Are Large Earthquakes Preferentially Triggered by Other Large Events?

Journal Article

Author(s):
Nandan, Shyam; Ouillon, Guy; Sornette, Didier

Are Large Earthquakes Preferentially Triggered by Other Large Events?

Shyam Nandan[1] ![ORCID], **Guy Ouillon**[2], and **Didier Sornette**[3,4]

[1]Swiss Seismological Service, ETH Zürich, Zürich, Switzerland, [2]Lithophyse, Nice, France, [3]Department of Management, Technology and Economics, ETH Zürich, Zürich, Switzerland, [4]Institute of Risk Analysis, Prediction and Management (Risks-X), Academy for Advanced Interdisciplinary Studies, Southern University of Science and Technology (SUSTech), Shenzhen, China

Abstract Fundamentally related to the ultraviolet (UV) divergence problem in physics, conventional wisdom in statistical seismology is that the smallest earthquakes, which are numerous and often go undetected, dominate the triggering of major earthquakes, making accurate forecasting of the latter difficult if not inherently impossible. Using the general class of epidemic type aftershock sequence (ETAS) models and rigorous pseudo-prospective experiments, we show that ETAS models featuring a specific magnitude correlation between triggered and triggering earthquakes and a magnitude-dependent Omori kernel significantly outperform simpler ETAS models, in which these features are absent. Using the best forecasting model, we then show that large events preferentially trigger large earthquakes. These findings have far-reaching implications for short-term and medium-term seismic risk assessment and the development of a deeper theory without UV cut-off that is locally self-similar.

Plain Language Summary Conventional wisdom in statistical seismology is that the smallest earthquakes, which are numerous and often go undetected, dominate the triggering of major earthquakes, making accurate forecasting of the latter difficult if not inherently impossible. Here, we demonstrate that a significant fraction of large events are triggered by previous large events. Being large, these triggering events are completely reported in available earthquake catalogs. Thus, forecasting large events preceded by a set of seismic precursors would not be spoiled by the noise created by the myriad of small events whose recording is incomplete or whose parameters are highly uncertain.

1. Introduction

The observational science of seismology developed with the classification of earthquakes as either foreshocks, mainshocks, or aftershocks. Foreshocks precede mainshocks, which themselves trigger aftershocks obeying the Omori law. Time-independent forecasting methods do not take advantage of these widely observed dynamics to issue statements about future seismicity. However, time-dependent forecasts use these features but rely on different fundamental hypotheses about their genetic origin. A first class of methods appear quite agnostic on this topics (such as the M8 algorithm of Keilis-Borok and Kossobokov (1990); the Region-Time-Length algorithm of Sobolev and Tyupkin (1997); the Relative Intensity of Holliday et al. (2005); and the Proportional Hazard Model of Faenza and Marzocchi (2010)). Some other approaches consider that small earthquakes are mainly passive markers of the stress field in space and time, so their influence on future seismicity is negligible. This is the underlying concept of many earthquake precursor models such as the characteristic earthquake (D. Schwartz, 1981; D. P. Schwartz & Coppersmith, 1984; Reid, 1910), the Mogi donut and precursory quiescence (Mogi, 1969; Wyss et al., 1996; Yamashita & Knopoff, 1989), b-value mapping (Imoto et al., 1990), the load-unload response ratio (LURR; Yin et al., 1995), or the pattern informatics method (PI; Rundle et al., 2002). The accelerated moment release (AMR) pattern may also belong to this category as the increase of seismic activity before a large event may simply be due to the increase of the size of the spatial domain where stress is sufficiently large to produce a constant background rate of events per unit area and time (King & Bowman, 2003); see also (Mignan, 2011) for a review and the non-critical precursory accelerating seismicity theory). At the other end of the spectrum, some proposed methods consider that all earthquakes play a role in triggering future events, whatever their respective sizes. Those triggering models include the short-term non-Poissonian earthquake clustering model (Ebel et al., 2007), the short-term earthquake precursor model (Gerstenberger et al., 2005), as well as the Every Earthquake a Precursor According to Scale model, an epidemic-type model where small events do not trigger larger

ones (Evison & Rhoades, 2001). Note that the AMR model may also belong to this class if the large earthquake is considered as a critical point so that all previous events are indeed building bricks of the stress field intensity and spatial correlations (Bowman et al., 1998). However, a more parsimonious physics-based concept has emerged in the last two decades that undistinguishable earthquakes trigger other earthquakes. In this understanding, a mainshock may just be an "aftershock" of a previous earthquake (then called a foreshock), which happens to be larger, and the above classification is nothing but an artificial ex-post construction. This new paradigm has been bolstered up by being the driving concept behind the most successful statistical forecasting models of the general class of epidemic type aftershock sequence (ETAS) models, in which relatively rare background events, powered by plate tectonics, cascade into multitudes of triggered events sharing the same physical properties and the same space-time-magnitude distribution laws, encapsulated in the so-called triggering kernels.

ETAS models belong to the class of self-excited Hawkes conditional point processes widely adopted in recent years to model bursty behavior in many systems (Bacry et al., 2015; Crane & Sornette, 2008; Hawkes, 1971; Kobayashi & Lambiotte, 2016; Ogata, 1988) and describe the conditional seismicity rate as a linear superposition of a stationary background rate, thought to be mainly controlled by the far-field loading dynamics, and a set of transient clusters that are triggered by all past events. In its standard form, the ETAS model describes the conditional seismicity rate of magnitude m events, $\lambda(t, x, y, m | \mathcal{H}_t)$, at any location (x, y) and time t as follows:

$$\lambda(t, x, y, m | \mathcal{H}_t) = \left[\mu + \sum_{i : t_i < t} g(t - t_i, x - x_i, y - y_i, m_i) \right] \times \beta e^{-\beta(m - M_0)} \tag{1}$$

where μ is the background intensity function, assumed to be independent of time, while $\mathcal{H}_t = \{(t_i, x_i, y_i, m_i) : t_i < t\}$ stands for the history of the process up to time t. The variables (t_i, x_i, y_i, m_i) correspond to the time, x-coordinate, y-coordinate, and magnitude of the ith earthquake in the catalog, respectively, while $g(t - t_i, x - x_i, y - y_i, m_i)$ is the triggering kernel, defined in Equation 2, quantifying the temporal and spatial influence of past events onto future events:

$$g(x, y, t - t_i, x - x_i, y - y_i, m_i) = K e^{a(m_i - M_0)} \frac{T_{norm} e^{-\frac{t - t_i}{\tau}}}{\{t - t_i + c_0\}^{p_0}} \frac{S_{norm}}{\left\{ (x - x_i)^2 + (y - y_i)^2 + d e^{\gamma(m_i - M_0)} \right\}^{1+\rho}} \tag{2}$$

M_0 is the magnitude of the smallest event able to trigger some other ones, while T_{norm} and S_{norm} ensure normalization of the temporal and spatial components of the triggering kernel, respectively. Equation 2 combines, in a standard way, the fertility law $K e^{a(m_i - M_0)}$ giving the number of events directly triggered by an earthquake of magnitude m_i, a time kernel based on the exponentially tapered Omori-Utsu law (Omori, 1894; Utsu & Ogata, 1995) for aftershocks and a spatial kernel.

The ubiquitous decreasing exponential Gutenberg-Richter (GR) law (Gutenberg & Richter, 1944), which quantifies the distribution of earthquake magnitudes with exponent β, is factored out in Equation 1. This factoring out makes the following two assumptions: (a) the magnitudes of background and triggered earthquakes are distributed according to the same GR law $e^{-\beta(m - M_0)}$ such that both spontaneous and triggered events differ only in their space and time rates, not in their respective physical origins; (b) the magnitude distribution of the triggered earthquakes does not depend on the magnitude of the earthquake that triggered them such that the sole memory of the size of past events is conveyed by their total productivity function. Thus, if the space-time rate of immediate future events can be forecasted, their magnitudes cannot be forecasted better than by a purely stochastic sampling of the GR law (Helmstetter & Sornette, 2003). In most of the literature on the ETAS model, the above two assumptions are tacitly or explicitly accepted. An interesting consequence of the standard ETAS formulation concerns the collective properties of triggering: if $a - \beta < 0$, most earthquakes are triggered by previous small events, and this has been the dominant paradigm until now in statistical models of earthquake occurrences, based on calibrations of the ETAS model and on other indirect approaches (Helmstetter, 2003; Helmstetter et al., 2005). Small earthquakes contribute little to regional seismic moment release or seismic slip along major faults but give a strong contribution to the redistribution of elastic stresses (Nandan et al., 2016) and, according to existing statistical models, to the triggering of subsequent earthquakes. In the context of the ETAS model, this requires the introduction of a UV cut-off to regularize the theory, in the form of a minimum magnitude M_0 below in which no triggering occurs (Sornette & Werner, 2005).

This study builds mainly upon the empirical evidences presented in Nandan, Ouillon, and Sornette (2019), which have called into question the validity of both the aforementioned assumptions about the magnitude distribution of background and triggered earthquakes. Inspired from Vere-Jones (2005) and Saichev and Sornette (2005), Nandan, Ouillon, & Sornette (2019) proposed the following modification to the ETAS model:

$$\lambda(t, x, y, m | \mathcal{H}_t) = \mu f_{bkg}(m) + \sum_{i: t_i < t} g(t - t_i, x - x_i, y - y_i, m_i) f_{aft}(m | m_i) \tag{3}$$

where the background intensity function μ and the triggering function $g(t - t_i, x - x_i, y - y_i, m_i)$ are defined by Equation 2. Thus, the modified model differs from the standard model by replacing Equation 1 with Equation 3, which involves a new form of the GR law, which is not more factored out in order to express the existence of magnitude correlations. This new form is made of several parts: (i) $f_{bkg}(m)$ quantifies the magnitude distribution of the background earthquakes, given by

$$f_{bkg}(m) = \beta_{bkg} e^{-\beta_{bkg}(m - M_0)} \tag{4}$$

(ii) $f_{aft}(m|m_i)$ generalizes the standard GR law and quantifies the conditional distribution of events of magnitude m directly triggered by an earthquake of magnitude m_i, according to (Nandan, Ouillon, & Sornette, 2019; Saichev & Sornette, 2005; Vere-Jones, 2005):

$$f_{aft}(m | m_i) = \begin{cases} C_1 e^{-\beta_1(m - M_0)} & \forall \quad m < m_i \\ C_2 e^{-\beta_2(m - M_0)} & \forall \quad m \geq m_i \end{cases} \tag{5}$$

The constants C_1 and C_2 are derived from the constraints that $f_{aft}(m|m_i)$ is a probability density function (PDF) so that $\int_{M_0}^{\infty} f_{aft}(m|m_i) \, dm = 1$, and $C_2 e^{-\beta_2(m_i - M_0)} = C_2 e^{-\beta_1(m_i - M_0)}$, a condition that ensures the continuity of the PDF for $m = m_i$.

Equation 5 implies that $f_{aft}(m|m_i)$ describes triggered earthquake magnitudes as clustered around that of their parent earthquake. Using the nested likelihood tests, Nandan, Ouillon, and Sornette (2019) established the superiority of this model compared to the standard ETAS model presented in Equation 1 and speculated that these correlations between magnitudes of the mainshocks and aftershocks can significantly improve the forecasting performance of the ETAS models.

In this paper, we have set forth the goal of comparing the forecasting performance of the ETAS model described in Equation 3 to the standard ETAS model (Equation 1), as it is by forecasting that one can best establish the worthiness of these modifications. We also build in part on the pseudo-prospective testing framework established by Nandan, Kamer et al. (2021), who showed the superiority of an ETAS model with magnitude-dependent Omori kernel over the ETAS model, which ignored the magnitude dependence. The magnitude dependence comes in the form of the linear dependence of the exponent of the Omori kernel and the exponential dependence of the regularizer on the mainshock magnitude. Thus, we also build on the findings of Nandan, Kamer et al. (2021) and conduct a horse race between four ETAS models. Using the best of the four models from the pseudo-prospective experiments, we investigate which magnitude range (small vs. large) dominates the triggering of large earthquakes.

This paper is organized as follows: We first describe the data used in this study in Section 2. The detailed description of the four competing models is presented in Section 3.1. In Section 3.2, we briefly describe the expectation maximization (EM) algorithm and its modifications for the parameter inversion of the competing models. Section 3.3 describes the settings used for the pseudo-prospective testing and the metrics used for model evaluation. Section 4 presents the results of the pseudo-prospective experiments and scrutinizes the collective properties of triggering within the best model in Section 5. Finally, Section 6 presents the conclusion of the analysis undertaken in this study.

2. Description of the Data Set

We use $\approx$1.2 million reported earthquakes (with $M \geq 1$) within 1975–2020 in the study region (co-ordinates of the spatial polygon in Table S1 of Supporting Information S1) surrounding California in the Advanced National Seismic System (ANSS) earthquake catalog. Nandan, Ram et al. (2021) (see their Figure S2 and section Text S2 in Supporting Information S1) have shown using several proxies that the catalog within this region and time can be considered to be reasonably complete above $M = 3$. Thus, we set the magnitude of completeness $M_c = 3$ in this study as well. We also bin the magnitude reported in the catalogs at 0.1 units. All earthquakes above M_c will be used to invert the parameters of the models, as well as to simulate future data sets. We shall also assume that M_c coincides with M_0. However, only events above a specified magnitude threshold M_t will be used to check the forecasts (see Section 3.3).

An important consideration when calibrating the ETAS model is the choice of the auxiliary and primary periods (Wang et al., 2010). The earthquakes in the auxiliary period only serve as sources. In contrast, the earthquakes in the primary period can act as both sources and targets during the ETAS model's calibration. Without this consideration, the calibration process would yield a disproportionate fraction of background earthquakes at the beginning of the catalog, as there are no or very few events to act as potential triggers. For the catalog used in this study, the earthquakes between 1975 and 1981 are taken as part of the auxiliary catalog. All earthquakes following 1981 are taken as part of the primary catalog.

3. Method

3.1. Description of the Models

Based on recent empirical evidences (Nandan, Kamer et al., 2021; Nandan, Ouillon, & Sornette, 2019; Nandan, Ram et al., 2021; Nichols & Schoenberg, 2014; Ouillon & Sornette, 2005; Ouillon et al., 2009; Sornette & Ouillon, 2005; Spassiani & Sebastiani, 2016; Tsai et al., 2012), we propose a suite of four models that generalize ETAS and are tested against it.

- *Model* 1 uses the conditional seismicity rate given by Equation 1 but extends the standard ETAS model with a space-varying background intensity function (Nandan, Ram et al., 2021; Zhuang et al., 2002) given by Equation 6, with the guiding idea that the future background earthquakes occur mostly in regions where the intensity of past background earthquakes has been high.

$$\mu(x, y) = T^{-1} \sum_{i=1}^{N} IP_i \pi^{-1} Q D^{2Q} \left((x - x_i)^2 + (y - y_i)^2 + D^2\right)^{-1-Q} \tag{6}$$

where IP_i is the independence probability of the earthquake. It is not known a priori but estimated iteratively using the EM algorithm used to invert the parameters (see Section 3.2). The proposed parameterization in Equation 6 should not be confused with the triggering part (Equation 2) of the ETAS model, which also involves a summation over previous earthquakes.

Model 1 uses a tapering of the Omori law, a minor but useful extension that has been shown to provide more stable and reliable estimations of the Omori exponent. Physically, the tapering part is supported by the long-time Maxwell type viscous relaxation of the visco-plasto-elastic Earth crust. It is predicted by physically based models, such as the multifractal stress activation model (Ouillon & Sornette, 2005; Ouillon et al., 2009; Sornette & Ouillon, 2005; Tsai et al., 2012) and the state-and-rate friction model (Dieterich, 1994).

- *Model* 2 is the same as Model 1, but with a modified time kernel

  $T_{\text{norm}} \{t - t_i + c(m_i)\}^{-p(m_i)} e^{-\frac{t-t_i}{\tau}}$, where $c(m_i) = c_0 10^{c_1 m_i}$ (Davidsen et al., 2015; Dieterich, 1994; Hainzl, 2016a; Narteau et al., 2005; Scholz, 1968; Shcherbakov et al., 2004) and $p = p_0 + p_1 m_i$ (Nandan, Kamer et al., 2021; Ouillon & Sornette, 2005; Ouillon et al., 2009; Sornette & Ouillon, 2005; Tsai et al., 2012). Thus, the regularizer and the exponent of the time kernel feature exponential and linear dependence on the magnitude of the mainshock, respectively.

