

Launching AoyuX: A 25-Year Pseudo-Prospective Earthquake Forecasting Experiment at the China Seismic Experimental Site

Jiawei Li¹, Qingyuan Zhang^{1,2,3,4}, Didier Sornette¹

¹Institute of Risk Analysis, Prediction and Management (Risks-X), Academy for Advanced Interdisciplinary Studies, Southern University of Science and Technology (SUSTech), Shenzhen, China.

²School of Systems Science, Beijing Normal University, Beijing, China.

³International Academic Center of Complex Systems, Beijing Normal University, Zhuhai, China.

⁴Department of Systems Science, Faculty of Arts and Sciences, Beijing Normal University, Zhuhai, China.

Correspondence to: Jiawei Li (lijw@cea-igp.ac.cn)

Jiawei Li, Qingyuan Zhang, and Didier Sornette contributed equally to this work

Abstract. Forecast models in statistical seismology are commonly evaluated with log-likelihood scores of the full distribution $P(n)$ of earthquake numbers, yet heavy tails and out-of-range observations (realized counts exceeding all simulations) can bias model ranking. We develop a tail-aware evaluation framework that estimates cell-wise $P(n)$ using adaptive Gaussian kernel density estimation and tests three strategies for handling out-of-range counts (power law extrapolation, fixed minimum probability, and exclusion). Using the AoyuX platform, we perform a ~ 25 -year (2000–2023) month-by-month pseudo-prospective forecast experiment in the China Seismic Experimental Site (CSES), comparing Epidemic-Type Aftershock Sequence (ETAS) model with a homogeneous background (ETAS _{μ}) to a spatially heterogeneous variant (ETAS _{$\mu(x,y)$}) across six spatial resolutions and five magnitude thresholds; each forecast window is supported by 100,000 simulations. Empirical probability density functions (PDFs) of counts per cell are well described by power laws with exponents $a \approx 1.40 \pm 0.21$ across all settings. Using previous theoretical results, this provides a robust estimate of the productivity exponent, $\alpha = 0.57 \pm 0.08$ using a b -value equal to 0.8, providing a valuable quantification of this key parameter in aftershock modeling. Model ranking is sensitive to how the tail of the full distribution $P(n)$ of earthquake counts is treated: power law extrapolation is both theoretically justified and empirically the most robust, while the apparent superiority of the fixed-probability approach in a few coarse low-magnitude cases can be attributed to noise. Cumulative information gain (CIG) shows that ETAS _{$\mu(x,y)$} outperforms ETAS _{μ} in data-rich configurations (e.g., $M = 3.0$ to 3.5 and whole-region tests), whereas in data-poor settings stochastic fluctuations dominate. A coefficient-of-variation analysis of per-window log-likelihood differences distinguishes genuine upward trends in CIG from noise-dominated fluctuations. By aligning a fat-tail-aware scoring methodology with an open testing platform, our work advances fair and statistically grounded assessment of earthquake forecasting models for the CSES and beyond.

1 Introduction

Advancing earthquake forecasting (probabilistic) and/or prediction (deterministic) benefits from a transparent and iterative research paradigm: build models that explain observed seismicity; test them prospectively or pseudo-prospectively against withheld or genuinely future data; use the resulting feedback to validate assumptions; and then retain, revise, or replace model components before repeating the cycle. Such a loop of construction, testing, correction, and refinement underpins cumulative progress in the field (Kagan & Jackson, 2000; Werner et al., 2011; Ogata, 2022). Crucial to this paradigm is rigorous, unbiased performance evaluation: determining, with statistical significance, whether a candidate model truly outperforms a designated benchmark (the null hypothesis, H_0) rather than relying on ad hoc or subjective criteria. The international Collaboratory for the

Study of Earthquake Predictability (CSEP) and RichterX global earthquake forecasting platform formalized this ethos with standardized tests, scoring rules, and community benchmarks, demonstrating both the value and the challenges of fair model comparison (Schorlemmer et al., 2007; 2010; Taroni et al., 2018; Kamer et al., 2021; Nandan et al., 2021a; Bayliss et al., 2022).

In practice, a widely adopted evaluation framework is to run pseudo-prospective experiments over consecutive, non-overlapping time windows. Genuine prospective forecasting, evaluation on data not yet observed, should be stressed, with pseudo-prospective experiments serving as a preparatory or complementary step when fully prospective deployment is not yet feasible. For each time window and for each model–benchmark pair, one computes the full probability distribution $P(n)$ of earthquake numbers n at the grid-cell level (given a chosen magnitude threshold M_i and spatial resolution) and then scores the model by the log-likelihood obtained as the logarithm of the products over all cells of the $P(n_i)$'s, where n_i are the observed counts in each i -th grid (see Section 2.2 for more details). Forecasting and scoring the distribution, rather than only its mean rate, has been shown to be both fair and robust, and it aligns naturally with aftershock-cascade physics (Ogata et al., 2013; Nandan et al., 2019a; 2019b). The approach is compatible with a range of models used operationally, including Epidemic-Type Aftershock Sequence (ETAS) model and its variants (Zhuang, 2011; Werner et al., 2011; Hainzl et al., 2024; Mizrahi et al., 2024; Han et al., 2025).

Two methodological issues arise when estimating the full distribution $P(n)$:

(1) Randomness. Because $P(n)$ is typically derived from a large but finite ensemble of simulated catalogs, sampling variability introduces artificial roughness into the distribution. To mitigate this effect, studies have traditionally adopted parametric count models, such as Poisson, negative-binomial or Beta-negative binomial approximations, or, more recently, non-parametric smoothers like adaptive kernel density estimators (Harte, 2017; Kagan, 2010; 2017; Molchan & Varini, 2024; Nandan et al., 2019a; Taroni et al., 2023; Werner et al., 2011). The use of the Poisson distribution is particularly natural when a forecast specifies only the mean rate, since, under the maximum entropy principle, the Poisson distribution is the most parsimonious choice: it imposes the fewest assumptions beyond the mean number of events. However, once a forecast conveys information beyond the mean rate, restricting the distribution to Poisson is no longer theoretically justified. In this context, non-parametric approaches, especially adaptive kernel density estimators, have emerged as effective tools for stabilizing likelihood-based evaluations while minimizing model-induced bias (Nandan et al., 2019a; 2021a; 2021b).

(2) Fat-tailed behavior. The frequency distribution of counts across cells is empirically fat-tailed, often close to a power law in its upper tail, reflecting clustered triggering and cascade dynamics (Kagan, 2017; Nandan et al., 2019a; Nandan et al., 2021a; see also Figure 6 in the present study as an example). This power law property is in fact predicted by theory for ETAS models of earthquake triggering (Saichev et al., 2005; Saichev & Sornette, 2006; 2007). Yet most evaluation pipelines, especially those relying on non-parametric smoothing, do not explicitly constrain or extrapolate the tail (Nandan et al., 2019a; Nandan et al., 2021b; Mizrahi et al., 2024; Han et al., 2025). In cells impacted by aftershocks of great earthquakes, observed counts can exceed the maximum of the simulated ensemble, requiring an out-of-range tail treatment (e.g., extrapolation vs. truncation). How such tail handling affects comparative model ranking has not been systematically quantified.

Increasing the number of simulated catalogs can, to some extent, mitigate the two issues mentioned above. However, simply enlarging the ensemble size is not a practical solution in real applications. Doing so would not only lead to a substantial waste of computational resources but also significantly prolong the running time, thereby raising the overall computational cost. More

importantly, even an arbitrarily large number of simulated catalogs cannot fully guarantee a perfect resolution of the underlying problems, as the fundamental limitations are methodological rather than merely statistical.

This study addresses that gap. We develop an evaluation framework that (i) estimates the cell-wise probability distribution $P(n)$ of earthquake numbers n with adaptive smoothing while (ii) explicitly incorporating fat-tail behaviour and principled tail-extrapolation rules for out-of-range observations, and (iii) assesses sensitivity of model ranking to these choices. Our analysis complements recent advances in short-term forecasting across diverse regions and model classes, including ETAS generalizations and alternative point-process formulations (Zhuang, 2011; Werner et al., 2011; Davis et al., 2024; Hainzl et al., 2024; Mizrahi et al., 2024; Han et al., 2025).

To enable transparent and repeatable testing at scale, we introduce AoyuX, an online research platform dedicated to earthquake predictability studies in China. Developed by the Institute of Risk Analysis, Prediction and Management (Risks-X) at the Southern University of Science and Technology (SUSTech) with the China Earthquake Networks Center and other institutions as partner organizations, AoyuX integrates state-of-the-art statistical seismology, centered on ETAS-type models that jointly forecast time–space–magnitude probabilities, with an extensible architecture for benchmarking, visualization, and (future) model-comparison challenges. The initial operational focus is the China Seismic Experimental Site (CSES) in Sichuan-Yunnan region, one of the nation’s most instrumented and seismically active regions (Li et al., 2022; Wu et al., 2019; Wu & Li, 2021a; 2021b; Wu, 2022). Since June 2024, AoyuX has issued monthly, quarterly, semi-annual, and annual forward forecast bulletins for the CSES (see Open Research), thereby establishing a sustained prospective and pseudo-prospective testing environment. Looking ahead, AoyuX will fuse geophysical observations with AI-assisted and physics-based components to support mid-term, short-term, and imminent forecasting, and to facilitate communication between experts and the public.

In the present study, we conducted a ~25-year month-by-month pseudo-prospective forecasting experiment for the China Seismic Experimental Site (CSES), using the ETAS model together with a spatially heterogeneous variant in which the background seismicity rate varies across space. We further evaluated the impact of innovative fat-tail treatments for out-of-range observations and investigated under which conditions the cumulative information gain (CIG) curves exhibit a clear long-term upward trend, or alternatively, when such a trend is obscured by stochastic fluctuations. By integrating a fat-tail-aware scoring methodology with a dedicated open testing platform, our work advances both the methodological foundations and the research infrastructure required for fair and statistically rigorous assessment of earthquake forecasting models in China and beyond.

