We thank the reviewers for their valuable comments and suggestions. All comments have been carefully considered and the manuscript has been revised accordingly. Concerning each point, we refer to the place where the manuscript has been revised. The corresponding revisions are colored in the annotated manuscript. Below, repetitions of the reviewer's comments are shown in black, while our comments and replies to the corresponding points are marked in blue. The page and line numbers refer to the manuscript with marked changes.

Reviewer 1

The manuscript develops a Bayesian ETAS framework in which the background rate is modeled via a latent Gaussian field using a log-Gaussian Cox process prior. Compared with existing approaches—such as logistic-Gaussian Cox process ETAS (e.g., Molkenthin et al.) and hierarchical or deep Gaussian process ETAS (e.g., Miur & Ross)—the present work adopts an SPDE–GMRF representation to approximate the Gaussian process prior and employs Laplace approximation to improve computational efficiency.

In my view, the proposed framework falls within the class of latent Gaussian ETAS models, and the main contribution lies in computational implementation—specifically, the SPDE–GMRF approximation and Laplace-based inference—rather than in a fundamentally new statistical model. The paper is potentially publishable, but several clarifications and additional analyses are needed to properly position the contribution and convincingly demonstrate its advantages.

We are very grateful to the first reviewer for her/his careful and constructive review of our manuscript, which helped us to clarify important points and significantly improve the paper. In the following, we show how we have met her/his recommendations:

Major Comments

1. Clarification of Model Contribution
The manuscript should more clearly distinguish between modeling innovation and computational implementation. While the paper presents the SPDE–ETAS formulation as a new model, the underlying structure remains that of a latent Gaussian ETAS framework. The key difference lies in replacing the Gaussian process prior with its SPDE–GMRF representation, resulting in a sparse precision matrix and improved computational scalability. The authors are encouraged to revise the presentation accordingly, emphasizing that the contribution is primarily computational rather than methodological.

We agree that our original presentation could have overstated the methodological novelty of the approach. As correctly pointed out, the proposed SPDE–ETAS framework does not modify the underlying latent Gaussian ETAS model, but rather provides a computationally efficient representation of the Gaussian process prior using its SPDE–GMRF formulation. We have revised the manuscript accordingly to clarify that the main contribution is computational. In particular, we now emphasize that the SPDE approach provides an equivalent representation of Gaussian processes with Matérn covariance functions, leading to a sparse precision matrix and improved scalability for Bayesian inference. The abstract and introduction have been updated to reflect this distinction more clearly, with the phrase '..we introduce a computational formulation of the Bayesian ETAS model..', and in the introduction 'consistent with existing latent Gaussian ETAS frameworks...' and 'As a result, the proposed approach does not modify the underlying statistical model, but provides a computationally efficient approximation that enables scalable Bayesian inference.'. We have replaced statements suggesting the introduction of a new model with wording that highlights the computationally efficient reformulation of an existing framework.

2. Positioning Relative to Existing Work
The relationship to existing GP-based ETAS formulations should be clarified more explicitly. In particular: (a) Logistic-Gaussian Cox process ETAS (e.g., Molkenthin et al.); (b) GP-based or hierarchical (deep) GP ETAS approaches (e.g., Miur & Ross); (c) Log-Gaussian Cox process formulations of spatial point processes. The manuscript would benefit from explicitly stating that the proposed approach shares the same latent Gaussian field structure, with the main distinction being the SPDE–GMRF approximation and associated computational advantages.

We fully agree that the proposed SPDE–ETAS framework shares the same underlying statistical structure as previously developed approaches based on Gaussian processes and Log-Gaussian Cox processes. In particular, the background rate is modeled as a Log-Gaussian Cox process (LGCP), where the log-intensity is driven by a latent Gaussian random field. This formulation is consistent with earlier GP-based ETAS approaches, including logistic-Gaussian Cox process models (e.g., Molkenthin et al., 2022) as well as hierarchical or deep Gaussian process models (e.g., Muir and Ross, 2023). In all these approaches, the key modeling assumption is the use of a latent Gaussian field to describe the spatial variability of the background rate. The main contribution of the present work lies not in modifying this statistical framework, but in its computational representation. Specifically, we replace the standard Gaussian process prior, which involves a dense covariance matrix, with its equivalent SPDE-based representation, leading to a Gaussian Markov random field (GMRF) with a sparse precision matrix (Lindgren et al., 2011). This reformulation preserves the Matérn covariance structure while enabling substantially improved computational scalability. To address this point, we have revised the manuscript to explicitly state that the proposed SPDE–ETAS approach should be interpreted as a computationally efficient implementation of an existing latent Gaussian ETAS framework, rather than as a fundamentally new statistical model in the abstract and introduction like we pointed out in the first comment. Additionally in the definition of Log-Gaussian Cox Process we added '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).'