- *Model* 3 is defined by the following equation for the conditional seismicity rate of events of magnitude m:

$$\lambda(t, x, y, m | \mathcal{H}_t) = \mu(x, y) f_{bkg}(m) + \sum_{i:t_i<t} g(t - t_i, x - x_i, y - y_i, m_i) f_{aft}(m|m_i) \tag{7}$$

where the background intensity function $\mu(x, y)$ is given by Equation 6 and the triggering function $g(t - t_i, x - x_i, y - y_i, m_i)$ is defined by Equation 2; f_{bkg} and f_{aft} are defined by Equation 4 and Equation 5, respectively.

- *Model* 4 combines all previous novel ingredients of Models 2 and 3, that is, it uses different magnitude distributions for triggered events and their triggers as in Model 3 and a time kernel dependent on the magnitude of the triggers as in Model 2.

3.2. Calibration of the Models

All four models are calibrated using the EM algorithm proposed by Veen and Schoenberg (2008), with some necessary modifications allowing for the inversion of the optimal space varying background rate for all models (Nandan, Ram et al., 2021), magnitude-dependent parameters of the time kernel (Nandan, Kamer et al., 2021) (Models 2 and 4), and trigger-dependent magnitude distribution (Nandan, Ouillon, & Sornette, 2019) (Models 3 and 4). In the following, we briefly describe the modifications brought to the EM process of inversion of Models 1–4:

1. We start with an initial guess of the independence probabilities, IP_i, for all the earthquakes in the primary catalog. This initial guess can be created by drawing a uniform random number between 0 and 1 corresponding to each earthquake. We also make an initial guess for the parameters of the triggering kernel as well as the smoothing parameters (D, Q) for the background kernel (Equation 6). The initial values of these parameters are chosen randomly within a prespecified range.

2. *E-step*: Using the current value of the parameters and the independence probabilities, we define the probability that the ith earthquake triggered the jth earthquake as follows:

$$P_{ij} = \frac{g(t_j - t_i, x_j - x_i, y_j - y_i, m_i)}{\mu(x_j, y_j) + \sum_{i:t_i < t_j} g(t_j - t_i, x_j - x_i, y_j - y_i, m_i)} \tag{8}$$

where $\mu(x_j, y_j) = T^{-1} \sum_{i \neq j} IP_i \pi^{-1} Q D^{2Q} ((x_j - x_i)^2 + (y_j - y_i)^2 + D^2)^{1+Q}$. The reason for the choice of the summation $\sum_{i \neq j}$ will become apparent in the following steps. The new estimates of the independence probabilities can be obtained as $IP_j = 1 - \sum_i P_{ij}$. We update the current estimates of independence probabilities with the new estimates.

3. *M-step* 1: Using the current estimates of all IP_i values, we define the PDF of the location of background earthquakes at the location of jth background earthquake as follows:

$$\mu_{PDF}(x_j, y_j) = \frac{\sum_{i \neq j} IP_i \pi^{-1} Q D^{2Q} ((x_j - x_i)^2 + (y_j - y_i)^2 + D^2)^{1+Q}}{\sum_{i \neq j} IP_i} \tag{9}$$

Using $\mu_{PDF}(x_j, y_j)$ and IP_j, we can define the complete data log-likelihood for the spatial distribution of the background earthquakes as follows:

$$LL_{bkg} = \sum_{j=1}^{N} IP_j \times \ln \mu_{PDF}(x_j, y_j) \tag{10}$$

LL_{bkg} can be optimized for the parameters D and Q to obtain their new estimates. Several caveats are important to consider:

a. The minimum value of D is set to the location error.
b. To obtain $\mu_{PDF}(x_j, y_j)$, the summation in the right-hand side of Equation 9 is done for $i \neq j$. Otherwise, the optimization of LL_{bkg} leads to D being very close to 0 and Q assuming very large values, which creates a Dirac function at the location of the jth earthquake, thus leading to the maximal possible value of LL_{bkg}. This artificial situation can be avoided if one uses all earthquakes except the jth earthquake to explain the background rate at its location, which amounts to a leave-one-out strategy.

4. *M-step* 2: Maximize LL_{trig}, defined in Equation 11, for the parameters (θ) of the triggering kernel.

$$LL_{trig} = \sum_i \{-\log(\Gamma(\psi_i + 1)) - G_i(\theta) + \psi_i \log(G_i(\theta))\} +$$
$$\sum_{ij} P_{ij}^k \log \{g(t_j - t_i, x_j - x_i, y_j - y_i, m_i)\} \tag{11}$$

where $\psi_i = \sum_{j \forall t_i < t_j} P_{ij}$ is the number of direct aftershocks of the ith earthquake and $G_i(\theta)$ is the expected number of offsprings triggered by an earthquake (t_i, x_i, y_i, m_i) within the study region S and the primary time period $[T_1, T_2]$ and is given by.

$$\int_{\max(t_i, T_1)}^{T_2} \int \int_S g(t - t_i, x - x_i, y - y_i, m_i) \, dx \, dy \, dt.$$

5. Update the current estimates of all the parameters to the new estimates obtained in steps 3 and 4.
6. Repeat steps 2–5 until convergence.

The parameter β for models 1 and 2 is obtained independently of the remaining parameters of the models, as the GR distribution can be factored out due to assumptions underlying these models (Section 1). We use the formula for binned magnitude (Tinti & Mulargia, 1987), magnitude bin (Δm) equal to 0.1 and all the magnitudes above $M_c = 3$ to obtain β.

For Models 3 and 4, we follow the same algorithm as above with the modification that the triggering probabilities are now defined as follows:

$$P_{ij} = \frac{g(t_j - t_i, x_j - x_i, y_j - y_i, m_i) f_{aft}(m_j | m_i)}{\mu(x_j, y_j) f_{bkg}(m_j) + \sum_{i:t_i < t_j} g(t_j - t_i, x_j - x_i, y_j - y_i, m_i) f_{aft}(m_j | m_i)} \tag{12}$$

For these two models, the parameters of f_{bkg} and f_{aft} defined in Equations 4 and 5 are jointly maximized on the M step 2 of the algorithm prescribed above. For the optimization of parameters of f_{bkg} and f_{aft}, an additional log-likelihood term, LL_+, is added to LL_{trig} defined in Equation 11:

$$LL_+ = \log(\beta_{bkg}) \sum_j IP_j - \beta_{bkg} \sum_j IP_j(m_j - M_0) + \sum_{ij} P_{ij} \log f_a(m_j | m_i, \beta_1, \beta_2) \tag{13}$$

where

$$\log f_a(m | m_i, \beta_0, \delta) = \begin{cases} -\log\left[\frac{2\delta}{\beta_0^2 - \delta^2}\left\{\left(\frac{\beta_0 + \delta}{2\delta}\right) e^{-(\beta_0 - \delta)M_0} - e^{-(\beta_0 - \delta)m_i}\right\}\right] - (\beta_0 - \delta)m & \forall \quad m \le m_i \\ -\log\left[\frac{2\delta}{\beta_0^2 - \delta^2}\left\{\left(\frac{\beta_0 + \delta}{2\delta}\right) e^{-(\beta_0 - \delta)M_0} - e^{-(\beta_0 - \delta)m_i}\right\}\right] + 2\delta m_i - (\beta_0 + \delta)m & \forall \quad m > m_i \end{cases} \tag{14}$$

with $\beta_0 = \frac{\beta_1 + \beta_2}{2}$ and $2\delta = \beta_2 - \beta_1$.

3.3. Testing Protocols

Following the norm of scientific epistemology, we propose to identify the best model for understanding seismic processes as the one that provides the best predictive ability. For this, we set up pseudo-prospective forecasting experiments using earthquakes reported in the ANSS catalog around the state of California, for which we estimated the completeness magnitude $M_c = 3$ (Section 2). In these experiments, we use the early part of the data to calibrate the models and leave the future data unseen to compare with the forecasts constructed using the first part of the data. Starting on 1 January 1990, we perform 380 pseudo prospective experiments. Testing periods have a fixed duration of 30 days, and all the testing periods are nonoverlapping.

The four competing models issue forecasts as simulated stochastic catalogs (location, occurrence time and magnitude of future events). Each model simulates 1 million catalogs for each of the testing periods to obtain high-precision distributions of forecasts, following the recommendation of (Nandan, Ouillon, Sornette, & Wiemer, 2019a, 2019b). We perform 380 million simulations to create the forecasts of a given model for all the 380 testing periods and each of the four models. The stochastic catalogs are used to construct the models' forecasts at any spatial resolution and magnitude threshold during the testing periods. In this work, the models are evaluated at four different spatial resolutions whose corresponding areas are equivalent to those of circular

regions with radii (R_{res}): 11, 22, 45, and 89 km, and four different target magnitude thresholds (M_t): 3.5, 4, 4.5, and 5. During a given testing period, competing models forecast the distribution of the number of earthquakes (of magnitude $\geq M_t$) in the triangular pixels of an equal area, which are used to divide the study region. Lower magnitude events are used in the calibration and simulations but are not part of the test sets. We then count the actual number of observed earthquakes $\geq M_t$ within each pixel during the time window defining the testing period. With these two pieces of information, the log-likelihood LL_A^i of Model A during the ith testing period is defined using the equation:

$$LL_A^i = \sum_{j=1}^{N} \ln \left[Pr_{A,j}^i \left(n_j^i \right) \right] \tag{15}$$

where $Pr_{A,j}^i$ is the PDF of the number of earthquakes $\geq M_t$ forecasted by Model A in pixel j during the ith testing period, while n_j^i is the observed number of such earthquakes in the same pixel and time period. Using the likelihoods LL_A^i and LL_B^i of two competing models, A and B, we can then define the information gain $\left(IG_{AB}^i \right)$ of Model A over Model B during the ith testing period as

$$IG_{AB}^i = LL_A^i - LL_B^i. \tag{16}$$

To quantify the robustness of model performance, such information gains are computed for different combinations of spatial resolution and magnitude threshold M_t. We then rank the models by computing the cumulative information gain (CIG) and mean information gain $\left(MIG = \frac{\Sigma_i IG_i}{380} \right)$ that they obtain in all the testing periods relative to each other. To assess the significance of the model performance, we test the null hypothesis that the mean information gains are equal to 0 against the alternative that they are significantly larger than 0 using the paired T-test. Additionally, we also quantify the significance of the model performance by testing the null hypothesis that the median information gains (Median IGs) are equal to 0 against the alternative that they are significantly larger than 0 using the paired Ranksum-test.

Finally, we use the consistency tests (N-, M-, and S-test) proposed by Savran et al. (2020) to quantify the consistency of the models with observed earthquakes during the testing periods. Note that these consistency tests are designed to judge how well a given model forecasts the numbers, magnitude, and spatial distribution of earthquakes during testing periods. Following Savran et al. (2020), we estimate the quantities γ_N, γ_M and γ_S corresponding to the N-, M-, and S-test, respectively, for each model for all the testing periods. We then check whether these quantities estimated from all the testing periods for a given model are uniformly distributed, as would be the case if the model is the data generating process.

4. Identification of the Best Model

4.1. Model Fits

We first quantify how well the competing models fit the data set. To quantify the goodness of fit, we use all earthquakes with $M \geq 3$ in the ANSS catalog between 1975 and 2020 for the calibration of the model, with the definition of the study region, auxiliary, and primary period specified in Section 2. Since the models differ in their specification of the magnitude distributions of the background and triggered earthquakes as well as the rate of decay of the triggered earthquakes, we focus on how well the competing models capture the distributions of the relevant random variables.

Figure 1; Figures S1 and S2 in Supporting Information S1 show the magnitude distributions of background and triggered earthquakes inferred using models 1–4. Figures 1a and 1d; Figures S1a and S2a in Supporting Information S1 show the empirical distributions of background earthquakes M_{bkg} using black markers. For each model, these empirical distributions are obtained using the magnitudes of the earthquakes in the catalog and the independence probabilities (IP's) assigned to them as a result of calibration of the competing models on the catalog. The solid black lines in these figures show the best-fit GR distribution, and the corresponding parameters are indicated in the relevant panels. We find that all models show the desirable self-consistency between the empirical and the underlying theoretical distributions, albeit with a slight variation of the exponent of the GR distribution.
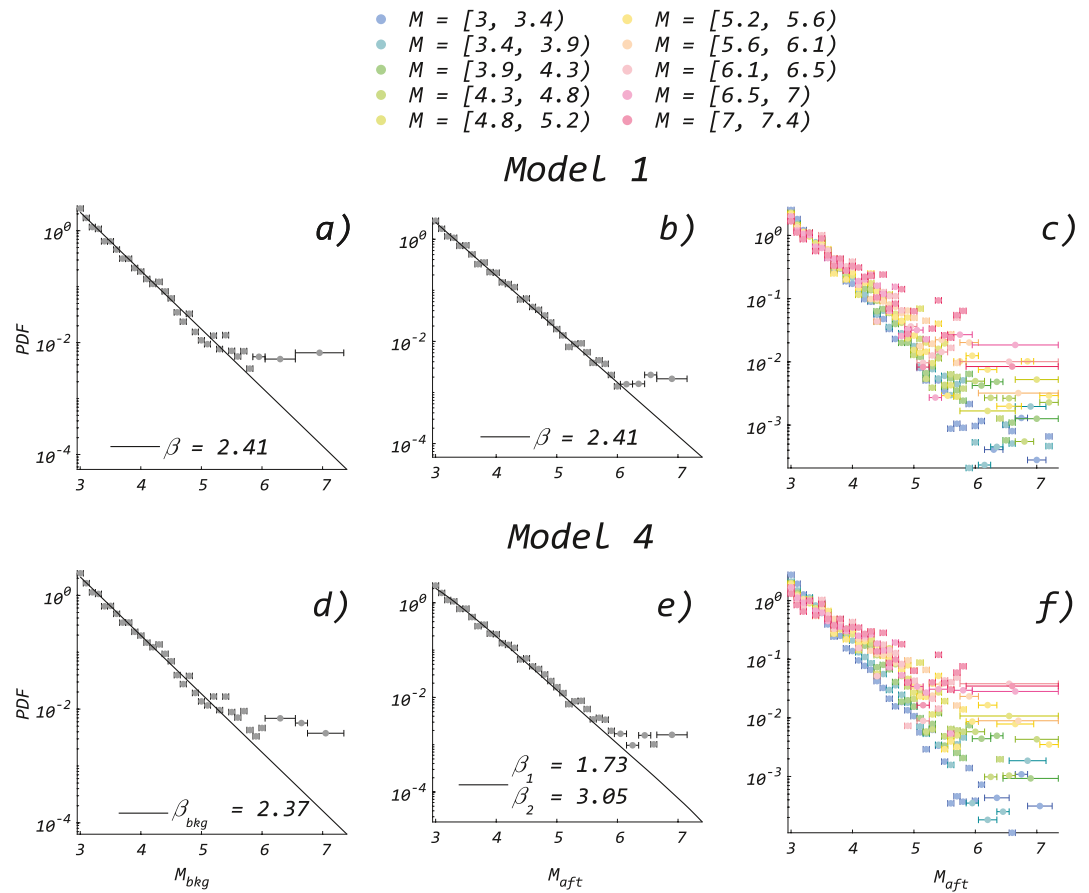
Figure 1. Magnitude distributions of background and triggered earthquakes inferred using Model 1 (a–c) and Model 4 (d–f). (a, d) Empirical distributions of the magnitudes of background earthquakes (M_{bkg}) are shown as black circles; solid line shows the best fit Gutenberg-Richter (GR) distribution with exponent $\beta = 2.41$ for Model 1 and $\beta_{bkg} = 2.37$ for Model 4. (b, e) Empirical distributions of the magnitudes of triggered earthquakes (M_{aft}) are shown as black circles. Solid lines in (b and e) show the best fitted GR distribution with exponent $\beta = 2.41$ and the renormalized best fitted conditional magnitude distribution specified in Equation 5 with parameters $\beta_1 = 1.73$ and $\beta_2 = 3.05$, respectively. (c, f) Empirical distributions (M_{aft}) for different mainshock magnitude bins specified on the top of panel c. The error bar sizes correspond to the bin sizes used for estimating the empirical probability density functions (PDFs).