2 Methods

2.1 The Epidemic-Type Aftershock Sequence (ETAS) model for earthquake forecasting

The Epidemic-Type Aftershock Sequence (ETAS) model is a widely used formulation of the self-exciting Hawkes point process, specifically adapted to characterize seismicity. Unlike Markovian processes, the ETAS model generally exhibits non-Markovian behavior when its temporal memory kernel deviates from an exponential form (Hawkes & Oakes, 1974; Kagan & Knopoff, 1981, 1987; Ogata, 1988, 1998; Kanazawa & Sornette, 2020; 2024). At any spatiotemporal location (x, y, t) , the model defines the expected seismicity rate, or intensity function $\lambda(x, y, t)$, which depends on the complete history of seismic events up to time t , denoted as H_t :

$$\begin{aligned}
\lambda(t, x, y \mid H_t) &= \mu + \sum_{i:t_i < t} F(m_i) T(t - t_i) S(x - x_i, y - y_i, m_i) \\
&= \mu + \sum_{i:t_i < t} K e^{\alpha(m - M_{co})} \cdot \frac{T_{\text{norm}} \cdot e^{-\frac{t-t_i}{\tau}}}{(t - t_i + c)^{1+\omega}} \cdot \frac{S_{\text{norm}}}{[(x - x_i)^2 + (y - y_i)^2 + d e^{\gamma(m_i - M_{co})}]^{1+\rho}}
\end{aligned} \tag{1}$$

Here, the intensity λ consists of a background rate μ and contributions from all prior earthquakes. Each past event i , characterized by its magnitude m_i and location (x_i, y_i, t_i) , influences future seismicity through a separable spatiotemporal triggering kernel, weighted by a fertility (or productivity) function $F(m_i)$, parameterized by constants K and α . The spatiotemporal triggering kernel is typically defined as the product of two components: a temporal kernel described by the Omori-Utsu law $T(t - t_i)$, with parameters c , ω , and τ , and a spatial kernel represented by a Green's function with parameters d , γ , and ρ , together describing the temporal and spatial dependencies of aftershock triggering. The normalization constants T_{norm} and S_{norm} ensure that the temporal and spatial kernels are valid probability density functions (PDFs). This formulation reflects a commonly adopted choice for the spatiotemporal triggering kernel in ETAS models (Nandan et al., 2021b; 2022), although alternative kernel forms have also been proposed in the literature (Helmstetter & Sornette, 2002; Felzer et al., 2002, 2003; Zhuang et al., 2002, 2005).

The ETAS model admits two conceptually distinct yet mathematically equivalent interpretations (Helmstetter & Sornette, 2002; Li et al., 2025a):

(a) Branching interpretation (Hawkes & Oakes, 1974; Kagan, 1991; Sornette & Werner, 2005). Each observed earthquake can be probabilistically linked to a parent event, forming a cascade of primary and secondary aftershocks in a tree-like structure. In this framework, background events follow a Poisson process, while each event may trigger its own sequence of aftershocks. Every tree configuration is associated with a probability that quantifies the likelihood of a specific parent-offspring association, given the observed catalog. These probabilities are derived from the ETAS intensity function $\lambda(x, y, t)$ (Zhuang et al., 2002).

(b) Collective influence interpretation (Helmstetter & Sornette, 2002). Each non-background event arises from the cumulative triggering effect of all preceding earthquakes. In this view, the intensity function $\lambda(x, y, t)$ represents the combined influence of historical seismicity at a given location and time, superimposed on the background rate.

The mathematical equivalence between these two interpretations was rigorously established by Hawkes & Oakes (1974). Sornette & Werner (2005) further offered an intuitive explanation for this equivalence, based on the combination of (i) the formulation of the ETAS seismicity rate expressed as a linear superposition of contributions from prior events and of (ii) the exponential form of the point process hazard rate. This dual perspective enables the ETAS model to function both as a generative engine of synthetic earthquake catalogs and as a predictive tool for time-dependent seismic hazard assessment.

In the present study, we employ two variants of the ETAS model that share an identical triggering kernel, with the sole difference being the treatment of the background seismicity rate (Nandan et al., 2021b; 2022):

(1) ETAS model with a constant background rate (ETAS $_{\mu}$). This variant, defined by Equation (1), assumes a spatially homogeneous background rate μ . We adopt it as the benchmark model in the present study. Its estimation involves determining the following set of model parameters: $\{\mu, K, \alpha, c, \omega, \tau, d, \gamma, \rho\}$.

(2) ETAS model with spatially variable background rate (ETAS $_{\mu(x,y)}$). In this augmented variant, the background rate $\mu(x, y)$ is treated as a nonparametric function that captures spatial heterogeneity of the background events. Its estimation follows a weighted kernel approach, as introduced by Nandan et al. (2021b, 2022):

$$\mu(x, y) = \frac{1}{T} \sum_{i=1}^N IP_i \cdot \frac{QD^{2Q}}{\pi} \frac{1}{[(x - x_i)^2 + (y - y_i)^2 + D^2]^{1+Q}} \quad (2)$$

where the summation is over all earthquakes in the catalog, weighted by their probability IP_i of being background events (Zhuang et al., 2002). The normalization by the catalog duration T ensures that $\mu(x, y)$ represents the seismicity rate per unit time, while the factor QD^{2Q}/π normalizes the kernel per unit area at location (x, y) . The choice of a power law kernel is motivated by its demonstrated superiority over the commonly used Gaussian kernel in previous studies (e.g., Helmstetter et al., 2007; Nandan et al., 2021a). Its estimation involves determining the following set of parameters: $\{Q, D, K, \alpha, c, \omega, \tau, d, \gamma, \rho\}$.

The parameter calibration of both variants is carried out using the extended Expectation-Maximization (EM) algorithm (Veen & Schoenberg, 2008). We refer readers to Nandan et al. (2021b, 2022) for full methodological details.

2.2 Pseudo-prospective forecasting experiments

The pseudo-prospective forecasting experiment is conducted over the testing period from January 1, 2000, to December 31, 2023. Each forecast window spans one calendar month, and the windows are non-overlapping. Specifically, for a given forecast window (e.g., a particular month in a given year), the earthquake catalog prior to that month is used to calibrate the model parameters. Based on the calibrated parameters, 100,000 synthetic earthquake catalogs are simulated for the forecast window. These synthetic catalogs are used to estimate the occurrence probability of the future earthquakes. The process is then repeated for each subsequent month in the testing period. To assess the spatial performance of the models, the study region is discretized using a geodesic triangular tessellation on the Earth's surface accounting for its spherical geometry, with six different spatial resolutions defined by equal-area triangles of approximately 195 km², 779 km², 3,117 km², 12,470 km², 49,883 km², and the entire CSES region as a single unit (approximately 780,000 km²). In addition, five different magnitude thresholds M are considered: 3.0, 3.5, 4.0, 4.5, and 5.0.

For each forecast window, magnitude threshold, and spatial resolution (space-time-magnitude bin), we compute the distribution of the number of simulated earthquakes in each triangular cell. The probability distribution $P(n)$, where n denotes the number of earthquakes, is then estimated using the kernel-based smoothing method described in the next section. Accordingly, the probability of observing n_i earthquakes in the i -th triangle is given by $P(n_i)$. Given a particular spatial resolution and magnitude threshold, the log-likelihood score of a model for a specific forecast window is defined as:

$$LL = \sum_{i=1}^N \log P(n_i) \quad (3)$$

where N is the number of spatial cells at the specified resolution. The information gain (IG) is defined as the difference in log-likelihood scores between $ETAS_{\mu(x, y)}$ and $ETAS_{\mu}$. A higher IG indicates superior performance of the spatially heterogeneous $ETAS_{\mu(x, y)}$ model relative to the benchmark $ETAS_{\mu}$ model, and vice versa.

2.3 Distribution of forecasted earthquake numbers

To obtain a smooth distribution of forecasted earthquake numbers and minimize uncertainties arising from the finite number of simulations, we adopt an adaptive Gaussian kernel density estimation (KDE) method to represent the distribution of simulated earthquake numbers within each grid cell, given a specific forecast window, magnitude threshold M , and spatial resolution. The adaptive kernel density estimation method is a nonparametric approach that adjusts to varying data density by allowing the

bandwidth to vary locally (Cranmer, 2001). Its key characteristic is that regions with higher data density are assigned narrower kernels, while sparser regions use wider kernels (Figure 1a), thereby improving estimation accuracy across the full distribution.

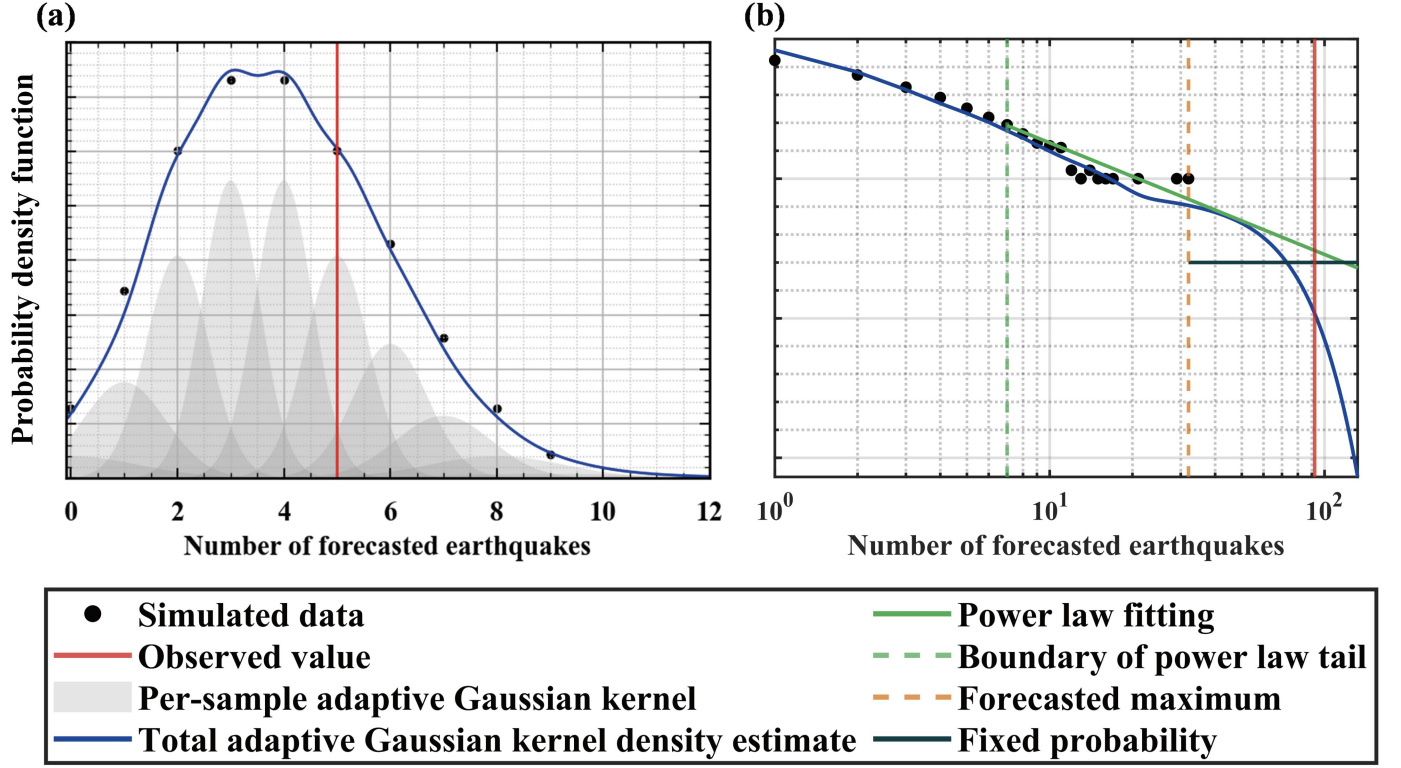


Figure 1. Illustration of the probability distribution for the number of forecasted earthquakes, shown for two scenarios: (a) where the observed value (red solid line) falls within the range of simulated values, and (b) where the observed value exceeds the maximum of the simulated distribution. The forecasted probability corresponding to the observed value n_i in the i -th triangle is denoted as $P(n_i)$. In the present study, simulated data (black solid circles) are generated from N_{sim} sets of synthetic catalogs ($N_{\text{sim}} = 100,000$). The per-sample adaptive Gaussian kernels (gray shaded areas), each with its own bandwidth, are summed to construct the total adaptive Gaussian kernel density estimate (blue solid line), which is used to present the overall probability distribution. When n_i (red solid line) exceeds the forecasted maximum (orange dashed line) in panel (b), we propose three alternative strategies for assigning $P(n_i)$: (1) fitting a power law function (green solid line) to the upper 30% of the simulated tail data (to the right of the green dashed line); (2) assigning a fixed probability value (black horizontal line), equal to $1/(N_{\text{sim}})^2$; or (3) excluding $P(n_i)$ altogether in such cases.

