



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

Fidel Ernesto Castro Morales¹, Antonio Marcos Batista do Nascimento¹, Marina Silva Paez², Daniele Torres Rodrigues³, and Carla de Moraes Apolinário¹

¹Department of Statistics, Federal University of Rio Grande do Norte, Campus Universitário – Lagoa Nova, Natal, 59078-970, Rio Grande do Norte, Brazil

²Department of Statistical Methods, Institute of Mathematics, Federal University of Rio de Janeiro, Avenida Athos da Silveira Ramos, 149 – Edifício do Centro de Tecnologia, Bloco C (Térreo) – Cidade Universitária, Rio de Janeiro, 21941-909, Rio de Janeiro, Brazil

³Department of Statistics, Federal University of Piauí, Av. Universitária, s/n – Ininga, Teresina, 64049-550, Piauí, Brazil

Correspondence: Fidel Ernesto Castro Morales (fidel.castro@ufrn.br)

Received: 30 May 2025 – Discussion started: 23 September 2025

Revised: 27 April 2026 – Accepted: 17 June 2026 – Published: 30 June 2026

Abstract. This paper proposes an innovative geostatistical model based on self-exciting Hawkes processes for modeling the Rnnmm-type extreme climate index, representing a novel contribution to the literature on climate extremes. The proposed approach generalizes non-homogeneous spatio-temporal Poisson models by incorporating temporal dependence between events through excitation functions, enabling the capture of clustering patterns commonly observed in precipitation time series. The model is formulated within a Bayesian framework, with parameter estimation performed via Markov Chain Monte Carlo (MCMC) methods. Spatial dependence is introduced through hierarchical Gaussian processes, allowing for interpolation in locations without observed data. The model is applied to the R20mm index (annual number of days with precipitation exceeding 20 mm) using data from northern Maranhão (Brazil) for the period 2013–2022. Cross-validation results demonstrate that the proposed model outperforms non-homogeneous Poisson models with and without seasonality in terms of predictive accuracy. Furthermore, the excitation parameters provide additional insights into the persistence and intensity of extreme events, revealing patterns not captured by conventional models. These findings highlight the model's potential to enhance the analysis of climate extremes in regions with high spatio-temporal variability in precipitation.

1 Introduction

The occurrence of extreme climate events, such as intense and prolonged rainfall, poses significant challenges in vulnerable regions, particularly in areas characterized by high climate variability. In the state of Maranhão, Brazil, for instance, the rainfall regime exhibits pronounced seasonality. During the rainy season, precipitation events exceeding 20 mm tend to occur frequently, with short intervals between them. As the dry season approaches, these intervals gradually increase until such events cease to be recorded. Capturing this dynamic and irregular behavior is essential for a more accurate understanding of climate extremes and for supporting mitigation and adaptation strategies in local contexts.

In this regard, climate indices developed by the Expert Team on Climate Change Detection and Indices (ETCCDI), such as the Rnnmm index, have proven to be valuable tools for measuring and monitoring climate extremes. The Rnnmm index is defined as the annual number of days in which daily precipitation exceeds nn mm, where nn is a user-defined threshold. Despite its usefulness, the statistical modeling of these indices still faces limitations, especially in regions affected by large-scale phenomena such as El Niño and La Niña, which irregularly influence seasonal precipitation cycles.

Recent studies, such as Morales and Vicini (2020) and Morales and Rodrigues (2023), have investigated the behav-

ior of extreme rainfall frequencies using the Rnnmm index. Morales and Vicini (2020) incorporated anisotropy into the spatial covariance structure of a spatiotemporal model based on inhomogeneous Poisson processes, originally proposed by Morales et al. (2017), thereby enhancing the model's ability to represent complex spatial patterns. Subsequently, Morales and Rodrigues (2023) developed a more comprehensive model that accounts for the high spatial and temporal variability of precipitation and captures regular seasonal rainfall cycles by incorporating a cyclic function into the intensity function. However, despite these advances, the approach proposed by Morales and Rodrigues (2023) has limitations in capturing temporal factors that influence the index, particularly in regions such as northeastern Brazil, where precipitation cycles are strongly affected by irregular large-scale climate phenomena, such as El Niño and La Niña, leading to significant deviations from expected seasonal patterns.