3. Role of KDE Baseline (Figure 2)
Kernel density estimation (KDE) provides a simple and computationally efficient baseline for background rate estimation. While less flexible than GP-based approaches, it is easy to implement and often performs reasonably well in practice. The authors should better discuss the trade-off between model flexibility and computational cost, and clarify under which conditions the proposed SPDE–ETAS approach offers clear advantages over simpler alternatives such as KDE.

This trade-off between computational simplicity and modeling capability is already discussed in the manuscript. In particular, the KDE–ETAS approach relies on maximum likelihood estimation, which yields only point estimates for the background rate and the triggering parameters. As a consequence, no uncertainty quantification is available for either component of the model. In contrast, the SPDE–ETAS framework adopts a fully Bayesian formulation, providing posterior distributions for both the background rate and the triggering parameters. This allows for the construction of credible intervals, which offer a more informative characterization of the model parameters and their associated uncertainties. Furthermore, the bandwidth, i.e., the choice of the nearest neighbor used for the KDE kernel, has to be predefined and is not a result of the optimization. Finally, as illustrated in Figures 2 and 4, the maximum likelihood estimates obtained from KDE–ETAS tend to deviate from the true underlying parameters.

We have now discussed this point in Synthetic test section, where we have added the following sentence to clarify this point: “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.” Furthermore, we added the sentence ”The value of n_p must be predefined and is not a result of the optimization” in Appendix D.

4. Evaluation of Practical Advantages (Figure 5)
Based on Figure 5, it is difficult to clearly distinguish the performance of GP-ETAS and SPDE-ETAS. The differences appear relatively small and may depend on implementation details. To convincingly demonstrate the advantage of the proposed method, the authors should provide a more comprehensive evaluation, including: (a) quantitative performance metrics (e.g., log-likelihood, predictive scores), (b) computational time comparisons, and, if possible, (c) theoretical or empirical complexity analysis. If predictive performance is comparable, then the primary justification for the method should be computational scalability rather than improved statistical accuracy.

We would like to clarify that the quantitative and computational comparisons are already included in the manuscript, but we agree that their interpretation can be made more explicit. First, we assess model fit using a normalized log-likelihood computed at each grid cell by comparing the posterior distributions obtained from both GP–ETAS and SPDE–ETAS with the corresponding true background values. This approach fully accounts for the posterior variability at each spatial location and provides a direct and robust comparison between the two methods (Figure 5a). While the overall performance of the two approaches is comparable, the results show a slight improvement in favor of SPDE–ETAS. Second, the computational time comparison is presented in Figure 5b, where it can be seen that SPDE–ETAS is significantly faster than GP–ETAS, with the difference increasing as the number of events grows. Finally, to further clarify the computational advantage, and following the reviewer’s suggestion, we have added an additional panel (Figure 5c) showing memory usage. This highlights the benefit of the SPDE–GMRF formulation, which relies on a sparse precision matrix and therefore requires substantially less memory than the dense covariance representation used in GP–ETAS. Overall, while both approaches provide comparable model fit, with a slight advantage for SPDE–ETAS, the results clearly show that the main strength of SPDE–ETAS lies in its computational scalability, both in terms of runtime and memory usage.

5. Inclusion of Memory Usage (Figure 5)
In addition to runtime comparisons, it would be highly informative to include memory usage as a performance metric. Since a key advantage of the SPDE–GMRF formulation lies in the sparse precision matrix representation, reduced memory consumption is expected to be a major benefit compared to GP-based approaches with dense covariance matrices. Including memory usage would provide a more complete and fair assessment of computational efficiency.

We thank you for your comment, we have now added panel (c) to Figure 5, comparing the memory usage of both models.

6. Need for Additional Baseline Comparisons (Section 4)

The empirical evaluation would benefit from a more comprehensive comparison with existing GP-ETAS (with full covariance structure), and KDE-based ETAS methods. Including these baselines would help clarify the practical gains of the SPDE–ETAS approach in terms of both accuracy and computational efficiency.

We agree that comparisons with GP-based ETAS models and KDE-based approaches are important to assess the practical performance of the proposed method. Regarding GP-based ETAS models with a full covariance structure, we would like to clarify that their application becomes computationally prohibitive for large datasets. In particular, the Gaussian process formulation requires the construction and inversion of a dense covariance matrix, leading to $\mathcal{O}(n^3)$ computational cost and $\mathcal{O}(n^2)$ memory requirements. In our experiments, we observed that for catalogs with several thousand events (around 10,000), the GP-based approach becomes infeasible due to memory limitations, preventing the completion of the inference procedure. This is now mentioned in the case study section, where we added "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". On the other hand, we have now also applied the KDE-ETAS with the different nearest neighbor settings: 5, 10, and 15. The corresponding results have been added to the text and provided in the new

7. Transparency of GP-ETAS Inference (Appendix C) The manuscript should include a clear description of the inference procedure for GP-ETAS, such as likelihood formulation, prior specification, computational method (e.g., MCMC or approximation). This is important for ensuring a fair and transparent comparison with the proposed SPDE–ETAS framework.