In the present study, the kernel density $f(n)$ for the number of simulated earthquakes n is given by:

$$f(n) = \frac{1}{N_{\text{sim}}} \sum_{j=1}^{N_{\text{sim}}} \frac{1}{h_j} K\left(\frac{n - n_j}{h_j}\right) \quad (4)$$

where N_{sim} is the number of synthetic catalogs (set to 100,000 in the present study), n_j is the number of earthquakes in the j -th simulated catalog, $K(\cdot)$ is the standard Gaussian kernel with zero mean and unit variance, h_j is the locally adaptive bandwidth for the j -th sample. To compute the pilot kernel density $\hat{f}(n)$, each h_j is initially set to a fixed bandwidth h_0 , which is selected to minimize the mean integrated squared error (MISE) following a rule-of-thumb bandwidth selection (Scott, 2015). The adaptive bandwidths h_j are then adjusted according to:

$$h_j = \frac{h_0}{f_0(n_j)} \quad (5)$$

where $\hat{f}(n_j)$ is the pilot density estimate at n_j . This formulation ensures narrower bandwidths in regions where simulated earthquake counts are concentrated and broader bandwidths where data are sparse. In effect, each simulation contributes a

Gaussian kernel in probability space, with its width adaptively scaled by the local data density—i.e., the number of synthetic catalogs yielding similar outcomes.

However, when the observed number of earthquakes within a grid cell—given a specific forecast window, magnitude threshold M_c , and spatial resolution—exceeds the maximum value obtained from the simulated catalogs, the adaptive kernel method may yield serious biased estimates of the probability $P(n)$ (see, for example, Figures 1 and 6 in the present study and Figure 1 in Nandan et al. (2019a)). To address this issue, we propose three alternative strategies for assigning $P(n)$ in such cases (Figure 1b):

- (1) **Power law extrapolation.** Fit a power law function to the upper 30% of the simulated tail data and use the extrapolated value as $P(n)$;
- (2) **Fixed minimum probability.** Assign a fixed probability value equal to $1/(N_{\text{sim}})^2$;
- (3) **Exclusion.** Exclude the corresponding $P(n)$ value entirely from the likelihood computation.

In the following sections, we evaluate the impact of these three strategies on model performance in forecasting experiments.

3 Seismicity in China Seismic Experimental Site (CSES)

Situated in a seismically active region (Li et al., 2020; 2021; 2022), the China Seismic Experimental Site (CSES) is designed to integrate dense, multidisciplinary observations to investigate the complex physical processes underlying earthquake generation (Wu et al., 2019; Wu & Li, 2021a; 2021b; Wu, 2022). By combining geological, geophysical, and geochemical data, the site enables a comprehensive investigation of seismic hazards. This integrative approach facilitates the development and validation of predictive models, thereby contributing significantly to earthquake forecasting efforts in China. The CSES region has long maintained a dense and high-quality seismic monitoring network, resulting in an extended record of reliable earthquake catalogs (Li et al., 2023; 2025c). As shown in Figure 2, the temporal–spatial distribution of completeness magnitude M_c indicates that the CSES catalog has been nearly complete for events with $M \geq 3$ since 2000, with a brief exception following the 2008 Wenchuan M_s 8.0 earthquake, when the completeness probability for $M \geq 3$ dropped to approximately 97.8% (Figure 2c). Based on this, we set the catalog cut-off magnitude at $M_{c0} = 3.0$.

The earthquake catalogs used in the present study were provided by the China Earthquake Networks Center (CENC). The catalog of earthquakes with $M \geq 3.0$ in the CSES region during 1970–2023 is available as Data Set S1 in Li et al. (2025a). The CSES catalog includes a mixture of magnitude types, primarily local magnitude (M_L) and surface-wave magnitude (M_s), which introduces potential complications due to magnitude-type heterogeneity. The magnitude bin size is set to 0.1. Temporally, the catalog is divided into two segments: (1) an auxiliary catalog from January 1, 1970, to December 31, 1984, used to avoid misclassifying early events in the primary catalog as background events in the ETAS model; and (2) the primary catalog, extending from January 1, 1985, to the day before the forecast window, used for calibrating the ETAS model parameters.

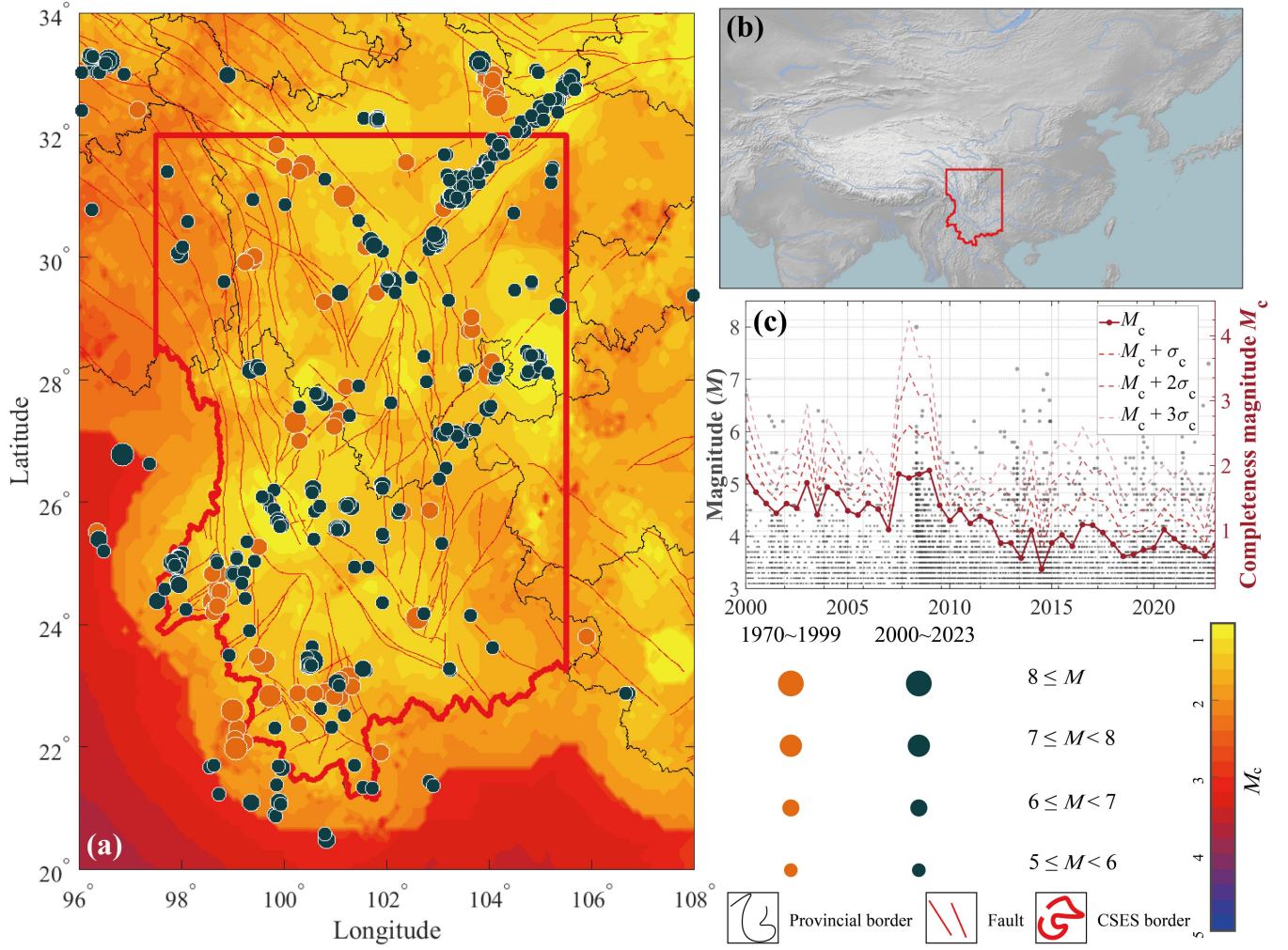


Figure 2. Spatiotemporal distribution of seismicity and completeness magnitude (M_c) in the China Seismic Experimental Site (CSES), located in the Sichuan-Yunnan region. (a) Spatial distribution of earthquakes from 1970 to 2023, completeness magnitude (M_c) since 2009, and active faults within the CSES. The background map of M_c is adapted from Li et al. (2023). Fault locations are taken from Deng et al. (2003). (b) Location of the CSES shown on a broader regional map. (c) Temporal evolution of M_c (right y-axis) and the magnitude-time (MT) plot of earthquakes with $M \geq 3$ (left y-axis) since 2000. The estimation of M_c was performed using a half-year step and one-year window. Red dashed lines with decreasing color indicate completeness probabilities of 84.2% ($M_c + \alpha$), 97.8% ($M_c + 2\alpha$), and 99.9% ($M_c + 3\alpha$), respectively. M_c and α are estimated by the method proposed by Wang et al. (2025) and Li et al., (2025b). The earthquake magnitudes shown in the magnitude-time plot have been perturbed by a uniformly distributed random error within the range $[-0.05, 0.05]$.

4 Results

In this section, we present two sets of results: (1) the calibrated parameters of the ETAS models, and (2) the outcomes of pseudo-prospective earthquake forecasting experiments based on the calibrated models.

4.1 Calibrated parameters of the ETAS model

To examine the temporal stability and evolution of ETAS model parameters, we calibrate both $\text{ETAS}_{\mu(x,y)}$ and the homogeneous counterpart ETAS_{μ} over a series of monthly forecast windows. For each calibration step, the primary catalog includes all events from January 1, 1985, up to the day preceding the forecast window. In addition, a 15-year auxiliary catalog covering the period from January 1, 1970, to December 31, 1984, is appended to provide historical context and reduce the likelihood of misclassifying earlier events as background earthquakes.

Figure 3 shows the temporal evolution of the estimated parameters for both $\text{ETAS}_{\mu(x,y)}$ and ETAS_{μ} . Overall, the two models yield

noticeably different parameter estimates, with $\text{ETAS}_{\mu(x,y)}$ identifying significantly more background events than ETAS_{μ} , primarily as a result of allowing for a spatially heterogeneous background seismicity, which is otherwise likely to be misclassified as triggered seismicity in ETAS_{μ} . Furthermore, from 2000 to 2023, the differences between the two models in several key parameters remain relatively stable over time. Specifically, the fertility law parameters ($\log_{10}K$ and α), the spatial kernel parameter ($\log_{10}d$), and the branching ratio h consistently exhibit similar difference amplitudes between $\text{ETAS}_{\mu(x,y)}$ and ETAS_{μ} . In contrast, the gap between the two models in terms of the cumulative number of background events and $\log_{10}\tau$ gradually widens over time. On the other hand, the differences in ρ , γ , $\log_{10}c$, and $1+\omega$ tend to decrease with time and, in some cases, nearly vanish. Notably, abrupt changes in several parameter time series coincide with the occurrence of major earthquakes in the region. Examples include the 2008 Wenchuan M_s 8.0 earthquake (May 12), the 2013 Lushan M_s 7.0 earthquake (April 20), and the 2017 Jiuzhaigou M_s 7.0 earthquake (August 8), among others.