To address these limitations, Morales (2023) proposed an alternative approach based on partitioning the analysis period, allowing different intensity functions to be defined for each partition of the inhomogeneous Poisson process. This method employs a priori specification of the intensity function parameters using state-space models (West and Harrison, 1997), providing greater flexibility in the modeling process. Additionally, the proposed approach incorporates anisotropy in the spatial covariance structure, further improving the representation of complex spatial patterns. By explicitly considering temporal variability, this strategy enhances the model's ability to capture fluctuations driven by global climate phenomena, resulting in a more refined understanding of extreme precipitation events over time.

Moreover, one of the major challenges in developing studies in this field is the limited spatial and temporal coverage of historical data, particularly in developing countries such as Brazil, where the density of rain gauges is below the threshold recommended by the World Meteorological Organization (Curtarelli et al., 2014). To overcome this limitation, researchers have increasingly relied on remote sensing estimates, such as those provided by satellite products. However, studies indicate that these estimates often underestimate extreme precipitation events. For example, Rodrigues et al. (2020) found that the 3B42V7 product from the Tropical Rainfall Measuring Mission satellite underestimated extreme events between 2000 and 2015 in northeastern Brazil. Similarly, dos Santos et al. (2022) analyzed the Integrated Multi-satellite Retrievals for GPM (IMERG) product from the Global Precipitation Measurement (GPM) mission and demonstrated that the quality of climate index estimates, including Rnnmm, depends on factors such as location and time period. Comparable findings were reported by Batista et al. (2024), who evaluated eight extreme precipitation indices estimated by IMERG in the Parnaíba basin.

Although the models proposed by Morales and Vicini (2020); Morales and Rodrigues (2023) and Morales (2023) have advanced the modeling of the Rnnmm index, several

aspects remain unexplored. For instance, the specific seasonal patterns of the rainfall regime in Maranhão have not yet been thoroughly examined. To bridge this gap, this study proposes an innovative geostatistical model based on self-exciting Hawkes processes (Hawkes, 1971), which enables a more accurate representation of the temporal and spatial dynamics of climate extremes, particularly in regional contexts such as Maranhão. Self-exciting Hawkes processes are a useful class of stochastic processes which can model point patterns exhibiting temporal clustering features, in which the occurrence of an event of interest increases the likelihood of observing new occurrences in the near future. This approach is crucial for improving the understanding of the frequency and intensity of extreme events, thereby contributing to the development of effective mitigation and adaptation strategies for climate change. Originally introduced by Hawkes in 1971, Hawkes processes and extensions have been extensively used in real-world applications in several different fields such as earthquakes (Ogata, 1988), violence (Holbrook et al., 2021) and social interactions (Miscouridou et al., 2018). We also refer to (Reinhart, 2018; Laub et al., 2025) for reviews on the subject.

2 The Model

We now formally define the geostatistical model considered in this article. Let $\mathcal{A} \subset \mathbb{R}^d$ denote the spatial domain of interest, with n fixed observation sites located at $s_1, \dots, s_n \in \mathcal{A}$. These sites represent the locations where data are collected over time. The data are assumed to originate from an underlying stochastic process, evolving as follows.

At each site s_j , for $j = 1, \dots, n$, we observe a continuous-time counting process $N_j(t)$, which keeps track of the number of events that have occurred at location s_j up to time t . The history of each process is denoted by $\mathcal{H}_{\Phi_j} = \{\mathcal{H}_{\Phi_j}(t), t \geq 0\}$, and it depends on a set of unknown parameters $\Phi_j = (\varphi_j, \varrho_j)$, where φ_j and ϱ_j are vectors whose specific roles will be defined later. The collection of all such parameters across sites is denoted by $\Phi = (\Phi_j)_{j=1}^n$.

Conditional on a realization of Φ , we assume that the counting processes $\{N_j(t)\}_{j=1}^n$ evolve independently according to self-exciting point processes, meaning that the occurrence of an event at any given time increases the likelihood of subsequent events occurring in the near future. This characteristic makes the model particularly suitable for representing extreme weather events, where an intense occurrence at a given location tends to trigger additional events within a short period. Specifically, each $N_j(t)$ follows a Hawkes process, a model designed to capture clustering patterns in the event occurrences.