The GP–ETAS framework used in this study follows previously published formulations, for which a detailed description of the likelihood, prior specification, and inference procedure is already available in the literature (Molkenthin et al., 2022). In the present work, our objective is not to introduce a new GP-based model, but rather to use an existing GP–ETAS formulation as a benchmark for comparison with the proposed SPDE–ETAS approach. For this reason, we have not repeated the full methodological development in the manuscript. Instead we only provide necessary information on GP-ETAS, which are needed to follow our study and refer the reader to the original references for a complete description, i.e. to Molkenthin et al. (2022), where all mathematical and computational ingredients are described. For clarity, we emphasize that we follow the same Bayesian inference procedure as in the referenced GP–ETAS framework, including the likelihood formulation and prior structure. This ensures that the comparison between GP–ETAS and SPDE–ETAS is conducted on a consistent basis, with the primary difference being the representation of the latent Gaussian field (dense covariance versus sparse precision matrix). The phrase 'The Bayesian scheme used in this paper for the GP-ETAS is similar to what has been done in Molkenthin et al., 2022' has been added to appendix B.

Reviewer 2

This work presents an efficient way to model the background density function using Gaussian processes, based on the representation of Lindgren et al. (2011). Although it shows a clear improvement in computational cost compared to Molkenthin et al. (2022), I believe there are important aspects of the work that require improvement, which are described below.

We are very grateful to the second reviewer for her/his very useful comments and suggestions, which helped us to improve the paper. We have addressed the recommendations as follows:

Abstract

1. In line 7 of page one, it is stated: "An alternative representation of the Gaussian Process is the Gaussian Markov Random Field, which offers a sparse precision matrix formulation that significantly reduces computational complexity." This statement may be imprecise, since not all Gaussian processes admit this representation, as it requires a Matérn covariance function. I suggest being more explicit about this condition.

Thank you for this comment. We have the line according to your comment and changed to 'An alternative representation of Gaussian processes with Matérn covariance functions can be obtained through Gaussian Markov Random Fields (GMRFs) via the SPDE approach of Lindgren et al. (2011), which provides a sparse precision matrix formulation and significantly reduces computational complexity'.

Introduction

1. In the last line of page 20, it is stated: "which reproduces the spatiotemporal clustering of earthquakes." However, instead of reproduces, I believe it would be more appropriate to use models, since, as follows from the model presented in the article, the approach of Ogata (1998) still requires refinement.

We have changed the word from 'reproduces' to 'models'.

2. In line 30 of page 2, it is stated: "In most applications, the ETAS model is fitted using maximum likelihood estimation." I believe it is important to include references and examples that support this statement.

We have added references in line 30 on page 2 : e.g., Ogata, 1998; Zhuang et al., 2002; Helmstetter and Sornette, 2002; Zhuang et al., 2004; Hainzl et al., 2024.

3. Similarly to the previous comment regarding the statement "The likelihood function is often relatively flat," I suggest adding references that support this claim.

We have added the two references (Veen and Schoenberg, 2008; Ross, 2021) to the text.

4. Regarding the statement "Bayesian inference offers an alternative framework in which uncertainty about model parameters is represented explicitly through the posterior distribution," it may be somewhat inaccurate, since the uncertainty being referred to corresponds specifically to posterior uncertainty. The prior distribution also reflects uncertainty about the model parameters.

We have changed the sentence to this "Bayesian inference offers an alternative framework in which uncertainty about model parameters is explicitly represented through probability distributions, with prior distributions encoding initial uncertainty and the posterior distribution reflecting updated uncertainty after incorporating the observed data through the likelihood."

5. In line 50 of page 2, the preprint (Kolev and Ross, 2020) is cited. However, I believe it is important to revise and update the reference.

We have updated the reference accordingly

SPDE-ETAS

1. The elements of Equation 1 are not properly defined. In particular, there is no reference explaining what x_i represents. Additionally, θ is defined ambiguously, and its relationship with θ_μ and θ_ϕ is not made explicit.

Thank you for reporting this, the elements of Eq. (1) are now better explained.

2. In Equation 1, $\mu(x \mid \theta_\mu)$ is introduced as if it were defined in (Ogata, 1998); however, this is not the case. A more appropriate reference would be would be: Zhuang, J., Ogata, Y., & Vere-Jones, D. (2002). Stochastic declustering of space-time earthquake occurrences. Journal of the American Statistical Association, 97(458), 369–380.

The reference (Zhuang et al. 2002) has been added here.

3. Following the previous comment, I suggest modifying the summation subscript and writing it as in Zhuang et al., since, as currently written, x_i remains fixed.

The summation has been changed according to Zhuang et al. 2002.