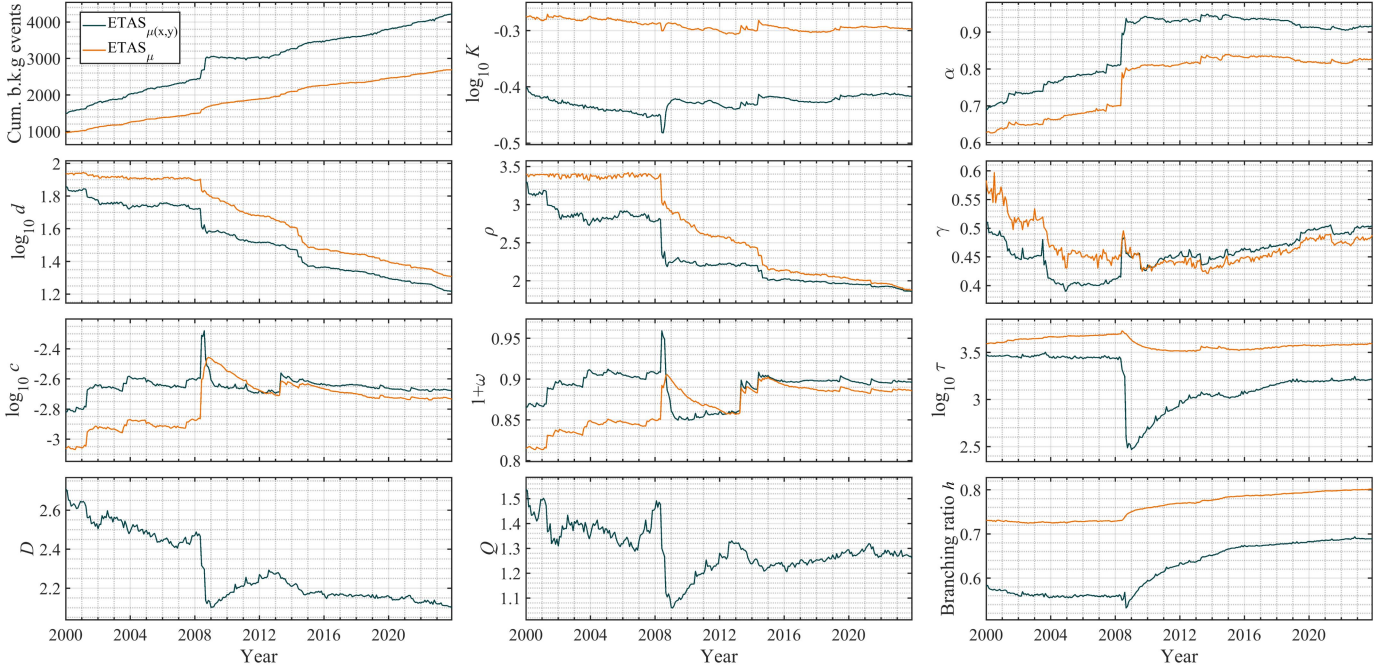


Figure 3. Temporal evolution of the calibrated parameters in the ETAS model with spatially variable background seismicity ($\text{ETAS}_{\mu(x,y)}$) and in the ETAS model with a constant background rate (ETAS_{μ}). For model calibration, the primary catalog spans from January 1, 1985, to the day before each monthly forecast window, and is preceded by a 15-year auxiliary catalog (from January 1, 1970, to December 31, 1984) to prevent misclassification of early events in the primary catalog as background events. Note that the parameters D and Q are specific to $\text{ETAS}_{\mu(x,y)}$ and are not included in ETAS_{μ} .

4.2 Cumulative information gain (CIG) of $\text{ETAS}_{\mu(x,y)}$ over ETAS_{μ}

The calibrated parameters prepared for each forecast window were used to generate synthetic earthquake catalogs based on the first of the two conceptually distinct yet mathematically equivalent interpretations of the ETAS model, that is, the branching interpretation (see the second paragraph of Section 2.1). For each forecast window, 100,000 synthetic catalogs were generated in this study. All simulated catalogs produced by the two ETAS models are openly available through Li et al. (2025d). Figure 4 presents the time series of cumulative information gain (CIG) of $\text{ETAS}_{\mu(x,y)}$ over ETAS_{μ} from 2000 to 2023. Results are shown for six spatial resolutions, five magnitude thresholds (M), and three strategies for assigning $P(n_i)$ when the observed value n_i in the i -th triangle exceeds the maximum value obtained from the simulations. The green, black, and orange curves correspond to strategies (1), (2), and (3) described in Section 2.3, respectively. We also provide, in the Open Research section, adaptive Gaussian kernel density fits for both models across the five magnitude thresholds for the full CSES region.

For $M_t = 3$ and 3.5 (i.e., the first and second columns of Figure 4 across all spatial resolutions) and for the entire CSES region (i.e., the last row of Figure 4 across all M_t), the CIG curves of $\text{ETAS}_{\mu(x,y)}$ over ETAS_μ remain consistently above zero and exhibit an overall upward trend over time. This indicates that $\text{ETAS}_{\mu(x,y)}$ outperforms the baseline ETAS_μ model in terms of forecasting skill under these settings. For other combinations of spatial resolution and M_t , however, the results are more ambiguous. The CIG curves often display substantial fluctuations (see discussion in the next section), and in many cases, the performance of $\text{ETAS}_{\mu(x,y)}$ does not significantly exceed that of ETAS_μ , reflecting the complex nature of seismicity in the study region. As with the parameter time series discussed in Section 4.1, many abrupt changes in CIG coincide with major earthquakes in the region, most notably the 2008 Wenchuan M_s 8.0 earthquake (May 12), which is discussed in detail in the next section.

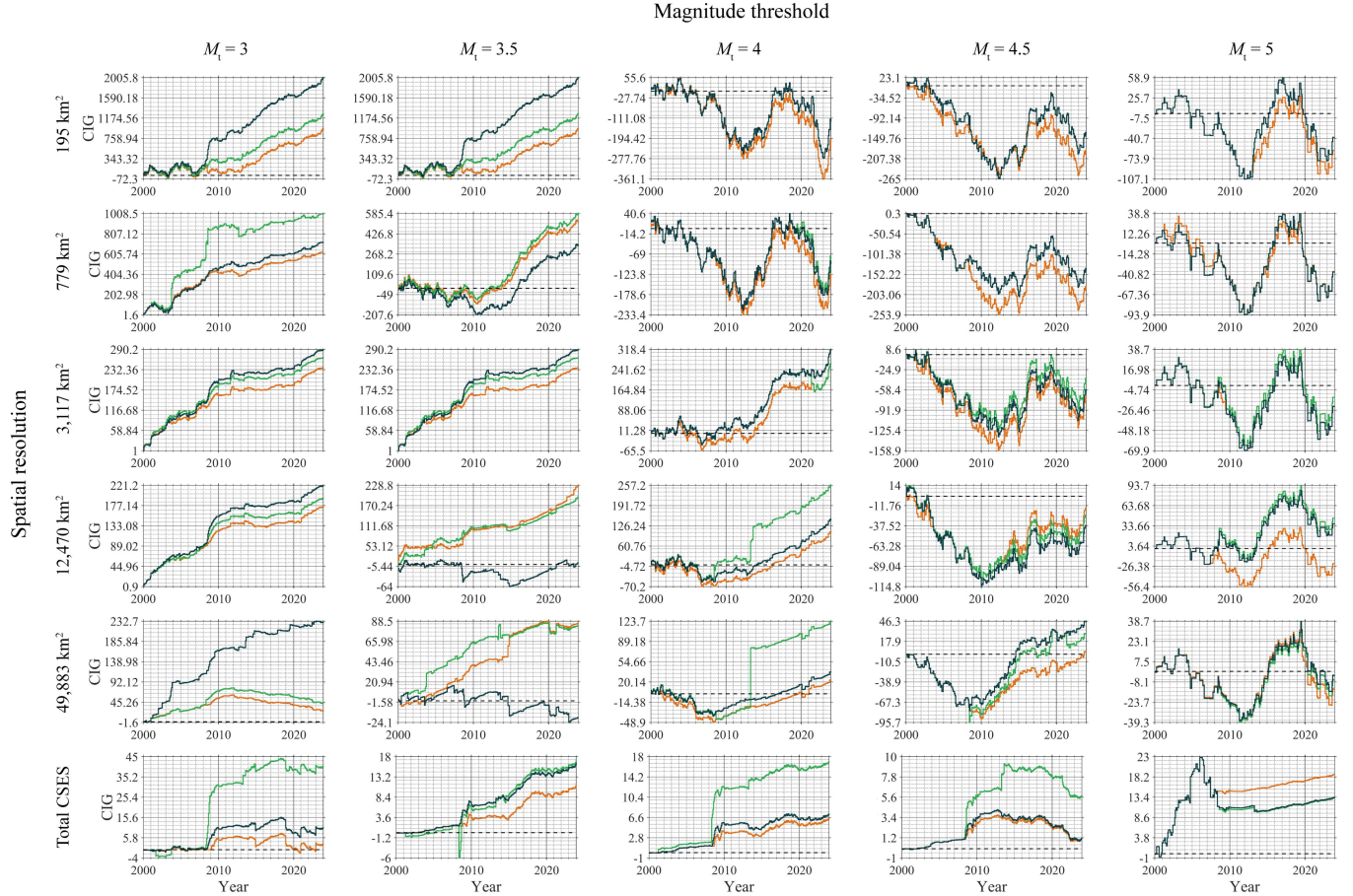


Figure 4. Time series of cumulative information gain (CIG) of $\text{ETAS}_{\mu(x,y)}$ over ETAS_μ in pseudo-prospective forecasting experiments at the CSES from 2000 to 2023. Results are shown for six spatial resolutions, five magnitude thresholds (M_t), and three strategies for assigning $P(n_i)$ when the observed value n_i in the i -th triangle exceeds the forecasted maximum. The green, black, and orange curves correspond to strategies (1), (2), and (3) described in Figure 1, respectively. The pseudo-prospective forecasting experiments use calendar months as time windows, resulting in a total of 288 time windows over the study period. The horizontal black dashed line indicates zero cumulative information gain.

Figure 4 also shows that the choice of strategy for assigning $P(n_i)$ when the observed value n_i exceeds the forecasted maximum can significantly influence the model evaluation results. For lower magnitude thresholds (e.g., $M_t = 3$ to 4), corresponding to the left-hand panels in Figure 4, the three strategies, power law extrapolation, fixed minimum probability, and exclusion, yield markedly different CIG curves. This highlights the sensitivity of model evaluation to how such rare cases are handled. In most instances, the power law extrapolation strategy results in the highest CIG values. However, there are exceptions, for example, under the combinations of spatial resolution 195, 3117, 12,470, and 49,883 km² with $M_t = 3$, the fixed minimum probability strategy consistently produces higher CIG curves than the power law extrapolation approach.

