doi:10.26443/seismica.v5i2.2585

SEISMICA

SPDE–ETAS: Fast and accurate Bayesian inference for the spatio-temporal epidemic type aftershock sequence (ETAS) model

Sofiane Taki-Eddine Rahmani * [1,2], **Gert Zöller** [1], **Sebastian Hainzl** [2,3], **Behnam Maleki Asayesh** [1,2]

[1]University of Potsdam, Institute for Mathematics, Karl-Liebknecht-Str.24/25, Potsdam-Golm, Germany, [2]GFZ Helmholtz Centre for Geosciences, Telegrafenberg, 14473 Potsdam, Germany, [3]University of Potsdam, Institute of Geosciences, Karl-Liebknecht-Str. 24-25 14476, Potsdam, Germany

Author contributions: *Conceptualization, Formal analysis, Methodology, Writing – original draft*: Sofiane Taki-Eddine Rahmani. *Writing – review & editing, Supervision, Methodology, Validation*: Gert Zöller, Sebastian Hainzl. *Writing – review & editing*: Behnam Maleki Asayesh.

Production Editor:
Gareth Funning
Handling Editor:
Giuseppe Petrillo
Copy & Layout Editor:
Nathan T. Stevens

Received:
January 30, 2026
Accepted:
June 22, 2026
Published:
August 7, 2026

Abstract The Epidemic-Type Aftershock Sequence model (ETAS) is a well-established point process framework for characterizing earthquake occurrences. Traditionally, ETAS parameters are estimated using maximum likelihood methods, where the background rate is smoothed by a Gaussian kernel density. More recently, a Bayesian formulation has been proposed, in which the background rate is modeled as a Gaussian Process prior, enabling a flexible semi-parametric estimation. However, this approach can become computationally demanding for large datasets due to the dense covariance structure of the Gaussian Process. An alternative representation of Gaussian processes with Matérn covariance functions can be obtained through Gaussian Markov Random Fields via the Stochastic Partial Differential Equation (SPDE) approach, which provides a sparse precision matrix formulation and significantly reduces computational complexity. Building on this connection, we introduce a computationally efficient formulation of the Bayesian Gaussian Process ETAS model, termed SPDE–ETAS model, where the Gaussian Process covariance matrix is replaced by a sparse precision matrix within a Log-Gaussian Cox Process framework. The precision matrix is constructed using the finite element method on a triangular mesh, and the log-prior of the Gaussian Markov Random Field is approximated via the Laplace approximation to accelerate computation. The model is first validated on synthetic datasets to assess its accuracy and computational efficiency compared to the kernel-based and Gaussian Process Epidemic-Type Aftershock Sequence approaches. It is then applied to the Italian earthquake catalog (1960–2025) to estimate the stationary background seismicity with its uncertainties, relevant for seismic hazard assessment.

1 Introduction

Point process models are widely used to characterize the occurrence of earthquakes in time and space. One of the most commonly used models is the Epidemic-Type Aftershock Sequence (ETAS) model (Ogata, 1998), which models the spatiotemporal clustering of earthquakes. The ETAS model has a wide range of applications, including earthquake forecasting, characterization of earthquake clusters, and understanding changes in earthquake patterns (e.g., Console and Murru, 2001; Zhuang et al., 2004, 2008; Werner et al., 2011; Lombardi, 2015; Ogata et al., 2019; Ebrahimian et al., 2022; Asayesh et al., 2023, 2025).

The ETAS model consists of two components. The first one is the background rate, which represents long-term, spontaneous earthquake activity related to tectonic processes. The second one is the triggering component, which models aftershocks generated by previous earthquakes. While the triggering mechanism is often well described by established parametric laws, such as the modified Omori law (Omori, 1894), the background rate is much harder to specify and remains a major source of uncertainty.

In most applications, the ETAS model is fitted using maximum likelihood estimation (e.g., Ogata, 1998; Zhuang et al., 2002; Helmstetter and Sornette, 2002; Zhuang et al., 2004; Hainzl et al., 2024). This approach provides a single estimate of the model parameters, which is then used for forecasting future seismicity. However, parameter estimation in the ETAS model is known to be difficult. The likelihood function is often relatively flat (Veen and Schoenberg, 2008; Ross, 2021), meaning that several parameter combinations can explain the observed data almost equally well. As a result, parameter estimates can be uncertain, and forecasts based on point estimates may be unreliable, even for large earthquake catalogs (Ross, 2021; Ebrahimian et al., 2022). Moreover, uncertainty quantification under the maximum likelihood framework is challenging and often relies on computationally expensive procedures such as bootstrap methods (Dutilleul et al., 2024).

Bayesian inference offers an alternative framework in which uncertainty about model parameters is explicitly represented through probability distributions,

*Corresponding author: rahmani@uni-potsdam.de

with prior distributions encoding initial uncertainty and the posterior distribution reflecting updated uncertainty after incorporating the observed data through the likelihood. In principle, this allows parameter uncertainty to be naturally propagated into earthquake forecasts (Ebrahimian et al., 2022; Molkenthin et al., 2024). In practice, however, Bayesian inference for the ETAS model is difficult. The posterior distribution is highly complex and exhibits strong parameter dependencies, making direct Markov chain Monte Carlo (MCMC) sampling inefficient. Standard random-walk Metropolis-Hastings algorithms mix poorly and do not scale well as the number of earthquakes increases (Ross, 2021).

One effective way to simplify Bayesian inference is to introduce a latent branching structure, in which each earthquake is classified as either a background event or a triggered aftershock. This representation reduces parameter correlations and leads to more stable and efficient inference algorithms (Ross, 2021).

Building on this idea, several Bayesian extensions of the ETAS model have been proposed that allow the background rate to be modeled flexibly. In particular, GP–ETAS models use Gaussian process priors to represent spatial (Molkenthin et al., 2022) or temporal variations (Muir and Ross, 2023) in the background intensity, whereas Dirichlet-ETAS models rely on Dirichlet process mixtures to capture spatial heterogeneity (Ross and Kolev, 2022). These approaches avoid strong parametric assumptions and can capture complex patterns in seismicity. However, Gaussian process–based models involve dense covariance matrices, and Dirichlet process formulations require repeated sampling over complex mixture structures. As a result, these methods become difficult to apply to large earthquake catalogs.

In this work, we propose an alternative Bayesian GP–ETAS model, termed Stochastic Partial Differential Equation ETAS (SPDE–ETAS), that retains a flexible, non-parametric background representation while improving computational efficiency. The spatial background seismicity is modeled within a Log-Gaussian Cox Process (LGCP) framework, consistent with existing latent Gaussian ETAS frameworks, but with the latent Gaussian field defined by a stochastic partial differential equation (SPDE). This construction leads to a Gaussian Markov random field (GMRF) with a sparse precision matrix, thereby substantially reducing computational costs (Lindgren et al., 2011). As a result, the proposed approach does not modify the underlying statistical model, but provides a computationally efficient approximation that enables scalable Bayesian inference. The SPDE representation is discretized using the finite element method on a triangular mesh, resulting in a GMRF approximation of the latent spatial field, whose posterior distribution is approximated using a Laplace approximation around its mode (Simpson et al., 2016). This approach allows the model to represent complex spatial patterns while remaining scalable to large datasets. Combined with a latent branching formulation of the ETAS model, the proposed approach enables efficient Bayesian inference for earthquake catalogs that are challenging to analyze using maximum likelihood estimation or GP-based Bayesian methods.

The remainder of this paper is organized as follows. We first introduce the SPDE–ETAS model and describe its numerical implementation with a definition of the LGCP and SPDE. Then, we illustrate the performance of the method using synthetic and real earthquake data and discuss the results.

2 SPDE–ETAS

The ETAS model (Ogata, 1998; Zhuang et al., 2002) is a Hawkes process — a self-exciting point process model (Hawkes, 1971) — that models earthquake occurrences in space, time, and magnitude. It decomposes seismicity into two components: spontaneous background events and triggered aftershocks, with a latent branching structure governing the causal relationships between events. The conditional intensity function of the ETAS model is

$$\lambda(t, \boldsymbol{x}, m \mid H_t, \theta_\mu, \theta_\varphi) = \mu(\boldsymbol{x} \mid \theta_\mu) + \sum_{\{i:t_i<t\}} \varphi(t - t_i, \boldsymbol{x} - \boldsymbol{x}_i, m_i \mid \theta_\varphi)\,. \quad (1)$$

Here, H_t denotes the history of the process up to time t, consisting of the set of past events $\{(t_i, \boldsymbol{x}_i, m_i)\}$, where t_i, $\boldsymbol{x}_i$, and m_i represent the occurrence time, spatial location, and magnitude of event i, respectively. The parameter vector θ denotes all model parameters and is decomposed as $\theta = (\theta_\mu, \theta_\varphi)$, corresponding to the background and triggering parameters, respectively. The first term, $\mu(\boldsymbol{x} \mid \theta_\mu)$, represents the background rate intensity, modeling spontaneous events in space, e.g. related to tectonic loading, with parameters θ_μ specified in more detail in Section 2.1. The second term, $\varphi(\cdot)$, is the *triggering function*, modeling aftershocks caused by previous events.

Event magnitudes m_i are independent and follow the Gutenberg–Richter distribution:

$$p_M(m \mid \beta) = \beta\, e^{-\beta(m-m_0)}, \quad m \geq m_0, \quad (2)$$

where $\beta = \ln(10)b$ is the rescaled Gutenberg-Richter b-value and m_0 is the minimum magnitude.

A typical parametrization of the triggering function with $\theta_\varphi = (K, \alpha, c, p, d, \gamma, q)$ is

$$\varphi(t - t_i, \boldsymbol{x} - \boldsymbol{x}_i, m_i \mid \theta_\varphi) = \kappa(m_i \mid K, \alpha)\, g(t - t_i \mid c, p)\, s(\boldsymbol{x} - \boldsymbol{x}_i, m_i \mid d, \gamma, q), \quad (3)$$

where the terms correspond to:

- *Productivity* $\kappa(m_i \mid K, \alpha) = K e^{\alpha(m_i - m_0)}$, controlling the expected number of aftershocks per event (also called the Utsu law, see Utsu, 1970);
- *Temporal decay* $g(t - t_i \mid c, p) = (t - t_i + c)^{-p}$, representing the modified Omori–Utsu law (Omori, 1894; Utsu, 1961), which characterizes the temporal decay of aftershocks after a mainshock at time t_i, with p being the power law exponent and c a time lag between the mainshock and the start of the power-law decay;

- *Spatial kernel*

$$s(\boldsymbol{x}-\boldsymbol{x}_i, m_i \mid d,\gamma,q) = \frac{q-1}{\pi\,\sigma_m(m_i)}\left[1+\frac{(\boldsymbol{x}-\boldsymbol{x}_i)(\boldsymbol{x}-\boldsymbol{x}_i)}{\sigma_m(m_i\mid d,\gamma)}\right]^{-q},$$

where $\sigma_m(m_i \mid d,\gamma) = d^2 10^{2\gamma m_i}$ models the spatial distribution of aftershocks around the parent event with parameters d, γ and q (Kagan, 2002; Ogata and Zhuang, 2006).