4. Equation 9 does not formally define a Gaussian field. As written, it states that Z follows a multivariate normal distribution; however, its dimension is not specified. Since the intention is for the support of Z to be a subset of $\mathbb{R}^2$, a single multivariate normal distribution is not sufficient to define the process. Instead, it should be clarified that the definition refers to the finite-dimensional distributions of Z. I suggest reviewing Definition 3 in Lindgren et al. for a formal definition and rewriting the definition accordingly.

We have changed the introduction of the Gaussian random field. Now it is introduced as follows: Formally, a Gaussian random field $z(x)$, 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 $x_1, \ldots, x_n \in \mathcal{D}$, the random vector $(z(x_1), \ldots, z(x_n))^\top$ follows a multivariate normal distribution:

$$z(x_1), \ldots, z(x_n))^\top \sim \mathcal{N}(\mu, \Sigma)$$

5. Rewrite Section 2.3 by explicitly presenting the likelihood under the augmented model and clearly specifying the prior distributions. Additionally, Equation 15 is imprecise, since the parameter vector follows the posterior distribution only in the limit as $k \to \infty$. Finally, a discussion on chain thinning is missing and should be included.

We agree that a more detailed presentation of the likelihood and prior distributions, as well as the Bayesian sampling scheme, is important for completeness. These elements are fully described in Appendix B, where we provide a detailed formulation of the augmented likelihood, prior distributions, and the sampling procedure. In the main manuscript, we deliberately chose to keep the presentation concise in order to maintain readability and accessibility, particularly for readers who may not be familiar with Bayesian inference methods. Regarding Equation (15), we acknowledge that the formulation could be interpreted as imprecise. We have clarified in the revised manuscript and modified Equation 15 and also added equation 16. Finally, we agree that chain thinning is an important aspect of MCMC diagnostics. We have added a brief discussion on this point in the revised manuscript in Appendix B.

Synthetic experiments

1. On line 173, instead of the phrase "to invert the spatial variations of background activity," I suggest using "to estimate the spatial background activity."

We have replaced "To invert the spatial variations of background activity" with "to estimate the spatial background activity."

2. What is the rationale for the parameters in line 181?

Now, we clarify that we selected values within the typical range of observations.

3. I don't understand the caption of Figure 1: "Uncertainty of the SPDE–ETAS estimate, expressed as the standard deviation $(q_{95} - q_5)/2$." Why not report probability intervals directly? Why approximate the standard deviation this way if it is not known a posteriori whether this approximation is meaningful?

We changed the name from "standard deviation" to "semi-interquantile 0.05, 0.95 distance" in the captions, text, and figures. The semi-interquantile 0.05, 0.95 distance has already been used, e.g., by Molkenthin et al., 2022.

4. Please check the difference between confidence interval and probability interval (or credible interval)

Thank you. We have clarified the difference throughout the whole manscript.

5. I suggest splitting Figure 1 into 2 or 3 separate figures.

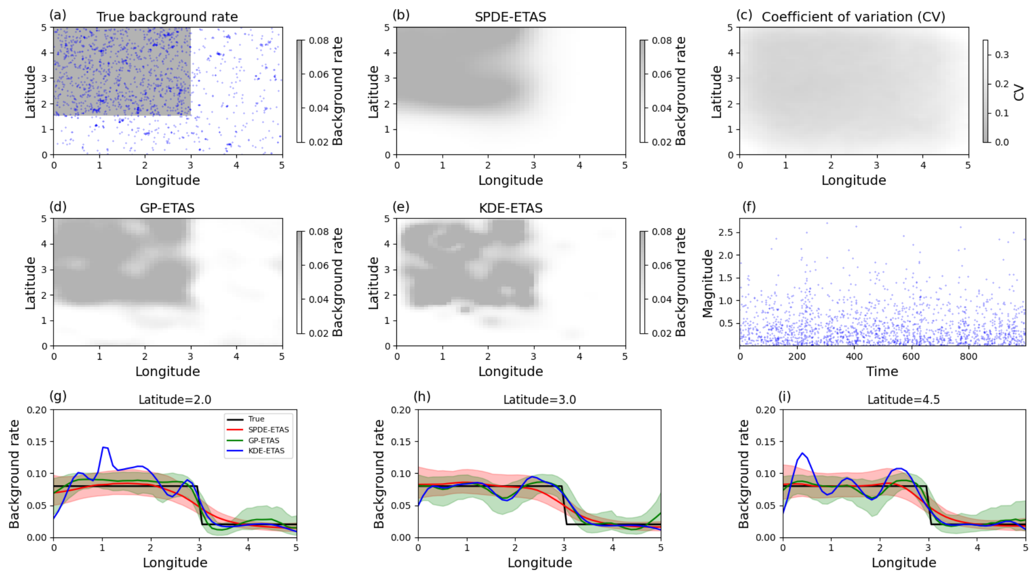
Unfortunately, the figures cannot be splitted since the journal Seismica accepts only a maximum of ten figures and we think that the figures are important enough to not be put in supplementary materials.