4.3 Bias from improper tail treatment

However, the apparent advantage of the CIG curves obtained using the fixed minimum probability strategy for assigning $P(n_i)$ is in fact illusory. When both models ($\text{ETAS}_{\mu(x,y)}$ and ETAS_μ) assign $P(n_i)$ based on the same fixed minimum probability, i.e., when the observed earthquake count exceeds the maximum simulated count of both models, their log-likelihood scores become identical, resulting in an information gain (IG) of zero. In contrast, when the observed count exceeds the maximum simulated value of only one model but not the other, the former adopts the fixed minimum probability, while the latter still uses the adaptive Gaussian kernel density estimate to assign $P(n_i)$. This asymmetric treatment introduces a strong bias in model performance. Such cases frequently occur in the months following large earthquakes (e.g., events with $M > 7$); examples are observed for $M = 3.0$ during June and November 2008 in the full CSES region (see Open Research section).

During these post-mainshock forecast windows, $\text{ETAS}_{\mu(x,y)}$ systematically generates more tail events in the probability distribution $P(n)$ than ETAS_μ . This difference arises because the spatially varying $\text{ETAS}_{\mu(x,y)}$ incorporates a heterogeneous background rate $\mu(x, y)$, which accounts for spatial clustering and variability in triggering intensity. Statistically, this produces a mixture of local Poisson-branching processes with different means, naturally leading to a heavier-tailed distribution. Physically, high- μ regions act as persistent triggering hotspots, sustaining multiple generations of aftershocks and resulting in abnormally large event counts within certain cells. Consequently, $\text{ETAS}_{\mu(x,y)}$ exhibits a broader and more asymmetric $P(n)$, consistent with the enhanced spatial heterogeneity of real seismicity.

In fact, the exclusion strategy for assigning $P(n_i)$ suffers from a similar problem, though its impact on CIG is less pronounced than that of the fixed minimum probability. Specifically, when both models assign probabilities under the exclusion rule (i.e., when the observed count exceeds the simulated maximum of both models), the IG of $\text{ETAS}_{\mu(x,y)}$ over ETAS_μ is zero. When the observed count exceeds the simulated range of only one model, the log-likelihood (LL) for that model cannot be computed, and thus the information gain (IG) for this forecast window becomes undefined. In practice, this window is omitted from the IG calculation, which explains why its impact on the overall CIG is less pronounced than that of the fixed minimum probability approach. In either case, whether using the fixed minimum probability or exclusion, the model performance is systematically biased following large earthquakes. Therefore, neither method should be recommended for evaluating IG in such contexts.

5 Discussion

In this section, we present the results of pseudo-prospective experiments conducted for May 2008 across the entire CSES region to quantitatively examine the impact of out-of-range earthquake counts on the calculation of information gain. Following this, we further explain the underlying causes of noisy cumulative information gain, highlighting both methodological and data-related factors.

5.1 An example of out-of-rang earthquake counts: The May 2008

Within the spatiotemporal window of the present study, the largest event is the $M = 8.0$ Wenchuan earthquake of May 12, 2008. This mainshock and its aftershocks produced an out-of-range situation, as illustrated in Panel (b) of Figure 1, under all magnitude thresholds (M) considered in the present analysis (see Figure 5). Figure 5 also presents the probability distributions of earthquake counts forecasted by ETAS_μ and $\text{ETAS}_{\mu(x,y)}$ for the same month (May 2008) across the entire CSES region. The results clearly demonstrate that, once the observed counts fall outside the simulated ensemble, the extrapolation based on

adaptive Gaussian kernel density estimation becomes severely biased and drastically underestimates the probability of the observed earthquake numbers. In contrast, power law distributions with fat-tailed behavior provide a more reliable extrapolation, yielding non-negligible probabilities for these extreme counts.

From Figure 5, the information gain (IG) of $\text{ETAS}_{\mu(x,y)}$ relative to ETAS_{μ} can be computed using Equation (3). The corresponding IG values are approximately 2, -2, -1, -3, and 1 for $M_t = 3, 3.5, 4, 4.5$, and 5, respectively. These results suggest that, at the time of the Wenchuan mainshock itself, the two models did not exhibit a consistently large performance gap across thresholds. However, the bottom row of Figure 4 reveals a critical point: when extrapolation is performed using power law tails, the IG of $\text{ETAS}_{\mu(x,y)}$ relative to ETAS_{μ} continues to rise in the months following May 2008. This indicates that $\text{ETAS}_{\mu(x,y)}$ systematically outperforms ETAS_{μ} in forecasting the Wenchuan aftershocks, leading to the marked upward shift observed in the cumulative information gain (CIG) curve in Figure 4.

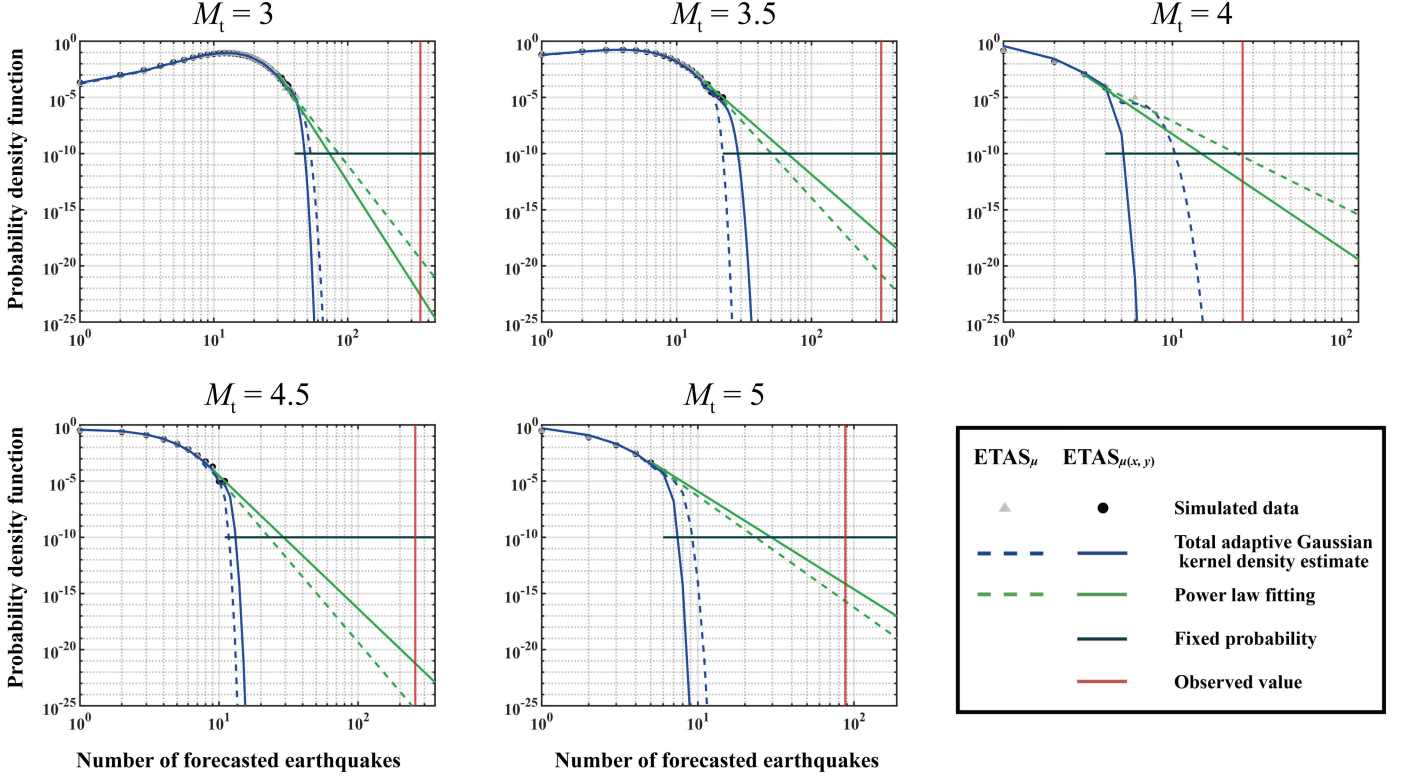


Figure 5. Probability distributions of earthquake counts forecasted by (dashed line) ETAS_{μ} and (solid line) $\text{ETAS}_{\mu(x,y)}$ for May 2008 across the entire CSES region, under different magnitude thresholds (M_t). The M_t 8.0 Wenchuan earthquake on May 12 and its aftershocks (marked by the red vertical line) generate an out-of-range situation as shown in Panel (b) of Figure 1, which corresponds to a single time-window slice taken from the last row of Figure 5. All plotting elements are consistent with those in Figure 1, except that the ETAS_{μ} data points are shown as solid gray triangles for visual distinction. The Open Research section provides links to access the forecasted probability density functions of both models for all forecast windows and magnitude thresholds across the full CSES region.

5.2 Quantitative rationale for the variability of CIG

Figure 4 shows that the CIG curves of $\text{ETAS}_{\mu(x,y)}$ relative to ETAS_{μ} lose their long-term upward trend and become increasingly erratic when the spatial resolution is coarsened and M_t is raised. Below we offer a quantitative rationale for this behaviour.

Let n be the number of earthquakes observed in a single grid cell at the chosen resolution r (expressed as the mesh area) during the i th forecast window for which there are N_r grid cells. The number of grid cells scales as $N_r \sim 1/r$. Let us assume that n follows a discrete probability law $P(n)$ with mean n_0 (> 0). Suppose models A and B also generate counts from $P(n)$ but with means n_A and n_B , respectively, subject only to the accuracy condition $|n_A - n_0| < |n_B - n_0|$ indicating that model A is better than model B. Define the per-cell log-likelihood difference during i th forecast window

$$\Delta LL_i = \log L_A^i(n) - \log L_B^i(n) \quad (6)$$

Since $|n_A - n_0| < |n_B - n_0|$, the average log-likelihood difference $\text{Mean}(\Delta LL)$ is positive, indicating that model A is the better fit. The accuracy of this statistic is assessed with the coefficient of variation (CV) of the log-likelihood difference:

$$\text{CV} = \frac{\text{Std}(\Delta LL)}{\text{Mean}(\Delta LL)} \quad (7)$$

where the mean and standard deviation are calculated over the set of N_r cells. The CV is a function of the resolution r via the dependence of the number of grid cells $N_r \sim 1/r$. The coarser the resolution r is, the larger is n_0 and the narrower is $P(n)$, but the smaller is N_r . Thus coarsening the resolution r improves the statistics per cell but deteriorates the statistical accuracy obtained by averaging over all N_r cells. This competition between in-cell and ensemble-cell statistics explains the complicated dependences of the CIG curves on the resolution and magnitude threshold shown in Figure 4.

