Manuscript ID: *egusphere-2025-2542*

Journal: ***EGUsphere***

Improvement of the **Rnnmm-type** climate index approach with a spatio-temporal model based on the Hawkes process

We thank the reviewer for the careful reading and constructive suggestions. Below we respond point by point. Reviewer comments are in *italics*; our replies are in blue. All mentioned edits will be incorporated into the revised manuscript.

Minor comments

1. *In the title, "Rnnmm type" should be "Rnnmm-type".*

   Done.

2. *The full form of IMERG is not introduced.*

   We now expand to "Integrated Multi-satellite Retrievals for GPM (IMERG)" at first use and use IMERG thereafter.

3. *I feel that presenting some tables in a horizontal fashion rather than in a vertical fashion (like now) would look better.*

   As suggested by the reviewer, we changed some of the tables to a horizontal format.

4. *Although the authors claim to provide an extensive cross-validation analysis, my concern is that they draw this conclusion based on a very limited dataset. Given that no extensive simulation has been shown, I think the word "extensive" should be toned down unless a larger number of spatial locations or a larger spatial domain is used.*

   We thank the reviewer for this helpful suggestion. We agree that the term "extensive" may be too strong given that the application is based on 20 monitoring stations. We have therefore revised the wording throughout the manuscript (Abstract and Conclusions), replacing "extensive cross-validation" with "leave-one-out cross-validation" (or equivalent wording that more precisely reflects the scope of the analysis).

   We emphasize that our evaluation is based on a leave-one-out cross-validation design across all available stations (20 folds), in which each station is sequentially excluded and predicted using the remaining data. The proposed model is compared against two Poisson-based alternatives, and predictive performance is assessed using MAD and MSE metrics.

   Regarding the number of locations, expanding the monitoring network is constrained both by data availability and by the computational cost of MCMC estimation, which increases substantially with the number of spatial locations. In the Brazilian environmental context, rain-gauge coverage is often limited in several regions, and a set of 20 quality-controlled stations already represents a realistic and informative regional network for applied studies.

Major comments

5. *While the article addresses an environmental statistics problem, the main contribution appears to be on the statistical side. However, the introduction cites only a few papers, mostly by the authors, and attempts to convince the reader that the existing literature is inappropriate for the problem discussed in the paper. Near the end of the Section, the authors simply claim to introduce an innovative self-exciting Hawkes process, without citing any papers or providing a proper literature review of their proposal. The introduction suggests the authors are introducing the self-exciting Hawkes process for the first time. The approach was proposed by Hawkes in the Biometrika paper "Spectra of some self-exciting and mutually exciting point processes" in 1971, more than fifty years ago. A clearly written paragraph including an appropriate literature survey on the self-exciting Hawkes process must be provided in the introduction section. Not only the Hawkes process, but also its spatio-temporal versions are common in the literature. For example, a review article on this topic is "A Review of Self-Exciting Spatio-Temporal Point Processes and Their Applications", by Alex Reinhart, published in Statistical Science in 2018. The authors should clearly indicate what their novel contribution is from a statistical perspective, or they should simply demonstrate the usefulness of an existing statistical method in the context of climate extremes.*

   We thank the reviewer for this careful and insightful comment. We agree that the introduction did not provide an adequate literature review on the self-exciting Hawkes process, and we appreciate the opportunity to improve this aspect of the manuscript. In the new version, we included in the introduction a clear paragraph providing an appropriate literature survey on the self-exciting Hawkes process and its extensions, in particular, we refer the papers "Spectra of some self-exciting and mutually exciting point processes", written by Hawkes (1971), and "A Review of Self-Exciting Spatio-Temporal Point Processes and Their Applications", by Alex Reinhart (2018), as suggested by the reviewer. We have also revised the text to clearly state the contribution of our work. We thank the reviewer again for this important suggestion, which has significantly improved the clarity and positioning of our contribution within the existing literature.

6. *Overall, the proposed model is a latent Gaussian model, where separate self-exciting Hawkes processes are used to model individual time series across locations, and then the potentially transformed spatially varying coefficients are modeled using Gaussian processes. In this approach, conditioning on the model coefficients, the data are modeled spatially independently. However, a convolution through a Gaussian process does not introduce extremal dependence. The authors can refer to a large statistical literature on spatial extremes in this regard. Hence, as a spatiotemporal model for inferring spatially varying coefficients, this approach may be better suited, but may not be from a spatial-extreme perspective.*

   We thank the reviewer for this conceptually important remark.

   We agree that the proposed model belongs to the class of latent Gaussian models, in which spatial dependence is introduced through Gaussian processes defined on the transformed coefficients, while the counting processes at each location are conditionally independent given these coefficients. In this sense, we acknowledge that this structure does not induce asymptotic extremal dependence in the classical sense considered in the spatial extremes literature (e.g., max-stable or Pareto process models).