Also, adjust the vertical margins of subfigures g, h, and i.

We have adjusted the vertical margins.

Additionally, I recommend considering the coefficient of variation to allow a better comparison of the uncertainty in b.

We computed the coefficient of variation for this synthetic example. However, we still believe that for our purposes the interquantile range provides a more illustrative representation of the credible interval than the coefficient of variation. Indeed, for low background rates, the coefficient of variation can take artificially large values, which may misleadingly suggest that the result is poor. Nevertheless, we have added the corresponding figure of the coefficient of variation here, and if the editor considers it necessary to include it in the manuscript, we will be happy to do so.

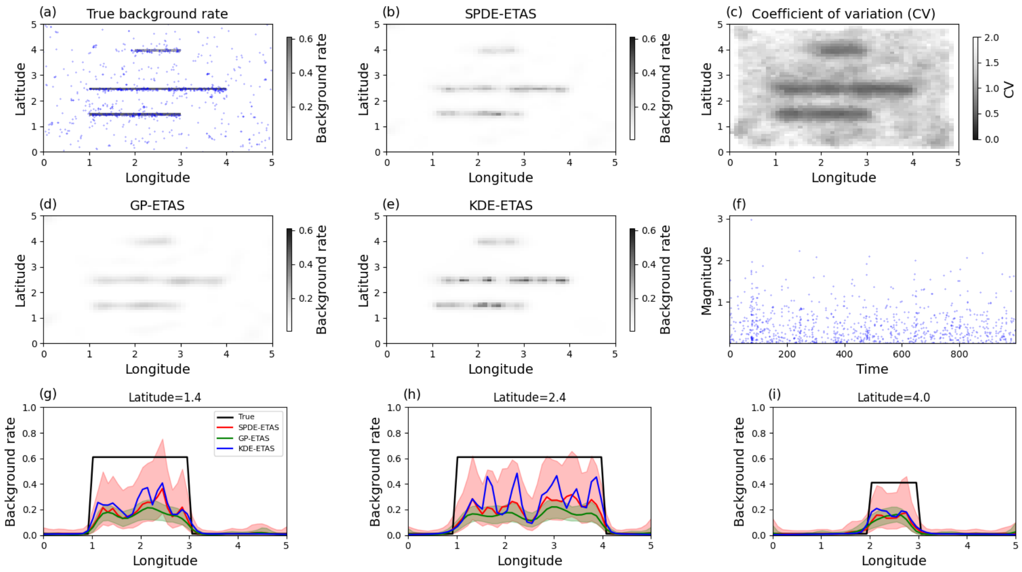


6. I suggest splitting Figure 1 into 2 or 3 separate figures. Also, adjust the horizontal margins of subfigures g, h, and i. Additionally, I recommend considering the coefficient of variation to allow a better comparison of the uncertainty in b. It may also be useful to consult: Bañales, I., Nishikawa, T., Ito, Y., & Aguilar-Velázquez, M. J. (2025). Estimating Inhomogeneous Spatio-Temporal Background Intensity Functions using Graphical Dirichlet Processes. arXiv preprint arXiv:2511.04974.

We assume that you are referring to Figure 3 for the second synthetic test. The same response as in the previous comment applies here. The corresponding figure of the coefficient of variation is provided below.

7. On line 221, it is stated: "The confidence band of the SPDE–ETAS tends to be wider than that of the GP–ETAS in regions with high background seismicity." I consider this an important point to highlight. As mentioned in Equation 12, you are using a homogeneous Matérn kernel, which can introduce significant biases in the estimation since it applies the same covariance weighting across different regions. Moreover, when $\nu \to \infty$, the Matérn covariance approaches a Gaussian covariance. Therefore, it is not clear why the SPDE intervals are much wider than those from the GP–ETAS. This could be due to the Laplace approximation, which might allow the acceptance of inadmissible values. I would greatly appreciate a more detailed discussion regarding this point.

We would like to clarify that the wider uncertainty bands observed for the SPDE–ETAS model are not a systematic feature. This behavior is mainly visible in the second synthetic test (Figure 3), particularly in regions with high background seismicity, while it is not obvious in the first synthetic test (Figure 2). In test 2, the SPDE–ETAS credible intervals are noticeably wider in regions of



high background intensity (see Figure 3(c) and (g–i)). Importantly, these wider intervals tend to better cover the true background rate, as shown in the cross-sections, 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 quantification of uncertainty in regions where the intensity is high and spatially structured, rather than as a drawback of the model. In contrast, in test 1, where the background field is smoother, both approaches produce comparable uncertainty levels. We agree that approximation effects, including the Laplace approximation, may influence the width of the intervals, although this cannot be fully disentangled in the present experiments. This explanation has been included in the second synthetic test.

8. Please provide the references you used to conceptualize Figure 5a, and explicitly describe how it was constructed. Are you using the MAP estimate or the posterior mean? Did you quantify the uncertainty? Why did you not compare the mean integrated squared error (MISE) directly?