Table 1 reports the CV computed solely from the orange CIG curves in Figure 4, i.e., using the exclusion strategy (see more details in Section 2.3). Under this strategy only the probabilities obtained from the adaptive-Gaussian kernels are retained; any cells whose observed counts fall outside the simulated range are discarded. This ensures that all CV values are derived from a single, internally consistent method, without the complications introduced by fixing minimum probability or power law extrapolation. The results show a clear pattern: for a given spatial resolution, a larger M_t lowers the average count n_0 (e.g., Figures 5 and 6); the mean of ΔLL therefore declines while its sampling variance grows, and the CV increases accordingly.

Table 1. Coefficient of variation (CV) calculated based on the cumulative information gain (CIG) results for $\text{ETAS}_{\mu(x, y)}$ relative to ETAS_μ in the China Seismic Experimental Site (CSES) region, evaluated for each combination of spatial resolution and magnitude threshold M_t . The CV values are calculated from the Exclusion strategy only, that is, from the orange CIG curves shown in Figure 4.

Spatial resolution	$M_t = 3$	$M_t = 3.5$	$M_t = 4$	$M_t = 4.5$	$M_t = 5$
195 km ²	7.28	15.73	22.92	16.02	34.62
779 km ²	3.87	7.19	28.18	13.38	48.38
3,117 km ²	3.08	4.64	9.05	27.25	68.67
12,470 km ²	2.62	3.42	13.03	93.82	54.27
49,883 km ²	10.90	4.72	25.78	21.50	64.33

5.3 Consistency of empirical scaling and ETAS-based theoretical predictions

Figure 6 demonstrates that the empirical probability density functions (PDFs) of earthquake counts per grid cell are well approximated by the power law $P(n) \sim n^{-(1+a)}$ across all tested magnitude thresholds and spatial resolutions. For each combination of spatial resolution and magnitude threshold, the mean and standard deviation of the fitted exponent a were estimated from 200 bootstrap resamples. Figure 7 then plots the histogram of the 25 mean a values shown in Figure 6 together with their Gaussian fit. The overall mean (1.40) and median (1.39) of these 25 mean a values are close to each other, as are their standard deviation (0.21) and interquartile range (0.21). This supports the parsimonious hypothesis of a common exponent $a = 1.40 \pm 0.21$.

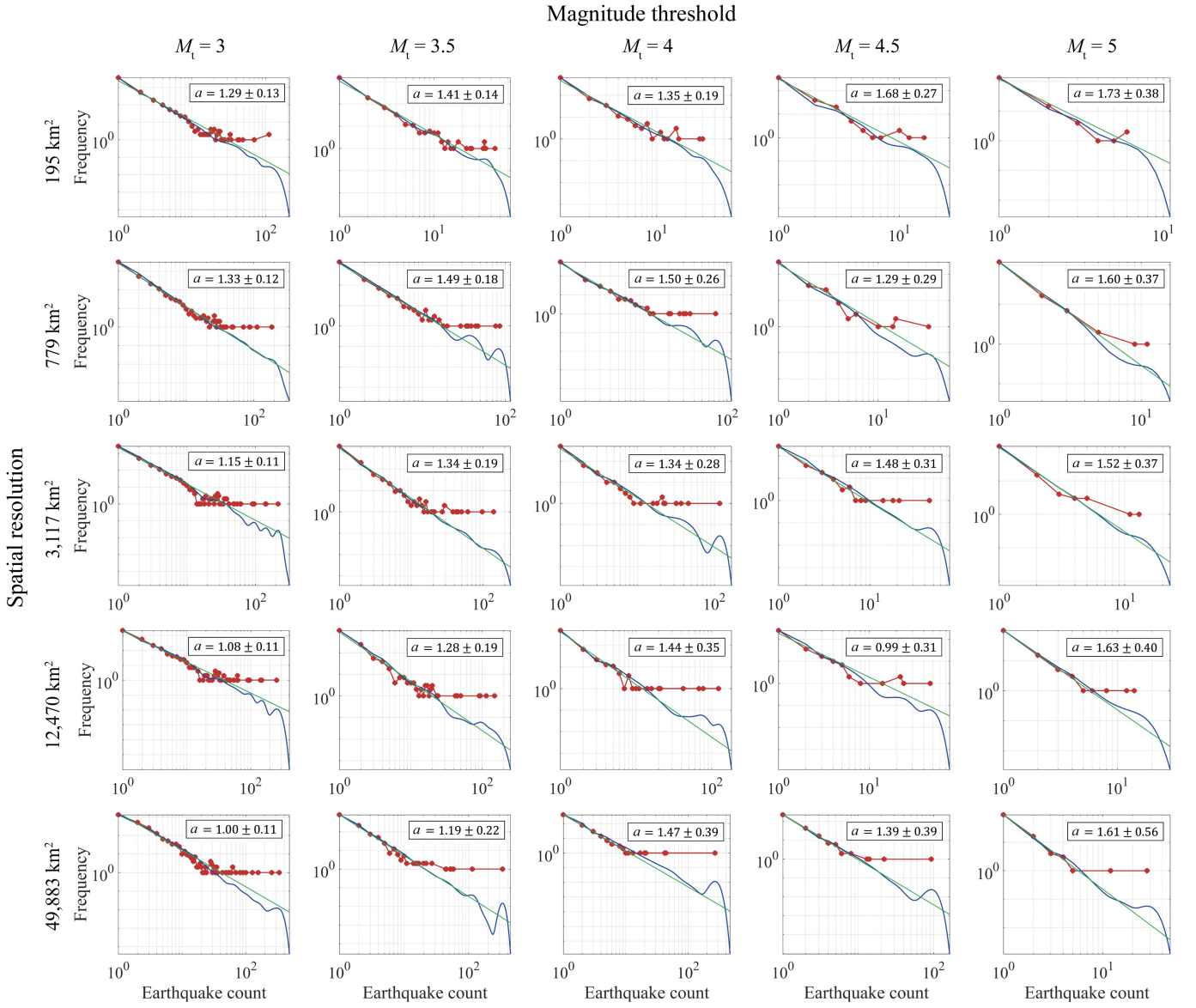


Figure 6. Frequency distribution function of earthquake count over the ensemble of grid cells, aggregated over all 288 forecast windows. Each point represents the number of grid cells (on the y -axis) that experienced a specific number of earthquakes (on the x -axis), for a given combination of spatial resolution and magnitude threshold. In other words, the curves show the empirical non-normalized probability density function of earthquake counts per cell. Consistent with Figure 1, the blue solid line represents the adaptive Gaussian kernel density estimate, while the green solid line shows the fitted power law function ($\sim n^{-(1+a)}$). For each combination of spatial resolution and magnitude threshold, the mean and standard deviation of the fitted exponent a were estimated from 200 bootstrap resamples.

Using the technique of generating probability functions applied to the ETAS model, Saichev et al. (2005) derived the theoretical distribution of the total number of aftershocks, which follows the power law $P(n) \sim n^{-(1+a)}$ in the non-critical regime (i.e., when its branching ratio h is not close to 1 and for earthquake counts $n > 1/(1 - h)^{g/(g-1)}$ where $g = b/\alpha$, b is the Gutenberg-Richter (GR) slope and α is the fertility (or productivity) exponent describing how the mean number of aftershocks triggered by a mainshock scales with its magnitude. The power law exponent a is related to the model parameters through $a = g = b/\alpha$. For smaller event counts $n < 1/(1 - h)^{g/(g-1)}$, the exponent a is renormalized to $a = 1/g$. The empirical observation of a single power law with an exponent $a > 1$ therefore indicates that the system operates in a subcritical regime, i.e., the branching ratio h is substantially below 1. This finding corroborates previous studies suggesting that the Earth's crust, on average, remains far from criticality (Nandan et al., 2021b; Li et al., 2025a). Saichev & Sornette (2006) further demonstrated that this same scaling law emerges as the asymptotic behavior of space-time windowed event-number distributions, implying that such power laws are a fundamental statistical property of branching-type seismicity models.

Moreover, the observed scaling exponent a provides an indirect way to infer the ETAS fertility (or productivity) exponent α . Given the theoretical relation $a = g := b/\alpha$, the empirical result $a = 1.40 \pm 0.21$ and the Gutenberg-Richter slope value $b = 0.80$ (see Li et al., 2025a) yield an estimated fertility exponent of $\alpha = b/a = 0.57 \pm 0.08$. This inferred value of α falls well within the physically plausible range reported by previous ETAS-based studies and reinforces the interpretation that seismic triggering efficiency increases moderately with magnitude but remains sublinear. In contrast, the parameter estimates obtained directly from the ETAS-based models are systematically higher (Figure 3): $\alpha \approx 0.7\text{--}1.0$ for ETAS $_{\mu(x, y)}$ and $\alpha \approx 0.6\text{--}0.9$ for ETAS $_{\mu(x, y)}$. Although these two ranges partially overlap with the theoretical estimate, the ETAS-derived values tend to be biased upward.

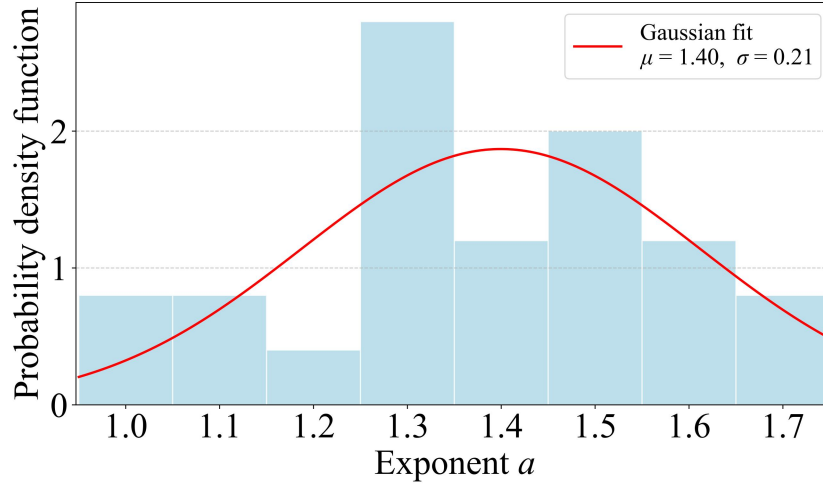


Figure 7. Histogram of the 25 mean power law exponents a shown in Figure 6 and their Gaussian fit, with mean 1.40, median 1.39, standard deviation 0.21 and interquartile range 0.21.

Several factors may explain this discrepancy. First, the theoretical derivation assumes an idealized, stationary branching process in the non-critical regime, while the ETAS fitting is affected by non-stationary background rates, spatial heterogeneity, and catalog incompleteness, all of which can inflate the apparent fertility. Second, the estimation of α in ETAS models is strongly coupled with that of other parameters, particularly the background rate μ and the spatiotemporal decay parameter ($1+\rho$ and $1+\omega$), leading to potential trade-offs that shift the fitted values upward. Third, the boundary and finite-size effects inherent in the CSES region (Li et al., 2025a), especially the presence of clustered aftershock sequences such as the 2008 Wenchuan earthquake, may enhance the observed triggering efficiency relative to the long-term theoretical expectation. Despite these differences, both approaches consistently yield $\alpha < b$, indicating that small earthquakes collectively dominate the overall triggering process. This imbalance implies that, while larger events are individually more efficient triggers, their cumulative contribution is outweighed by the far more numerous small events. This is a manifestation of the so-called ultraviolet divergence problem in the ETAS framework. Because direct empirical estimation of the fertility exponent α is notoriously difficult, often hindered by limited temporal completeness and strong parameter correlations, our indirect inference offers a robust and independent constraint. It provides empirical support for theoretical expectations derived from branching process theory and anchors a key ETAS parameter on a statistically grounded foundation directly linked to the observed scaling behavior of earthquake count distributions.