Figures 1b and 1e; Figures S1b and S2b in Supporting Information S1 show the empirical distributions of triggered earthquakes M_{aft} using black markers. The empirical distributions are obtained using the magnitudes of all earthquakes in the catalog weighted by $1 - \text{IP}_j$, where IP_j is the independence probability of the *jth* earthquake. For models 1 and 2, the black lines show the best-fit GR distribution with parameter β. On the other hand, for models 3 and 4, the black lines show the renormalized best-fit conditional magnitude distribution specified in Equation 5. We again find that all the models show the desirable consistency between the empirical and underlying theoretical distributions. It is worth noticing that even though magnitude distributions of triggered earthquakes in models 3 and 4 feature a substantial kink (with $\beta_2 \approx 1.7\beta_1$), the renormalization considering the relative contributions of triggering earthquakes in the entire magnitude range leads to a distribution with a shape very similar to a standard GR distribution.

Figures 1c and 1f; Figures S1c and S2c in Supporting Information S1 show the empirical distributions of triggered earthquakes M_{aft} for different mainshock magnitude bins. We notice that the empirical distributions for all the models show a systematic dependence on the mainshock magnitude. While this behavior is expected for models 3 and 4, it points to the hidden inconsistency in the magnitude distribution of triggered earthquakes for models 1 and 2, which only becomes apparent upon conditioning on the mainshock magnitude. Figure 2; Figures S3 and S4 in Supporting Information S1 illustrate this effect more clearly with separate panels for different mainshock magnitude bins. For models 1 and 2 (Figures 2a–2j; Figure S3 in Supporting Information S1, respectively), the
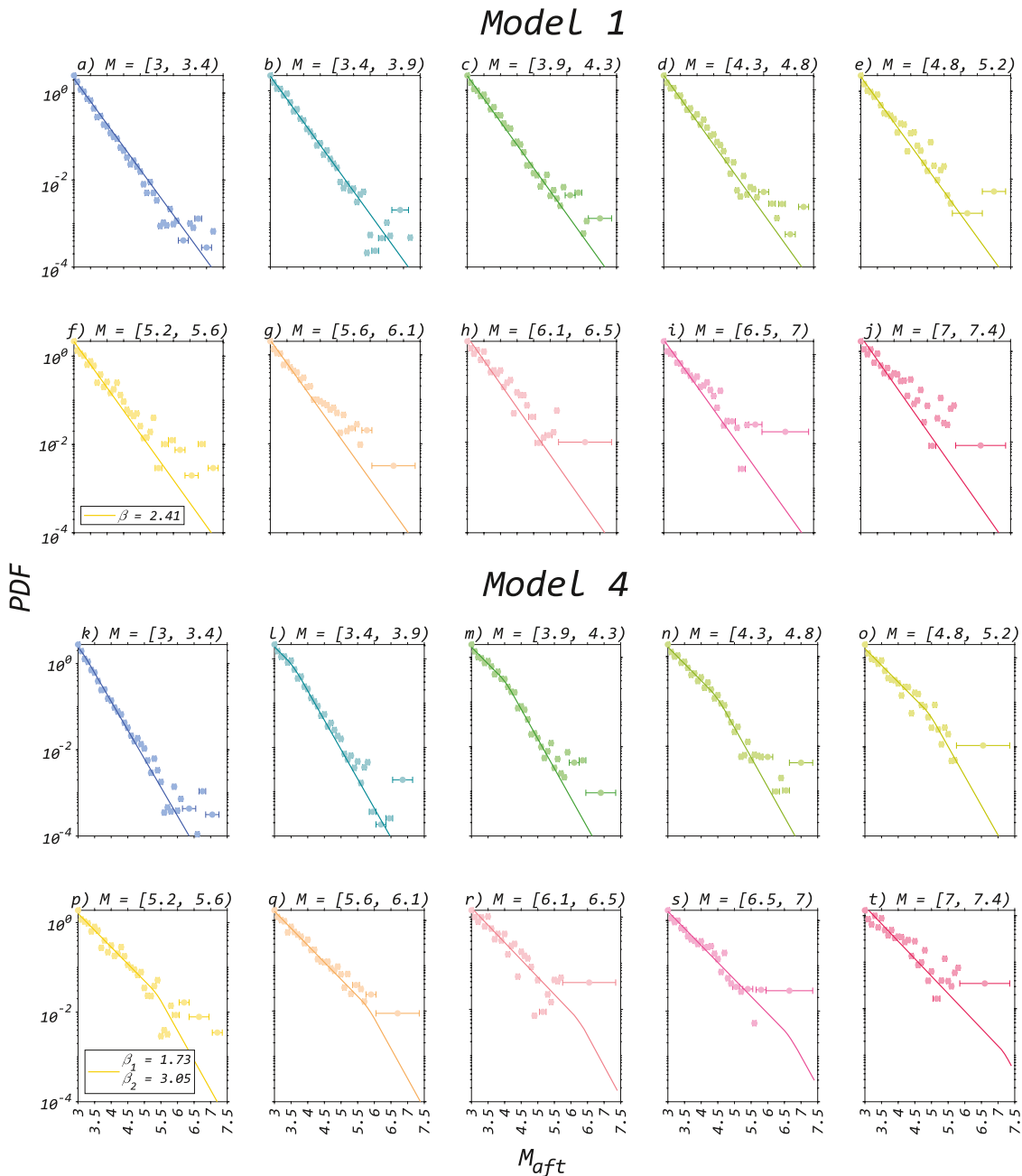
Figure 2. Empirical and best fitted magnitude distributions of triggered earthquakes, inferred using Model 1 (a–j) and Model 4 (k–t), in 10 mainshock magnitude bins indicated on the top of each panel. Solid lines in (a–j) show the best fit Gutenberg-Richter distribution with exponent $\beta = 2.41$. Solid lines in (k–t) show the renormalized best fitted kinked GR distribution (Equation 5) with parameters $\beta_1 = 1.73$ and $\beta_2 = 3.05$ for the corresponding mainshock magnitude bins. The width of the error bars corresponds to the bin size used for estimating the empirical probability density functions (PDFs).

underlying unconditional GR distribution produces a biased fit to the conditional empirical magnitude distributions of the triggered earthquakes. On the other hand, for models 3 and 4 (Figures 2k–2t; Figure S4 in Supporting Information S1, respectively), the conditional empirical distributions for the different mainshock magnitude bins exhibit clear kinks, precisely at the corresponding mainshock magnitude bins. The renormalized kinked GR distribution (Equation 5) captures this effect. As expected, both branches of the kinked GR distribution are visible for the intermediate mainshock magnitude bins. In contrast, the smaller and larger mainshock magnitude bins exhibit only the upper and lower branches of the kinked GR distribution.
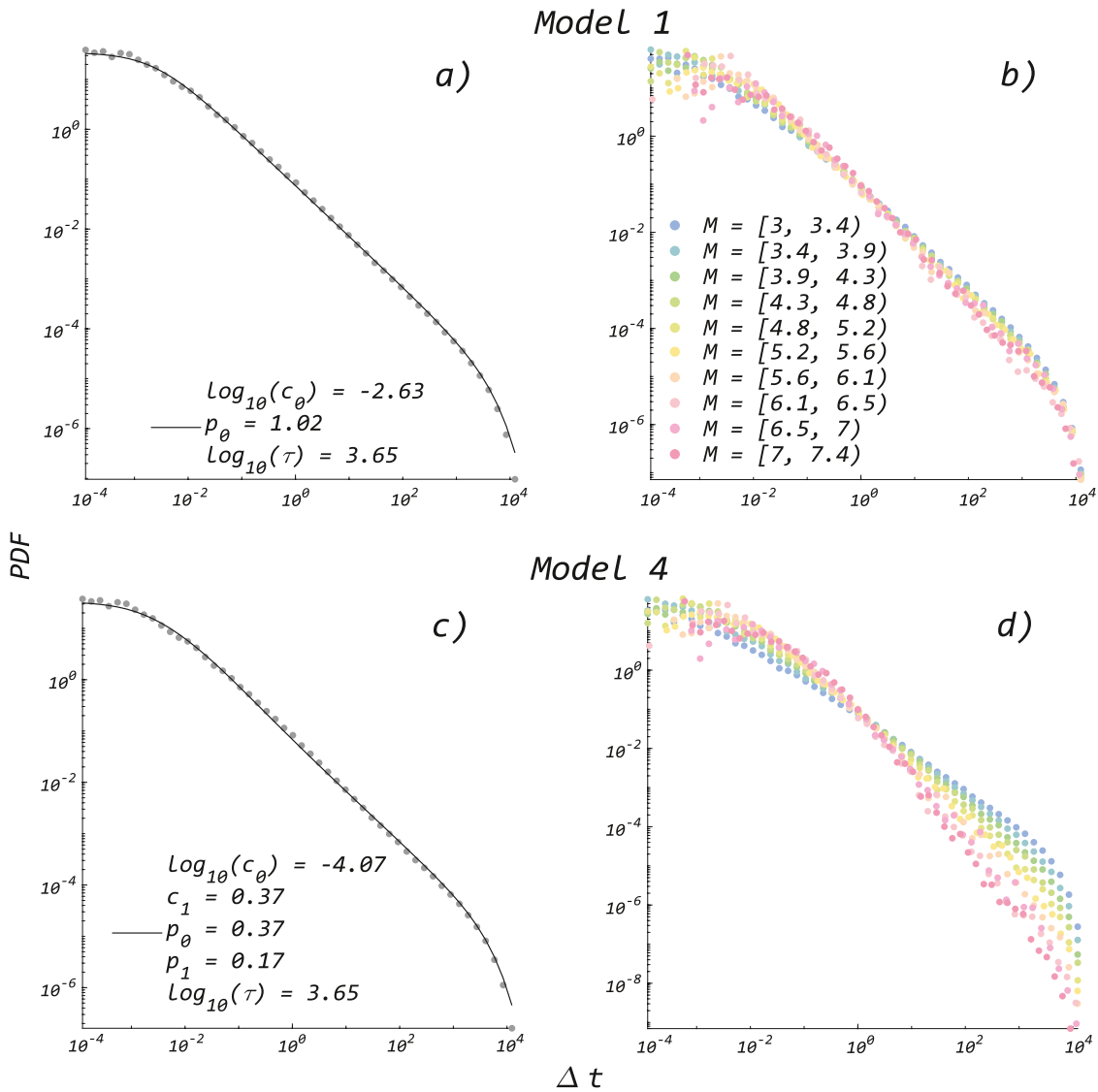
Figure 3. Distribution of the times of the triggered earthquakes since their triggers (Δt) inferred using Model 1 (a–b) and Model 4 (c–d). (a, c) Empirical distributions of (Δt) are shown as black circles. Solid lines in panels a and c, respectively, show the best fitted exponentially tapered Omori-Utsu kernel, $T_{\mathrm{norm}}\{t - t_i + c_0\}^{-p_0} e^{-\frac{t-t_i}{\tau}}$, and the renormalized modified time kernel $T_{\mathrm{norm}}\{t - t_i + c(m_i)\}^{-p(m_i)} e^{-\frac{t-t_i}{\tau}}$, where $c(m_i) = c_0 10^{c_1 m_i}$ and $p = p_0 + p_1 m_i$. (b, d) Empirical distributions of (Δt) for different mainshock magnitude bins specified in panel b.

Figures 3a and 3c; Figures S5a and S6a in Supporting Information S1 show the empirical distributions of (Δt), obtained using $\Delta t_{ij} (= t_i - t_j)$ between all causal pairs of earthquakes weighted by the corresponding triggering probability (P_{ij}). For models 1 and 2, the black lines show the best fit of the exponentially tapered Omori-Utsu kernel, $T_{\mathrm{norm}}\{t - t_i + c_0\}^{-p_0} e^{-\frac{t-t_i}{\tau}}$. On the other hand, for models 3 and 4, the black lines show the renormalized modified time kernel.

$T_{\mathrm{norm}}\{t - t_i + c(m_i)\}^{-p(m_i)} e^{-\frac{t-t_i}{\tau}}$, where $c(m_i) = c_0 10^{c_1 m_i}$ and $p = p_0 + p_1 m_i$. All the theoretical distributions underlying all the models are in agreement with the empirical distributions. Furthermore, despite a clear magnitude dependence of the regularizer and the exponent of the time kernel for models 2 and 4, the renormalization accounting for relative contributions of triggering earthquakes in the entire magnitude range leads to a distribution with the shape very similar to that of the exponentially tapered Omori-Utsu distribution.
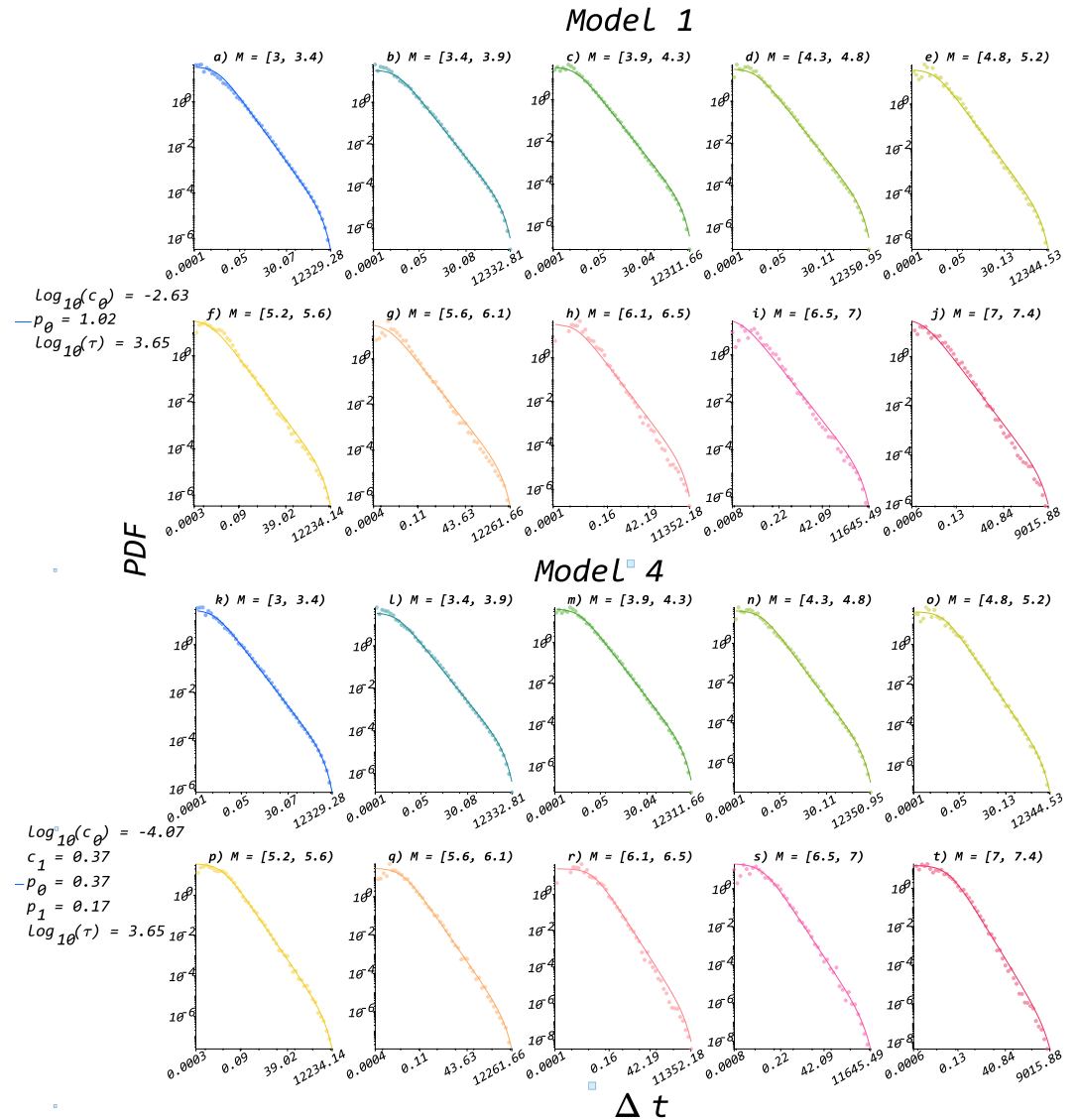
Figure 4. Empirical and the best fitted distribution of times of the triggered earthquakes since their triggers (Δt), inferred using Model 1 (a–j) and Model 4 (k–t), in 10 mainshock magnitude bins indicated on the top of each panel. Solid lines in (a–j) show the best fitted exponentially tapered Omori-Utsu kernel, $T_{\mathrm{norm}}\{t - t_i + c_0\}^{-p_0} e^{-\frac{t-t_i}{\tau}}$. Solid lines in (k–t) show the renormalized best fitted modified time kernel $T_{\mathrm{norm}}\{t - t_i + c(m_i)\}^{-p(m_i)} e^{-\frac{t-t_i}{\tau}}$, where $c(m_i) = c_0 10^{c_1 m_i}$ and $p = p_0 + p_1 m_i$, in a given mainshock magnitude bin.