The likelihood of an event occurring at site s_j at time t is governed by the conditional intensity function $\lambda_j(t | \Phi)$, which is defined as:

$$\lambda_j(t | \Phi) = g_j(t | \Phi) + \sum_{t_{j,k} < t} \mu_j(t - t_{j,k} | \Phi). \quad (1)$$

Here, the function $g_j(t | \Phi)$ represents a baseline rate of events, independent of past occurrences. This can be interpreted as the background intensity, describing how frequently events would occur if there were no interactions between them. The second term, $\sum_{t_{j,k} < t} \mu_j(t - t_{j,k} | \Phi)$, accounts for the influence of past events.

The quantity $t_{j,k}$ denotes the occurrence time of the k -th event observed at site s_j , with all such events forming the ordered sequence $\{t_{j,1}, t_{j,2}, \dots\}$. Each past event at time $t_{j,k}$ contributes an additional increase to the intensity at time t , according to the function $\mu_j(t - t_{j,k} | \Phi)$. This function, often called the excitation kernel, controls how strongly and for how long past events influence the likelihood of future occurrences.

To precisely characterize this excitation mechanism, we assume that the function $\mu_j(\tau | \Phi)$ follows an exponential decay:

$$\mu_j(\tau | \Phi) = \mu(\tau | \varphi_j) = \alpha_j e^{-\beta_j \tau}, \quad \tau \geq 0, \quad (2)$$

where $\varphi_j = (\alpha_j, \beta_j)$, with $\alpha_j > 0$, $\beta_j > 0$, and $\alpha_j < \beta_j$. This formulation ensures that the influence of past events diminishes over time, capturing the natural decay in their impact on future occurrences.

Simultaneously, the background intensity function $g_j(t | \Phi)$ is assumed to follow the parametric form:

$$g_j(\tau | \Phi) = g(\tau | \varphi_j) = \gamma_j \eta_j \tau^{\eta_j - 1}, \quad \tau \geq 0, \quad (3)$$

where $\varphi_j = (\gamma_j, \eta_j)$ are site-specific parameters that control the baseline occurrence of events independently of past observations. This parametric form provides flexibility in capturing different temporal patterns of extreme events.

The model specified by the excitation function in Eq. (2) and the background rate in Eq. (3) will be referred to as the *Weibull-Hawkes model*, reflecting the Weibull-distributed baseline hazard function and the Hawkes process governing self-exciting dynamics.

2.1 Model's Prior Distributions

To fully specify the Weibull-Hawkes model in a Bayesian framework, we must assign prior distributions to its parameters. These priors reflect our initial beliefs about the parameter values before incorporating observed data and play a fundamental role in the inference process. A well-chosen prior structure not only provides regularization, but also ensures that the model remains flexible enough to capture the underlying patterns in the data.

In this work, we assign prior distributions to the parameters β_j , α_j , γ_j , and η_j for $j = 1, \dots, n$. The choice of these

priors is guided by both computational considerations and prior knowledge about their plausible values.

The excitation parameter β_j is particularly relevant, as it controls the decay rate of the self-excitation function and influences the temporal clustering of events. The selection of its prior is crucial for the stability and convergence of the estimation algorithm. Some studies suggest using a uniform prior $U(0, a)$ for a suitable choice of constant $a > 0$. However, in this work, we opt for a hierarchical formulation that introduces additional flexibility while preserving interpretability. Specifically, we assume that β_j follows the hierarchical structure:

$$\beta_j = aZ_j, \quad \text{where} \quad (4)$$

$$Z_j \sim \text{Beta}(v\tau, (1 - \tau)v), \quad (5)$$

$$\tau \sim \text{Beta}(a_\tau, b_\tau),$$

where $\text{Beta}(\cdot, \cdot)$ denotes the Beta distribution, and a , v , a_τ , and b_τ are known constants. This hierarchical formulation allows for variability across locations while maintaining control over the range of β_j values.

For the background rate γ_j and the excitation parameter α_j , we assign priors to their logarithmic transformations to ensure positivity and improve numerical stability. Specifically, we define:

$$W_j = \log(\gamma_j), \quad U_j = \log(\alpha_j), \quad M_j = \log(\eta_j), \quad (6)$$

for each $j = 1, \dots, n$. Instead of assuming independent priors for these transformed parameters, we introduce a spatial dependency structure by modeling them as realizations of Gaussian processes:

$$\begin{aligned} W(\cdot) &\sim \mathcal{GP}(\mathbf{x}_W(\cdot)' \Psi_W, \sigma_W^2 \rho_{\phi_W}(\cdot, \cdot)), \\ U(\cdot) &\sim \mathcal{GP}(\mathbf{x}_U(\cdot)' \Psi_U, \sigma_U^2 \rho_{\phi_U}(\cdot, \cdot)), \\ M(\cdot) &\sim \mathcal{GP}(\mathbf{x}_M(\cdot)' \Psi_M, \sigma_M^2 \rho_{\phi_M}(\cdot, \cdot)), \end{aligned} \quad (7)$$

where \mathbf{x}_k represents a vector of covariates, Ψ_k is a vector of unknown regression coefficients, σ_k^2 is a scale parameter, and $\rho_{\phi_k}(\cdot, \cdot)$ is a spatial correlation function depending on the parameter ϕ_k , for $k \in \{W, U, M\}$. This formulation ensures that sites closer to each other exhibit more similar parameter values, allowing spatial smoothing.

By standard properties of Gaussian processes, the prior distributions of the transformed parameters are given by:

$$\begin{aligned} \mathbf{W} &\sim N(\mathbf{X}_W \Psi_W, \Sigma_W), \\ \mathbf{U} &\sim N(\mathbf{X}_U \Psi_U, \Sigma_U), \\ \mathbf{M} &\sim N(\mathbf{X}_M \Psi_M, \Sigma_M), \end{aligned} \quad (8)$$

where \mathbf{X}_k is a matrix of covariates, and Σ_k is the covariance matrix with entries:

$$\Sigma_k[i, j] = \sigma_k^2 \rho_{\phi_k}(s_i, s_j). \quad (9)$$

To model spatial dependencies, we assume an exponential correlation function for all processes W , U , and M , such that:

$$\rho_{\phi_k}(s_i, s_j) = \exp(-\phi_k |s_i - s_j|), \tag{10}$$

where $|\cdot|$ denotes the Euclidean distance between locations s_i and s_j . This choice ensures that correlation decays smoothly as the distance between sites increases.

To complete the prior specification, we assign hyperpriors to the unknown parameters governing the Gaussian process structure. Specifically, for each $k \in \{W, U, M\}$, we assume:

$$\begin{aligned} \Psi_k &\sim N(\mathbf{m}_k, \mathbf{C}_k), \\ \sigma_k^2 &\sim \text{IG}(a_{\sigma_k}, b_{\sigma_k}), \\ \phi_k &\sim \text{Gamma}(a_{\phi_k}, b_{\phi_k}), \end{aligned} \tag{11}$$

where \mathbf{m}_k , \mathbf{C}_k , a_{σ_k} , b_{σ_k} , a_{ϕ_k} , and b_{ϕ_k} are known hyperparameters that control the mean, variance, and spatial correlation range of each Gaussian process.

To formally describe the full prior distribution of the Weibull-Hawkes model, let:

$$\Theta = (\beta, \mathbf{W}, \mathbf{U}, \mathbf{M}, \Psi_W, \Psi_U, \Psi_M, \sigma_W^2, \sigma_U^2, \sigma_M^2, \phi_W, \phi_U, \phi_M). \tag{12}$$

The joint prior distribution factorizes as:

$$p(\Theta) = p(\beta) p(\mathbf{W}) p(\mathbf{U}) p(\mathbf{M}) \prod_{k=W,U,M} p(\Psi_k) p(\sigma_k^2) p(\phi_k). \tag{13}$$

This factorization assumes that the priors for the excitation parameter vector β , the transformed background and excitation parameters \mathbf{W} , \mathbf{U} , and \mathbf{M} , and the hyperparameters governing the Gaussian processes are mutually independent. That is, we assume that knowledge about one group of parameters does not inform or constrain the prior distribution of another.

Although the assumption of prior independence simplifies computations and facilitates inference, it may not always be strictly valid in practical applications. In some cases, incorporating dependencies between priors through hierarchical structures or copula models could improve the model's flexibility. However, for the purposes of this work, we adopt the independence assumption to maintain tractability while still allowing for spatial dependencies to be captured through the Gaussian process priors on \mathbf{W} , \mathbf{U} , and \mathbf{M} .

This completes the prior specification for the Weibull-Hawkes model, establishing a structured Bayesian framework that integrates parameter uncertainty while accounting for spatial dependencies.

3 Bayesian Inference

In this section, we outline the Bayesian inference procedure used to estimate the parameters of the Weibull-Hawkes

model. Our objective is to derive the posterior distribution of the parameter set Θ , given the observed event times t .

