquake sequence. *J. Geophys. Res.: Solid Earth* **126**, 2 (2021).
50. Ross, G. J. Bayesian estimation of the ETAS model for earthquake occurrences. *Bull. Seism. Soc. Am.* **111**(3), 1473–1480 (2021).
51. Bach, C. & Hainzl, S. Improving empirical aftershock modeling based on additional source information. *J. Geophys. Res. Solid Earth* **117**, B4 (2012).
52. Asayesh, B. M., Zafarani, H., Hainzl, S. & Sharma, S. Effects of large aftershocks on spatial aftershock forecasts during the 2017–2019 western Iran sequence. *Geophys. J. Int.* **232**(1), 147–161 (2022).
53. Papadimitriou, C., Beck, J. L. & Katafygiotis, L. S. Updating robust reliability using structural test data. *Probabilist. Eng. Mech.* **16**(2), 103–113 (2001).
54. Beck, J. L. & Au, S. K. Bayesian updating of structural models and reliability using Markov Chain Monte Carlo simulation. *J. Eng. Mech. ASCE* **128**(4), 380–391 (2002).
55. Ebrahimiyan, H. *et al.* Site-specific probabilistic seismic hazard analysis for the western area of Naples Italy. *Bull. Earth. Eng.* **17**(9), 4743–4796 (2019).
56. Convertito, V. *et al.* Time-dependent seismic hazard analysis for induced seismicity: The case of St Gallen (Switzerland). *Geothermal Field. Energies* **14**(10), 2747 (2021).
57. Daley, D. J., & Vere-Jones, D. *An introduction to the theory of point processes: volume I: elementary theory and methods* (Springer New York 2003).
58. Schorlemmer, D., Gerstenberger, M. C., Wiemer, S., Jackson, D. D. & Rhoades, D. A. Earthquake likelihood model testing. *Seism. Res. Lett.* **78**(1), 17–29 (2007).
59. Zechar, J. D., Gerstenberger, M. C. & Rhoades, D. A. Likelihood-based tests for evaluating space–rate–magnitude earthquake forecasts. *Bull. Seism. Soc. Am.* **100**(3), 1184–1195 (2010).
60. Vernant, P. *et al.* (2004). Present-day crustal deformation and plate kinematics in the Middle East constrained by GPS measurements in Iran and northern Oman. *Geophys. J. Int.* **157**(1), 381–398 (2004)
61. Haynes, S. J. & McQuillan, H. Evolution of the Zagros suture zone, southern Iran. *Geolog. Soc. Am. Bull.* **85**(5), 739–744 (1974).
62. Blanc, E. P., Allen, M. B., Inger, S. & Hassani, H. Structural styles in the Zagros simple folded zone Iran. *J. Geolog. Soc.* **160**(3), 401–412 (2003).
63. Ahmadi, A. & Bazargan-Hejazi, S. 2017 Kermanshah earthquake; lessons learned. *J Injury and Violence Research* **10**(1), 1 (2018).
64. Fathian, A. *et al.* Complex co-and postseismic faulting of the 2017–2018 seismic sequence in western Iran revealed by InSAR and seismic data. *Remote Sensing Env.* **253**, 112224 (2021).
65. Zhuang, J. & Ogata, Y. Properties of the probability distribution associated with the largest event in an earthquake cluster and their implications to foreshocks. *Phys. Rev. E* **73**, 046134 (2006).
66. Lolli, B. & Gasperini, P. Aftershocks hazard in Italy part I: Estimation of time-magnitude distribution model parameters and computation of probabilities of occurrence. *J. Seism.* **7**(2), 235–257 (2003).
67. Reasenber, P. A. & Jones, L. M. Earthquake hazard after a main shock in California. *Science* **243**(4895), 1173–1176 (1989).
68. Reasenber, P. A. & Jones, L. M. Earthquake aftershocks: update. *Science* **265**, 1251–1252 (1994).
69. Eberhart-Phillips, D. Aftershocks sequence parameter in New Zealand. *Bull. Seism. Soc. Am.* **88**, 1095–1097 (1998).
70. Molkenhuth, C., Donner, C., Reich, S., Zöller, G., Hainzl, S., Holschneider, M., & Opper, M. (2022). GP-ETAS: Semiparametric Bayesian inference for the spatio-temporal epidemic type aftershock sequence model. *Statistics and Computing* **32**, 2, 1–25 (2022).
71. Schoenberg, F. P. Facilitated estimation of ETAS. *Bull. Seism. Soc. Am.* **103**(1), 601–605 (2013).
72. Utsu, T., Ogata, Y. & Matsu'ura, R. S. The centenary of the Omori formula for a decay law of aftershock activity. *J. Phys. Earth.* **43**, 1–33 (1995).
73. Vere-Jones, D. & Davies, R. B. A statistical survey of earthquakes in the main seismic region of New Zealand. *New Zealand J Geology Geophys.* **9**(3), 251–284 (1966).
74. Hainzl, S. & Christophersen, A. Testing alternative temporal aftershock decay functions in an ETAS framework. *Geophys. J. Int.* **210**(2), 585–593 (2017).
75. Console, R., Murru, M. & Lombardi, A. M. Refining earthquake clustering models. *J. Geophys. Res.* **108**, 468 (2003).
76. Helmstetter, A., Kagan, Y. Y. & Jackson, D. D. Comparison of short-term and time-independent earthquake forecast models for Southern California. *Bull. Seism. Soc. Am.* **96**, 90–106 (2006).
77. Wells, D. L. & Coppersmith, K. J. New empirical relationships among magnitude, rupture length, rupture width, rupture area, and surface displacements. *Bull. Seism. Soc. Am.* **84**(4), 974–1002 (1994).
78. Grimm, C., Käser, M., Hainzl, S., Pagani, M. & Küchenhoff, H. Improving earthquake doublet frequency predictions by modified spatial trigger kernels in the epidemic-type aftershock sequence (ETAS) model. *Bull. Seism. Soc. Am.* **112**(1), 474–493 (2022).
79. Kagan, Y. Y. Short-term properties of earthquake catalogs and models of earthquake source. *Bull. Seism. Soc. Am.* **94**(4), 1207–1228 (2004).
80. de Arcangelis, L., Godano, C., Lippiello, E. The overlap of aftershock coda waves and short-term postseismic forecasting. *J. Geophys. Res.* **123**, 7 (2018).
81. Lippiello, E., Cirillo, A., Godano, C., Papadimitriou, E. & Karakostas, V. Post seismic catalog incompleteness and aftershock forecasting. *Geosciences* **9**, 8 (2019).
82. Hainzl, S. Rate-dependent incompleteness of earthquake catalogs. *Seism. Research Letters* **87**(2A), 337–344 (2016).
83. Hainzl, S. ETAS-Approach accounting for short-Term incompleteness of earthquake catalogs. *Bull. Seism. Soc. Am.* **112**, 494–507 (2022).