Figures 3b and 3d; Figures S5b and S6b in Supporting Information S1 show that the empirical distributions of (Δt) show a systematic dependence on the mainshock magnitude. While this mainshock dependence is expected for models 2 and 4, it points to the hidden inconsistency in the time distribution of the triggered earthquakes in models 1 and 3. This inconsistency is only revealed upon conditioning on the mainshock magnitude. Figure 4; Figures S7 and S8 in Supporting Information S1 illustrate this effect more clearly with separate panels for different mainshock magnitude bins, where we clearly see that the underlying mainshock-magnitude-independent time kernel for models 1 and 3 produce a biased fit to the conditional empirical distributions of Δt of the triggered earthquakes. On the other hand, the mainshock-magnitude-dependent time kernel for models 2 and 4 explains the magnitude dependence of the empirical distributions well.

As the models differ in complexity, better fits to the data alone cannot ensure the superiority of the models in terms of forecasting performance, which can be better distinguished by pseudo-prospective experiments (Sections 3.3 and 4.4.2). However, they do highlight the self-consistency or inconsistencies of the models.
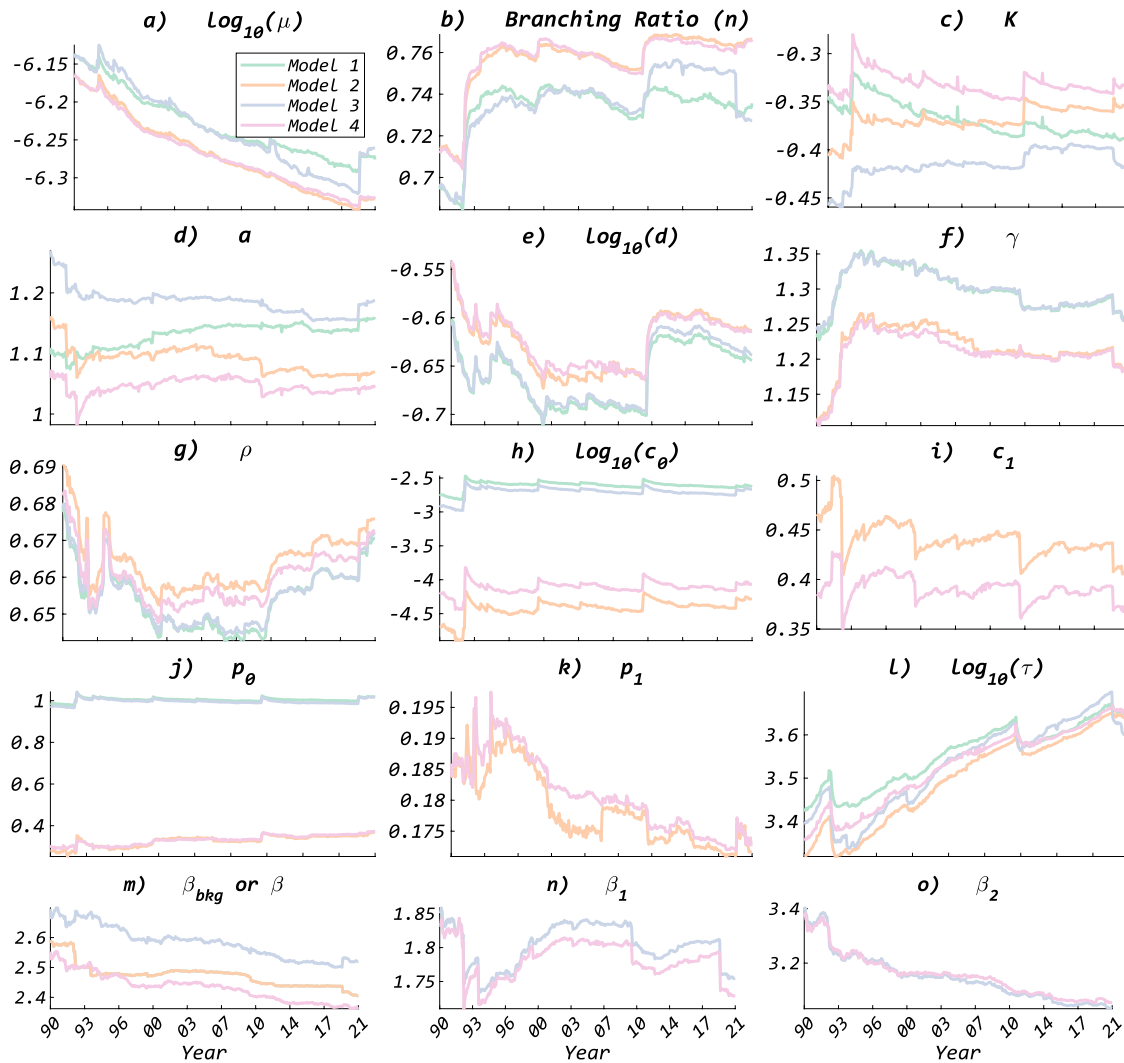
Figure 5. Time series of parameters of the four models represented using four colors. The parameters c_1 and p_1 are only present in Models 2 and 4, and the parameters β_1 and β_2 are only present in Models 3 and 4.

4.2. Time-Dependence of the Estimated Parameters for the Four Models

Figure 5 presents the time series obtained from the calibration of the four models on the training sets associated with each of the 380 successive testing periods. The size of these training sets is progressively increasing as more and more earthquakes are added to the training earthquake catalog.

Both the regularizer and the exponent of the time kernel show a strong dependence on the magnitude m of the trigger in Models 2 and 4. For instance, the regularizer increases as $10^{-3.94+0.34m}$ and $10^{-3.91+0.33m}$ for Models 2 and 4, respectively. It increases more slowly than the rupture length, suggesting that this characteristic time does not scale with rupture duration. For the final training period, the exponent of the Omori law increases as $0.46 + 0.15\,m$ for both models, that is, in a manner very similar to previous results (Nandan, Kamer et al., 2021; Ouillon & Sornette, 2005; Ouillon et al., 2009; Sornette & Ouillon, 2005; Tsai et al., 2012) reported for many regional and global catalogs.

Models 3 and 4 infer a substantial kink in the magnitude distribution of triggered events, with the kink appearing at the magnitude of the trigger (Equation 5). Below that magnitude, the magnitudes of the triggered earthquakes follow an exponential distribution with an exponent (≈ 1.75 and 1.73 for models 3 and 4, respectively, for the final training period) that is much smaller than that of the background earthquakes (≈ 2.52 and 2.37 for models 3 and 4, respectively, over the same period). The magnitudes of triggered events larger than their trigger also follow

an exponential distribution but with an exponent ($\approx$3.03 and 3.05 for models 3 and 4, respectively, for the final training period) comparatively much larger than that of the background earthquakes. For models 1 and 2, we show the time series of β (defined in the standard GR law $\beta e^{-\beta(m-M_0)}$ in Equation 1), which characterizes the magnitude distribution of both the background and triggered earthquakes. As all models use the same training catalog, these estimates of β are also valid for Models 3 and 4. For the final training period, $\beta = 2.39$ is quite close to the average estimate of β_{bkg} for Model 4 and only marginally smaller than the estimate of β_{bkg} for Model 3.

Despite the substantial kink in the magnitude distribution of the triggered earthquakes, it can be shown that the frequency-magnitude distribution of all earthquakes simulated using these models shows no discernible difference from the standard Gutenberg Richter (GR) law. For this purpose, we simulate the catalogs using the parameters inferred for Model 4. The time series of the parameters can be found in Figure 5. We select the values of the parameters for the last testing period, that is, corresponding to the use of the entirety of the catalog. Figure S9 in Supporting Information S1 shows the frequency-magnitude distribution of all earthquakes simulated using Model 4 on a log-linear scale. For reference, we show a straight line depicting an exponential GR distribution. It is evident from the figure that the renormalized magnitude distribution of all earthquakes in the simulated catalog behaves like a standard GR law. This is in agreement with the rigorous and analytical derivation made in Saichev and Sornette (2005), which shows that the two branches of the kinked GR distribution (Equation 5) in addition to a GR distribution for background events (with a different b-value) still give by "renormalization over all triggered cascades" a pure GR distribution with b-value equal to the background b-value. Intuitively, summing kinked GR distributions over all earthquake generations that span a large set of magnitudes erases the observability of the kinked GR distributions conditional on the magnitude of the triggering earthquakes. Nandan, Ouillon, and Sornette (2019) have shown that the underlying hidden kinked GR distribution can be retrieved when using the EM (expectation maximization) algorithm and conditioning on the magnitude of the triggering earthquake.

We also note that the inclusion of magnitude dependence in the parameters of the ETAS model does not lead to a dramatic change in the criticality parameter (branching ratio), in contrast with the speculation in Zhuang et al. (2013).

All parameters show variations with time to some extent. A major source of this time variation is a progressive increase in the size of the training catalog, leading to a converging trend in some of the parameters such as μ, the βs, branching ratio, and p_1. Note that the estimate of the branching ratio is obtained empirically as the fraction of earthquakes identified as triggered events. In the time series of some of the parameters such as K, a, c_0, c_1, p_0 and so on, one can also notice a tendency for sudden, yet small, jumps. These jumps are associated with some prominent events in the earthquake catalog, including the Landers, Hector Mine, El-Major earthquakes, etc.

4.3. An Example Forecast

Before presenting the results of pseudo-prospective experiments, we illustrate how the forecasts of the models are specified during a 30-day long testing period. For example, we take the testing period following the 2019 M7.1 Ridgecrest earthquake, that is, between 28 July 2019 and 27 August 2019. Figure 6a shows the map of the seismicity rate forecasted by Model 1 during this testing period and the location of $M \geq 3.5$ earthquakes that occurred during this period (black stars). Although the seismicity rate map is a convenient way to illustrate the forecasts, the actual forecasts for all the models are specified by 1 million stochastic catalogs for a given testing period. These stochastic catalogs can then be used to create the distribution of the number of simulated earthquakes (N_{sim}) in a simulation for any space-magnitude bin combination. For instance, Figure 6b shows the distribution of N_{sim} in the entire testing region for four magnitude thresholds M_t at which the forecasts will be evaluated in Section 4.4. For comparison, we also show the distribution of N_{sim} forecasted by a Poisson distribution, with the mean rate being equal to the average number of earthquakes generated by the model in all the simulations (dashed lines). It is evident that the Poisson distribution fails to approximate the distribution of N_{sim} and decays much faster. So, even though the model assigns a much larger probability to the observed counts (circular markers) of earthquakes during the testing period, a much smaller probability will get assigned to the observations if the Poissonian approximation were used. Since the log of these probabilities (or log-likelihoods) is used for scoring, the model gets an improper scoring under the Poissonian approximation. Thus, in this study, we use the empirical "number distribution" for different combinations of spatial resolutions and magnitude thresholds to evaluate the models. These empirical distributions can be obtained from the stochastic catalogs and are processed as in Nandan, Ouillon, Sornette, and Wiemer (2019a) using adaptive kernel-density techniques to obtain a smooth distribution
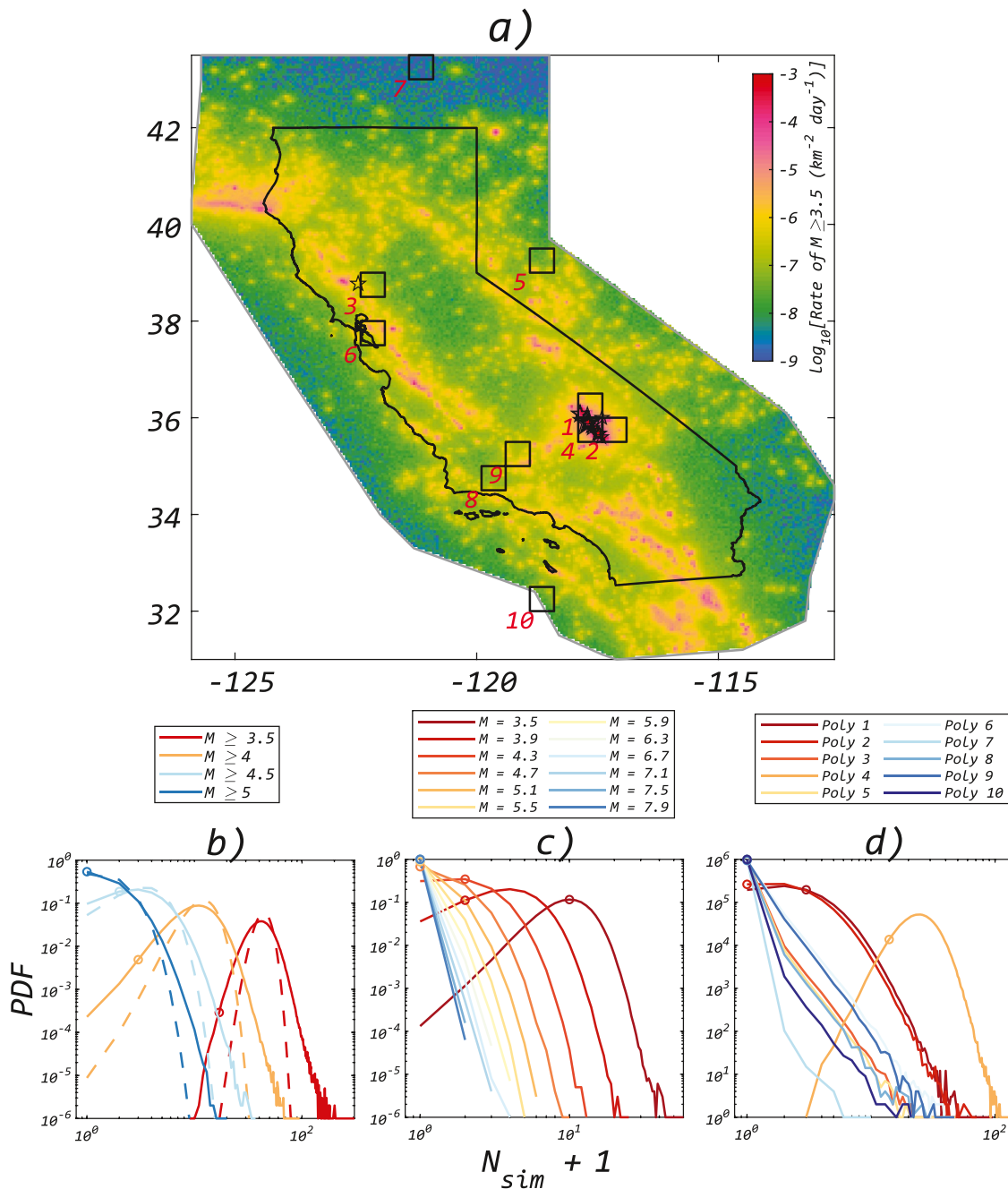
Figure 6. An illustration of seismicity forecasted by the competing epidemic type aftershock sequence models. (a) Seismicity rate forecasted by Model 1 after the M7.1 Ridgecrest earthquake (28 July 2019–27 August 2019). Black stars show the location of $M \geq 3.5$ earthquakes that occurred during the forecasted period. Black squares show the location of the 10 randomly chosen polygons (with areas 0.5×0.5 deg^2 each) for which the distribution of earthquake numbers (N_{sim}) forecasted by the model in 1 million simulations are shown in panel d. (b) Solid curves show the distribution of N_{sim} forecasted by the model in the entire study region for four magnitude thresholds ($M \geq 3.5, 4, 4.5, 5$) of the target catalog. Dashed curves show the distribution of N_{sim} forecasted by a Poisson distribution, with the mean rate being equal to the average number of earthquakes generated by the model in all the simulations. (c) Distribution of N_{sim} forecasted in the entire study region in different magnitude bins with bin size 0.1. (d) Distribution of N_{sim} forecasted in the 10 polygons shown in panel a. Circular markers in panels b–d show the observed counts of earthquakes during the testing period for different magnitude thresholds (b), magnitude bins (c) and spatial polygons (d).

of N_{sim} without "holes." Figures 6c and 6d show the empirical distribution of N_{sim} forecasted for different magnitude bins with 0.1 unit bin size in the entire study region and the 10 polygons (0.5×0.5 deg^2) shown in panel 6a, respectively.