Let $L(\Theta | t)$ denote the likelihood function of Θ , conditional on the observed data t . Due to the model's construction, the likelihood factorizes across the n observation sites as:

$$L(\Theta | t) = \prod_{j=1}^n L_j(\Theta | t_j), \tag{14}$$

where $L_j(\Theta | t_j)$ represents the likelihood contribution from site s_j . The likelihood at each site is given by:

$$L_j(\Theta | t_j) = \left[\prod_{i=1}^{m_j} \lambda_j(t_{j,i} | \Theta) \right] \exp(-\Lambda_j(T_j | \Theta)), \tag{15}$$

where $\lambda_j(t | \Theta)$ denotes the conditional intensity function of the process at site s_j , evaluated at time t , and $\Lambda_j(T | \Theta)$ represents the integrated intensity over the observation window $[0, T]$, corresponding to the expected number of events up to time T .

For the Weibull-Hawkes model, the conditional intensity function is given by:

$$\lambda_j(t | \Theta) = \gamma_j \eta_j t^{\eta_j - 1} + \sum_{t_{j,i} < t} \alpha_j e^{-\beta_j(t - t_{j,i})}, \tag{16}$$

where the first term corresponds to the background Weibull rate, while the second term captures the influence of past events at site s_j . The integrated intensity is expressed as:

$$\Lambda_j(T_j | \Theta) = \gamma_j T_j^{\eta_j} + \frac{\alpha_j}{\beta_j} \sum_{k=1}^{n_j(T_j)} \left(1 - e^{-\beta_j(T_j - t_{j,k})} \right), \tag{17}$$

where $n_j(T_j)$ denotes the number of observed events at site s_j up to time T_j .

Bayesian inference is conducted by combining the likelihood function with the prior distribution to obtain the posterior distribution of the parameters. Using Bayes' theorem, the posterior is proportional to:

$$p(\Theta | t) \propto L(\Theta | t) p(\Theta), \tag{18}$$

where $p(\Theta)$ is the prior distribution specified in the previous section.

Given the complexity of the likelihood function and the hierarchical structure of the model, exact analytical inference is intractable. Therefore, we employ Markov chain Monte Carlo (MCMC) methods to sample from the posterior distribution. Specifically, a Metropolis-Hastings algorithm or a Gibbs sampler can be used to efficiently explore the high-dimensional posterior space. These methods not only provide point estimates for the parameters but also allow for uncertainty quantification, yielding credible intervals for key parameters such as β_j , α_j , γ_j , and η_j in the Weibull-Hawkes model.

3.1 Estimation Scheme

We employ a Markov Chain Monte Carlo (MCMC) algorithm to estimate the model parameters. The estimation process iteratively updates the parameters by sampling from their respective full conditional distributions. To ensure reliable posterior inference, we discard the initial samples (burn-in period) and retain only the post-convergence samples for posterior estimation.

The full estimation scheme is detailed in Algorithm 1.