The link between the background and triggered events can be characterized by an unobserved *latent branching structure* (Hawkes and Oakes, 1974), where each event i has an unobserved parent indicator B_i and is typically modeled by discrete integer values according to:

$$B_i = \begin{cases} 0, & \text{if event } i \text{ is a background event,} \\ j>0, & \text{if event } i \text{ is triggered by event } j. \end{cases} \tag{4}$$

Background events occur according to a Poisson process with intensity $\mu(\boldsymbol{x})$ and serve as roots of clusters. Each past event j independently generates offspring events (aftershocks) according to a Poisson process with rate $\varphi(t-t_j, \boldsymbol{x}-\boldsymbol{x}_j \mid m_j)$. Although B_i is not observed, one can compute the probability that an event is either a background event

$$P_{i0} = \frac{\mu(\boldsymbol{x}_i \mid \theta_\mu)}{\lambda(t_i,\boldsymbol{x}_i \mid H_{t_i},\theta_\mu,\theta_\varphi)}, \tag{5}$$

or triggered by a specific parent:

$$P_{ij} = \frac{\varphi(t_i-t_j,\boldsymbol{x}_i-\boldsymbol{x}_j \mid m_j)}{\lambda(t_i,\boldsymbol{x}_i \mid H_{t_i},\theta_\mu,\theta_\varphi)}. \tag{6}$$

This latent structure provides a probabilistic decomposition of the catalog into background and triggered events, highlighting the cascading nature of seismicity.

2.1 Log-Gaussian Cox Process (LGCP)

While the ETAS model relies on well-defined parametric formulations of the triggering kernel, the spatial background component may display complex variability that is difficult to capture with a simple parametric form. To overcome this limitation, the background rate can instead be modeled as a realization of a latent random field, allowing for greater flexibility in representing heterogeneous spatial structures. In this work, we adopt a hierarchical formulation in which the background intensity is treated as a stochastic process rather than a fixed but unknown function.

A natural extension in this direction is the Cox process, or doubly stochastic Poisson process (Cox, 1955), where the intensity function itself is modeled as a stochastic process. Among the different formulations, the *log-Gaussian Cox process* (LGCP) is particularly appealing: it assumes that the log-intensity is driven by a Gaussian Random Field (GRF), which enables a non-parametric and spatially correlated description of the seismic background activity.

For a homogeneous Poisson process, the intensity λ can be expressed as

$$\lambda = \exp(\mu_0), \tag{7}$$

where μ_0 represents the average log-intensity of the process. In the case of an inhomogeneous Poisson process, a spatially varying GRF $z(\boldsymbol{x})$ can be added to account for residual spatial correlation among points:

$$\lambda(\boldsymbol{x}) = \exp\{\mu_0 + z(\boldsymbol{x})\}. \tag{8}$$

Here, $\mu_0 + z(\boldsymbol{x})$ serves as a linear predictor, representing the log-intensity of the point process as an additive combination of a global mean level and a spatially structured GRF (Cox, 1955). This formulation is consistent with existing Gaussian process-based ETAS approaches, including logistic-Gaussian Cox process models and hierarchical GP-based formulations, which similarly rely on a latent Gaussian field (Molkenthin et al., 2022).

The latent component $z(\boldsymbol{x})$ is naturally modeled as a GRF, a powerful mathematical framework for representing latent or unobserved spatial processes (Cressie and Johannesson, 2008; Stein, 1999; Chiles and Delfiner, 2012). In this context, the GRF captures the spatial correlation structure underlying the LGCP background intensity.

Formally, a Gaussian random field $z(\boldsymbol{x})$, defined on a spatial domain $\mathcal{D} \subset \mathbb{R}^2$, is characterized by its finite-dimensional distributions. That is, for any finite collection of locations $\boldsymbol{x}_1,\dots,\boldsymbol{x}_n \in \mathcal{D}$, the random vector $(z(\boldsymbol{x}_1),\dots,z(\boldsymbol{x}_n))^\top$ follows a multivariate normal distribution:

$$(z(\boldsymbol{x}_1),\dots,z(\boldsymbol{x}_n))^\top \sim \mathcal{N}(\mu,\Sigma), \tag{9}$$

where μ is the mean vector and Σ is the covariance matrix. Estimating Σ directly can be computationally expensive for large spatial datasets (Banerjee et al., 2003). Instead, a standard correlation function can be used to define the covariance between points based on their spatial distance. A common choice is the Matérn covariance function (Guttorp and Gneiting, 2006):

$$\mathrm{Cov}_M(\boldsymbol{x}_i,\boldsymbol{x}_j) = \sigma^2 \mathrm{Corr}_M(\boldsymbol{x}_i,\boldsymbol{x}_j), \tag{10}$$

where σ^2 is the variance, and Corr_M is the Matérn correlation function. For spatial locations $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$, the Matérn correlation is defined as

$$\mathrm{Corr}_M(\boldsymbol{x}_i,\boldsymbol{x}_j) = \frac{2^{1-\nu}}{\Gamma(\nu)}\left(\kappa\|\boldsymbol{x}_i-\boldsymbol{x}_j\|\right)^\nu K_\nu\left(\kappa\|\boldsymbol{x}_i-\boldsymbol{x}_j\|\right), \tag{11}$$

where $\|\boldsymbol{x}_i - \boldsymbol{x}_j\|$ is the Euclidean distance between points, K_ν is the modified Bessel function of the second kind, and κ and ν are parameters controlling the range and smoothness of the correlation. In many applications, ν is fixed, simplifying the correlation to depend mainly on κ and the distance between points. Smaller values of κ correspond to stronger long-range spatial dependence. The background parameter vector θ_μ introduced in Section 2 is specified within the LGCP formulation as consisting of the mean log-intensity μ_0 and the parameters governing the Gaussian random field $z(\boldsymbol{x})$, in particular the variance σ^2 and the range parameter κ.

Although the parameters σ^2 and κ fully describe the Gaussian field, computing the full covariance matrix can be computationally demanding. By assuming that the field is locally dependent—meaning that only nearby points are correlated—the covariance matrix becomes sparse. This leads to an approximation of the GRF by a Gaussian Markov random field (GMRF). Lindgren et al. (2011) provided a framework linking GRFs with Matérn covariance to GMRFs, enabling efficient computation even when the true spatial correlation is long-range.

2.2 Stochastic Partial Differential equation (SPDE)

Lindgren et al. (2011) showed that a GMRF can be obtained as a finite-dimensional approximation to a GRF with a Matérn covariance function, the field is defined as the solution of a stochastic partial differential equation (SPDE), in a studied domain D, of the form:

$$(\kappa^2 - \Delta)^{\frac{\alpha}{2}}(\tau u) = W \;, \tag{12}$$

where Δ is the Laplacian operator, W is Gaussian white noise, and τ is given by

$$\tau^2 = \frac{\Gamma(\nu)}{\sigma^2 \kappa^{2\nu} (4\pi)^{d/2} \Gamma(\nu + d/2)}, \tag{13}$$

and the fractional power α is linked to the smoothness parameter by

$$\alpha = \nu + \frac{d}{2}. \tag{14}$$

For 1D cases, the SPDE can be solved by the finite difference method. However, for higher-dimensional problems such as seismicity distributions, it is solved with the finite element method. Following Lindgren et al. (2011), the finite element discretization leads to a sparse precision matrix representation of the GMRF (Edelsbrunner, 2001; Henderson et al., 2002; Hjelle and Dæhlen, 2006). The full finite element discretization, including mesh construction, basis functions, and the explicit form of the mass and stiffness matrices, is detailed in Appendix A.

2.3 Bayesian framework

Posterior inference is performed using a blockwise MCMC scheme that largely follows the methodology of Ross (2021) and Molkenthin et al. (2022). The main difference lies in the treatment of the spatially varying background rate, which is here modeled as a LGCP and inferred jointly with its parameters through a GMRF representation and a Laplace approximation.

The MCMC algorithm updates the model components sequentially in three weakly dependent blocks:

- the branching structure B,
- the background rate $\mu(x, y)$, through joint updates of the latent field z and parameters θ_μ using a Laplace approximation,
- the triggering parameters θ_φ.

At iteration k, the state of the Markov chain is

$$\theta_k = (B_k, z_k, \theta_{\mu,k}, \theta_{\varphi,k}), \tag{15}$$

and updates are generated using Metropolis–Hastings steps targeting the joint posterior distribution

$$\pi(B, z, \theta_\mu, \theta_\varphi \mid Y), \tag{16}$$

where $Y = \{(t_i, x_i, y_i, m_i)\}_{i=1}^N$ denotes the observed earthquake catalogue, consisting of the occurrence times t_i, spatial coordinates (x_i, y_i), and magnitudes m_i of the N recorded events. The transition kernels are constructed such that this posterior distribution is the stationary distribution of the Markov chain. For completeness, the full conditional distributions, likelihood formulations, and computational details of each MCMC update step are provided in Appendix B.

3 Synthetic experiments

To demonstrate the potential of our approach to estimate the spatial background activity, we conducted three tests by comparing three models: SPDE–ETAS, GP–ETAS (see Appendix C for details), and Kernel Density Estimation ETAS (KDE–ETAS; see Appendix D; Zhuang et al., 2002). For the first two experiments, all three models are considered, whereas only the SPDE–ETAS and GP-–ETAS models are used in the third experiment.

3.1 Area-like source

The first experiment was conducted by generating a synthetic ETAS-based catalog over a region $[0.0, 5.0] \times [0.0, 5.0]$ degrees for a period of 1000 days, with two zones characterized by different background rates defined as

$$\mu(x) = \begin{cases} 0.08, & x \in [0, 3] \times [1.5, 5.0], \\ 0.01, & \text{otherwise}. \end{cases} \tag{17}$$

For the aftershock parameters, we selected values within the typical range of observations (Ogata, 1998; Zhuang et al., 2004), namely $K = 0.01$, $\alpha = 1.70$, $c = 0.01$ days, $p = 1.20$, $D = 0.015$ km, $\gamma = 0.20$, and $q = 2.00$. Earthquakes magnitudes are sampled from a Gutenberg-Richter relation with $b = 1.0$, i.e. $\beta = 2.30$, and $m \geq 0$. The corresponding branching ratio is $n = 0.42$, which represents the mean number of offsprings generated by a parent (Helmstetter and Sornette, 2002). This configuration ensures that aftershocks are generated in both space and time, resulting in a total of 1458 events. The spatial and temporal distribution is shown in Figs. 1a and f.