4.4. Model Evaluations

4.4.1. Performance of Model 1 Compared to a Spatially and Temporally Homogeneous Poisson Process (STHPP)

We first compare Model 1 to a spatially and temporally homogeneous Poisson process (STHPP) to establish it as a reliable benchmark against which all the other models will be evaluated. As the name suggests, the STHPP model forecasts the rate of future earthquakes as being homogeneous in space and uniform in time. The rate forecasted by the STHPP is estimated using the data in the training period and is given by

$$\lambda = \frac{N\,(\geq M_t)}{A \times T} \tag{17}$$

where λ is the average rate of earthquakes with magnitudes larger than M_t per day per km^2, A is the area of the study region, and T is the time duration of the training catalog. M_t is equal to the testing magnitude threshold. For the current study region, $A = 961, 238$ km^2. The value of T depends on the end time of the training data set (which is also the starting time of the testing period). For the first testing period, $T = 3, 286$ days (≈ 9 years, considering training data between 1 January 1981, and 1 January 1990), increasing by 30 days as the training period becomes larger and larger. Having obtained λ from a given training period, the forecast of the STHPP model for the following testing period (of a duration of 30 days) is prescribed as a mean rate in all the equal-area pixels as $\lambda_{\text{pixel}} = \lambda \times 30 \times A_{\text{pixel}}$, where A_{pixel} is the area of triangular pixels with which the study region is tiled. For the ith testing period, the performance of the model is estimated using the Poissonian log-likelihood as follows:

$$LL^i_{STHPP} = \sum_j n^i_j \ln \lambda^i_{pixel} - \lambda^i_{pixel} - \ln n^i_j! \tag{18}$$

where index i stands for the testing period, while j stands for the pixel's index. The information gain of Model 1 over STHPP in the ith testing period is simply $IG_i = LL^i_{Model1} - LL^i_{STHPP}$. We then obtain the information gain per earthquake (IGPE) that Model 1 obtains over STHPP as $IGPE = \frac{\sum_i IG_i}{\sum_i N_i(\geq M_t)}$, where $N_i(\geq M_t)$ is the observed number of earthquakes with magnitude larger than M_t during the ith testing period. Figure S10 in Supporting Information S1 shows the IGPE (colored bars) and its 95% confidence interval (orange error bars obtained using bootstrapping the IGs obtained from the 380 testing periods). Since the models are compared at four different spatial resolutions (indicated on the x-axis) and four testing magnitude thresholds (indicated by colors of the bars), we get four groups of four bars. Figure S10 in Supporting Information S1 leads to the following conclusions:

1. At all spatial resolutions, Model 1 significantly outperforms the STHPP with IGPE ranging from 1.72 ($R_{\text{res}} = 89$ km, $M_t = 5$) to 3.7 ($R_{\text{res}} = 11$ km, $M_t = 3.5$), with *p-values* resulting from pairwise T-tests being below the computer precision.
2. With increasing spatial resolution (i.e., decreasing R_{res}), Model 1 obtains higher IGPE over the STHPP.
3. At all spatial resolutions, although there is no general trend of a drop in performance with increasing M_t, Model 1 features a consistently lower IGPE at $M_t = 5$ compared to its IGPE for other M_t's. However, we cannot present evidence whether IGPE could continue decreasing at larger M_t values, as the number of simulated events would be too small to reliably estimate the performance of the forecasts. For that purpose, one would have to drastically increase the number of simulated catalogs.

The IGPEs of Model 1 over the STHPP translate into a probability gain per earthquake ($=e^{\text{IGPE}}$) ranging between 5.62 and 42.52. These numbers indicate how likely Model 1 is to explain an individual earthquake relative to the STHPP. These numbers show that Model 1 is a much superior model to STHPP and can thus act as a strong benchmark against which all other models can be evaluated.

4.4.2. Relative Performance of Models in Pseudo-Prospective Forecasting Experiments

In this section, we compare the performance of the models relative to Model 1 and each other. Model 1 is used as the null model as it is a robust benchmark whose superiority against the standard ETAS model with a spatially homogeneous background rate has already been established in Nandan, Ram et al. (2021) and recalled in the previous subsection. Figures 7a–7p show the time series of the CIG that models 2–4 obtain relative to Model 1. In total, we did the evaluations of the models relative to Model 1 at 16 settings (resulting from a combination of four
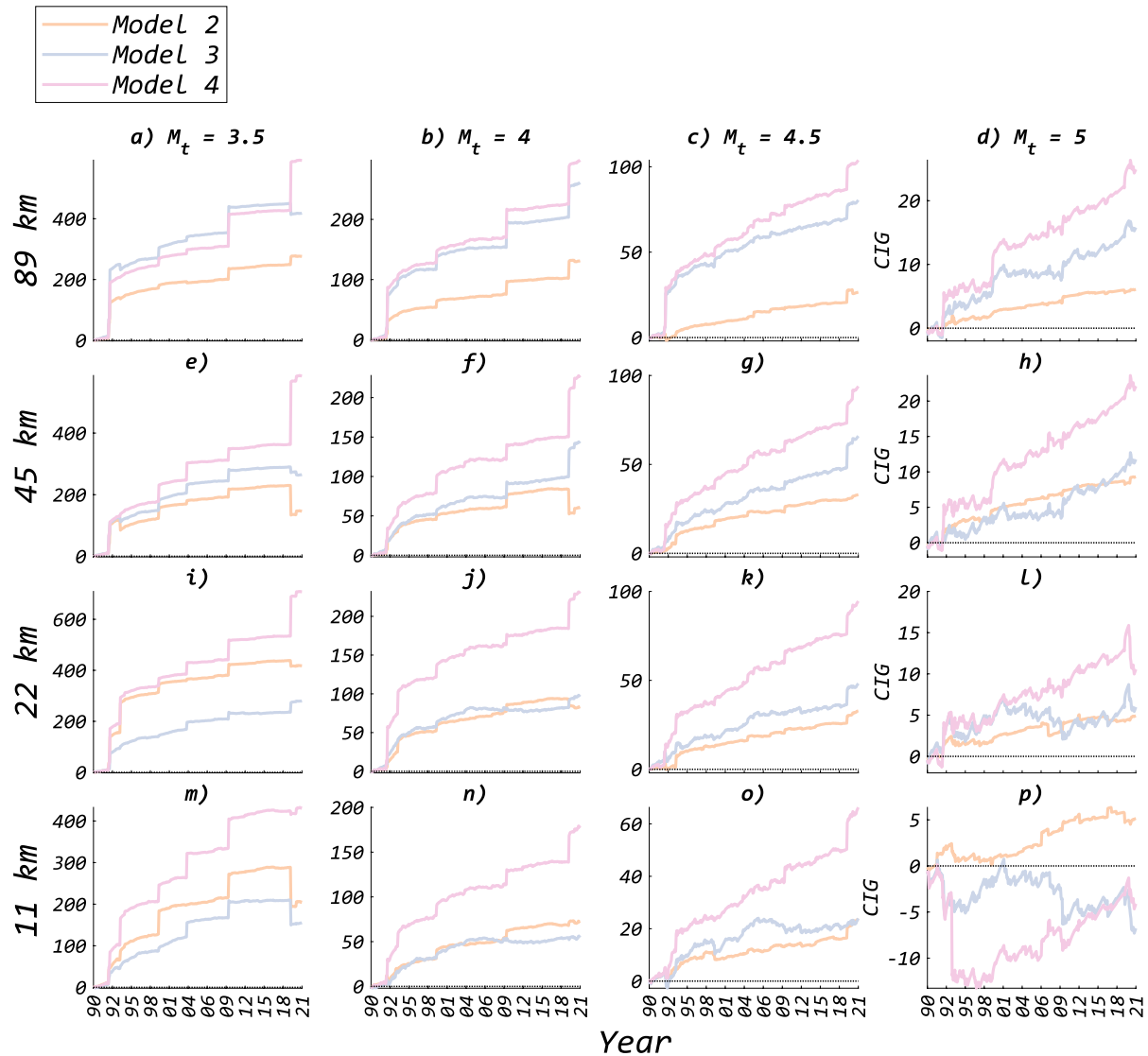
Figure 7. (a–p) Time series of cumulative information gain (CIG) that Models 2–4 obtain with respect to Model 1 for 16 different testing settings corresponding to different panels.

different M_t thresholds and four different spatial resolutions). We can observe from Figures 7a–7p that models 2–4 outperform Model 1 in nearly all testing settings (see Section 4.4.3 for a discussion of the deterioration of the models' performance with larger M_t and larger spatial resolutions).

To quantify the significance of the model performance, we use two metrics (mean information gain (MeanIG) and median information gain (MedianIG)), which are computed from the 380 values of information gains obtained for each individual testing period for a given test setting. The statistical significance of MeanIG and MedianIG are evaluated using the T-test and Ranksum test, respectively. We can observe from Figures 8a–8h that models 2–4 also significantly outperform Model 1 in nearly all testing settings. Furthermore, one can observe that Model 4 is the best, as it outperforms all the other three models in 15 out of 16 test settings (see Section 3.3, Figures 7 and 8).

One could object that the kinked GR law underlying models 3 and 4 is nothing but the manifestation of short-term aftershock incompleteness (STAI) (Hainzl, 2016b; Helmstetter et al., 2006) observed after relatively large
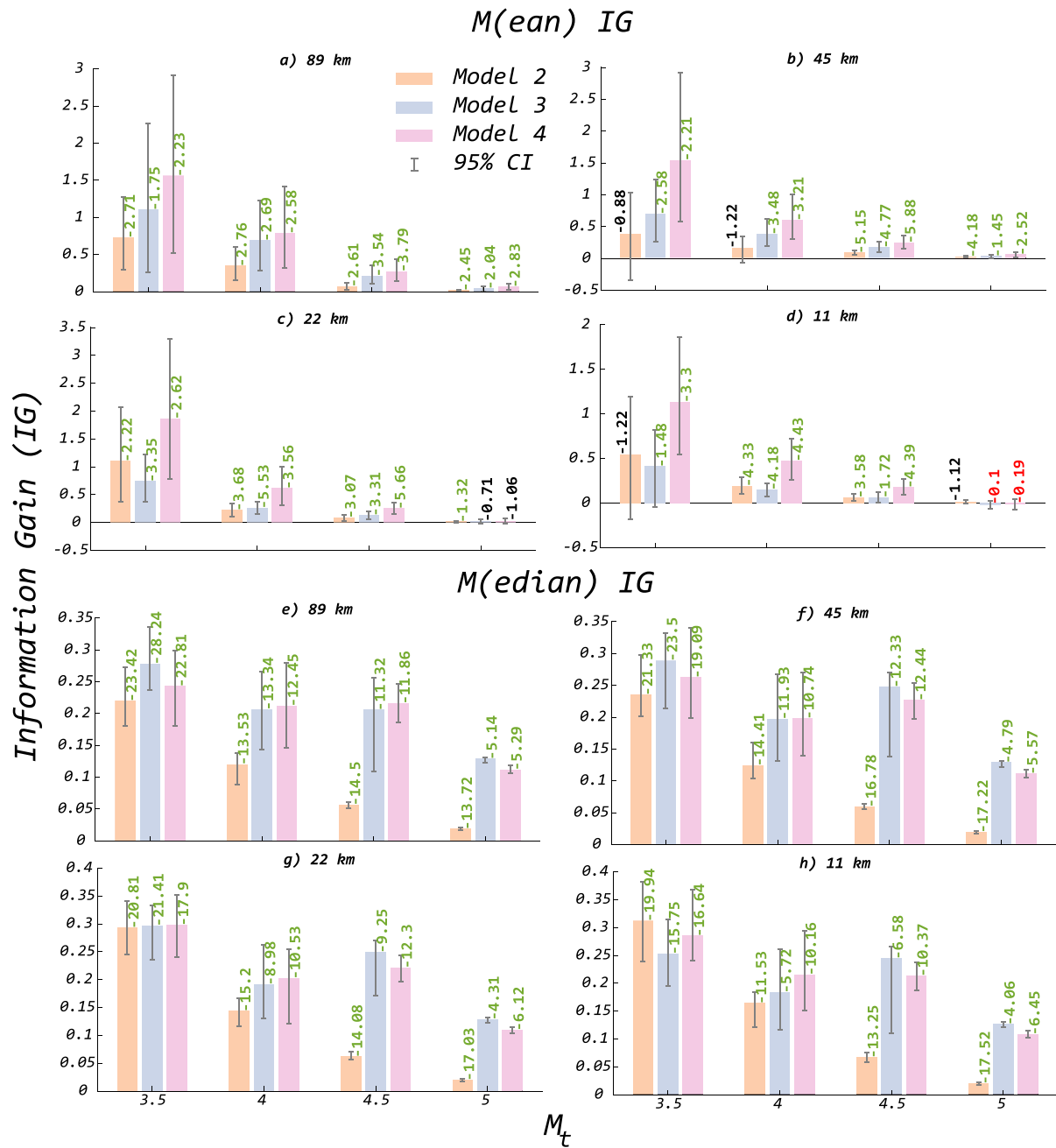
Figure 8. (a–d) Mean information gain (MeanIG) that Models 2–4 obtain with respect to Model 1 for 16 different testing settings. The four panels correspond to the spatial resolutions at which models were compared. The *x*-axis shows the four magnitude thresholds (M_t) of the target catalog at which the testing was performed. Error bars show the 95 confidence interval of MeanIG obtained using bootstrapping. The number above each bar is the log(*p*-value) of the test of the null hypothesis that each of the MeanIGs are equal to 0 against the alternative that they are significantly larger than 0, using the paired T-test. Green colored numbers indicate that the null hypothesis can be rejected (at a significance level of 0.05). Black and red colored numbers, respectively, indicate that the null hypothesis is not rejected, but the alternative model is more informative than the null model and vice-versa. (e–h) Same as panels (a–d) but using the Median information gain (MedianIG). The null hypothesis tested is that the MedianIGs are equal to 0 against the alternative that they are significantly larger than 0, using paired Ranksum test.

earthquakes. Thus, by capturing the STAI in the form of kinked GR law, Model 4 outperforms all the competing models. The following pieces of evidence allow us to reject this objection:

1. Although it could be argued that STAI could play a very dominant role as more and more small events have been detected after the template-matching techniques (MFT) has been applied to pick up small earthquakes (Ross et al., 2019), the matched filtering approaches seem to detect additional events mostly in magnitude
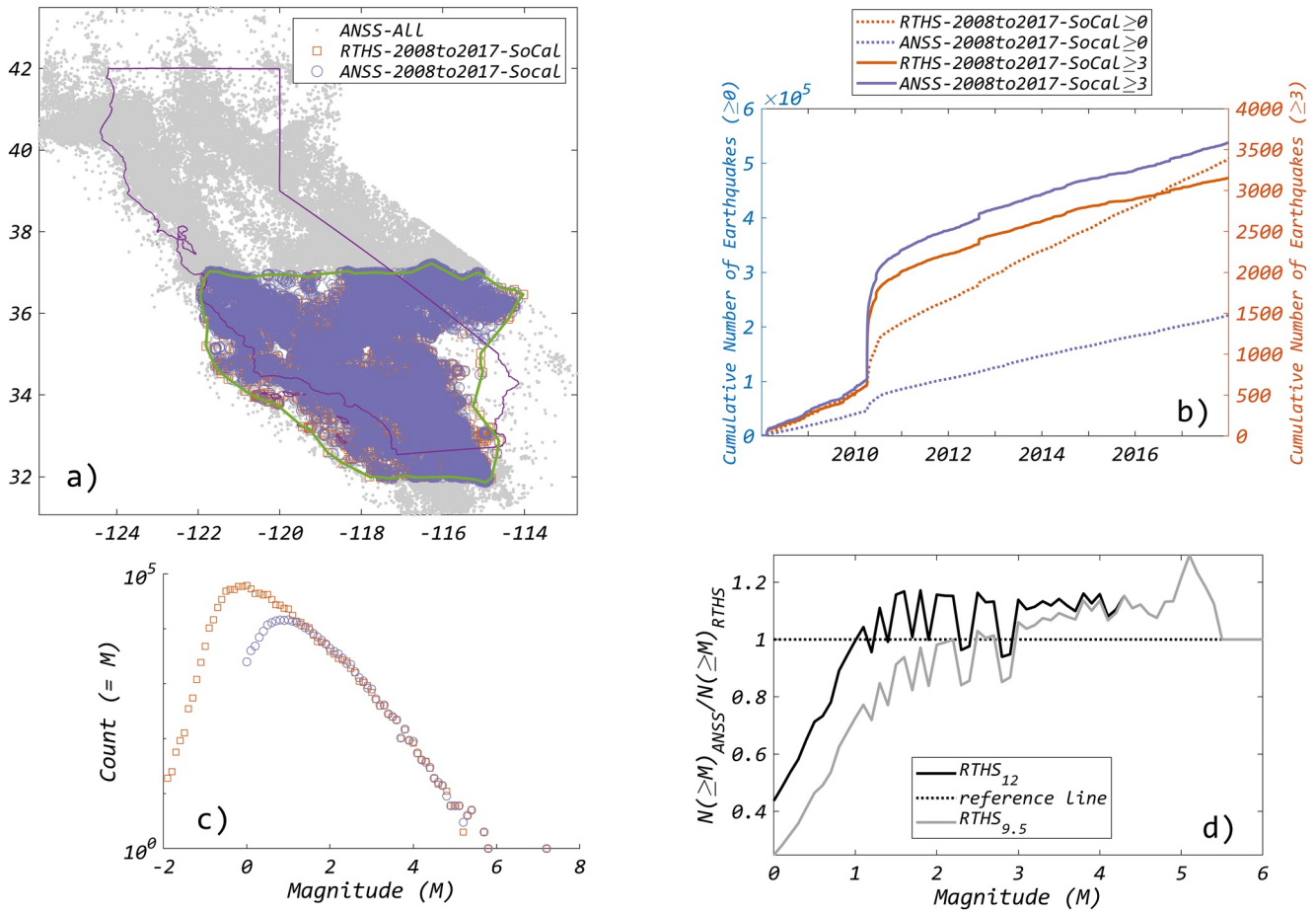
Figure 9. (a) Spatial distribution of all earthquakes in the Advanced National Seismic System (ANSS) catalog are shown as gray dots. Orange squares show the location of earthquakes reported in the Ross-Trugman-Hauksson-Shearer (RTHS) catalog between the years 2008 and 2017 in Southern California (region delineated using a green line). Violet circles show the location of earthquakes in the ANSS catalog between 2008 and 2017 in the green zone. (b) Dashed and solid lines show the cumulative number of earthquakes ($M \geq 0$ and $M \geq 3$, respectively) in the green zone in the RTHS (orange) and the ANSS (violet) catalog in the period 2008 to 2017. (c) Frequency magnitude distribution of earthquakes in the RTHS (orange) and ANSS (violet) catalog in the green zone from 2008 to 2017. (d) The ratio of the number of earthquakes $\geq M$ in the ANSS and the RTHS (black) catalog in the green zone between 2008 and 2017 as a function of magnitude (M). The Gray curve shows the same ratio but for the RTHS catalog, which features a lower detection threshold for event detection (qtm_final_9.5dev.hypo). The subscripts show the distinction between the two RTHS catalogs.

ranges smaller than those used in this study. We demonstrate this by comparing the number of earthquakes reported in the ANSS earthquake catalog and the Ross-Trugman-Hauksson-Shearer (RTHS) catalog (Ross et al., 2019) reported for Southern California (region delineated using the green line) in Figure 9a. The RTHS catalog has been obtained using the template matching approach (Ross et al., 2019). Note that there are two catalog versions: *qtm_final_9.5dev.hypo* and *qtm_final_12dev.hypo* (see the acknowledgment section for the link to the data sets). We selected the latter because it contains the highest confidence events and corresponds to a detection threshold of 12.0 times the median absolute deviation. The *qtm_final_12dev.hypo* catalog is referred to as the RTHS catalog hereafter. The location of all earthquakes in the RTHS catalog are shown as orange squares in Figure 9a. As these earthquakes are only reported within the green zone (Figure 9a), we also restricted the ANSS catalog to the same zone for the following comparisons. Furthermore, the RTHS catalog is only reported for the period between 2008 and 2017, including 2008 and 2017, so the ANSS catalog was also restricted to the same period. In Figure 9b, we show the cumulative number of earthquakes ($M \geq 0$ and $M \geq 3$) for the RTHS and the restricted ANSS as dashed and solid lines, respectively. We find that the RTHS and the restricted ANSS catalogs report $\approx$505,000 and $\approx$220,000 $M \geq 0$ earthquakes, that is, the RTHS catalog reports nearly $\approx$2.3 times more $M \geq 0$ events than the ANSS catalog. However, for $M \geq 3$, the restricted ANSS catalog reports $\approx$3,560 earthquakes while the RTHS catalog reports $\approx$3,150 earthquakes,

with an RTHS to ANSS ratio of 0.88. This implies that using the sophisticated template matching approach (Ross et al., 2019), the RTHS catalog mainly adds below $M = 3$, which is the bulk magnitude of completeness threshold considered in this study. In Figure 9c, we further show the frequency magnitude distribution (FMD) of the RTHS and the ANSS catalogs. The figure clearly shows that the FMDs obtained from the two catalogs overlap up to a magnitude threshold $M = 1.2$, below which the fraction of earthquakes reported in the RTHS catalog shows a systematic increase. Finally, Figure 9d conveys the same information as Figure 9c differently by showing the ratio of the number of earthquakes $\geq M$ in the ANSS and the RTHS (distinguished as RTHS_12 in Figure 9d) catalog as a function of magnitude (M). The ANSS catalog tends to report about 11% more earthquakes than the RTHS catalog above magnitude 3. This could be attributed to our choice to use the higher confidence qtm_final_12dev.hypo catalog instead of lower confidence qtm_final_9.5dev.hypo catalog. We verify this using the lower confidence qtm_final_9.5dev.hypo catalog (RTHS_9.5) in Figure 9d. We find that, although the ratio of the numbers of earthquakes $\geq M$ in the ANSS and the RTHS_9.5 catalog is lower than when considering the RTHS_12 catalog, the effect is mostly below $M = 3$. At $M = 3$, the ANSS catalog still reports about 9% more earthquakes than the RTHS_9.5 catalog. Despite the improvements offered by the template matching approach in identifying the missing earthquakes, the method remains susceptible to STAI, which would imply that both the ANSS and RTHS catalogs would have missing $M \geq 3$ events. Thus, the ratio of the number of events in the two catalogs alone cannot indicate whether the ANSS catalog is complete above $M = 3$.

2. To assess the severity of incompleteness at different magnitude thresholds, Mizrahi et al. (2021) use the extension of the EM algorithm to infer the self-consistent estimation of ETAS parameters and high-frequency time series of the magnitude of incompleteness. This method allows inferring the fraction of incomplete events at different magnitude thresholds M_t. This incompleteness fraction decreases from $\approx 6.2\%$ for $M_t = 2.5$ to $\approx 1.3\%$ for $M_t = 3$ to $\approx 0.3\%$ for $M_t = 3.5$, in the case of the Californian catalog for almost the same time period. Thus, it is clear that STAI is not a significant factor if we consider a reasonably large enough bulk magnitude of completeness, such as $M_c = 3$, as done in this study.

3. In Nandan, Ouillon, and Sornette (2019), it was shown using synthetic tests that STAI cannot result in a kinked GR law. The authors first simulated a perfect synthetic catalog using the ETAS model in these tests. These catalogs were subjected to incompleteness filters (Helmstetter et al., 2006), leading to catalogs exhibiting STAI. Nandan, Ouillon, and Sornette (2019) found that after inversion, these incomplete catalogs do not feature the kinked GR law distribution as in the case of real catalogs. Note that the study of Nandan, Ouillon, and Sornette (2019) pertained to the same study region, nearly identical duration of the catalog, as well as the same bulk magnitude of completeness as in this study.

4. Model 4 not only outperforms Model 1 and Model 2 in forecasting $M \geq 3.5$ earthquakes but also at higher magnitude thresholds such as 4.5 and 5. If the kinked GR law underlying Model 4 was indeed the manifestation of STAI, the model would have lost its edge over Model 1 and 2 when forecasting earthquakes $M \geq 4.5$ and $M \geq 5$, which is not the case (see Figures 7 and 8).

We conclude that the outstanding performance of Model 4 is due to two crucial properties: a time Omori kernel dependent on the magnitude of the trigger and a kinked GR law conditional on the magnitude of the trigger. The first property has been predicted theoretically and later confirmed by various empirical tests. It is due to the interplay between long-range elastic stress transfer and effective thermal activation of rupture (Nandan, Kamer et al., 2021; Ouillon & Sornette, 2005; Ouillon et al., 2009; Sornette & Ouillon, 2005; Tsai et al., 2012). The second property is also a theoretical prediction when imposing a scale-invariant property generalizing the ETAS model so that the minimum cut-off magnitude M_0 is pushed to $-\infty$ (Nandan, Ouillon, & Sornette, 2019; Saichev & Sornette, 2005; Vere-Jones, 2005), that is, the minimum rupture scale is vanishing. The first empirical confirmation of the kinked GR law was reported by Nandan, Ouillon, and Sornette (2019), who proposed a mechanical interpretation using the fact that the symmetry of the deformation tensor at any scale tends to mimic the orthorhombic symmetry of the loading stress field. Thus, each time an event occurs on a fault and creates a monoclinic strain perturbation, another event of similar size tends to be induced on a conjugate fault to reestablish the orthorhombic symmetry.

4.4.3. Remarks on the Relative Performance of Models

1. Figures 7a–7p show that the performance of the models decreases with the increasing magnitude threshold of the target catalog. However, such a decrease is not surprising considering that the CIG reported in the figure are not normalized by the number of earthquakes above different M_t's. As the smaller earthquakes occur more

frequently than larger ones, a fairer comparison for models at different magnitude thresholds can be made if the gains are further normalized by the frequency of events in the testing period. Given the b-value of ≈ 1 (Figure 1), M3.5+ earthquakes are ≈ 32 times more frequent than M5+ earthquakes, making the per-event information gains comparable for these two end-member settings. With the same heuristics, per-event gains are similar for all the testing magnitude settings.

2. The sudden decrease in net CIG of these models at the highest spatial resolution is likely due to model assessments based on empirical earthquake number distributions. As the spatial resolution increases, the empirical distributions become increasingly unrepresentative of the underlying distribution owing to the monotonically decreasing likelihood of simulated earthquakes sampling a given pixel (Saichev & Sornette, 2006). Increasing the magnitude threshold has a similar impact on the sampling. Thus, we see the strongest deterioration in the performance of the models in Figures 7l, 7o, and 7p. One possibility to avoid these artifacts resulting from under-sampling of the underlying distribution is to increase the number of simulations. However, since we have already performed nearly 1 million simulations, increasing the number of simulations to determine the true performance at higher resolutions and magnitude thresholds will be a major computational challenge. If the magnitude threshold is increased from 3.5 to 5, a similar sampling of a given pixel as with $M_t = 3.5$ would require a 32-fold increase in simulations (assuming a b-value of 1). Thus, if 1 million simulations were sufficient for $M_t = 3.5$, 32 million simulations need to be performed for $M_t = 5$.

3. Figures 7a–7p shows that the growth in the CIG of the models is punctuated as a result of the temporal clustering of earthquakes.

4.4.4. Consistency Tests

Figure 10 shows the cumulative distribution function (CDF) of quantities γ_N, γ_M and γ_S defined in Savran et al. (2020), which are used to measure the consistency of the models in the N-, M-, and S-test, respectively. These quantities are expected to follow a uniform distribution when evaluated over multiple testing periods if the observations are perfectly consistent with the models.

Using the N-test, we measure the consistency of the four models at the four M_t's. The number of earthquakes forecasted by the model for a given testing period can be deemed consistent at the 95% confidence level if $0.025 < \gamma_N < 0.975$. We find that the numbers of earthquakes forecasted by the models are consistent with the observations for most of the testing periods. Even for Model 1 and $M_t = 3.5$, which has the largest inconsistency, the fraction of testing periods in which $0.025 < \gamma_N < 0.975$ is about 0.75. We also find that the fraction of testing periods for which the numbers forecasted by the models are consistent increases as the threshold M_t of the testing catalog is increased. Given the definition of γ_N in Savran et al. (2020), the shape of the CDF of γ_N from all the testing periods reveals that the models tend to over-forecast the overall number of earthquakes. Furthermore, this tendency to over-forecast progressively decreases with increasing M_t. We also note that models 3 and 4 have the least tendency to over-forecast as the corresponding CDFs of γ_N are closest to a uniform distribution.

γ_M and γ_S defined in Savran et al. (2020) are such that they become undefined when the observed or the simulated catalog has 0 events in some bins. As the occurrence of 0 events in some bins of the testing catalog as well as of the simulated catalogs becomes more prevalent with increasing M_t, the M and the S tests in their current formulation in Savran et al. (2020) cannot be used to properly judge the consistency of the models, especially at large M_t values. Addressing these deficiencies is beyond the scope of this work and will be considered in a future study. In our present work, we have restricted ourselves to assessing the consistency of the models using the M and S tests only for $M_t = 3.5$, as it is in this case alone that we never encounter this issue of 0 earthquake occurrence in a testing period.

From the CDF of γ_M and γ_S, we can conclude that the forecasted magnitude and the spatial distributions of earthquakes by all four models have nearly the same level of (in)consistency with respect to the observed distributions. For both γ_M and γ_S, we find a significant departure from the perfect uniform distribution, although the deviation seems stronger for γ_M. In conclusion, the tests based on γ_N, γ_M and γ_S (Savran et al., 2020) lack the power to distinguish the different models, which justifies our emphasis on the tests based on pseudo-prospective forecasts, which turn out to be much more discriminative and also directly relevant for assessing the dynamical evolution of seismic hazards.
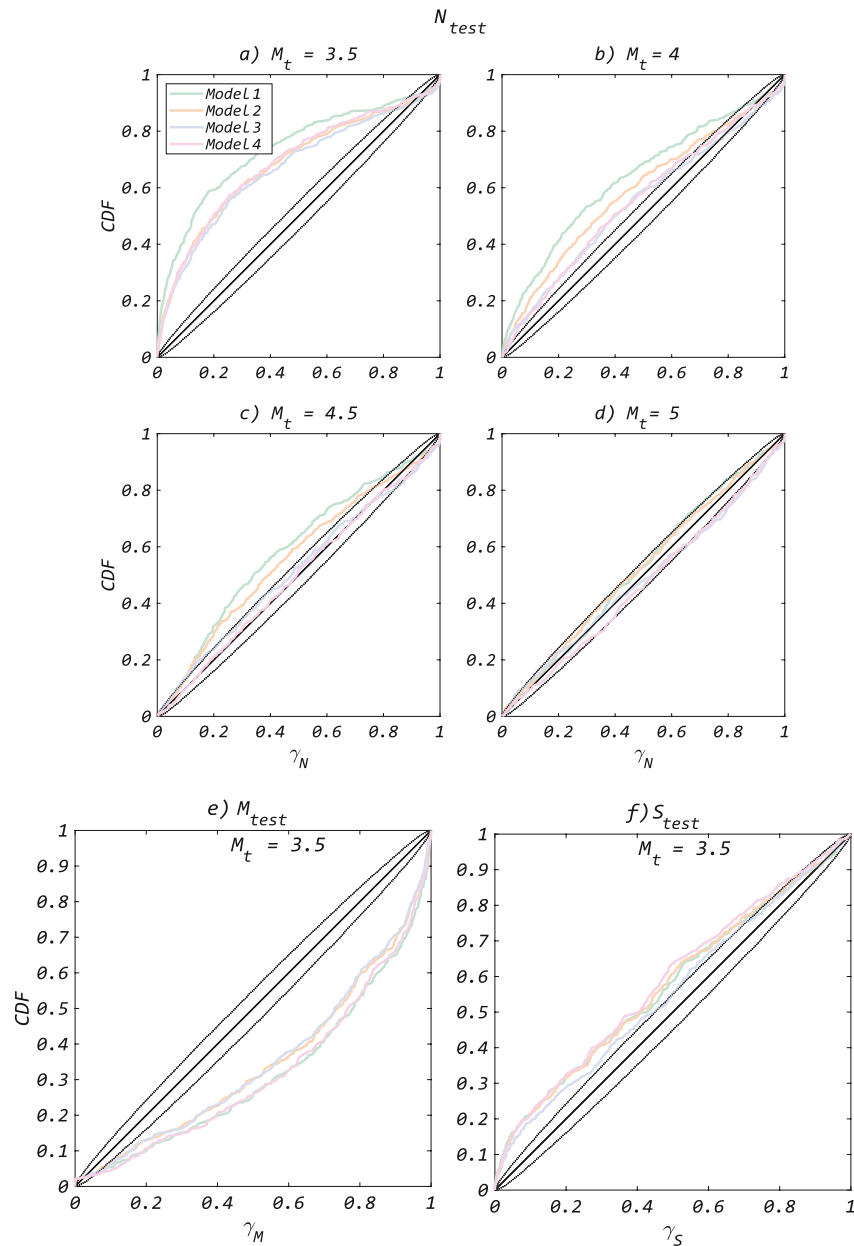
Figure 10. (a–d) Cumulative distribution function (CDF) of quantities γ_N, γ_M, and γ_S defined in Savran et al. (2020) used to measure the consistency of the models in the N-, M-, and S-test, respectively. Each colored curve corresponds to a competing model. The solid and dashed black curves depict the CDF of perfect uniform distribution and its 95% confidence intervals, given 380 data points. The N-test is conducted for four magnitude thresholds (M_t = 3.5, 4, 4.5, 5), while the M- and S-test are conducted only at M_t = 3.5, for reasons explained in Section 4.4.4.