Algorithm 1 MCMC Sampling Scheme for Parameter Estimation.

```

Input: Initial values  $\Theta^{(0)}$ , total number of iterations  $K$ , burn-in period  $B$ 
1 for  $k = 1$  to  $K$  (total iterations) do
    /* Step 1: Sample the regression coefficients */
2 for  $X \in \{W, U, M\}$  do
3     Sample  $\Psi_X^{(k)}$  from:
4      $\Psi_X^{(k)} \sim N(\mathbf{A}_X, \mathbf{B}_X)$ , where  $\mathbf{A}_X = (\mathbf{m}_X \mathbf{C}_X^{-1} + \mathbf{X}'\Sigma_X^{-1}\mathbf{X})\mathbf{B}_X$   $\mathbf{B}_X = (\mathbf{C}_X^{-1} + \mathbf{X}'\Sigma_X^{-1}\mathbf{X})^{-1}$ 
    /* Step 2: Sample the variance parameters */
5 for  $X \in \{W, U, M\}$  do
6     Sample  $\sigma_X^{2(k)}$  from:
7      $\sigma_X^{2(k)} \sim \text{IG}\left(\frac{n}{2} + a_{\sigma_X}, \frac{1}{2}(\mathbf{X} - \mathbf{X}_X \Psi_X^{(k)})' \mathbf{R}_X^{-1} (\mathbf{X} - \mathbf{X}_X \Psi_X^{(k)}) + b_{\sigma_X}\right)$ 
    /* Step 3: Sample the spatial dependence parameters */
8 for  $X \in \{W, U, M\}$  do
9     Sample  $\phi_X^{(k)}$  from:
10     $P(\phi_X | \Theta_{-\phi_X}^{(k)}) \propto \phi_X^{a_{\phi_X} - 1} |\Sigma_X|^{-\frac{1}{2}}$ 
11     $\times \exp\left[-\frac{1}{2}(\mathbf{X} - \mathbf{X}_X \Psi_X^{(k)})' \Sigma_X^{-1} (\mathbf{X} - \mathbf{X}_X \Psi_X^{(k)}) - b_{\phi_X} \phi_X\right]$ 
    /* Step 4: Sample the latent variables */
12 for  $X \in \{W, U, M\}$  do
13     Sample  $\mathbf{X}^{(k)}$  from:
14      $p(\mathbf{X} | \Theta_{-\mathbf{X}}^{(k)}) \propto L^k(\Theta^k | \mathbf{t}) \exp\left[-\frac{1}{2}(\mathbf{X} - \mathbf{X}_X \Psi_X^{(k)})' \Sigma_X^{-1} (\mathbf{X} - \mathbf{X}_X \Psi_X^{(k)})\right]$ 
    /* Step 5: Sample the auxiliary parameter */
15 for  $j = 1, \dots, n$  do
16     Sample  $\beta_j^{(k)}$  from:
17      $p(\beta_j | \Theta_{-\beta_j}^{(k)}) \propto L_j^k(\Theta^k | \mathbf{t}_j) \beta_j^{v\tau} (a - \beta_j)^{(1-\tau)v}$ 
    /* Step 6: Remove burn-in period */
18 Discard initial  $B$  samples  $\{\Theta^{(1)}, \dots, \Theta^{(B)}\}$ 
19 Retain samples  $\{\Theta^{(B+1)}, \dots, \Theta^{(K)}\}$  for posterior estimation
Output: Posterior samples  $\{\Theta^{(B+1)}, \dots, \Theta^{(K)}\}$  after burn-in
    
```

4 Interpolation

In this section, we address the problem of estimating the conditional intensity function, $\lambda_*(t)$, at a new location s^* , where no events have been previously recorded. Since the model assumes that event occurrences are spatially correlated, we can leverage information from the observed locations $\{s_1, s_2, \dots, s_n\}$ to make predictions at unobserved sites.

This is achieved by applying spatial interpolation techniques to the components of the intensity function.

Specifically, we consider interpolation for the background intensity function $g(\tau | \varphi)$ and the excitation function $\mu(\tau | \varrho)$. In Sect. 4.1, we outline the procedure for interpolating $g(\tau | \varphi)$ at the new location; in Sect. 4.2, we extend the interpolation framework to $\mu(\tau | \varrho)$, ensuring that excitation dynamics are spatially consistent; finally, in Sect. 4.3, we describe how these interpolated components are combined to estimate the full conditional intensity function $\lambda_*(t)$ over continuous time.

4.1 Interpolation of the function $g(\tau | \varphi)$

At the new location s^* , the background intensity function is given by $g(\tau | \varphi_*) = \gamma_* \eta_* \tau^{\eta_* - 1}$, where the parameter vector $\varphi_* = (\gamma_*, \eta_*)$ must be inferred from the observed locations $\{\varphi_1, \varphi_2, \dots, \varphi_n\}$.

Recall that we defined $W_j = \log(\gamma_j)$ and $M_j = \log(\eta_j)$ as the logarithmic transformations of γ_j and η_j . Since these transformed parameters follow Gaussian process priors, the joint distribution of the observed values W_1, W_2, \dots, W_n and the unobserved value W_* at s^* is Gaussian. The same holds for M_* .

The posterior predictive distribution for W_* is given by:

$$\mathbf{W}_* | W_1, \dots, W_n \sim N(\mu_*, \Sigma_*), \text{ where} \tag{19}$$

$$\mu_* = \mathbf{x}_W(s_*)' \Psi_W + \Sigma_W(s_*, s_{1:n}) \Sigma_W^{-1} (\mathbf{W} - \mathbf{x}_W(s_{1:n})' \Psi_W), \tag{20}$$

$$\Sigma_* = \Sigma_W(s_*, s_*) - \Sigma_W(s_*, s_{1:n}) \Sigma_W^{-1} \Sigma_W(s_*, s_{1:n})'. \tag{21}$$

A similar expression holds for \mathbf{M}_* , where Σ_M and \mathbf{x}_M replace Σ_W and \mathbf{x}_W , respectively.