We then proceed to the estimation of the background rate and triggering parameters using the SPDE–ETAS, GP–ETAS, and KDE–ETAS models, and compare the results obtained for the same synthetic catalog. For the GP–ETAS model, a regular grid is used as introduced in Molkenthin et al. (2022) to discretize the domain into

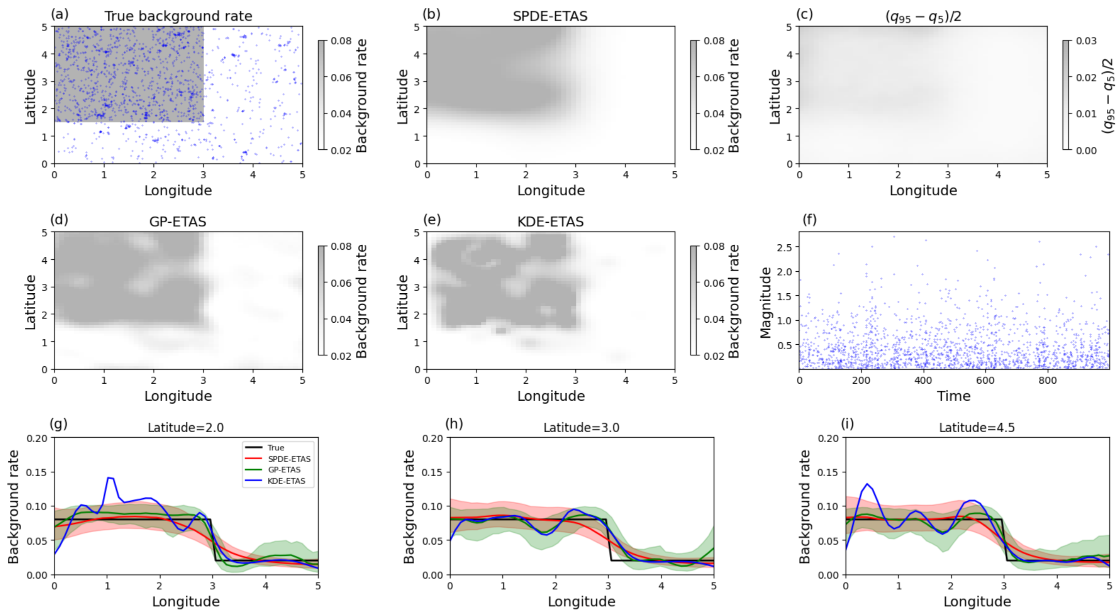

Figure 1 Synthetic catalog and experimental results of the first synthetic test. (a) True spatial background intensity with the spatial distribution of the synthetic events (blue points). (b) Median background intensity estimated using the SPDE–ETAS model. (c) Uncertainty of the SPDE–ETAS estimate, expressed as semi-interquartile 0.05, 0.95 distance $(q_{95} - q_5)/2$. (d) Background intensity estimated using the GP–ETAS model. (e) Background rate estimated using the KDE–ETAS method. (f) Magnitude–time plot of the synthetic catalog. (g), (h), and (i) Cross-sections at latitudes 2.0, 3.0, and 4.0, respectively. The black line represents the true background rate, the blue line the background rate estimated with KDE–ETAS, the red line and shaded band the background rate and its 90% credible interval from SPDE–ETAS, and the green line and shaded band the background rate and its credible interval from GP–ETAS.

2500 grid cells. The prior distributions for all parameters are set to finite uniform prior distributions, indicating no specific prior knowledge. For the SPDE–ETAS model, the domain is discretized using a triangular mesh, with smaller and more densely distributed elements in regions of higher seismic activity. The finite element method is then employed to solve the SPDE and to estimate both the mass and stiffness matrices, which are used to construct the precision matrix. The priors for both the LGCP parameters and the triggering parameters are set to uniform distributions. The Laplace approximation is used to compute the mode of the posterior distribution of the GMRF. Concerning the KDE–ETAS, the minimum bandwidth is set to 0.05 degrees with the number of nearest neighbors fixed to $n_p = 15$ (Zhuang et al., 2002).

The results for the background rate estimated using the SPDE–ETAS, GP–ETAS, and KDE–ETAS models are shown in panels b, c, d, and e of Fig. 1. Both SPDE–ETAS and GP–ETAS medians outperform the classical KDE–ETAS, as the background rate in regions of highest intensity is well recovered and exhibits a smooth spatial structure, in contrast to the strong local fluctuations observed in the KDE–ETAS estimates.

A more detailed comparison, including the associated uncertainty, is provided by the cross-sections along latitudes 2.0, 3.0, and 4.5 (panels g, h, and i of Fig. 1), which intersect areas of high background activity. In these cross-sections, the credible bands of the SPDE–ETAS and GP–ETAS models cover most of the ground truth and remain spatially smooth, whereas the KDE–ETAS estimates exhibit pronounced variability and provide no credible intervals.

Concerning the triggering parameters (Fig. 2), both SPDE–ETAS and GP–ETAS also provide uncertainty distributions, unlike the point estimations of KDE–ETAS. The posterior distributions obtained with the Bayesian approaches consistently cover the ground-truth parameter values, highlighting the advantage of Bayesian inference, which provides full posterior uncertainty quantification. Another disadvantage of KDE is that the nearest neighbor used for the bandwidth must be predefined (see Appendix D) and is not the result of an optimization procedure, which can lead to over-smoothing or under-smoothing if an inappropriate nearest neighbor parameter is selected. It is also worth noting that, for GP–ETAS, the true values of the parameters K, c, q, and γ lie close to the medians of the corresponding posterior distributions (Fig. 2a, b, f, and g). The true number of background events is also well covered by the posterior distributions of both SPDE–ETAS and GP–ETAS, with a slightly better performance observed for GP–ETAS (Fig. 2h).

3.2 Fault-like sources

The second experiment is designed to be more realistic with respect to earthquake occurrence, features three

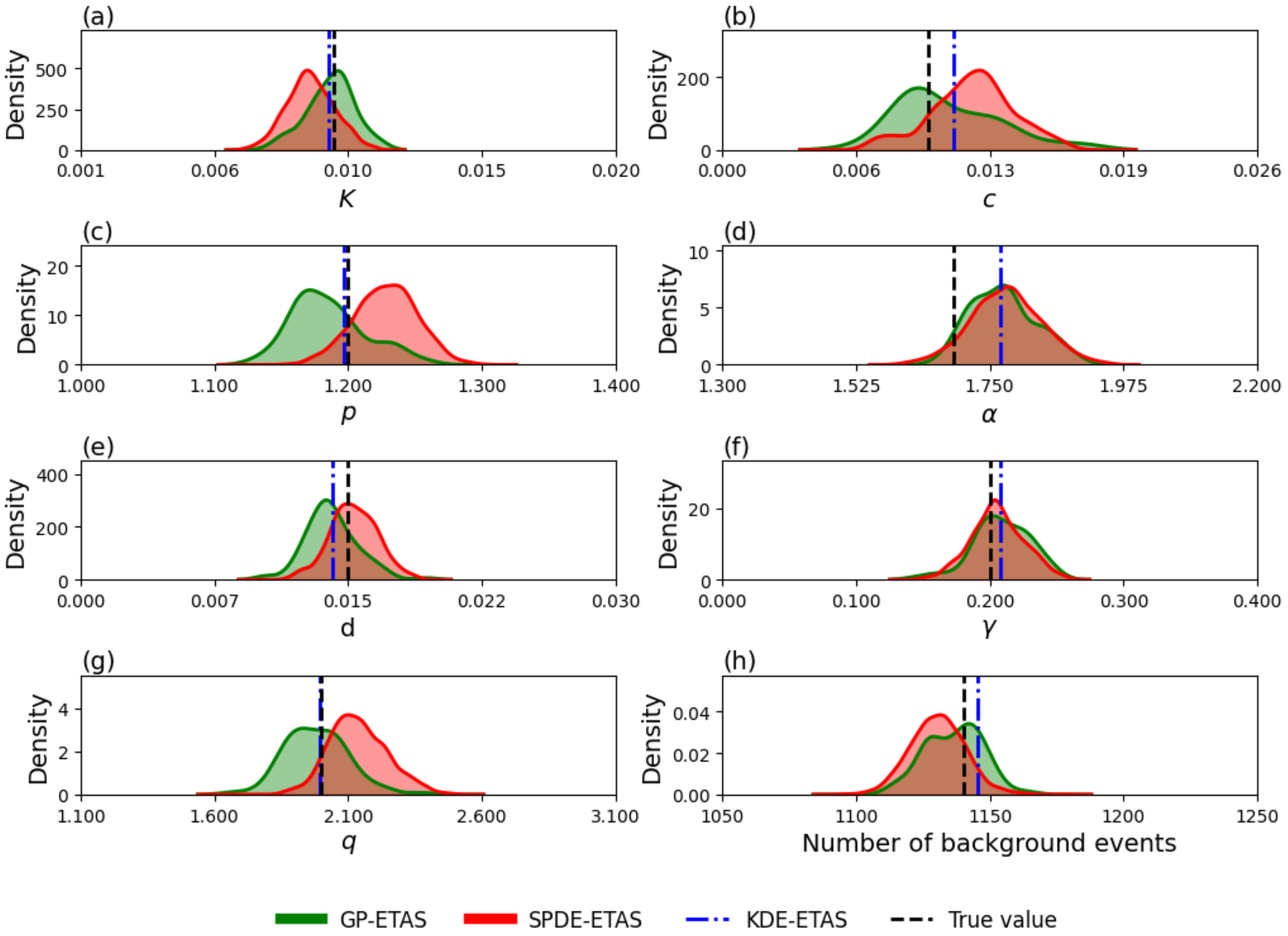

Figure 2 (a)–(g) Triggering parameter estimates, and (h) estimated number of background events for the first test case. In all panels, the black dashed line indicates the true values, the blue dashed line shows the KDE–ETAS estimates, the red shaded area represents the SPDE–ETAS posterior distributions, and the green shaded area represents the GP–ETAS posterior distributions.

elongated stripes representing fault zones with individual lengths and background rates. The synthetic catalog is generated over the region $[0.0, 5.0] \times [0.0, 5.0]$ for a period of 1000 days, with the background rate defined as

$$\mu(x) = \begin{cases} 0.6, & x \in [1.0, 3.0] \times [1.4, 1.5], \\ 0.6, & x \in [1.0, 4.0] \times [2.4, 2.5], \\ 0.4, & x \in [2.0, 3.0] \times [4.0, 4.1], \\ 0.01, & \text{otherwise}, \end{cases}$$

The aftershock parameters in this experiment are set to $K = 0.01$, $\alpha = 1.80$, $c = 0.01$, $p = 1.20$, $D = 0.015$, $\gamma = 0.50$, and $q = 1.80$. Earthquake magnitudes are generated in the same manner as in the first test, using the same b-value, with $m \geq 0$, resulting in a total number of 965 generated events.

The synthetic catalog and the inversion results are shown in Fig. 3. The credible band of the SPDE–ETAS tends to be wider than that of the GP–ETAS in regions with high background seismicity, with noticeable fluctuations, as also reflected in the semi-interquantile 0.05, 0.95 distance plot (Fig. 3b, c, d, g, h, and i). In certain areas, the SPDE–ETAS even reaches the maximum intensity. In contrast, the KDE–ETAS exhibits strong fluctuations and fails to reach the maximum intensity (Fig. 3e, g, h, and i).

It is important to note that this wider credible interval for the SPDE–ETAS are not a systematic feature. This behavior is mainly observed in the second synthetic test, particularly in regions with high background seismicity, while it is not observed in the first synthetic test, where both approaches yield comparable uncertainty levels. In the second test, the wider SPDE–ETAS credible intervals tend to better cover the true background rate, as illustrated in the cross-sections (Fig. 3g–i), where the SPDE–ETAS bands more consistently include the true signal compared to the GP–ETAS. We therefore interpret this behavior as a more conservative and robust quantification of uncertainty in regions with high and spatially structured intensity, rather than as a limitation of the model.