6 Conclusions

Advancing short-term earthquake forecasting has always been a core objective of the emerging China Seismic Experimental Site (CSES), led by China Earthquake Administration (CEA). A critical step in that agenda is the rigorous testing of scientific hypotheses on earthquake predictability. To that end, we have developed AoyuX, a bespoke short-term predictability platform tailored to the CSES region, which is currently in its initial implementation stage. AoyuX embeds the ETAS model, one of the few frameworks that simultaneously delivers time, space, magnitude, and probability information and that synthesises the current state of seismological knowledge. By embedding two variants of the ETAS model, ETAS_μ with a homogeneous background rate and $\text{ETAS}_{\mu(x, y)}$ with a spatially variable rate, we carried out about 25 years of monthly pseudo-prospective experiments (2000-2023).

The calibration phase reveals systematic contrasts between the two variants. $\text{ETAS}_{\mu(x, y)}$ consistently assigns more events to the background than ETAS_μ , reflecting the added flexibility of a non-parametric background field. Several parameter time series exhibit abrupt shifts that coincide with major regional earthquakes, indicating that large events can instantaneously reshape the triggering environment as well as be associated with transient data issues such as earthquake incompleteness. Forecast assessments, expressed through cumulative information gain (CIG), show a clear hierarchy under data-rich conditions: for the full CSES domain and for $M_t = 3.0$ to 3.5, $\text{ETAS}_{\mu(x, y)}$ maintains a persistent positive CIG over ETAS_μ . However, at finer spatial grids or larger magnitude thresholds the average cell count declines, sampling variance grows, and the CIG curves become highly volatile. A coefficient-of-variation analysis confirms that, once data scarcity pushes CV above ~ 10 , noise overwhelms deterministic skill and the apparent advantage of $\text{ETAS}_{\mu(x, y)}$ vanishes. Model ranking also depends sensitively on how “out-of-range” cell counts, observed counts exceeding all simulated values, are handled. Among the strategies tested, power law extrapolation consistently provides the most theoretically justified and empirically robust results, in line with both analytical predictions of branching-process theory and the clear scaling patterns shown in Figure 6 and earlier studies (Saichev et al., 2005; Saichev & Sornette, 2006; 2007). While in a few coarse-grained low-magnitude configurations, the fixed minimum probability rule produced slightly higher CIG values, these are best interpreted as artifacts of noise rather than genuine model superiority. Both the fixed-minimum probability and exclusion strategies introduce systematic biases in model evaluation following large earthquakes. Consequently, these approaches are not suitable for reliable assessment of information gain during periods of intense seismic activity. Taken together, the evidence strongly supports power law extrapolation as the most appropriate framework for handling extreme counts, especially in cells affected by large, cluster-forming earthquakes, which generate the fat-tailed distributions expected from aftershock cascades. This highlights the necessity of treating such extreme outliers not as anomalies to be truncated but as integral signatures of the underlying triggering dynamics.

As a bonus, our analysis yields a surprisingly precise and robust estimate of the fertility exponent, $\alpha = 0.57 \pm 0.08$, derived from well-behaved empirical distributions of earthquake counts and firmly supported by ETAS theory, thus providing one of the very few reliable quantifications of this key but elusive parameter.

In sum, $\text{ETAS}_{\mu(x, y)}$ provides demonstrable skill over its homogeneous counterpart in regimes where observational counts are plentiful. Where data are sparse, however, sampling noise, parameter instability, and treatment of extreme clusters become dominant concerns. Future work should therefore focus on adaptive strategies for rare-event handling and on hybrid schemes that combine ETAS-style triggering with physics-based constraints, further enhancing the operational value of AoyuX for short-term earthquake forecasting in the CSES and analogous natural laboratories worldwide.

Acknowledgments

The authors would like to thank Yang Zang from the China Earthquake Networks Center for his assistance in accessing the Chinese earthquake catalog data, as well as Prof. Jiancang Zhuang, Prof. Zhongliang Wu and Prof. Changsheng Jiang for their valuable suggestions, and Xinyi Wang for assistance in computing catalog completeness using the GR-AEReLU method. This work is partially supported by the Guangdong Basic and Applied Basic Research Foundation (Grant No. 2024A1515011568) and the Center for Computational Science and Engineering at Southern University of Science and Technology.

Data availability

The earthquake catalogs used in this study were obtained from the National Earthquake Science Data Center of China (https://data.earthquake.cn/datashare/report.shtml?PAGEID=earthquake_zhengshi), which is publicly accessible; however, user registration is required to access the catalog information. The catalog of earthquakes with magnitudes ≥ 3.0 in the CSES region during 1970–2023 is provided as Data Set S1 in Li et al. (2025a). The simulated earthquake catalogs generated by the two ETAS models can also be accessed through Li et al. (2025d). In addition, the forward monthly, quarterly, semi-annual, and annual AoyuX earthquake forecast reports for the CSES since June 2024 are available at Li (2025). Links to the forecasted probability density functions of both models for all forecast windows and magnitude thresholds across the CSES region are available at Li (2025a; 2025b; 2025c; 2025d; 2025e).

References

- Bayliss, K., Naylor, M., Kamranzad, F., and Main, I.: Pseudo-prospective testing of 5-year earthquake forecasts for California using inlabru, *Nat. Hazards Earth Syst. Sci.*, 22, 3231–3246, doi:10.5194/nhess-22-3231-2022, 2022.
- Cranmer, K.: Kernel estimation in high-energy physics, *Comput. Phys. Commun.*, 136, 198–207, doi:10.1016/S0010-4655(01)00145-6, 2001.
- Davis, L., Baeumer, B., and Wang, T.: A fractional Hawkes process model for earthquake aftershock sequences, *J. R. Stat. Soc. C*, 73, 1185–1202, doi:10.1093/jrsssc/qlae011, 2024.
- Deng, Q., Zhang, P., Ran, Y., Yang, X., Min, W., and Chu, Q.: Basic characteristics of active tectonics of China, *Sci. China Ser. D*, 46, 356–372, doi:10.1360/03yd9032, 2003.
- Felzer, K. R., Abercrombie, R. E., and Ekström, G.: Secondary aftershocks and their importance for aftershock forecasting, *B. Seismol. Soc. Am.*, 93, 1433–1448, doi:10.1785/0120020229, 2003.
- Felzer, K. R., Becker, T. W., Abercrombie, R. E., Ekström, G., and Rice, J. R.: Triggering of the 1999 MW 7.1 Hector Mine earthquake by aftershocks of the 1992 MW 7.3 Landers earthquake, *J. Geophys. Res.*, 107, 2190, doi:10.1029/2001JB000911, 2002.
- Hainzl, S., Kumazawa, T., and Ogata, Y.: Aftershock forecasts based on incomplete earthquake catalogues: ETASI model application to the 2023 SE Türkiye earthquake sequence, *Geophys. J. Int.*, 236, 1609–1620, doi:10.1093/gji/ggae154, 2024.
- Han, M., Mizrahi, L., and Wiemer, S.: Towards a harmonized operational earthquake forecasting model for Europe, *Nat. Hazards Earth Syst. Sci.*, 25, 991–1012, doi:10.5194/nhess-25-991-2025, 2025.
- Harte, D. S.: Probability distribution of forecasts based on the ETAS model, *Geophys. J. Int.*, 210, 90–104, doi:10.1093/gji/ggx150, 2017.

- Hawkes, A. G. and Oakes, D.: A cluster process representation of a self-exciting process, *J. Appl. Probab.*, 11, 493–503, doi:10.1017/s0021900200096273, 1974.
- Helmstetter, A., Kagan, Y. Y., and Jackson, D. D.: High-resolution time-independent grid-based forecast for $M \geq 5$ earthquakes in California, *Seismol. Res. Lett.*, 78, 78–86, doi:10.1785/gssrl.78.1.78, 2007.
- Helmstetter, A. and Sornette, D.: Subcritical and supercritical regimes in epidemic models of earthquake aftershocks, *J. Geophys. Res.*, 107, 2237, doi:10.1029/2001JB001580, 2002.
- Kagan, Y. Y.: Likelihood analysis of earthquake catalogues, *Geophys. J. Int.*, 106, 135–148, doi:10.1111/j.1365-246x.1991.tb04607.x, 1991.
- Kagan, Y. Y.: Statistical distributions of earthquake numbers: Consequence of branching process, *Geophys. J. Int.*, 180, 1313–1328, doi:10.1111/j.1365-246X.2009.04487.x, 2010.
- Kagan, Y. Y.: Earthquake number forecasts testing, *Geophys. J. Int.*, 211, 335–345, doi:10.1093/gji/ggx324, 2017.
- Kagan, Y. Y. and Knopoff, L.: Stochastic synthesis of earthquake catalogs, *J. Geophys. Res.*, 86, 2853–2862, doi:10.1029/jb086ib04p02853, 1981.
- Kagan, Y. Y. and Knopoff, L.: Statistical short-term earthquake prediction, *Science*, 236, 1563–1567, doi:10.1126/science.236.4808.1563, 1987.
- Kagan, Y. Y. and Jackson, D. D.: Probabilistic forecasting of earthquakes, *Geophys. J. Int.*, 143, 438–453, doi:10.1046/j.1365-246x.2000.01267.x, 2000.
- Kamer, Y., Nandan, S., Ouillon, G., Hiemer, S., and Sornette, D.: Democratizing earthquake predictability research: introducing the RichterX platform, *Eur. Phys. J. Spec. Top.*, 230, 451–471, doi:10.1140/epjst/e2020-000259-3, 2021.
- Kanazawa, K. and Sornette, D.: Field master equation theory of the self-excited Hawkes process, *Phys. Rev. Res.*, 2, 033442, doi:10.1103/PhysRevResearch.2.033442, 2020.
- Kanazawa, K. and Sornette, D.: Standard form of master equations for general non-Markovian jump processes: The Laplace-space embedding framework and asymptotic solution, *Phys. Rev. Res.*, 6, 023270, doi:10.1103/PhysRevResearch.6.023270, 2024.
- Li, J.: lijiawei098/lijiawei098.github.io: AoyuX 1.0 (V1.0) [Dataset], Zenodo, doi:10.5281/zenodo.17481512, 2025.
- Li, J.: Figure 5(a). Probability distributions of earthquake counts forecasted by $ETAS_{\mu}$ and $ETAS_{\mu}(x, y)$ for all forecast windows across the entire CSES region, under magnitude threshold $M_t = 3.0$ [Dataset], figshare, doi:10.6084/m9.figshare.30371602.v1, 2025a.
- Li, J.: Figure 5(b). Probability distributions of earthquake counts forecasted by $ETAS_{\mu}$ and $ETAS_{\mu}(x, y)$ for all forecast windows across the entire CSES region, under magnitude threshold $M_t = 3.5$ [Dataset], figshare, doi:10.6084/m9.figshare.30372214.v2, 2025b.
- Li, J.: Figure 5(c). Probability distributions of earthquake counts forecasted by $ETAS_{\mu}$ and $ETAS_{\mu}(x, y)$ for all forecast windows across the entire CSES region, under magnitude threshold $M_t = 4.0$ [Dataset], figshare, doi:10.6084/m9.figshare.30372271.v1, 2025c.
- Li, J.: Figure 5(d). Probability distributions of earthquake counts forecasted by $ETAS_{\mu}$ and $ETAS_{\mu}(x, y)$ for all forecast windows across the entire CSES region, under magnitude threshold $M_t = 4.5$ [Dataset], figshare, doi:10.6084/m9.figshare.30372298.v1, 2025d.
- Li, J.: Figure 5(e). Probability distributions of earthquake counts forecasted by $ETAS_{\mu}$ and $ETAS_{\mu}(x, y)$ for all forecast windows across the entire CSES region, under magnitude threshold $M_t = 5.0$ [Dataset], figshare, doi:10.6084/m9.figshare.30372319.v1, 2025e.