Here, we use the log-likelihood evaluated at each grid cell, where the true value is compared with the posterior distribution for both models. This approach allows us to take into account the full posterior distribution and to provide a more comprehensive comparison with the true underlying values than using only the MAP or mean values of the posterior. The log-likelihood is computed over the same number of grid cells for both models and normalized by the total number of cells to ensure an objective comparison. We have slightly modified the text for clarification.

Case study

1. Add references to support "as regions with fluids or magma are likely to exhibit higher background seismicity," in the same way for the paragraph on line 263.

We have added references.

B.2

1. On line 349, explicitly mention that reference is being made to an inhomogeneous Poisson process only in space.

We have changed the phrase from 'We consider an inhomogeneous Poisson process with a spatial intensity given by the LGCP' to 'We consider an inhomogeneous Poisson process in space, with an intensity given by the LGCP.'

2. What is the difference between A and $\tilde{A}$?

In the revised manuscript, we have clarified that A is the projection matrix onto the observation locations (event positions), whereas $\tilde{A}$ is the projection matrix onto the quadrature points used for numerical integration of the spatial domain.

3. On line 365, describe in more detail how Q is defined; similarly, on line 380, specify the conditional being referred to.

We have clarified in the revised manuscript that the precision matrix Q is derived from the finite element discretization of the SPDE representation of the Matérn field, and we now explicitly refer to Appendix A for details. We have also specified that the Laplace approximation is applied to the conditional posterior distribution of the latent field given both the parameters (κ, τ) and the observed data.

We thank the reviewers for their valuable comments and suggestions. All comments have been carefully considered, and the manuscript has been revised accordingly. Concerning each point, we refer to the place where the manuscript has been revised. The corresponding revisions are colored in the annotated manuscript. Below, repetitions of the reviewer's comments are shown in black, while our comments and replies to the corresponding points are marked in blue. The page and line numbers refer to the manuscript with marked changes.

Reviewer 1

Overall, the authors have revised the manuscript carefully in response to my comments. I do not fully agree with one point in their response to the third major comment—specifically, the statement that "the KDE–ETAS approach relies on maximum likelihood estimation, which yields only point estimates for the background rate and the triggering parameters." However, this issue does not materially affect the revised manuscript itself. I therefore recommend that the paper be accepted for publication.

We thank the reviewer for their careful reading of the revised manuscript and for their positive assessment. Regarding the remaining point, our statement was intended to highlight that, unlike Bayesian estimation, the KDE–ETAS approach does not provide a fully probabilistic framework for uncertainty quantification. We thank the reviewer for their constructive feedback and recommendation for acceptance.

Reviewer 2

The manuscript presents important improvements over the previous version; however, there are still methodological aspects that should be clarified and corrected before the work is ready for publication.

We are very grateful to the second reviewer for her/his very useful comments and suggestions, which helped us to improve the paper. We have addressed the recommendations as follows:

1. The sentence "we introduce a computationally efficient formulation of the Bayesian 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" is misleading, since there is no single canonical "Bayesian ETAS model". For instance, the formulation proposed by Ross and Kolev does not assume a Gaussian Process prior.

We agree that there is no single canonical "Bayesian ETAS model." Our statement was intended to refer specifically to Bayesian GP-ETAS formulations based on a Log-Gaussian Cox Process framework. To avoid any ambiguity, we will revise the sentence as follows: "We introduce a computationally efficient formulation of a 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."

2. In your answer for the first reviewer you said "As correctly pointed out, the proposed SPDE–ETAS framework does not modify the underlying latent Gaussian ETAS model, but rather provides a com-

putationally efficient representation of the Gaussian process prior using its SPDE–GMRF formulation." nevertheless in the paper you wrote "In this work, we propose an alternative Bayesian ETAS model, termed Stochastic Partial Differential Equation ". I agree with the first reviewer that the improvements are primarily computational rather than methodological. However, statements such as the one mentioned above make the presentation of the approach confusing, since the manuscript suggests that a new model is being proposed. Later in the text, the manuscript uses the expression "As a result, the proposed approach does not modify the underlying statistical model, but provides a computationally efficient approximation" which, in line with my previous paragraph, makes it unclear and somewhat contradictory whether the authors are actually proposing a new model or not.

We agree that our contribution is primarily computational rather than methodological, and that the current wording may be misleading. In particular, we have revised the sentence "In this work, we propose an alternative Bayesian ETAS model, termed Stochastic Partial Differential Equation ..." by replacing "Bayesian ETAS model" with "Bayesian GP-ETAS model," to clarify that we do not introduce a new statistical model but rather a computationally efficient representation of the same underlying model.

3. In this new version of the manuscript, the definition of θ is clearer. However, the definitions of θ_ϕ and θ_μ remain unclear, at least within Section 2. These quantities are only defined in Appendix B, yet there is no reference to this in Section 2. In my opinion, Section 2 should be written with more mathematical rigor in order to clearly define all introduced objects.