5. What Magnitude Range Dominates Triggering?

5.1. Relative Contributions of Events of Different Magnitudes to Triggering

As Model 4 provides the best available description of seismicity, we are now in position to reexamine from its vantage the claim that small earthquakes dominate the triggering of large earthquakes (Helmstetter, 2003; Helmstetter et al., 2005), which is deeply embedded in the structure of the standard ETAS model and its generalization with a spatially heterogenous background rate in Model 1. We address this question by computing the relative contribution $F(m|M)$ to the triggering of events of magnitude larger than M by events of magnitude m, for

various M's, that Model 4 predicts. For the application to California, we take the average values of the parameters inverted using Model 4 on the ANSS catalog: $a = 1.1$, $\beta_1 = 1.73$, $\beta_2 = 3.05$, $\beta_{bkg} = 2.4$, and $M_0 = 3$.

Denoting m (resp. M) as the magnitude of the triggering (resp. triggered) earthquake, for Model 4, the average number $N_{>M}^{m>M}(m)$ of events of magnitudes larger than M triggered by an event of magnitude m, with $m > M$, is given by

$$N_{>M}^{m>M} = K e^{a(m-M_0)} \left[1 - \frac{\beta_2 \left(1 - e^{-\beta_1(M-M_0)} \right)}{1 - \left(1 - \frac{\beta_1}{\beta_2} \right) e^{-\beta_1(m-M_0)}} \right] \qquad \forall\ m > M \qquad (19)$$

which is Equation S6 in Text S1 of Supporting Information S1, with the correspondence of notations $m_{aft} \to M$ and $m_{main} \to m$.

The number $N_{>M}^{m \leq M}(m)$ of events of magnitudes larger than M triggered by an event of magnitude m, with $m \leq M$, is given by

$$N_{>M}^{m \leq M}(m) = K\beta_1 \frac{e^{(a+\beta_2-\beta_1)(m-M_0)}}{1 - \left(1 - \frac{\beta_1}{\beta_2} \right) e^{-\beta_1(m-M_0)}} e^{-\beta_2(M-M_0)} \qquad \forall\ m \leq M \qquad (20)$$

which is Equation S9 in Text S1 of Supporting Information S1, with the correspondence of notations $m_{aft} \to M$ and $m_{main} \to m$.

Combining the GR distribution of the background events with these "renormalized" productivity laws for triggered events of magnitudes larger than M, the expected contribution of the triggers of magnitudes falling in the interval $[M_0 + i\Delta m, M_0 + (i+1)\Delta m)$ to the triggering of events with magnitude $\geq M$ is given by

$$E\left(M_0 + i\Delta m, M_0 + (i+1)\Delta m | M\right) = \begin{cases} \int_{M_0+i\Delta m}^{M_0+(i+1)\Delta m} N_{>M}^{m>M}(m)\, \beta_{bkg} e^{-\beta_{bkg}(m-M_0)} dm & \forall\ m > M \\ \int_{M_0+i\Delta m}^{M_0+(i+1)\Delta m} N_{>M}^{m \leq M}(m)\, \beta_{bkg} e^{-\beta_{bkg}(m-M_0)} dm & \forall\ m \leq M \end{cases} \qquad (21)$$

Then, the fraction $F(m|M)$ of events (of magnitudes $\geq M$) triggered by events of magnitude in the interval $(m, m + \Delta m)$ is obtained dividing Equation 21 by the total contribution of earthquakes of all the magnitudes, that is, by normalizing Equation 21 by its sum over all triggers' magnitude intervals. Figure 11a shows the resulting $F(m|M)$ for different values of M. Except for M values close to the lower boundary M_0, one can observe the existence of a secondary peak that becomes dominant for larger values of M: the main triggers of large events of magnitude $\geq M$ are events of magnitude M.

The larger the value of M, the more pronounced this phenomenon. This result for $F(m|M)$ is fundamentally different from the prediction of the standard ETAS model, which is recovered for $M = M_0 = 3$ and represented by the blue curve: in this case, Equation 19 reduces to $N_{>M}^{m>M}(m) = K e^{a(m-M_0)}$ which, together with the first expression in Equation 21, recovers the standard ETAS dependence $\sim K e^{(a-\beta_{bkg})(m-M_0)}$. Thus, in contrast to the prediction of the standard ETAS model, earthquakes are mostly triggered by earthquakes of similar or larger magnitudes, implying a fundamental change of paradigm in earthquake forecasting. The previous paradigm that small earthquakes dominate triggering only holds for triggered earthquakes of magnitudes close to the triggering threshold M_0.

5.2. Probability of an Event With Magnitude M to Be Triggered by a Previous Event of Magnitude m

The message of Figure 11a can be complemented by considering the probability $P(m|M)$ that a given event of magnitude M has been triggered by a previous event of magnitude m, which reads, using Bayes theorem:

$$P(m|M) = \frac{P(M|m)P(m)}{\int_{M_0}^{+\infty} P(M|m)P(m)\, dm} \qquad (22)$$
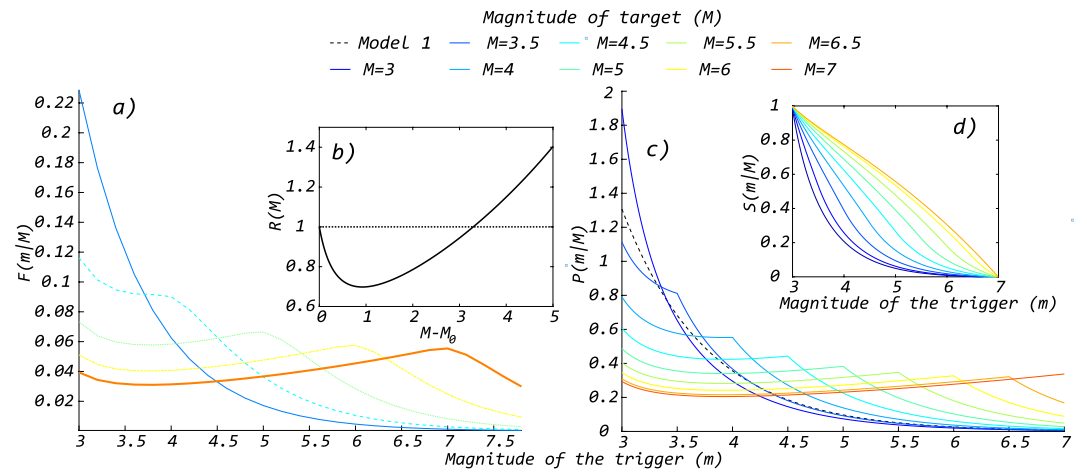
Figure 11. (a) Fraction $F(m|M)$ (obtained from Equation 21 by normalization) of events (of magnitudes $\geq M$) triggered by events of magnitude in the interval $(m, m+ \Delta m)$; Different colors correspond to different values of M, indicated in the legend; The curves are obtained for the parameters of Model 4 given by $K = 0.45$, $a = 1.1$, $\beta_1 = 1.73$, $\beta_2 = 3.05$, $\beta = 2.4$, $M_0 = 3$ and $\Delta m = 0.2$, which best fit the Advanced National Seismic System (ANSS) Californian catalog. (b) Dependence of the ratio $R(M)$ given in Equation 26 as a function of $M - M_0$. (c) Conditional probability that an event of magnitude M has been triggered by an event of magnitude m, $P(m|M)$. The black dashed line corresponds to Model 1. (d) Conditional probability that an event of magnitude M has been triggered by an event of magnitude m or larger, $S(m|M)$. In both panels, different colors correspond to different values of magnitude of target (M), indicated in the legend; The curves are obtained for the parameters of Model 4 given by which best fit the ANSS Californian catalog.

where $P(M|m) \sim K e^{a(m-M_0)} f_{aft}(M|m)$, $f_{aft}(M|m)$ is a kinked GR law with parameters β_1 and β_2, defined using Equations S1 and S2 in Text S1 of Supporting Information S1, and $P(m) = f_{bkg}(m)$ is a standard GR law (Equation 4 in Section 3.1) with parameter β_{bkg}.

In the case where $m < M$, we have

$$P(m|M) \sim \beta_1 \beta_2 e^{-\beta_2 (M-M_0)} \left[e^{(\beta_{bkg}-a-\beta_2)(m-M_0)} \beta_1 \beta_2 \left(e^{\beta_1 (m-M_0)} - 1 \right) \right]^{-1} \tag{23}$$

from which we get, neglecting positive pre-factors:

$$\frac{d}{dm} P(m|M) \sim \beta_2 (a + \beta_2 - \beta_1 - \beta_{bkg}) e^{\beta_1 (m-M_0)} - (\beta_2 - \beta_1)(a + \beta_2 - \beta_{bkg}) \tag{24}$$

For the parameters calibrated on the Californian catalog with Model 4, this gives $\frac{d}{dm} P(m|M) \sim 0.62 e^{1.6(m-3)} - 2.7$. Thus, $P(m|M)$ is an increasing function of m for $m > 3.9$, and a decreasing function for $m < 3.9$.

In the case where $m > M$, we have

$$P(m|M) \sim \beta_1 \beta_2 e^{-\beta_1 (M-M_0)} \left[\beta_1 e^{(\beta_{bkg}-a-\beta_1)(m-M_0)} + \beta_2 e^{(\beta_{bkg}-a)(m-M_0)} \left(1 - e^{-\beta_1 (m-M_0)} \right) \right]^{-1} \tag{25}$$

After a little algebra and omitting all positive pre-factors, we finally find that $\frac{d}{dm} P(m|M) \sim \beta_2 (a - \beta_{bkg}) e^{\beta_1 (m-M_0)} - (\beta_2 - \beta_1)(a + \beta_1 - \beta_{bkg})$ which takes the form $\frac{d}{dm} P(m|M) \sim -4 e^{1.6(m-3)} - 0.45$, and is always negative.

Putting these results together and assuming $M_0 = 3$, $P(m|M)$ is first decreasing for $3 < m < 3.9$, then increasing for $3.9 < m < M$, and then decreasing again for $m > M$. For $3 < M < 3.9$, $P(m|M)$ is monotonically decreasing with m for all m's. This implies that, apart from contributions at magnitudes close to M_0, the most likely triggering event of an earthquake of magnitude M has itself a magnitude M.

Figure 11c shows the behavior of $P(m|M)$, which progressively delocalizes toward large magnitude m as M increases. For comparison, the pdf corresponding to the ETAS model (Model 1), which does not depend on M, is plotted using a black dashed line.

Integrating $P(m|M)$ with m, we can estimate $S(m|M)$, the share of events with magnitude larger than m that trigger an event of magnitude M (see Figure 11d). Consider for instance the median value $S(m|M) = 0.50$; then the corresponding triggering magnitude increases from $m = 3.4$ for $M = 3$, to $m = 5.24$ for $M = 7$. Thus, forecasting correctly half of the triggered events of magnitude M requires detecting properly only the largest events if M is large. However, it must be kept in mind that forecasting the background events of similar magnitude M requires additional ingredients or mechanisms beyond Model 4, which assumes them as Poissonian in time.

5.3. Relative Propensity of an Event With Magnitude M to Trigger Another Event of Magnitude M

Using Equation 23, we obtain a further diagnostic

$$\frac{P(m = M|M)}{P(m = M_0|M)} = \frac{\beta_1 e^{-(\beta_{bkg}-a-\beta_2)(M-M_0)}}{\left(\beta_1 + \beta_2 \left(e^{\beta_1(M-M_0)} - 1\right)\right)} := R(M) \tag{26}$$

which quantifies the relative contributions of events of magnitude M to those of magnitude close to the lower threshold M_0 in their propensity to trigger an event of magnitude M. Figure 11b plots the dependence of this ratio $R(M)$ as a function of M. There is a cross-over magnitude M^* such that $R(M) < 1$ for $M < M^*$ and $R(M) > 1$ for $M > M^*$. Thus, the triggering of earthquakes with $M < M^*$ is dominated by previous events with magnitude M_0, while the triggering of earthquakes with $M > M^*$ is dominated by previous events with magnitude M. This provides another confirmation of the major difference of Model 4 compared with the conclusions drawn using the standard ETAS model, where the triggering of events of any magnitude range is dominated by events with magnitude M_0, since $a - \beta_{bkg} < 0$ holds true in general.

5.4. Discussion

In our calibration of the models, we considered that $M_0 = M_c$, that is, the minimum magnitude M_c above which all events are thought to have been detected is equal to the minimum magnitude M_0 at which earthquakes can trigger other earthquakes. However, the true M_0 is likely much smaller than M_c (Sornette & Werner, 2005). Since events with $M < M^*$ are either unpredictable background events or are mainly triggered by unobserved events, it follows that the performance of any statistical forecasting technique can be expected to be low for such a magnitude range. On the other hand, provided that $M^* > M_c$, events of magnitude $M > M^*$ are either background events or triggered by observable events of similar magnitude. Our best Model 4 should thus perform better on that magnitude range. One should, however, keep in mind that the qualitative behavior of $P(m|M)$ in other areas may be different, depending on the local set of exponents (a, β_1, β_2, and β_{bkg}). For California, Figure 11b shows that $M^* \simeq M_0 + 3.3$.

The productivity exponent (a) defined in the first term $Ke^{a(m_i-M_0)}$ of Equation 2 is underestimated when the anisotropy of the spatial organization of the seismicity is not considered during calibration of the parameters of the ETAS model (Guo et al., 2019; Hainzl et al., 2008; Zhuang et al., 2019). This underestimation of a then translates into a stronger apparent dominance of the smaller earthquakes in the overall triggering process. In contrast, if the anisotropy of the stress distribution due to an earthquake is accounted for during ETAS calibration, somewhat larger values of a are obtained. Thus, Helmstetter (2003) preferred to use the value of a estimated through nonparametric methods. The value of the productivity exponent reported in Helmstetter (2003) is 1.84 (compared to $\approx$1.1 reported in the present study). This former exponent is in line with what is usually obtained when anisotropy is accounted for. As Helmstetter (2003) was published at a time when the community was still unaware of the intricacies of conditional magnitude distribution of the aftershocks (as considered here), the natural conclusion at the time was that small earthquakes dominate triggering. Even with a larger a value obtained by accounting for anisotropy, our present finding of secondary peaks shown in Figure 11a remains unchanged. This is because these peaks result from the nature of the conditional magnitude distribution of aftershocks, which have been hitherto not considered in the question about the dominance of small versus large earthquakes in triggering. If anything, accounting for anisotropy increases the fertility exponent and thus tends to decrease the contribution of small earthquakes relative to large earthquakes, strengthening the main findings of this study.

6. Conclusion

We have compared four different stochastic models of seismicity and found that, for California, the model with parameters dependent on the magnitude of the triggering shock for the Omori time kernel and the magnitude distribution of triggered events (Model 4) offers the best performance in forecasting future seismicity. The ingredients of the best model can be rationalized from the norm and symmetry of the fluctuations of the stress and deformation tensors.