- Li, J., Böse, M., Wyss, M., Wald, D. J., Hutchison, A., Clinton, J. F., Wu, Z., Jiang, C., and Zhou, S.: Estimating rupture dimensions of three major earthquakes in Sichuan, China, for early warning and rapid loss estimates, *B. Seismol. Soc. Am.*, 110, 920–936, doi:10.1785/0120190179, 2020.
- Li, J., Böse, M., Feng, Y., and Yang, C.: Real-time characterization of finite rupture and its implication for earthquake early warning: application of FinDer to existing and planned stations in Southwest China, *Front. Earth Sci.*, 9, 699560, doi:10.3389/feart.2021.699560, 2021.
- Li, J., Mignan, A., Sornette, D., and Feng, Y.: Predicting the future performance of the planned seismic network in Chinese mainland, *Seismol. Res. Lett.*, 94, 2698–2711, doi:10.1785/0220230084, 2023.
- Li, J., Sornette, D., Wu, Z., Zhuang, J., and Jiang, C.: Revisiting seismicity criticality: A new framework for bias correction of statistical seismology model calibrations, *J. Geophys. Res.*, 130, e2024JB029337, doi:10.1029/2024JB029337, 2025a.
- Li, J., Wang, X., and Sornette, D.: Unifying the Gutenberg-Richter law with probabilistic catalog completeness, *arXiv*, doi:10.48550/arXiv.2506.16849, 2025b.
- Li, J., Sornette, D., and Feng, Y.: Estimating earthquake early warning effectiveness via blind zone sizes: a case study of the new seismic network in Chinese mainland, *Nat. Hazards*, 121, 7783–7809, doi:10.1007/s11069-025-06809-1, 2025c.
- Li, J., Zhang, Q., and Sornette, D.: Simulated monthly earthquake catalogs (2000–2023) for the China Seismic Experimental Site (CSES) [Dataset], *Zenodo*, doi:10.5281/zenodo.17480120, 2025d.
- Li, J., Wyss, M., Wu, Z., and Zhou, S.: Estimated casualty risk for disaster preparation in five scenario great earthquakes, Sichuan-Yunnan region, China, in: *China Seismic Experimental Site: Theoretical Framework and Ongoing Practice*, edited by: Li, Y.-G., Zhang, Y., and Wu, Z., Springer, Singapore, 171–196, doi:10.1007/978-981-19-0951-2_8, 2022.
- Mizrahi, L., Nandan, S., and Wiemer, S.: Embracing data incompleteness for better earthquake forecasting, *J. Geophys. Res.-Sol. Ea.*, 126, e2021JB022379, doi:10.1029/2021JB022379, 2021.
- Mizrahi, L., Dallo, I., van der Elst, N. J., Christophersen, A., Spassiani, I., Werner, M. J., and Wiemer, S.: Developing, testing, and communicating earthquake forecasts: Current practices and future directions, *Rev. Geophys.*, 62, e2023RG000823, doi:10.1029/2023RG000823, 2024.
- Molchan, G. and Varini, E.: The strongest aftershock in seismic models of epidemic type, *Geophys. J. Int.*, 236, 1440–1454, doi:10.1093/gji/ggae100, 2024.
- Nandan, S., Ouillon, G., Sornette, D., and Wiemer, S.: Forecasting the full distribution of earthquake numbers is fair, robust, and better, *Seismol. Res. Lett.*, 90, 1650–1659, doi:10.1785/0220190020, 2019a.
- Nandan, S., Ouillon, G., Sornette, D., and Wiemer, S.: Forecasting the rates of future aftershocks of all generations is essential to develop better earthquake forecast models, *J. Geophys. Res.-Sol. Ea.*, 124, 8404–8425, doi:10.1029/2019JB017817, 2019b.
- Nandan, S., Kamer, Y., Ouillon, G., Hiemer, S., and Sornette, D.: Global models for short-term earthquake forecasting and predictive skill assessment, *Eur. Phys. J. Spec. Top.*, 230, 425–449, doi:10.1140/epjst/e2020-000259-3, 2021a.
- Nandan, S., Ram, S. K., Ouillon, G., and Sornette, D.: Is seismicity operating at a critical point?, *Phys. Rev. Lett.*, 126, 128501, doi:10.1103/physrevlett.126.128501, 2021b.
- Nandan, S., Ouillon, G., and Sornette, D.: Are large earthquakes preferentially triggered by other large events?, *J. Geophys. Res.-Sol. Ea.*, 127, e2022JB024380, doi:10.1029/2022jb024380, 2022.
- Ogata, Y.: Statistical models for earthquake occurrences and residual analysis for point processes, *J. Am. Stat. Assoc.*, 83, 9–27, doi:10.1080/01621459.1988.10478560, 1988.
- Ogata, Y.: Space-time point-process models for earthquake occurrences, *Ann. I. Stat. Math.*, 50, 379–402, doi:10.1023/a:1003403601725, 1998.

- Ogata, Y., Katsura, K., Falcone, G., Nanjo, K., and Zhuang, J.: Comprehensive and topical evaluations of earthquake forecasts in terms of number, time, space, and magnitude, *B. Seismol. Soc. Am.*, 103, 1692–1708, doi:10.1785/0120120064, 2013.
- Ogata, Y.: Prediction and validation of short-to-long-term earthquake probabilities in inland Japan using the hierarchical space–time ETAS and space–time Poisson process models, *Earth Planets Space*, 74, 110, doi:10.1186/s40623-022-01670-x, 2022.
- Saichev, A. and Sornette, D.: Power law distribution of seismic rates: theory and data analysis, *Eur. Phys. J. B*, 49, 377–401, doi:10.1140/epjb/e2006-00089-9, 2006.
- Saichev, A. and Sornette, D.: Power law distributions of seismic rates, *Tectonophysics*, 431, 7–13, doi:10.1016/j.tecto.2006.05.023, 2007.
- Saichev, A., Helmstetter, A., and Sornette, D.: Power-law distributions of offspring and generation numbers in branching models of earthquake triggering, *Pure Appl. Geophys.*, 162, 1113–1134, doi:10.1007/s00024-004-2661-8, 2005.
- Schorlemmer, D., Gerstenberger, M. C., Wiemer, S., Jackson, D. D., and Rhoades, D. A.: Earthquake likelihood model testing, *Seismol. Res. Lett.*, 78, 17–29, doi:10.1785/gssrl.78.1.17, 2007.
- Schorlemmer, D., Zechar, J. D., Werner, M. J., Field, E. H., Jackson, D. D., Jordan, T. H., and the RELM Working Group: First results of the regional earthquake likelihood models experiment, *Pure Appl. Geophys.*, 167, 859–876, doi:10.1007/s00024-010-0081-5, 2010.
- Scott, D. W.: Multivariate density estimation: theory, practice, and visualization, John Wiley & Sons, New York, NY, USA, 2015.
- Sornette, D. and Werner, M. J.: Apparent clustering and apparent background earthquakes biased by undetected seismicity, *J. Geophys. Res.*, 110, B09303, doi:10.1029/2005jb003621, 2005.
- Taroni, M., Marzocchi, W., Schorlemmer, D., Werner, M. J., Wiemer, S., Zechar, J. D., and Euchner, F.: Prospective CSEP evaluation of 1-day, 3-month, and 5-yr earthquake forecasts for Italy, *Seismol. Res. Lett.*, 89, 1251–1261, doi:10.1785/0220170197, 2018.
- Taroni, M., Spassiani, I., Laskin, N., and Barani, S.: How many strong earthquakes will there be tomorrow?, *Front. Appl. Math. Stat.*, 9, doi:10.3389/fams.2023.1234567, 2023.
- Veen, A. and Schoenberg, F. P.: Estimation of space-time branching process models in seismology using an EM-type algorithm, *J. Am. Stat. Assoc.*, 103, 614–624, doi:10.1198/016214508000000148, 2008.
- Wang, X., Li, J., Feng, A., and Sornette, D.: Estimating magnitude completeness in earthquake catalogs: A comparative study of catalog-based methods, *J. Geophys. Res.-Sol. Ea.*, 130, e2025JB031441, doi:10.1029/2025JB031441, 2025.
- Werner, M. J., Helmstetter, A., Jackson, D. D., and Kagan, Y. Y.: High-resolution long-term and short-term earthquake forecasts for California, *B. Seismol. Soc. Am.*, 101, 1630–1648, doi:10.1785/0120090142, 2011.
- Wu, Z.: Chapter 1: Seismic experimental sites: Challenges and opportunities, in: *China Seismic Experimental Site: Theoretical Framework and Ongoing Practice*, edited by: Li, Y., Zhang, Y., and Wu, Z., Higher Education Press, Beijing, 1–12, 2022.
- Wu, Z., Zhang, Y., and Li, J.: Coordinated distributed experiments (CDEs) applied to earthquake forecast test sites, in: *Earthquake and Disaster Risk: Decade Retrospective of the Wenchuan Earthquake*, edited by: Li, Y.-G., Springer, Singapore, 107–115, doi:10.1007/978-981-99-4507-9_7, 2019.
- Wu, Z. and Li, L.: China seismic experimental site (CSES): A systems engineering perspective, *Earthq. Sci.*, 34, 192–198, doi:10.29382/eqs-2021-0006, 2021a.
- Wu, Z. and Li, L.: China seismic experimental site (CSES): A system science perspective, *J. Geol. Soc. India*, 97, 1551–1555, doi:10.1007/s12594-021-1912-y, 2021b.

- Zhuang, J.: Next-day earthquake forecasts for the Japan region generated by the ETAS model, *Earth Planets Space*, 63, 207–216, doi:10.5047/eps.2011.01.002, 2011.
- Zhuang, J., Chang, C. P., Ogata, Y., and Chen, Y. I.: A study on the background and clustering seismicity in the Taiwan region by using point process models, *J. Geophys. Res.*, 110, B05S18, doi:10.1029/2004jb003157, 2005.
- Zhuang, J., Ogata, Y., and Vere-Jones, D.: Stochastic declustering of space-time earthquake occurrences, *J. Am. Stat. Assoc.*, 97, 369–380, doi:10.1198/016214502760046925, 2002.