The aftershock triggering parameters are well covered by the posterior distributions of the SPDE–ETAS compared to the GP–ETAS, especially for the p-value (Fig. 4a-g). Similar observations can be done on the number of background events, where the SPDE–ETAS outperforms the GP–ETAS (Fig. 4h).

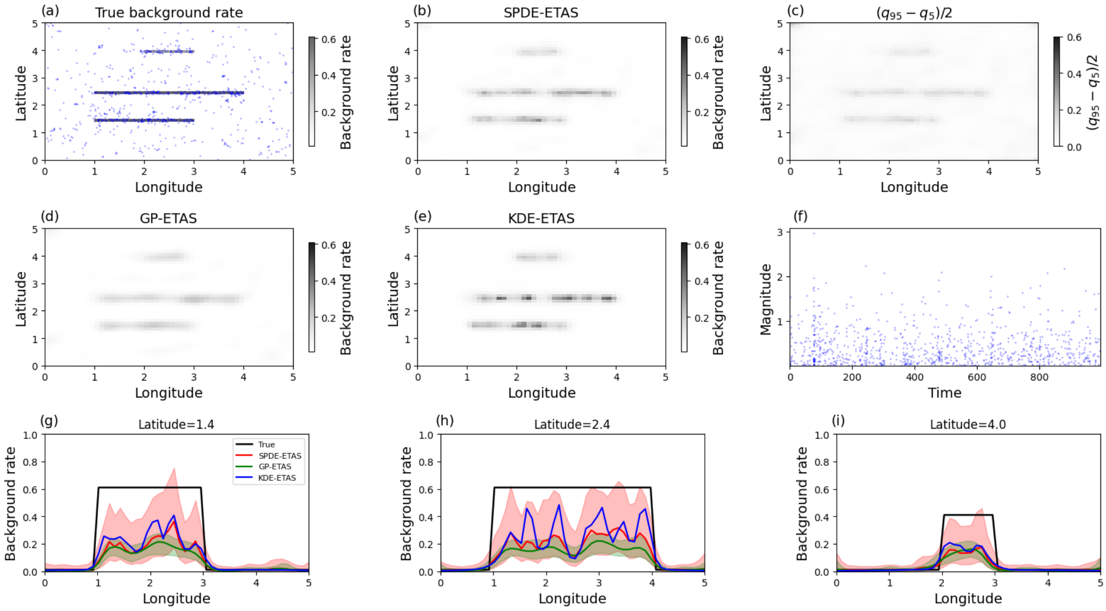

Figure 3 Similar figure as Fig. 1 for the second synthetic test. Here, panels (g), (h), and (i) show cross-sections at latitudes 1.4, 2.4, and 4.0, respectively.

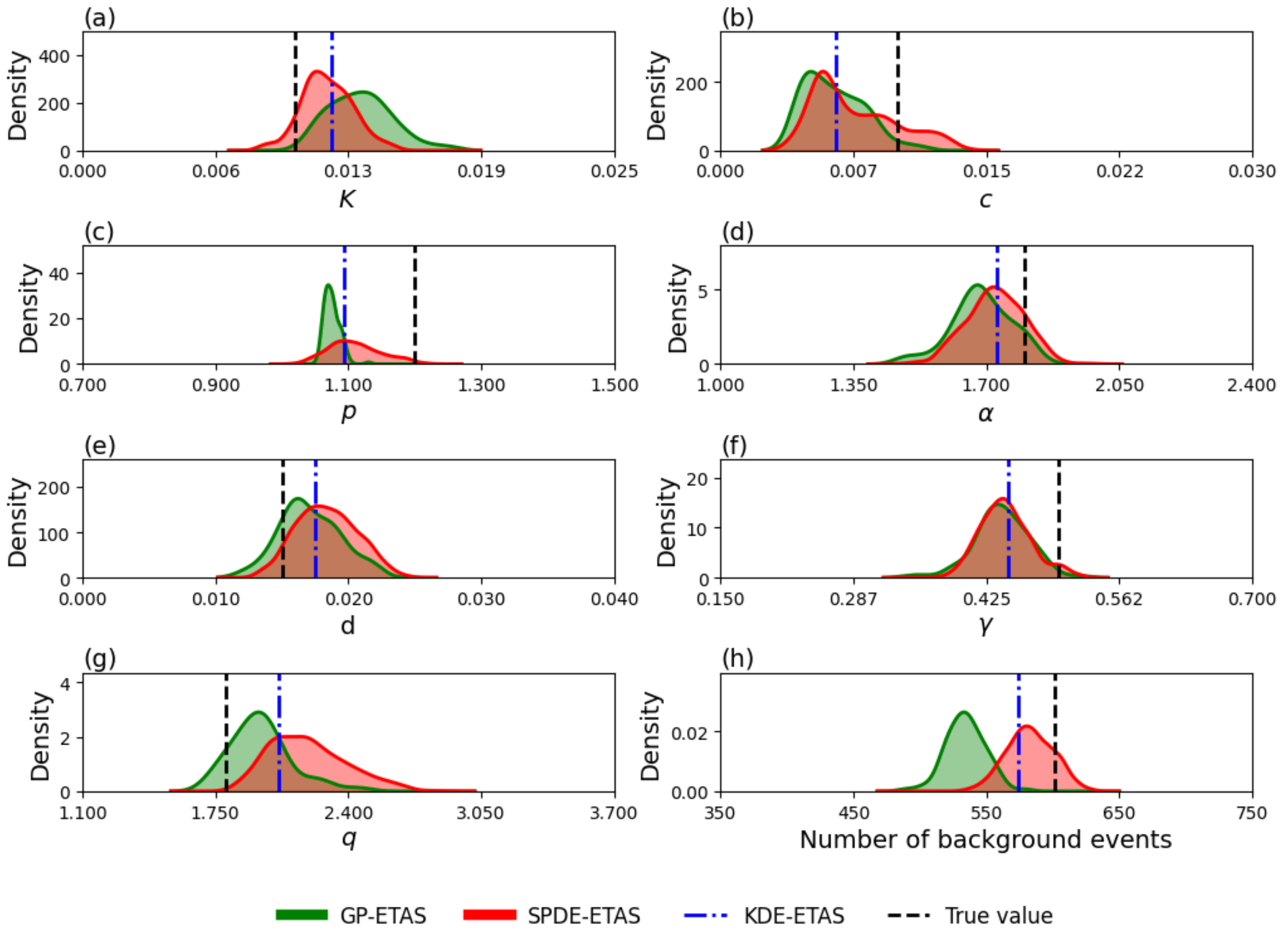

Figure 4 (a)–(g) Triggering parameter estimates, and (h) estimated number of background events for the second test.

3.3 Computational efficiency and accuracy

The third synthetic experiment focuses on the computational efficiency (speed and memory usage), and accuracy of both the SPDE–ETAS and GP–ETAS models. In each case, the background rate is increased to generate a larger number of events. The setup and the aftershock parameters are kept the same as in the second experiment, and a total of eleven synthetic catalogs are generated, ranging from 185 to 5200 events. The estimation procedure is then applied to all catalogs following the same approach as in the first two experiments for both the SPDE–ETAS and GP–ETAS models. The computations were performed on a high-performance computing cluster using a single node equipped with an Intel Xeon Gold 5115 CPU (2.40 GHz, up to 3.20 GHz), with 20 physical cores (2 sockets $\times$ 10 cores per socket, 40 threads) and 512 GB of RAM.

The results are shown in Fig. 5. As a measure of the accuracy, we calculate the normalized sum of the log-likelihood at N equidistant grid points, where the posterior pdf of the background rate is evaluated at the value of the true rate, i.e., $\mathcal{LL} = (1/N)\sum_{n=1}^{N} \ln[\mathrm{pdf}(\mu_{\mathrm{true}}(\boldsymbol{x}_n)]$. A water level of 10^{-4} was imposed to replace regions where the true background rate falls to zero in the probability density function. Larger $\mathcal{LL}$ values indicate a higher model accuracy. The result is presented in Fig. 5a. SPDE–ETAS outperforms GP–ETAS for most sample sizes, except for 493 and 5020 events where GP–ETAS performs slightly better. However, in regions with high background rate intensity, as highlighted in the inset panel, SPDE–ETAS clearly outperforms GP–ETAS for all sample sizes. Fig. 5b, and c show the computational efficiency and memory usage of both models for different numbers of events. As expected, due to the GMRF approximation for the latent random field, which keeps the precision matrix sparse, SPDE–ETAS is considerably faster and requires substantially less memory than GP–ETAS, even for large catalogs.

4 Case study: Italian region

As a real-data application, we choose the Italian catalog to show the results of the SPDE–ETAS model. The dataset comprises more than 514,465 events spanning from 1960 to 2025, over a spatial domain of $[37°\mathrm{N}, 47°\mathrm{N}] \times [7°\mathrm{E}, 18.5°\mathrm{E}]$. The data were obtained from the website of the National Institute of Geophysics and Volcanology of Italy (https://horus.bo.ingv.it, INGV), which provides a homogenized catalog (Lolli et al., 2020). For the fit, we selected the 14,802 events with magnitudes $M_w \geq 3.0$ in the period from 1980 to 2025, to ensure completeness of the fitted data (Fig. 6).

The priors and number of iterations used for the inference are the same as those used for the synthetic tests, except that no comparison with other models is performed in this case because SPDE–ETAS has been found to be superior in the synthetic tests, and the data set is too large for the GP–ETAS model. The domain is discretized using a triangular mesh, with a maximum triangle area set to 100 km^2 to ensure higher density in regions with a large number of events. The runtime for the Bayesian inference on this catalog is approximately 5 hours, which is efficient given the size of the dataset. For comparison, we also apply the KDE-ETAS method to estimate the parameters for three nearest-neighbor settings: 5, 10, and 15. However, we could not include GP–ETAS for this large dataset and long observation period, as its computational cost becomes too high for catalogs with around 10,000 events and above, as observed in our experiments. The results for the background rate estimation are shown in Fig. 7 on a logarithmic scale, including the median background rate (panel a) and the 5% and 95% quantiles (panels c and d, respectively). The highest background rate intensity is clearly located in the Etna volcanic area, which is consistent with expectations, as regions with fluids or magma are likely to exhibit higher background seismicity (e.g., Zhuang et al., 2002; Hainzl and Ogata, 2005; Lombardi et al., 2010). The estimated background rate also aligns well with the seismic hazard map of Italy MPS19 (Fig. 7b) (Meletti et al., 2019), where areas with the highest background rate correspond to regions with the highest probability of exceeding specified ground-motion thresholds according to the map.

The estimated triggering parameters are shown in Figs. 8a–g and Table 1. Almost all KDE estimates obtained with different nearest-neighbor values fall within the credible intervals of the SPDE–ETAS model. The parameters are shown in Figs. 8h–j. The range scale (κ) is small, indicating strong correlations between points at small distances, which is related to short-distance fluctuations of the background rate (Fig. 8h), and the value of τ, which controls the precision of the background rate, indicates limited to moderate amplitudes of background intensity variations (Fig. 8i).