84. Jalayer, F., Iervolino, I. & Manfredi, G. Structural modeling uncertainties and their influence on seismic assessment of existing RC structures. *Struct. Saf.* **32**(3), 220–228 (2010).
85. Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, E. Equations of state calculations by fast computing machines. *J. Chem. Phys.* **21**(6), 1087–1092 (1953).
86. Hastings, W. K. Monte-Carlo sampling methods using Markov chains and their applications. *Biometrika* **57**(1), 97–109 (1970).
87. Ogata, Y. On Lewis' simulation method for point processes. *IEEE Trans. Inform. Theory* **27**(1), 23–31 (1981).

Acknowledgements

This work has been partially supported by PRIN-2017 MATISSE (*Methodologies for the Assessment of anthropogenic environmental hazard: Induced Seismicity by Sub-surface geo-resources Exploitation*) project No. 20177EPPN2 funded by the Italian Ministry of Education and Research, and by the Iranian National Science Foundation Project PEGAH (2019-2021) “*Applied framework for quasi real-time and robust forecasting of aftershock hazard and damage*”. These supports are gratefully acknowledged. This work was carried out in the framework of the Memorandum of Understanding (MoU) between l'Università degli Studi di Napoli Federico II (UNINA) and the International Institute of Earthquake Engineering and Seismology (IIEES, Tehran, Iran), 2020–2023.

Author contributions

H.E. implemented the workflow and performed the analyses. F.J. developed the workflow. H.E. and F.J. contributed in an equal manner to the drafting of the manuscript. B.M.A. and H.Z. provided the data and earthquake catalog. B.M.A. estimated the background seismicity. S.H. provided insights and recommendations about the identification of completeness magnitude, choice of spatial kernel, choice and implementation of statistical tests, and presentation and interpretation of the results. All authors have contributed to the drafting of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-24080-1>.

Correspondence and requests for materials should be addressed to H.E.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022