   However, the primary objective of this work is not to model joint extremal dependence in the asymptotic multivariate extreme-value framework, but rather to model the temporal frequency of local extreme events (Rnnmm-type indices) and to allow spatially structured inference and interpolation of the associated parameters. The focus of the model is therefore on temporal clustering (via the Hawkes process) and spatial smoothing of coefficients, aiming to improve predictive performance and regional interpretability of climate extreme indices.

In this context, the proposed framework is particularly well suited for inferring spatially varying coefficients and for prediction over monitoring networks, even though it does not belong to the specific class of spatial-extreme models developed to capture asymptotic extremal dependence.

To avoid potential ambiguity, we will revise the Introduction and Conclusions to clarify this methodological positioning and explicitly delineate the scope of the proposed model with respect to the broader literature on spatial extremes.

7. *The authors choose gamma priors for the variance-related hyperparameters, while an inverse-gamma prior would be conjugate. The justification for choosing a non-conjugate prior should be provided. In the algorithm, the authors mention Step 2 as GI. What does it mean? Inverse-gamma? If so, the usual notation is IG.*

   We thank the reviewer for pointing out this important issue.

   First, we acknowledge that there was an inconsistency in notation in the manuscript. In Equation (11), the prior for the variance parameters σ_k^2 was incorrectly stated as Gamma, whereas the implemented and intended prior is the Inverse-Gamma distribution. This has now been corrected throughout the manuscript for consistency. Additionally, in Algorithm 1, the notation "GI" was a typographical error and has been replaced by the standard notation "IG" (Inverse-Gamma).

   In our model, for each $k \in \{W, U, M\}$, the variance parameters follow

   $$\sigma_k^2 \sim \mathrm{IG}(a_{\sigma_k}, b_{\sigma_k}),$$

   while the spatial dependence parameters follow

   $$\phi_k \sim \mathrm{Gamma}(a_{\phi_k}, b_{\phi_k}).$$

   The use of the Inverse-Gamma prior for σ_k^2 is motivated by its conjugacy with the Gaussian likelihood structure arising from the Gaussian process prior specification. This choice leads to closed-form full conditional distributions and facilitates efficient Gibbs sampling within the MCMC algorithm. Such a specification is standard in hierarchical spatial models (see, for example, Banerjee et al., 2014; Diggle and Ribeiro, 2007).

   For the spatial range parameters ϕ_k, conjugacy is not available under the exponential correlation function. The Gamma prior was chosen because it has support on the positive real line and provides sufficient flexibility to control prior mass over plausible correlation ranges. This specification is also commonly adopted in spatial Gaussian process modeling.

   The manuscript has been revised to ensure that all prior specifications are clearly and consistently stated.

8. *The manuscript does not include details on the MCMC diagnostics. Besides, no simulation study has been shown. While I agree that EGUsphere is an environmental sciences-focused*

journal, such details should be provided in the supplement, as the main focus of the manuscript is statistical modeling. Given that there are only 20 locations, I am highly curious about the MCMC chains for the spatial dependence parameters.

We thank the reviewer for this detailed and highly relevant comment. We fully agree that, given the methodological nature of the manuscript, the inclusion of formal MCMC diagnostics and a simulation study would substantially strengthen the statistical foundation of the work, particularly with respect to the inference of spatial dependence parameters under networks with a limited number of locations.

We are currently implementing additional analyses aimed at formally assessing convergence of the MCMC algorithm. These include running multiple independent chains, computing the potential scale reduction factor ($\hat{R}$), and generating trace plots and autocorrelation functions for the key spatial parameters (ϕ_W, ϕ_U, ϕ_M), as well as for the associated variance components.

In addition, we are structuring a simulation study designed to evaluate parameter recovery and predictive performance under different spatial configurations. Specifically, we consider scenarios with 20 locations (reflecting the empirical application) and a denser network with 49 locations over a unit square domain. Due to the hierarchical and spatial structure of the proposed model, these analyses involve substantial computational effort and additional code development, particularly when multiple replications and independent chains are considered.

Our intention is to incorporate these investigations rigorously in the Supplement of the revised manuscript in order to fully address the reviewer's concerns without compromising methodological quality. We would therefore appreciate guidance from the Editorial Office regarding an appropriate timeline for the submission of the revised version that includes these additional analyses.

Additional References Cited in the Responses

Banerjee, S., Carlin, B. P., and Gelfand, A. E.: *Hierarchical Modeling and Analysis for Spatial Data*, 2nd edn., Chapman and Hall/CRC, Boca Raton, FL, 2014.

Diggle, P. J. and Ribeiro, P. J.: *Model-based Geostatistics*, Springer Series in Statistics, Springer, New York, 2007.

Gelfand, A. E., Diggle, P. J., Fuentes, M., and Guttorp, P.: *Handbook of Spatial Statistics*, Chapman and Hall/CRC, Boca Raton, FL, 2010.