5 Discussion and conclusion

Reliable estimates of the spatially varying background rate are essential for both long-term seismic hazard assessment and operational earthquake forecasting. Traditional hazard assessments rely on two separate steps: First, aftershocks are removed using a specific declustering method. Second, earthquake rates are calculated by zonation or smoothing of the declustered earthquake locations. In contrast, SPDE–ETAS offers an approach that consistently estimates the spatially varying background rate. Moreover, the Bayesian formulation naturally quantifies the uncertainty of the background rate, providing credible intervals that support both robust hazard evaluation and operational forecasting.

We have demonstrated that the proposed SPDE–ETAS model provides a fast and efficient approach for characterizing the spatially varying background rate within the Bayesian ETAS framework. The background seismicity is modeled using a LGCP with a GMRF, obtained through the SPDE approach and discretized using a finite element method on a triangular mesh. The resulting sparse precision matrix enables efficient inference via Laplace approximation, making the model computationally attractive for large-scale applications.

Three synthetic experiments were conducted to as-

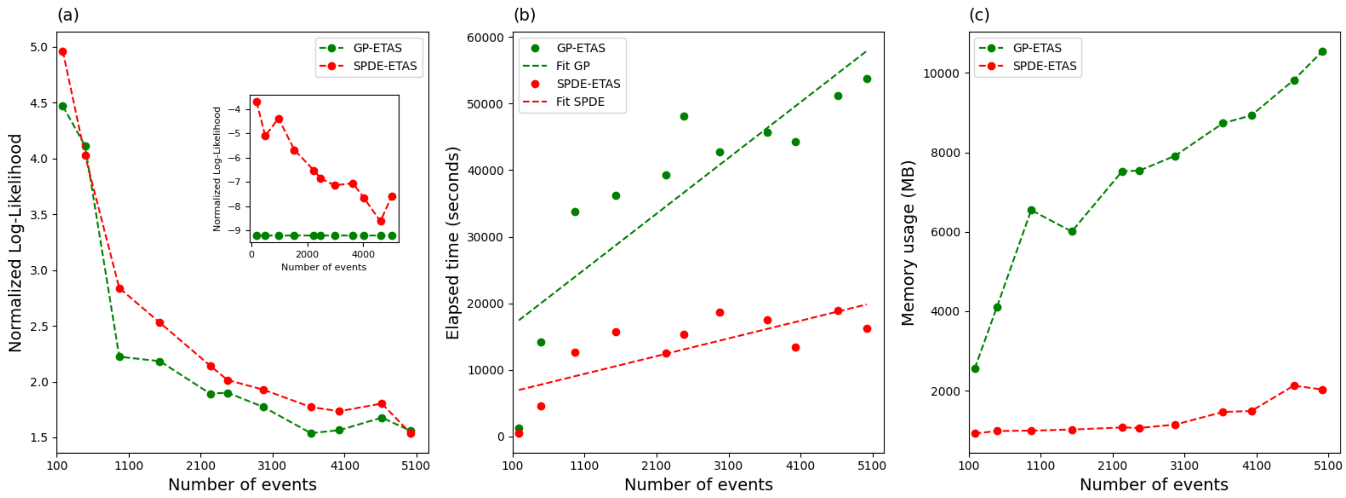

Figure 5 (a) Mean log-likelihood for equidistant grid cells, where the posterior distribution is compared to the true background rate. The SPDE–ETAS results are shown with the red dashed line with points, and the GP–ETAS results with the green dashed line with points. The inset panel shows the sum of the log-likelihoods, normalized by the number of grid cells, for the grid points corresponding only to regions of high background rate. (b) Computational efficiency of the GP–ETAS and SPDE–ETAS models. Red circles indicate the elapsed time for SPDE–ETAS, and green circles for GP–ETAS. The red dashed line shows the linear fit for SPDE–ETAS, and the green dashed line shows the fit for GP–ETAS. (c) Memory usage in Megabytes (MB) as a function of the number of events for GP–ETAS and SPDE–ETAS, represented by the green and red dashed lines with points, respectively.

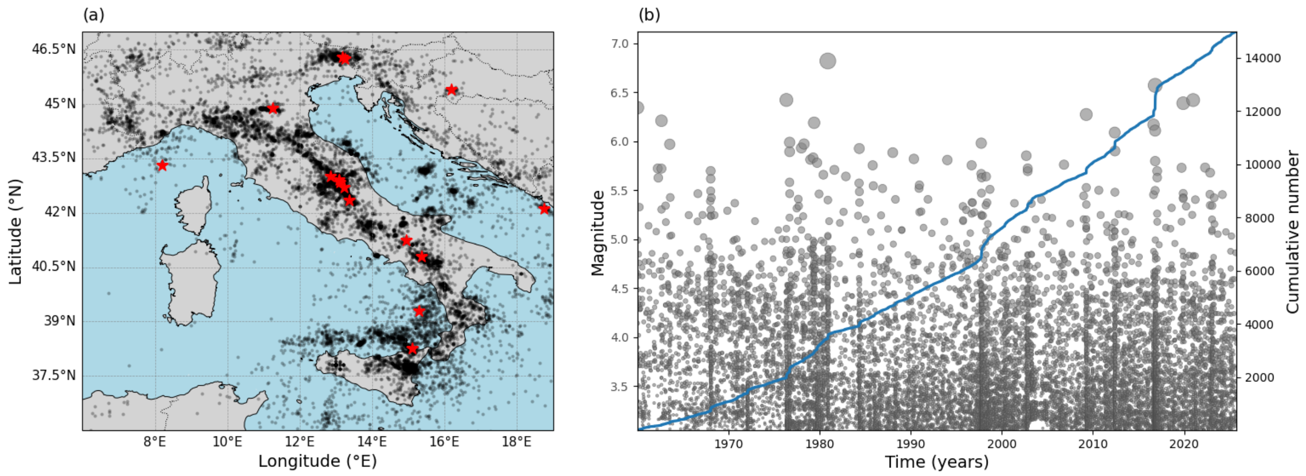

Figure 6 (a) Spatial distribution of earthquakes from 1960 to 2025 with magnitudes above 3.0. Red stars indicate events with magnitude greater than or equal to 6.0. (b) Magnitude-time plot of the earthquake catalog, where the symbol size reflects the magnitude. The blue line represents the cumulative number of earthquakes over time.

Method	K	α	p	c	d	γ	q
SPDE	0.019	1.30	1.18	0.015	0.10	0.23	1.50
Credible interval	[0.018, 0.021]	[1.26, 1.36]	[1.17, 1.20]	[0.012, 0.016]	[0.08, 0.11]	[0.22, 0.25]	[1.48, 1.53]
KDE (5)	0.018	1.27	1.17	0.015	0.10	0.25	1.50
KDE (10)	0.018	1.28	1.16	0.014	0.11	0.24	1.50
KDE (15)	0.019	1.27	1.17	0.014	0.11	0.25	1.51

Table 1 Estimated triggering parameters for SPDE–ETAS with credible intervals and KDE-ETAS with different nearest-neighbor values.

sess the performance of the SPDE–ETAS model. The results show a clear improvement over the classical KDE–ETAS approach and a substantial reduction in computational cost compared to GP–ETAS, while achieving estimation accuracy that is comparable and better for background rate inference using both LGCP and SPDE formulations. In particular, for large-scale spatial variations, SPDE–ETAS performs similarly to GP–ETAS, as

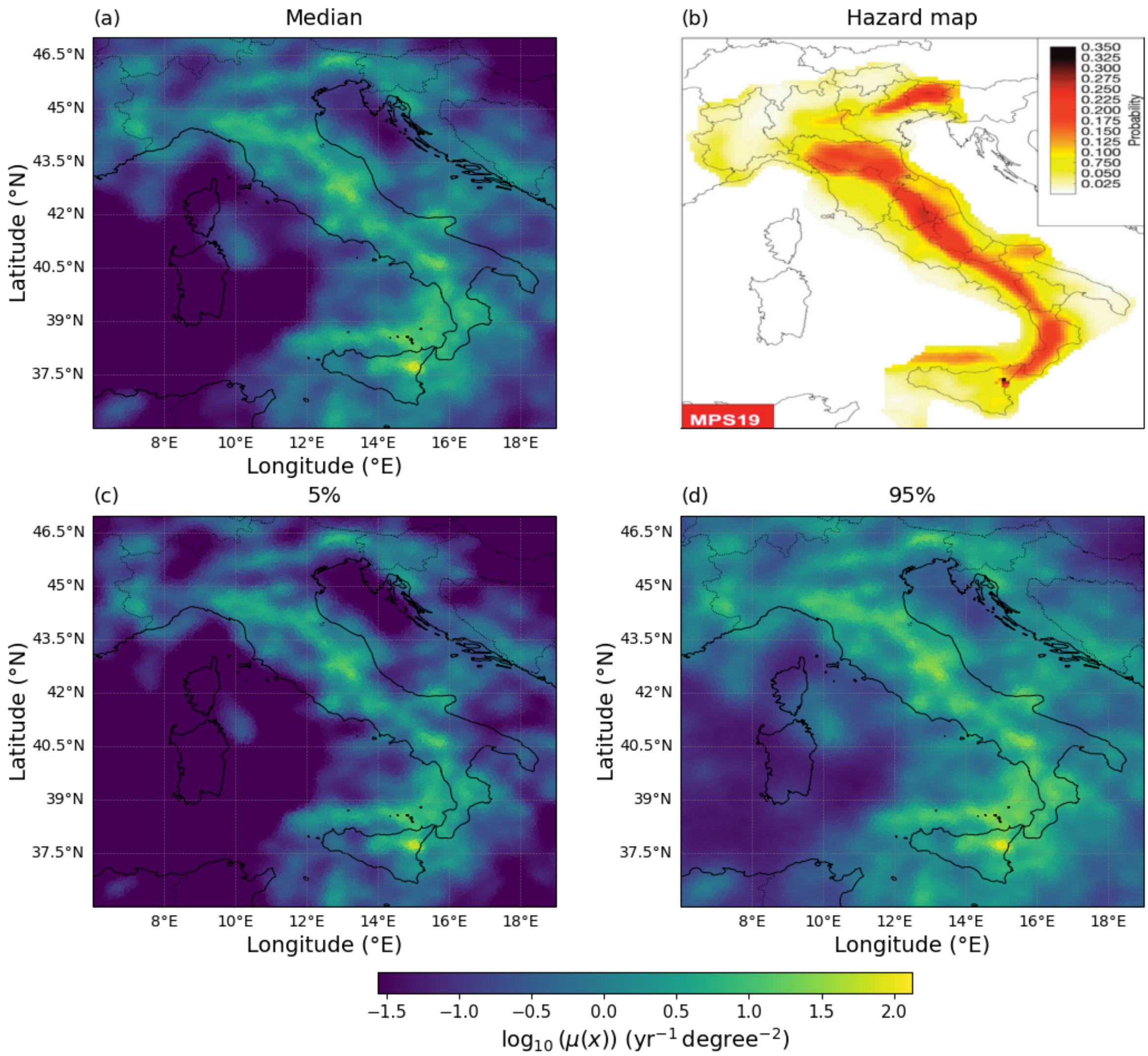

Figure 7 (a) Median background rate (log scale) estimated with SPDE–ETAS, expressed per year and degree^2. (b) The Italian seismic hazard map MPS19 (Meletti et al., 2019). (c) 5% quantile of the background rate estimation. (d) 95% quantile of the background rate estimation.

demonstrated in the first synthetic test. However, for small-scale spatial changes, such as the striped structures in the second synthetic test and as observed in the third one, SPDE–ETAS provides superior performance. This improvement is primarily due to the adaptive triangular mesh, which allows local refinement and better captures fine-scale heterogeneities. In contrast, GP–ETAS relies on a regular mesh, which limits its ability to resolve such localized variations unless a much finer global resolution is adopted, which is infeasible due to computational costs. In addition, the proposed approach allows for a robust quantification of uncertainties in both the triggering parameters and the background rate. A further advantage of the SPDE formulation lies in the intuitive interpretation of its parameters, which are directly linked to physically meaningful quantities such as correlation range and marginal variance.