Model 4 predicts that earthquakes up to a cross-over magnitude $M*$ are preferably triggered by the smallest possible triggering shocks; in contrast, earthquakes with magnitude M above $M*$ are preferably triggered by events with a similar magnitude M. Thus, forecasting small events is inherently unfeasible due to a lack of observations of their (small magnitude) triggers. In contrast, a significant fraction of large events is triggered by large events. Being large, these triggering events are observed, and their catalog is close to complete for the global Earth seismicity. Thus, forecasting large events preceded by a set of seismic precursors (Mignan, 2011) would not be spoiled by the noise created by the myriad of small events whose recording is incomplete or whose parameters are highly uncertain. Given that only large events are relevant for risk assessment and protection, our results suggest that modeling the spatial anisotropy of sources of large spatial dimensions defines the next important step toward a successful operational earthquake forecasting system (as large earthquakes cannot be considered as point sources).

Our improved ability to forecast applies only to the subset of triggered earthquakes, while background events remain statistically unpredictable according to the investigated models. To improve, a better description of the background (non-triggered) part of the seismic catalogs is needed. This has been partially achieved in the present set of models by introducing a spatially heterogeneous structure for the set of background events. But there is no time dependence. A substantial increase of forecasting power should focus on modeling possible time variations of local or regional strain fields and other geophysical fields that could be correlated with seismicity (Freund et al., 2021).

Data Availability Statement

The data used for this analysis are available through the website https://earthquake.usgs.gov/earthquakes/search/ (U.S. Geological Survey, Earthquake Hazards Program, 2017) and https://service.scedc.caltech.edu/ftp/QTMcatalog/.

References

Bacry, E., Mastromatteo, I., & Muzy, J.-F. (2015). Hawkes processes in finance. *Market Microstructure and Liquidity*, *1*(01), 1550005. https://doi.org/10.1142/s2382626615500057

Bowman, D., Ouillon, G., Sammis, C., Sornette, A., & Sornette, D. (1998). An observational test of the critical earthquake concept. *Journal of Geophysical Research*, *103*(B10), 24359–24372. https://doi.org/10.1029/98jb00792

Crane, R., & Sornette, D. (2008). Robust dynamic classes revealed by measuring the response function of a social system. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(41), 15649–15653. https://doi.org/10.1073/pnas.0803685105

Davidsen, J., Gu, C., & Baiesi, M. (2015). Generalized Omori–Utsu law for aftershock sequences in southern California. *Geophysical Journal International*, *201*(2), 965–978. https://doi.org/10.1093/gji/ggv061

Dieterich, J. (1994). A constitutive law for rate of earthquake production and its application to earthquake clustering. *Journal of Geophysical Research*, *99*(B2), 2601–2618. https://doi.org/10.1029/93jb02581

Ebel, J. E., Chambers, D. W., Kafka, A. L., & Baglivo, J. A. (2007). Non-Poissonian earthquake clustering and the hidden Markov model as bases for earthquake forecasting in California. *Seismological Research Letters*, *78*(1), 57–65. https://doi.org/10.1785/gssrl.78.1.57

Evison, F., & Rhoades, D. (2001). Model of long-term seismogenesis.

Faenza, L., & Marzocchi, W. (2010). The proportional hazard model as applied to the CSEP forcasting area in Italy. *Annals of Geophysics*, *53*(3), 77–84.

Freund, F., Ouillon, G., Scoville, J., & Sornette, D. (2021). Earthquake precursors in the light of peroxy defects theory: Critical review of systematic observations. *The European Physical Journal - Special Topics*, *230*(1), 7–46. https://doi.org/10.1140/epjst/e2020-000243-x

Gerstenberger, M. C., Wiemer, S., Jones, L. M., & Reasenberg, P. A. (2005). Real-time forecasts of tomorrow's earthquakes in California. *Nature*, *435*(7040), 328–331. https://doi.org/10.1038/nature03622

Guo, Y., Zhuang, J., & Ogata, Y. (2019). Modeling and forecasting aftershocks can be improved by incorporating rupture geometry in the ETAS model. *Geophysical Research Letters*, *46*(22), 12881–12889. https://doi.org/10.1029/2019gl084775

Gutenberg, B., & Richter, C. F. (1944). Frequency of earthquakes in California. *Bulletin of the Seismological Society of America*, *34*(4), 185–188. https://doi.org/10.1785/bssa0340040185

Hainzl, S. (2016a). Apparent triggering function of aftershocks resulting from rate-dependent incompleteness of earthquake catalogs. *Journal of Geophysical Research: Solid Earth*, *121*(9), 6499–6509. https://doi.org/10.1002/2016jb013319

Hainzl, S. (2016b). Rate-dependent incompleteness of earthquake catalogs. *Seismological Research Letters*, *87*(2A), 337–344. https://doi.org/10.1785/0220150211

Hainzl, S., Christophersen, A., & Enescu, B. (2008). Impact of earthquake rupture extensions on parameter estimations of point-process models. *Bulletin of the Seismological Society of America*, *98*(4), 2066–2072. https://doi.org/10.1785/0120070256

Hawkes, A. G. (1971). Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, *58*(1), 83–90. https://doi.org/10.1093/biomet/58.1.83

Helmstetter, A. (2003). Is earthquake triggering driven by small earthquakes? *Physical Review Letters*, *91*(5), 058501. https://doi.org/10.1103/physrevlett.91.058501

Helmstetter, A., Kagan, Y. Y., & Jackson, D. D. (2005). Importance of small earthquakes for stress transfers and earthquake triggering. *Journal of Geophysical Research*, *110*(B5), B05S08. https://doi.org/10.1029/2004jb003286

Helmstetter, A., Kagan, Y. Y., & Jackson, D. D. (2006). Comparison of short-term and time-independent earthquake forecast models for southern California. *Bulletin of the Seismological Society of America*, *96*(1), 90–106. https://doi.org/10.1785/0120050067

Helmstetter, A., & Sornette, D. (2003). Predictability in the ETAS model of interacting triggered seismicity. *Journal of Geophysical Research*, *108*, 2482. https://doi.org/10.1029/2003JB002485

Holliday, J. R., Nanjo, K. Z., Tiampo, K. F., Rundle, J. B., & Turcotte, D. L. (2005). Earthquake forecasting and its verification. *Nonlinear Processes in Geophysics*, *12*(6), 965–977. https://doi.org/10.5194/npg-12-965-2005

Imoto, M., Hurukawa, N., & Ogata, Y. (1990). Three-dimensional spatial variations of b-value in the Kanto area, Japan. *Zishin*, *43*(3), 321–326. https://doi.org/10.4294/zisin1948.43.3_321

Keilis-Borok, V. I., & Kossobokov, V. (1990). Times of increased probability of strong earthquakes diagnosed by algorithm M8 in Japan and adjacent territories. *Journal of Geophysical Research*, *95*(B8), 12413–12422. https://doi.org/10.1029/jb095ib08p12413

King, G. C., & Bowman, D. D. (2003). The evolution of regional seismicity between large earthquakes. *Journal of Geophysical Research*, *108*(B2), 2096. https://doi.org/10.1029/2001jb000783

Kobayashi, R., & Lambiotte, R. (2016). *Tideh: Time-dependent hawkes process for predicting retweet dynamics*. In *Tenth international AAAI conference on web and social media*.

Mignan, A. (2011). Retrospective on the Accelerating Seismic Release (ASR) hypothesis: Controversy and new horizons. *Tectonophysics*, *505*(1–4), 1–16. https://doi.org/10.1016/j.tecto.2011.03.010

Mizrahi, L., Nandan, S., & Wiemer, S. (2021). Embracing data incompleteness for better earthquake forecasting. *Earth and Space Science Open Archive*, *25*. https://doi.org/10.1002/essoar.10507003.1

Mogi, K. (1969). Some features of recent seismic activity in and near Japan (2), activity before and after large earthquakes. *Bulletin of the Earthquake Research Institute*, *47*, 395–417.

Nandan, S., Kamer, Y., Ouillon, G., Hiemer, S., & Sornette, D. (2021). Global models for short-term earthquake forecasting and predictive skill assessment. *The European Physical Journal - Special Topics*, *230*(1), 425–449. https://doi.org/10.1140/epjst/e2020-000259-3

Nandan, S., Ouillon, G., & Sornette, D. (2019). Magnitude of earthquakes controls the size distribution of their triggered events. *Journal of Geophysical Research: Solid Earth*, *124*(3), 2762–2780. https://doi.org/10.1029/2018jb017118

Nandan, S., Ouillon, G., Sornette, D., & Wiemer, S. (2019a). Forecasting the full distribution of earthquake numbers is fair, robust, and better. *Seismological Research Letters*, *90*(4), 1650–1659. https://doi.org/10.1785/0220180374

Nandan, S., Ouillon, G., Sornette, D., & Wiemer, S. (2019b). Forecasting the rates of future aftershocks of all generations is essential to develop better earthquake forecast models. *Journal of Geophysical Research: Solid Earth*, *124*(8), 8404–8425. https://doi.org/10.1029/2018jb016668

Nandan, S., Ouillon, G., Woessner, J., Sornette, D., & Wiemer, S. (2016). Systematic assessment of the static stress triggering hypothesis using interearthquake time statistics. *Journal of Geophysical Research: Solid Earth*, *121*(3), 1890–1909. https://doi.org/10.1002/2015jb012212

Nandan, S., Ram, S. K., Ouillon, G., & Sornette, D. (2021). Is seismicity operating at a critical point? *Physical Review Letters*, *126*(12), 128501. https://doi.org/10.1103/physrevlett.126.128501

Narteau, C., Shebalin, P., & Holschneider, M. (2005). Onset of power law aftershock decay rates in southern California. *Geophysical Research Letters*, *32*(22), L22312. https://doi.org/10.1029/2005gl023951

Nichols, K., & Schoenberg, F. P. (2014). Assessing the dependency between the magnitudes of earthquakes and the magnitudes of their aftershocks. *Environmetrics*, *25*(3), 143–151. https://doi.org/10.1002/env.2268

Ogata, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical Association*, *83*(401), 9–27. https://doi.org/10.1080/01621459.1988.10478560

Omori, F. (1894). On the after-shocks of earthquakes. *The Journal of the College of Science, Imperial University of Tokyo, Japan*, *7*, 111–200.

Ouillon, G., & Sornette, D. (2005). Magnitude-dependent Omori law: Theory and empirical study. *Journal of Geophysical Research*, *110*(B4), B04306. https://doi.org/10.1029/2004jb003311

Ouillon, G., Sornette, D., & Ribeiro, E. (2009). Multifractal Omori law for earthquake triggering: New tests on the California, Japan and worldwide catalogues. *Geophysical Journal International*, *178*(1), 215–243. https://doi.org/10.1111/j.1365-246x.2009.04079.x

Reid, H. F. (1910). *The mechanics of the earthquake*. In *The California earthquake of April 18, 1906, report of the state earthquake investigation commission*.

Ross, Z. E., Trugman, D. T., Hauksson, E., & Shearer, P. M. (2019). Searching for hidden earthquakes in southern California. *Science*, *364*(6442), 767–771. https://doi.org/10.1126/science.aaw6888

Rundle, J., Tiampo, K., Klein, W., & Martins, J. S. (2002). Self-organization in leaky threshold systems: The influence of near-mean field dynamics and its implications for earthquakes, neurobiology, and forecasting. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(1), 2514–2521. https://doi.org/10.1073/pnas.012581899

Saichev, A., & Sornette, D. (2005). Vere-Jones' self-similar branching model. *Physical Review*, *72*(5), 056122. https://doi.org/10.1103/physreve.72.056122

Saichev, A., & Sornette, D. (2006). Power law distribution of seismic rates: Theory and data. *European Physical Journal B: Condensed Matter and Complex Systems*, *49*(3), 377–401. https://doi.org/10.1140/epjb/e2006-00075-3

Savran, W. H., Werner, M. J., Marzocchi, W., Rhoades, D. A., Jackson, D. D., Milner, K., et al. (2020). Pseudoprospective evaluation of UCERF3-ETAS forecasts during the 2019 Ridgecrest sequence. *Bulletin of the Seismological Society of America*, *110*(4), 1799–1817. https://doi.org/10.1785/0120200026

Scholz, C. (1968). Microfractures, aftershocks, and seismicity. *Bulletin of the Seismological Society of America*, *58*(3), 1117–1130.

Schwartz, D. (1981). Characteristic earthquakes on intraplate normal faults. *Earthquake Notes*, *52*, 71.

Schwartz, D. P., & Coppersmith, K. J. (1984). Fault behavior and characteristic earthquakes: Examples from the Wasatch and San Andreas fault zones. *Journal of Geophysical Research*, *89*(B7), 5681–5698. https://doi.org/10.1029/jb089ib07p05681

Shcherbakov, R., Turcotte, D. L., & Rundle, J. B. (2004). A generalized Omori's law for earthquake aftershock decay. *Geophysical Research Letters*, *31*(11), L11613. https://doi.org/10.1029/2004gl019808

Sobolev, G., & Tyupkin, Y. S. (1997). Low-seismicity precursors of large earthquakes in Kamchatka. *Volcanology and Seismology*, *18*.

Sornette, D., & Ouillon, G. (2005). Multifractal scaling of thermally activated rupture processes. *Physical Review Letters*, *94*(3), 038501. https://doi.org/10.1103/physrevlett.94.038501

Sornette, D., & Werner, M. J. (2005). Constraints on the size of the smallest triggering earthquake from the epidemic-type aftershock sequence model, Båth's law, and observed aftershock sequences. *Journal of Geophysical Research*, *110*(B8), B08304. https://doi.org/10.1029/2004jb003535

Spassiani, I., & Sebastiani, G. (2016). Magnitude-dependent epidemic-type aftershock sequences model for earthquakes. *Physical Review*, *93*(4), 042134. https://doi.org/10.1103/physreve.93.042134

Tinti, S., & Mulargia, F. (1987). Confidence intervals of b values for grouped magnitudes. *Bulletin of the Seismological Society of America*, *77*(6), 2125–2134.

Tsai, C.-Y., Ouillon, G., & Sornette, D. (2012). New empirical tests of the multifractal Omori law for Taiwan. *Bulletin of the Seismological Society of America*, *102*(5), 2128–2138. https://doi.org/10.1785/0120110237

U.S. Geological Survey, Earthquake Hazards Program. (2017). Advanced national seismic system (ANSS) comprehensive catalog of earthquake events and products: Various. https://doi.org/10.5066/F7MS3QZH

Utsu, T., Ogata, Y., S, R., & Matsu'ura (1995). The centenary of the Omori formula for a decay law of aftershock activity. *Journal of Physics of the Earth*, *43*(1), 1–33. https://doi.org/10.4294/jpe1952.43.1

Veen, A., & Schoenberg, F. P. (2008). Estimation of space–time branching process models in seismology using an em–type algorithm. *Journal of the American Statistical Association*, *103*(482), 614–624. https://doi.org/10.1198/016214508000000148

Vere-Jones, D. (2005). A class of self-similar random measure. *Advances in Applied Probability*, *37*(4), 908–914. https://doi.org/10.1017/s0001867800000598

Wang, Q., Jackson, D. D., & Zhuang, J. (2010). Missing links in earthquake clustering models. *Geophysical Research Letters*, *37*(21), L21307. https://doi.org/10.1029/2010gl044858

Wyss, M., Shimazaki, K., & Urabe, T. (1996). Quantitative mapping of a precursory seismic quiescence to the Izu–Oshima 1990 (M6. 5) earthquake, Japan. *Geophysical Journal International*, *127*(3), 735–743. https://doi.org/10.1111/j.1365-246x.1996.tb04052.x

Yamashita, T., & Knopoff, L. (1989). A model of foreshock occurrence. *Geophysical Journal International*, *96*(3), 389–399. https://doi.org/10.1111/j.1365-246x.1989.tb06003.x

Yin, X.-C., Chen, X.-Z., Song, Z.-P., & Yin, C. (1995). A new approach to earthquake prediction: The Load/Unload Response Ratio (LURR) Theory. *Pure and Applied Geophysics*, *145*(3), 701–715. https://doi.org/10.1007/bf00879596

Zhuang, J., Murru, M., Falcone, G., & Guo, Y. (2019). An extensive study of clustering features of seismicity in Italy from 2005 to 2016. *Geophysical Journal International*, *216*(1), 302–318.

Zhuang, J., Ogata, Y., & Vere-Jones, D. (2002). Stochastic declustering of space-time earthquake occurrences. *Journal of the American Statistical Association*, *97*(458), 369–380. https://doi.org/10.1198/016214502760046925

Zhuang, J., Werner, M. J., & Harte, D. S. (2013). Stability of earthquake clustering models: Criticality and branching ratios. *Physical Review*, *88*(6), 062109. https://doi.org/10.1103/physreve.88.062109