To improve clarity and ensure that all parameters are properly defined within Section 2, we now explicitly introduce both θ_μ and θ_φ at their first occurrence, while keeping a concise presentation. In particular, we have revised the text as follows: "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 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." Furthermore, we add "with $\theta_\varphi = (K, \alpha, c, p, d, \gamma, q)$" when specifying the triggering function. Then, in section 2.1, we added "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 $z(x)$, in particular the variance σ^2 and the range parameter κ." This modification ensures that all introduced quantities are clearly defined in Section 2 while maintaining a concise and readable exposition.

4. Section 2.1 is not entirely clear to me. In particular, the introduction of Cox processes is somewhat confusing. Based on the citations to Zhuang and Ogata, I understood that μ was assumed to be fixed but unknown, and that a log-Gaussian process prior was used to estimate it. While the proposed methodology can indeed be applied to estimate a Cox process, and although there is a close relationship between the likelihoods of both models, I believe it is important to clearly and formally distinguish the modeling framework being considered. In particular, the manuscript should clarify whether the uncertainty is epistemic, in which case the prior is introduced to represent uncertainty about a fixed but unknown μ, or whether a hierarchical model is being assumed in which μ itself is

random and estimated through a Bayesian procedure, implying intrinsic (ontological) uncertainty.

I believe this issue could be clarified by rewriting Section 2 with a more formal presentation and by stating the posterior distribution explicitly. Alternatively, if the intention is to leave these details in Appendix B, the exposition should still be written with enough rigor to ensure that the methodology is fully reproducible.

We understand the concern regarding the distinction between a fixed but unknown background rate and a hierarchical Cox process formulation. In our approach, the background intensity is modeled within a log-Gaussian Cox process (LGCP) framework, meaning that $\mu(x) = \exp\{\mu_0 + z(x)\}$ is treated as a stochastic process driven by a latent Gaussian field $z(x)$. This corresponds to a hierarchical Bayesian formulation, where the uncertainty reflects intrinsic variability in the spatial background rate. To clarify this point, we have revised Section 2.1 by explicitly stating that we adopt a hierarchical formulation in which the background intensity is modeled as a stochastic process rather than a fixed but unknown function. We would also like to emphasize that the full hierarchical structure of the model is explicit in Appendix B, as written in equation 21.

$$\pi(B, z, \theta_\mu, \theta_\varphi \mid Y),$$

with $\theta_\mu = (\mu_0, \kappa, \tau)$, and the latent field prior

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

which fully specifies the Bayesian model and ensures reproducibility.

5. I appreciate that in Equation 3 the notation for the summation was changed to a form similar to that used by Zhuang. However, the notation should still include $\{\dots\}\{\dots\}$ to emphasize that the summation runs over that specific set, as written in Zhuang's formulation.

We assume that this suggestion applies to the summation in Eq. (1) in our manuscript. In order to clarify the domain of the summation, we added the curly brackets as in Eq. (4) in Zhuang et al. (2002).

6. In Equation 9, an opening parenthesis is missing.

This has been fixed.

7. In line 181 of the revised manuscript, some quantities are referred to as hyperparameters. However, the term "hyperparameter" is usually reserved for parameters that are fixed a priori in order to define the prior distribution. In contrast, in Appendix B, line 35, prior distributions are assigned to quantities that had previously been called hyperparameters, which becomes confusing. This issue could be clarified by formally defining the posterior distribution and the overall modeling framework. For further discussion on the distinction between parameters and hyperparameters, I suggest consulting Bayesian Nonparametric Data Analysis by Peter Müller and Fernando A. Quintana. The section 2.3 Bayesian framework is still confusing for me. What is z in equation 15? Is it a realization of the Gaussian Process?

We agree that the terminology regarding "hyperparameters" may lead to confusion. In the spatial statistics and SPDE literature, quantities such as (κ, τ) are commonly referred to as hyperparameters, as they govern the covariance structure of the latent Gaussian field. However, in a Bayesian hierarchical framework, these quantities are assigned prior distributions and inferred from the data, and therefore can be more appropriately viewed as model parameters. To avoid ambiguity, we have revised the manuscript to clarify this distinction and ensure consistent terminology throughout. Regarding the latent field z, it represents a realization of a Gaussian random field (approximated as a GMRF through the SPDE formulation), as explained before, and is treated as an unknown quantity to be inferred within the Bayesian framework.

8. In the caption of Figure 1, you describe the semi–interquantile distance from 0.05 to 0.95 using q95–q5, whereas it should be written as q0.95–q0.05, as it is written by Molkenthin et al.. In addition, you include a division by 22, but it is unclear why this factor is introduced. After reviewing Molkenthin et al., according to your response letter, I could not find any mention of the use of the factor 1/2. Moreover, dividing this distance by 22 does not appear to have a clear relationship with the posterior probability associated with the quantity of interest.