The application to a complete Italian earthquake catalog spanning a large spatial domain and a long time period highlights the practical utility of the SPDE–ETAS model. The approach is computationally efficient and scalable, while preserving good estimation accuracy, and is therefore well suited for large seismic databases.

Several natural extensions of the present framework can be envisaged. A spatio-temporal background rate can be incorporated in a straightforward manner within the LGCP framework (Soriano-Redondo et al., 2019; D'Angelo et al., 2022; Taylor et al., 2013), by combining two GMRFs: a spatial component obtained by solving the SPDE in two dimensions using the finite element method, as done in this study, and a temporal component obtained by solving a one-dimensional SPDE using finite difference methods (Muir and Ross, 2023). Such an extension would be particularly relevant for the characterization of seismic swarms, where aseismic transients often induce pronounced time-dependent variations in the background rate and may influence aftershock triggering parameters (Marsan et al., 2013; Hainzl et al., 2013). More recently, alternative nonparametric approaches, such as the estimation of inhomogeneous spatio-temporal background intensity functions using graphical Dirichlet processes, have also been proposed and could be incorporated within the ETAS

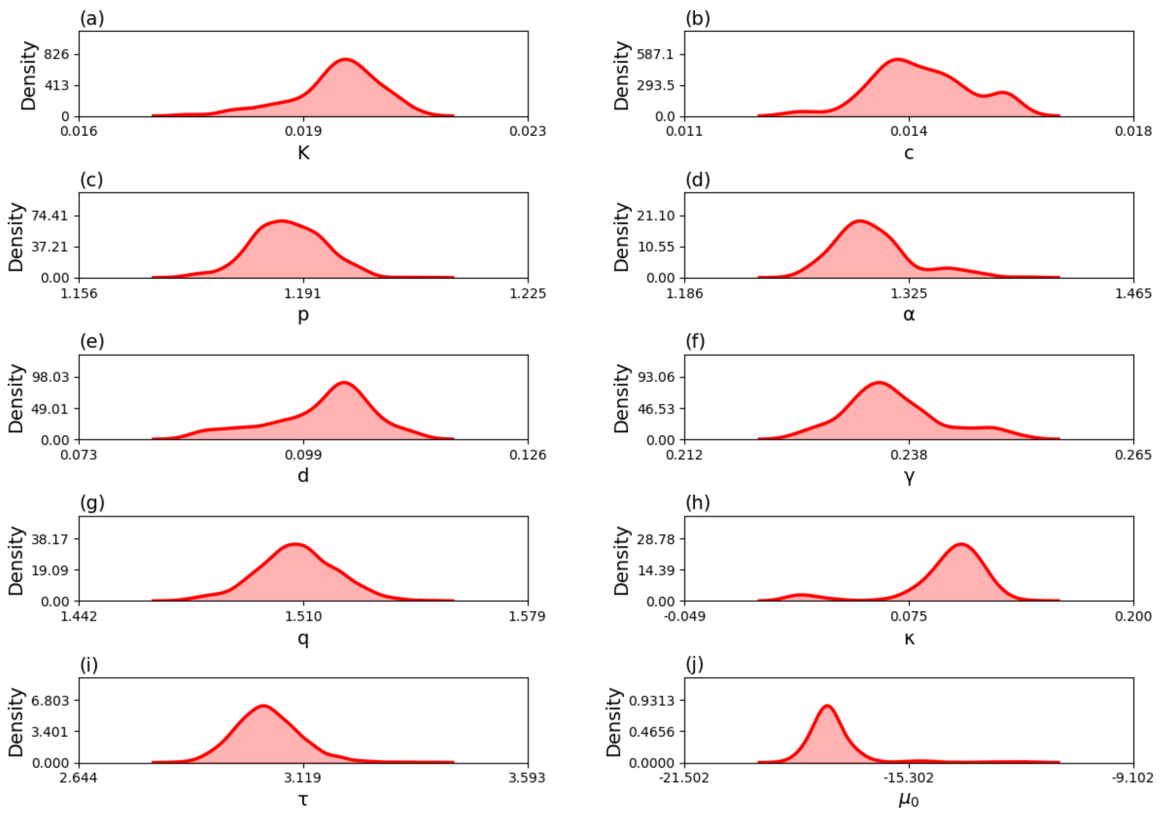

Figure 8 Parameter estimations for Italy: (a)–(g) Posterior distributions of the triggering parameters. (h)–(j) Posterior distributions of the LGCP parameters.

framework as flexible priors for the background component (Bañales et al., 2026).

Additional improvements could include the explicit modeling of short-term catalog incompleteness, which is known to bias parameter estimation due to the missing detection of small events following large earthquakes (Hainzl, 2021), as well as the incorporation of anisotropic aftershock kernels (Asayesh et al., 2025). The latter is particularly relevant for large fault-related events, where aftershock distributions often exhibit strong directional dependence (Grimm et al., 2021, 2022).

In summary, the SPDE–ETAS model provides a flexible, computationally efficient, and physically interpretable framework for modeling spatio-temporal seismicity patterns within a Bayesian setting, while enabling a rigorous quantification of uncertainties.

Acknowledgements

We thank two anonymous reviewers for their constructive and insightful comments, which greatly helped to improve the manuscript. This study has been funded by the Deutsche Forschungsgemeinschaft (DFG)- Project-ID 318763901 - SFB1294, project B04.

Data and code availability

The data used in the synthetic experiment were generated by the authors. The Italian earthquake catalog was downloaded from https://horus.bo.ingv.it. Figure 7b is reproduced from (Meletti et al., 2019) under the CC BY 4.0 license. The code for computing the SPDE–ETAS model is available at https://github.com/sofiane-T-rahmani/SPDE-ETAS.

Competing interests

The authors declare that they have no competing interests.

Appendix

A Finite Element Method (FEM)

We discretize the domain $\Omega \subset \mathbb{R}^2$ using a triangular mesh with nodes at the observation points and additional points added to ensure adequate mesh quality. On this mesh, the SPDE (eq. 12) is approximated via the finite element method. Using piecewise linear basis functions $\{\phi_i\}$, we define the mass and stiffness matrices

$$C_{ij} = \int_\Omega \phi_i \phi_j \, dx, \quad G_{ij} = \int_\Omega \nabla\phi_i \cdot \nabla\phi_j \, dx. \tag{18}$$

For different α, the precision matrix of the GMRF representation is computed recursively:

$$Q_{1,\kappa^2} = K_{\kappa^2}, \quad Q_{2,\kappa^2} = K_{\kappa^2} C^{-1} K_{\kappa^2}, \quad Q_{\alpha,\kappa^2} = K_{\kappa^2} C^{-1} Q_{\alpha-2,\kappa^2} C^{-1} K_{\kappa^2}, \quad \alpha = 3, 4, \ldots \tag{19}$$

with

$$K_{\kappa^2} = G + \kappa^2 C \tag{20}$$

This formulation naturally accounts for different smoothness levels through α while maintaining the finite element structure. Although, simple and efficient, this version is only applicable if α is integer constraining the analysis for integer smoothing values only.

B Bayesian sampling

We consider an ETAS model with a latent branching structure B, a spatially varying background rate $\mu(\boldsymbol{x})$ modeled as a LGCP (eq. 8), through the GMRF latent field z, and parameters $\theta_\mu = (\mu_0, \kappa, \tau)$, and triggering parameters $\theta_\varphi = (K, \alpha, c, p, q, d, \gamma)$.

Our MCMC scheme consists of sequentially sampling the parameters in three blocks: the branching structure B, the background rate $\mu(\boldsymbol{x})$, via sampling the GMRF latent field z and the hyperparameters (μ_0, κ and τ), and the triggering parameters θ_φ, which are only weakly dependent conditional on the other blocks. Let $\theta_k = [B_k, z_k, \theta_{\mu,k}, \theta_{\varphi,k}]$ denote the current state. The MCMC proceeds by iteratively generating updates via Metropolis–Hastings (MH) sampling, with transition kernels targeting the joint posterior

$$\pi(B, z, \theta_\mu, \theta_\phi \mid Y), \tag{21}$$

such that this posterior is the stationary distribution of the Markov chain, where π denotes the posterior distribution.

The Markov chain is constructed such that its stationary distribution is the posterior distribution. In practice, posterior samples are obtained after a burn-in period and are used to approximate the target distribution. The MCMC algorithm was run for 1100 iterations, including a burn-in period of 100 iterations. To reduce autocorrelation in the Markov chain, thinning was applied by retaining every 20th sample for both the SPDE latent field and the triggering parameters. The resulting samples were used for posterior inference.

B.1 Sample branching structure

For each event i, we introduce a latent variable $B_i \in \{0, 1, \ldots, i-1\}$, indicating whether the event is a parent event ($B_i = 0$) or triggered by a previous event $j > 0$. Conditional on $[z_k, \theta_{\mu,k}]$ and the triggering parameters $\theta_{\varphi,k}$, the exact conditional posterior of B_i is

$$P(B_i = j \mid Y, z_k, \theta_{\mu,k}, \theta_{\varphi,k}) = \frac{w_{ij}}{\lambda_i}, \tag{22}$$

where

$$w_{ij} = \begin{cases} \mu(x_i, y_i) = \exp\{\mu_0 + z_k(\boldsymbol{x}_i|\kappa,\tau)\}, & j = 0, \\ \kappa(m_j|k_k, \alpha_k) \cdot g(t_i - t_j|c_k, p_k) \cdot s(\boldsymbol{x}_i - \boldsymbol{x}_j|d_k, q_k, \gamma_k), & j > 0, \end{cases} \tag{23}$$

and $\lambda_i = \sum_{l=0}^{i-1} w_{il}$. Each B_i can therefore be sampled independently from this discrete distribution, yielding a full posterior over all possible declustering realizations of the catalog.