At each iteration of the MCMC algorithm, after drawing samples of \mathbf{W}_* and \mathbf{M}_* , we obtain the interpolated parameters $\gamma^* = \exp(\mathbf{W}^*)$ and $\eta^* = \exp(\mathbf{M}^*)$ at the new location s^* , conditioned on the observed values \mathbf{W} and \mathbf{M} . Consequently, in each iteration, we compute the estimated background intensity function $g(t | \varphi^*)$ at s^* , allowing us to extend the model to locations where no events have been previously recorded.

4.2 Interpolation of the function $\mu(\tau | \varrho)$

Interpolating the excitation function $\mu(\tau | \varrho)$ presents a unique challenge compared to the background intensity function $g(\tau | \varphi)$. Unlike $g(\tau | \varphi)$, which depends on continuous parameters that can be directly interpolated using Gaussian processes, $\mu(\tau | \varrho)$ is event-driven. It depends on the discrete event times at each location, making it more sensitive to local temporal clustering patterns. As a result, standard interpolation techniques are not directly applicable, since neighboring locations may exhibit different past event histories, affecting the excitation dynamics.

To address this issue, we introduce a spatially weighted sampling method that probabilistically selects nearby locations, ensuring that the interpolated excitation function at an unobserved site reflects the rainfall dynamics of its surroundings. The key idea is that locations close to each other tend to share similar excitation structures. Therefore, by borrowing information from neighboring sites while incorporating spatial variability, we can construct a reasonable estimate of $\mu(\tau | \boldsymbol{\varrho})$ at unobserved locations.

The method consists of two main steps, which are detailed in Algorithm 2:

1. *Spatially weighted selection of reference locations.* A location is chosen from the set of observed sites, with a probability that decreases with distance from the interpolation site s^* . This ensures that closer locations contribute more to the estimate of $\mu(\tau | \boldsymbol{\varrho})$ at s^* , while more distant locations have a lower influence.
2. *Reconstruction of the excitation function.* Once a reference location is selected, its event times are used to reconstruct the excitation function at s^* , incorporating the sampled parameters from the MCMC iterations.

To formalize the procedure, let \mathbf{S} be an $n \times 2$ matrix representing the spatial coordinates of the n monitoring stations, and let $A_n = \{1, \dots, n\}$ be the set of indices corresponding to the rows of \mathbf{S} . Given a new location s^* where we wish to interpolate $\mu_*(\tau | \boldsymbol{\varrho})$, we proceed as follows at each iteration of the MCMC algorithm:

Algorithm 2 Sampling from $\mu_*(\tau | \boldsymbol{\varrho}^{(k)})$ at an interpolation location s^* .

Input: Set of monitoring stations \mathbf{S} , distance threshold R , weighting parameter q

/* Step 1: Spatially weighted selection of a reference location */

19 Define the set of candidate reference locations:

$$A_n^* = \{j \in A_n; \|s^* - \mathbf{S}[j, \cdot]\| \leq R\}.$$

Assign weights to each candidate location based on inverse distance weighting:

$$P_j = \frac{\|s^* - \mathbf{S}[j, \cdot]\|^{-q}}{\sum_{i \in A_n^*} \|s^* - \mathbf{S}[i, \cdot]\|^{-q}}.$$

Sample a reference location s_j from A_n^* with probability proportional to P_j .

/* Step 2: Estimation of the excitation function μ_* */

20 Let $\{t_{j,k}\}_{k=1}^{n_j}$ be the set of observed event times at the sampled reference location s_j . Compute the excitation function at s^* using:

$$\mu_*(\tau | \boldsymbol{\varrho}_j^{(k)}) = \sum_{k=1}^{n_j} \alpha_j^{(k)} e^{-\beta_j^{(k)}(\tau - t_{j,k})},$$

for $\tau > t_{j,k}$.