Regarding the notation, we note that both expressions q_{95} and $q_{0.95}$ are equivalent and commonly used to denote the 95th percentile, and exist in Figure 3 of Molkenthin et al. (2022). Concerning the division by 2 (there is no division by 22), the quantity $(q_{0.95} - q_{0.05})/2$ corresponds to the semi-interquantile range, which provides a measure of dispersion around the central tendency. This normalization is commonly used to obtain a quantity comparable to a symmetric uncertainty interval, and is also used in Molkenthin et al. (2022, Figure 3).

9. You mention the number of CPUs used, but you do not specify their model or clock speed. I believe including this information would provide greater clarity and context for the reported computational results.

We now mention that all computations were performed on a machine equipped with an Intel Xeon Gold 5115 CPU (2.40 GHz), with 20 cores (2 sockets $\times$ 10 cores per socket).

10. You did not include any references addressing my comment from point 8 in the previous version. I find the proposed comparison somewhat unintuitive, since a higher likelihood does not necessarily imply a better fit. I would suggest considering a metric such as the MISE, providing references that justify the comparison methodology you propose, or alternatively considering the use of Bayes factors instead of relying directly on the likelihood.

We agree that likelihood-based comparisons do not always directly reflect model fit in a general sense. However, in our specific setting, the comparison is performed in a controlled framework where the true underlying background rate is known. This allows us to directly assess the posterior estimates on each grid cell and evaluate how well each model recovers the true spatial structure. In this context, the likelihood provides a consistent criterion for comparing the posterior fit to the known ground truth. For this reason, we prefer to retain this comparison, as it offers a direct and interpretable measure of how accurately the models recover the true background rate.

11. In line 401 of the revised manuscript with tracked changes, the text states "with an intensity given by the LGCP". However, my understanding is that the intensity was modeled as a Gaussian process with a Matérn covariance structure, rather than as an LGCP itself. I believe it is very important to rewrite this section with more mathematical rigor and precision.

In our framework, the background intensity is modeled within a log-Gaussian Cox process (LGCP) formulation, in which the intensity is defined as

$$\mu(x) = \exp\{\mu_0 + z(x)\},$$

where $z(x)$ is a latent Gaussian random field with a Matérn covariance structure. In practice, this field is approximated by a Gaussian Markov random field (GMRF) through the SPDE representation. We acknowledge that the phrasing "with an intensity given by the LGCP" may be misleading and revised this sentence to clarify that the LGCP provides a hierarchical construction of the intensity rather than being the intensity itself.

12. Although, as mentioned in the response to Reviewer 1, you rely on previously established literature, "For this reason, we have not repeated the full methodological development in the manuscript.". I believe it is important for the paper to be self-contained, particularly regarding the definition of the model. At present, I am unable to fully understand several of the modeling choices or reproduce the methodology based solely on the manuscript. For example, the manuscript states that ϕ_i are piecewise linear basis functions; however, from the current exposition I am unable to evaluate $\phi_1(x_i)$, since the explicit form of these functions is not provided. In addition, the definitions of A and $\tilde{A}$ remain ambiguous, possibly because both are evaluated at $\phi_i(x_j)$.

We agree that the manuscript should be sufficiently self-contained to ensure full reproducibility of the proposed methodology. We would like to clarify that the comment regarding previously established literature referred to the GP–ETAS formulation, and not to the SPDE-based model developed in this work. In the current manuscript, the SPDE–GMRF construction and the associated inference procedure are described in Appendix B. However, we acknowledge that some elements, such as the definition of the basis functions ϕ_j and the projection matrices A and $\tilde{A}$, may not have been sufficiently explicit for readers unfamiliar with finite element representations. To address this, we have revised the manuscript by explicitly defining the basis functions ϕ_j as piecewise linear functions associated with the vertices of the triangular mesh, equal to 1 at vertex j, 0 at all other vertices, and linearly interpolated within each triangle. We also clarify that their values at any point correspond to the barycentric coordinates within the triangle containing that point. In addition, we have clarified that $\boldsymbol{\mu}$ denotes the discretized background rate intensity at the mesh vertices, and we have refined the definitions of the projection matrices A and $\tilde{A}$, emphasizing that A evaluates the basis functions at the observation locations, while $\tilde{A}$ evaluates them at the quadrature points used for numerical integration. These additions ensure that the methodology can be fully understood and reproduced directly from the manuscript.

13. Recently, the work Estimating Inhomogeneous Spatiotemporal Background Intensity Functions Using Graphical Dirichlet Processes by Bañales et al. was published, presenting an important extension of Bayesian modeling for seismicity that I believe could contribute meaningfully to the discussion of your work, at least in the context of future research directions.

The reference, Bañales et al., 2026, has been added to the conclusion section in the paragraph, where we discuss future work and potential extensions of the proposed methodology.