B.2 Sample background rate

Conditionally on the branching structure B_{k+1}, the parent events

$$S_0 = \{i : B_i = 0\} \tag{24}$$

We consider an inhomogeneous Poisson process in space, with a background intensity modeled within an LGCP framework(eq. 8). The GMRF $z(\boldsymbol{x})$ is approximated using the SPDE representation (eq. 12) over a triangular mesh, with piecewise linear basis functions ϕ_j associated with each vertex of the mesh (Fig. 9). These basis functions are equal to 1 at vertex j, 0 at all other vertices, and linearly interpolated within each triangle. Their values at any point correspond to the barycentric coordinates within the triangle containing that point.

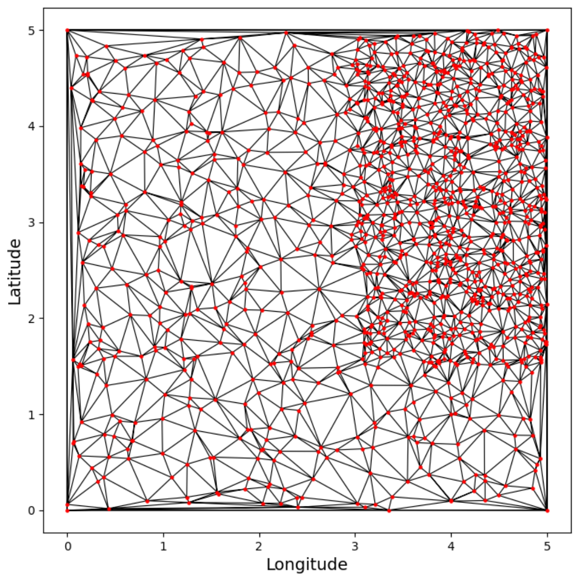

Figure 9 Illustration of a triangular mesh applied to a synthetic catalog

For an observed parent event at location $\boldsymbol{x}$, the background rate is evaluated by interpolation as

$$\mu(\boldsymbol{x}) = \sum_{j=1}^{m} \mu_j \phi_j(\boldsymbol{x}_i). \tag{25}$$

Let A denote the projection matrix with entries $A_{ij} = \phi_j(\boldsymbol{x}_i)$, so that $\mu(\boldsymbol{x}_i) = (A\mu(\boldsymbol{x}))_i$, where $\boldsymbol{\mu}$ denotes the discretized intensity at the mesh vertices.

The conditional log-likelihood of the background process is then

$$\ell_{\text{bg}}(z) = \sum_{i \in S_0} (\log(A\mu(\boldsymbol{x}))_i) - \int_{\boldsymbol{x}} \int_T \exp\{\mu_0 + z(\boldsymbol{x})\}\, d\boldsymbol{x}dt. \tag{26}$$

The spatial integral term is approximated numerically using a quadrature rule defined on the triangular mesh. Let $\{\boldsymbol{x}_l\}_{l=1}^{p}$ denote the quadrature points (e.g. triangle barycentres), and let α_l be the associated area weights, corresponding to the areas of the mesh triangles. The integral is approximated as

$$\int_{\boldsymbol{x}} \exp\{\mu_0 + z(\boldsymbol{x})\}\, d\boldsymbol{x} \;\approx\; \sum_{l=1}^{p} \alpha_l \exp\{\mu_0 + z(\boldsymbol{x}_l)\}. \tag{27}$$

Let $\tilde{A}$ denote the quadrature projection matrix with entries $\tilde{A}_{lj} = \phi_j(\boldsymbol{x}_l)$, which projects the background rate intensity onto the quadrature points $\boldsymbol{x}_l$ used for numerical integration. The log-likelihood can therefore be written as

$$\ell_{\text{bg}} = \sum_{i \in S_0} (\log(A\mu(\boldsymbol{x}))_i) - T \sum_{l=1}^{p} \alpha_l \exp\{(\tilde{A}(\mu_0 + z))_l\}, \tag{28}$$

where T is the studied time period.

The latent field z is assigned a GMRF prior

$$z \sim \mathcal{N}\left(0, Q^{-1}(\kappa, \tau)\right), \tag{29}$$

with log-density

$$\log \pi(z \mid \kappa, \tau) = -\frac{1}{2} z^\top Q(\kappa, \tau) z + \frac{1}{2} \log|Q(\kappa, \tau)| + \frac{N}{2}(2\pi), \tag{30}$$

where the precision matrix $Q(\kappa, \tau)$ is derived from the finite element discretization of the SPDE representation of the Matérn field (see Appendix A), $\log|Q(\kappa, \tau)|$ being the log determinant of the lower triangle of the Cholesky decomposition, and N the number of vertices. Combining likelihood and prior, the conditional log-posterior is

$$\begin{aligned} \log \pi(z \mid B_{k+1}, \theta_{\mu,k}, Y) = \\ \sum_{i \in S_0} (A\mu(x,y))_i - \\ T \sum_{l=1}^{p} \alpha_l \exp\{(\tilde{A}(\mu_0 + z))_l\} - \\ \frac{1}{2} z^\top Q z + \frac{1}{2} \log|Q(\kappa, \tau)| + \frac{N}{2}(2\pi) \end{aligned} \tag{31}$$

Since this posterior distribution is non-Gaussian, we apply a Laplace approximation around its mode

$$\hat{z} = \arg\max_{z} \log \pi(z \mid B_{k+1}, \theta_{\mu,k}, Y), \tag{32}$$

leading to the Gaussian approximation

$$\pi(z \mid B_{k+1}, \theta_{\mu,k}, Y) \;\approx\; \mathcal{N}\left(\hat{z}, H^{-1}\right), \tag{33}$$

where the negative Hessian at the mode is given by

$$H = Q + \tilde{A}^\top \operatorname{diag}\left(T\alpha_l \exp\{(\tilde{A}(\mu_0 + \hat{z}))_l\}\right) \tilde{A}. \tag{34}$$

At iteration k, the SPDE parameters are sampled independently from uniform prior distributions,

$$\begin{aligned} \kappa_k &\sim \mathcal{U}(\kappa_{\min}, \kappa_{\max}), \\ \tau_k &\sim \mathcal{U}(\tau_{\min}, \tau_{\max}), \end{aligned} \tag{35}$$

Conditionally on (κ_k, τ_k) and the observed data, the latent field z_k is updated using a Laplace approximation of its conditional posterior distribution, as described above.

B.3 Sample triggering parameters

Conditional on B_{k+1}, the triggering parameters $(K, \alpha, c, p, d, q, \gamma)$ can be sampled from their respective conditional posteriors. Let $S_j = \{i : B_i = j\}$ denote the offspring of event j.

Productivity parameters (K, α)**:**

$$\ell_{K,\alpha} = \sum_{j=1}^{N} \Big(|S_j| \log \kappa(m_j|K, \alpha) - \kappa(m_j|K, \alpha) G(T - t_j|c, p) \Big), \tag{36}$$

where G is the integral over time of the Omori-Utsu function.

Temporal kernel parameters (c, p)**:**

$$\begin{aligned} \ell_{c,p} = \sum_{j=1}^{N} \Big(\sum_{i \in S_j} \log g(t_i - t_j|c, p) \\ - \kappa(m_j|K, \alpha) G(T - t_j|c, p) \Big). \end{aligned} \tag{37}$$

Spatial kernel parameters d, q, γ**:** Define $(x_i', y_i') = (x_i - x_{B_i}, y_i - y_{B_i})$ for each offspring i. Then the conditional log-likelihood is

$$\ell_{d,q\gamma} = \sum_{i:B_i>0} \log s(\boldsymbol{x}_i'|d, q, \gamma). \tag{38}$$

C GP–ETAS

In GP–ETAS (Molkenthin et al., 2022), the background intensity $\mu(\boldsymbol{x})$ is modeled nonparametrically using a Gaussian process (GP) prior:

$$\mu(\boldsymbol{x}) = \bar{\lambda}\, \sigma(f(\boldsymbol{x})), \tag{39}$$

where $\sigma(z) = 1/(1+e^{-z})$ is the logistic sigmoid ensuring $\mu(\boldsymbol{x}) \in [0, \bar{\lambda}]$, and $f(\boldsymbol{x}) \sim \mathcal{GP}(m(\boldsymbol{x}), k(\boldsymbol{x}, \boldsymbol{x}' \mid \nu))$ is a latent GP capturing spatial correlations.

The GP defines a joint Gaussian distribution over any finite set of locations $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_K\}$:

$$[f(\boldsymbol{x}_1), \ldots, f(\boldsymbol{x}_K)]^\top \sim \mathcal{N}([m(\boldsymbol{x}_1), \ldots, m(\boldsymbol{x}_K)]^\top, \mathbf{K}), \tag{40}$$

with a mean vector $m(\boldsymbol{x})$, and covariance matrix $\mathbf{K}$ given by $k(\boldsymbol{x}_i, \boldsymbol{x}_j \mid \nu)$.

This formulation allows $\mu(\boldsymbol{x})$ to be learned directly from the data in a Bayesian way without specifying a functional form a priori, while the GP prior combined with data augmentation leads to conjugacy and simplifies posterior inference (Rasmussen, 2013). The Bayesian scheme used in this paper for the GP–ETAS is identical to that introduced in Molkenthin et al., 2022, with the same number of iterations as for SPDE–ETAS, to ensure fair comparison.

D KDE–ETAS

The background intensity $\mu(\boldsymbol{x})$ is estimated using a kernel density estimator with variable bandwidth (Zhuang et al., 2002), according to:

$$\mu_{\text{kde}}(\boldsymbol{x}) = \frac{1}{T} \sum_{i=1}^{N_D} p_i^0 \, k_{d_i}(\boldsymbol{x} - \boldsymbol{x}_i), \tag{41}$$

where T is the observation period, p_i^0 is the background probability of event i, and $k_{d_i}(\cdot)$ is a Gaussian kernel with bandwidth

$$d_i = \max\{d_{\min}, r_{i,n_p}\}, \tag{42}$$

with r_{i,n_p} the distance to the n_p nearest neighbors. The value of n_p must be predefined and is not a result of the optimization. Typical choices are $d_{\min} \in [0.02, 0.05]$ degrees and $n_p \in [10, 100]$ (Zhuang et al., 2002).

References

Asayesh, B. M., Hainzl, S., and Zöller, G. Depth-Dependent Aftershock Trigger Potential Revealed by 3D-ETAS Modeling. *Journal of Geophysical Research: Solid Earth*, 128(6), 2023. doi: 10.1029/2023jb026377.

Asayesh, B. M., Hainzl, S., and Zöller, G. Improved Aftershock Forecasts Using Mainshock Information in the Framework of the ETAS Model. *Journal of Geophysical Research: Solid Earth*, 130 (2), Feb. 2025. doi: 10.1029/2024jb030287.

Banerjee, S., Carlin, B. P., and Gelfand, A. E. *Hierarchical modeling and analysis for spatial data*. Chapman and Hall/CRC, 2003. doi: 10.1201/9781003401728.

Bañales, I., Nishikawa, T., Ito, Y., and Aguilar-Velázquez, M. J. Estimating Inhomogeneous Spatiotemporal Background Intensity Functions Using Graphical Dirichlet Processes. *Mathematical Geosciences*, May 2026. doi: 10.1007/s11004-026-10301-0.

Chiles, J.-P. and Delfiner, P. *Geostatistics: modeling spatial uncertainty*. John Wiley & Sons, 2012. doi: 10.1002/9781118136188.

Console, R. and Murru, M. A simple and testable model for earthquake clustering. *Journal of Geophysical Research: Solid Earth*, 106(B5):8699–8711, May 2001. doi: 10.1029/2000jb900269.

Cox, D. R. Some Statistical Methods Connected with Series of Events. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 17(2), 1955. doi: 10.1111/j.2517-6161.1955.tb00188.x.

Cressie, N. and Johannesson, G. Fixed Rank Kriging for Very Large Spatial Data Sets. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 70(1):209–226, Jan. 2008. doi: 10.1111/j.1467-9868.2007.00633.x.

Dutilleul, P., Genest, C., and Peng, R. Bootstrapping for parameter uncertainty in the space–time epidemic-type aftershock sequence model. *Geophysical Journal International*, 236(3): 1601–1608, Jan. 2024. doi: 10.1093/gji/ggae003.

D'Angelo, N., Adelfio, G., Abbruzzo, A., and Mateu, J. Inhomogeneous spatio-temporal point processes on linear networks for visitors' stops data. *The Annals of Applied Statistics*, 16(2), 2022. doi: 10.1214/21-aoas1519.

Ebrahimian, H., Jalayer, F., Maleki Asayesh, B., Hainzl, S., and Zafarani, H. Improvements to seismicity forecasting based on a Bayesian spatio-temporal ETAS model. *Scientific Reports*, 12(1), Dec. 2022. doi: 10.1038/s41598-022-24080-1.

Edelsbrunner, H. *Geometry and Topology for Mesh Generation*. Cambridge University Press, May 2001. doi: 10.1017/cbo9780511530067.

Grimm, C., Käser, M., Hainzl, S., Pagani, M., and Küchenhoff, H. Improving Earthquake Doublet Frequency Predictions by Modified Spatial Trigger Kernels in the Epidemic-Type Aftershock Sequence (ETAS) Model. *Bulletin of the Seismological Society of America*, 112(1), 2021. doi: 10.1785/0120210097.

Grimm, C., Hainzl, S., Käser, M., and Küchenhoff, H. Solving three major biases of the ETAS model to improve forecasts of the 2019 Ridgecrest sequence. *Stochastic Environmental Research and Risk Assessment*, 36(8):2133–2152, Apr. 2022. doi: 10.1007/s00477-022-02221-2.

Guttorp, P. and Gneiting, T. Studies in the history of probability and statistics XLIX On the Matérn correlation family. *Biometrika*, 93 (4):989–995, Dec. 2006. doi: 10.1093/biomet/93.4.989.

Hainzl, S. ETAS-Approach Accounting for Short-Term Incompleteness of Earthquake Catalogs. *Bulletin of the Seismological Society of America*, 112(1), 2021. doi: 10.1785/0120210146.

Hainzl, S. and Ogata, Y. Detecting fluid signals in seismicity data through statistical earthquake modeling. *Journal of Geophysical Research: Solid Earth*, 110(B5), Mar. 2005. doi: 10.1029/2004jb003247.

Hainzl, S., Zakharova, O., and Marsan, D. Impact of Aseismic Transients on the Estimation of Aftershock Productivity Parameters. *Bulletin of the Seismological Society of America*, 103(3), 2013. doi: 10.1785/0120120247.

Hainzl, S., Kumazawa, T., and Ogata, Y. Aftershock forecasts based on incomplete earthquake catalogues: ETASI model application to the 2023 SE Türkiye earthquake sequence. *Geophysical Journal International*, 236(3):1609–1620, Jan. 2024. doi: 10.1093/gji/ggae006.

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

Hawkes, A. G. and Oakes, D. A cluster process representation of a self-exciting process. *Journal of Applied Probability*, 11(3), 1974. doi: 10.2307/3212693.

Helmstetter, A. and Sornette, D. Subcritical and supercritical regimes in epidemic models of earthquake aftershocks. *Journal of Geophysical Research: Solid Earth*, 107(B10):ESE 10–1–ESE 10–21, 2002. doi: 10.1029/2001JB001580.

Henderson, R., Shimakura, S., and Gorst, D. Modeling spatial variation in leukemia survival data. *Journal of the American Statistical Association*, 97(460):965–972, 2002. doi: 10.1198/016214502388618753.

Hjelle, Ø. and Dæhlen, M. *Triangulations and applications*. Springer Science & Business Media, 2006. doi: 10.1007/3-540-33261-8.

Kagan, Y. Y. Aftershock Zone Scaling. *Bulletin of the Seismological Society of America*, 92(2):641–655, Mar. 2002. doi: 10.1785/0120010172.

Lindgren, F., Rue, H., and Lindström, J. An Explicit Link between Gaussian Fields and Gaussian Markov Random Fields: The Stochastic Partial Differential Equation Approach. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 73(4): 423–498, Aug. 2011. doi: 10.1111/j.1467-9868.2011.00777.x.

Lolli, B., Randazzo, D., Vannucci, G., and Gasperini, P. The Homogenized Instrumental Seismic Catalog (HORUS) of Italy from 1960 to Present. *Seismological Research Letters*, 91(6), 2020. doi: 10.1785/0220200148.

Lombardi, A. M. Estimation of the parameters of ETAS models by Simulated Annealing. *Scientific Reports*, 5(1), Feb. 2015. doi: 10.1038/srep08417.

Lombardi, A. M., Cocco, M., and Marzocchi, W. On the Increase of Background Seismicity Rate during the 1997-1998 Umbria-Marche, Central Italy, Sequence: Apparent Variation or Fluid-Driven Triggering? *Bulletin of the Seismological Society of America*, 100(3):1138–1152, May 2010. doi: 10.1785/0120090077.

Marsan, D., Prono, E., and Helmstetter, A. Monitoring Aseismic Forcing in Fault Zones Using Earthquake Time Series. *Bulletin of the Seismological Society of America*, 103(1):169–179, Feb. 2013. doi: 10.1785/0120110304.

Meletti, C., Marzocchi, W., D'Amico, V., Lanzano, G., Luzi, L., Martinelli, F., Pace, B., Rovida, A., Taroni, M., Visini, F., and Group, M. W. The new Italian seismic hazard model (MPS19). *Annals of Geophysics*, 2019. doi: 10.4401/ag-8579.

Molkenthin, C., Donner, C., Reich, S., Zöller, G., Hainzl, S., Holschneider, M., and Opper, M. GP-ETAS: semiparametric Bayesian inference for the spatio-temporal epidemic type aftershock sequence model. *Statistics and Computing*, 32(2), Mar. 2022. doi: 10.1007/s11222-022-10085-3.

Molkenthin, C., Zöller, G., Hainzl, S., and Holschneider, M. Bayesian Earthquake Forecasting Using Gaussian Process Modeling: GP-ETAS Applications. *Seismological Research Letters*, 95(6): 3532–3544, Oct. 2024. doi: 10.1785/0220240170.

Muir, J. B. and Ross, Z. E. A deep Gaussian process model for seismicity background rates. *Geophysical Journal International*, 234 (1):427–438, Feb. 2023. doi: 10.1093/gji/ggad074.

Ogata, Y. Space-Time Point-Process Models for Earthquake Occurrences. *Annals of the Institute of Statistical Mathematics*, 50(2), 1998. doi: 10.1023/a:1003403601725.

Ogata, Y. and Zhuang, J. Space–time ETAS models and an improved extension. *Tectonophysics*, 413(1-2):13–23, Feb. 2006. doi: 10.1016/j.tecto.2005.10.016.

Ogata, Y., Katsura, K., Tsuruoka, H., and Hirata, N. High-resolution 3D earthquake forecasting beneath the greater Tokyo area. *Earth, Planets and Space*, 71(1), Nov. 2019. doi: 10.1186/s40623-019-1086-7.

Omori, F. On the aftershocks of earthquakes. *J. Coll. Sci*, 7(2): 111–120, 1894.

Rasmussen, J. G. Bayesian Inference for Hawkes Processes. *Methodology and Computing in Applied Probability*, 15(3): 623–642, Sept. 2013. doi: 10.1007/s11009-011-9272-5.

Ross, G. J. Bayesian Estimation of the ETAS Model for Earthquake Occurrences. *Bulletin of the Seismological Society of America*, 111(3):1473–1480, May 2021. doi: 10.1785/0120200198.

Ross, G. J. and Kolev, A. A. Semiparametric Bayesian forecasting of SpatioTemporal earthquake occurrences. *The Annals of Applied Statistics*, 16(4), Dec. 2022. doi: 10.1214/21-aoas1554.

Simpson, D., Illian, J. B., Lindgren, F., Sørbye, S. H., and Rue, H. Going off grid: computationally efficient inference for log-Gaussian Cox processes. *Biometrika*, 103(1):49–70, Feb. 2016. doi: 10.1093/biomet/asv064.

Soriano-Redondo, A., Jones-Todd, C. M., Bearhop, S., Hilton, G. M., Lock, L., Stanbury, A., Votier, S. C., and Illian, J. B. Understanding species distribution in dynamic populations: a new approach using spatio-temporal point process models. *Ecography*, 42(6): 1092–1102, Mar. 2019. doi: 10.1111/ecog.03771.

Stein, M. L. *Interpolation of Spatial Data: Some Theory for Kriging*. Springer Science & Business Media, 1999. doi: 10.1007/978-1-4612-1494-6.

Taylor, B. M., Davies, T. M., Rowlingson, B. S., and Diggle, P. J. lgcp An R Package for Inference with Spatio-Temporal Log-Gaussian Cox Processes. *Journal of Statistical Software*, 52(4), 2013. doi: 10.18637/jss.v052.i04.

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

Utsu, T. Aftershocks and earthquake statistics (1): Some parameters which characterize an aftershock sequence and their interrelations. *Journal of the Faculty of Science, Hokkaido University. Series 7, Geophysics*, 3(3):129–195, 1970.

Veen, A. and Schoenberg, F. P. Estimation of Space–Time Branching Process Models in Seismology Using an EM–Type Algorithm. *Journal of the American Statistical Association*, 103(482), 2008. doi: 10.1198/016214508000000148.

Werner, M. J., Helmstetter, A., Jackson, D. D., and Kagan, Y. Y. High-Resolution Long-Term and Short-Term Earthquake Forecasts for California. *Bulletin of the Seismological Society of America*, 101 (4):1630–1648, Aug. 2011. doi: 10.1785/0120090340.

Zhuang, J., Ogata, Y., and Vere-Jones, D. Stochastic Declustering of Space-Time Earthquake Occurrences. *Journal of the American Statistical Association*, 97(458), 2002. doi: 10.1198/016214502760046925.

Zhuang, J., Ogata, Y., and Vere-Jones, D. Analyzing earthquake clustering features by using stochastic reconstruction. *Journal of Geophysical Research: Solid Earth*, 109(B5), May 2004. doi: 10.1029/2003jb002879.

Zhuang, J., Christophersen, A., Savage, M. K., Vere-Jones, D., Ogata, Y., and Jackson, D. D. Differences between spontaneous and triggered earthquakes: Their influences on foreshock probabilities. *Journal of Geophysical Research: Solid Earth*, 113(B11), Nov. 2008. doi: 10.1029/2008jb005579.