Output: Updated value of $\mu_*(\tau | \boldsymbol{\varrho}^{(k)})$ for interpolation at s^*

As outlined in Algorithm 2, this approach ensures that the interpolated excitation function preserves both spatial and temporal structure in rainfall events. The weighting scheme favors closer locations, capturing local variability while preventing excessive influence from distant sites. Additionally, the exponential decay function naturally reflects the temporal influence of past events, maintaining consistency in excitation dynamics across space.

4.3 Interpolation of the function $\lambda_*(t)$

With the interpolated background intensity function $g_*(\tau | \boldsymbol{\varphi}_*)$ and excitation function $\mu_*(\tau | \boldsymbol{\varrho}_*)$ at the new location s^* , we now estimate the full conditional intensity function $\lambda_*(t)$. Since $\lambda_*(t)$ governs the occurrence rate of events at s^* over time, its interpolation must reflect both the spatial structure of the background intensity and the temporal clustering effects captured by the excitation function.

Following the model specification, the conditional intensity function at s^* is given by:

$$\lambda_*(t | \boldsymbol{\varphi}_*, \boldsymbol{\varrho}_*) = g_*(t | \boldsymbol{\varphi}_*) + \sum_{k=1}^{n_*} \mu_*(t - t_{*,k} | \boldsymbol{\varrho}_*), \quad (22)$$

where:

– $g_*(t | \boldsymbol{\varphi}_*) = \gamma_* \eta_* t^{\eta_* - 1}$ is the interpolated background intensity function at s^* .

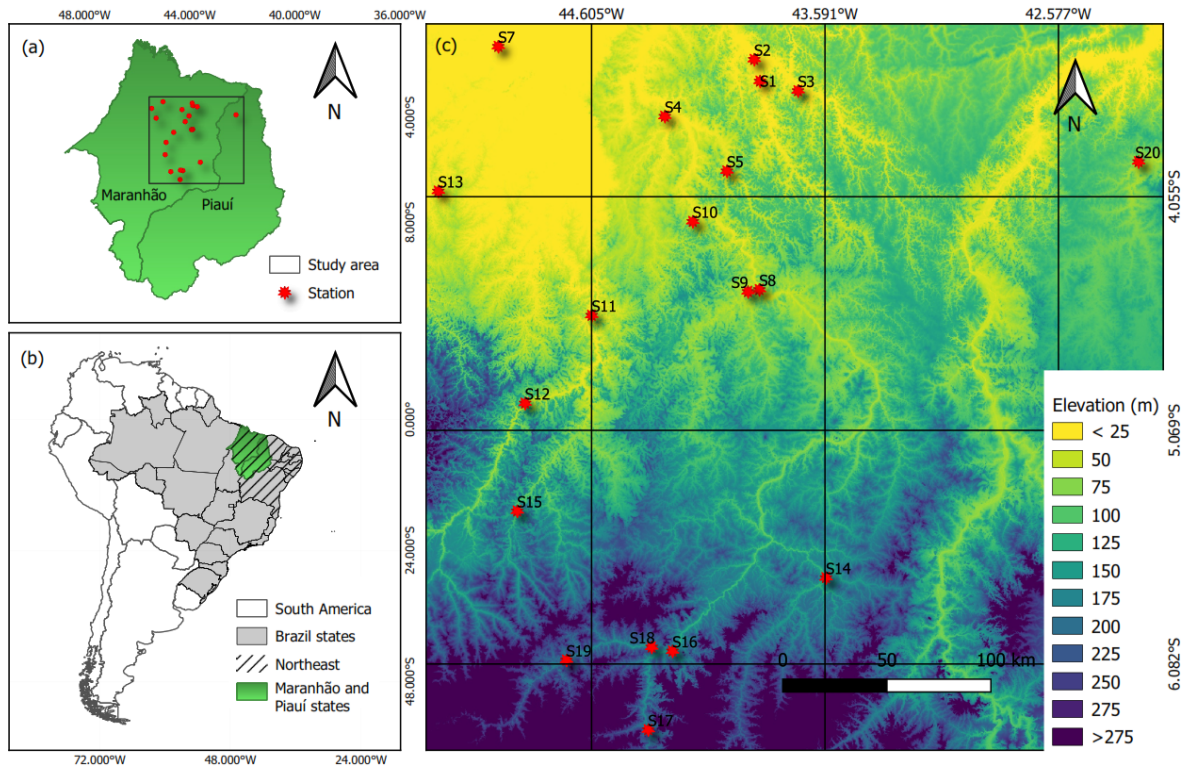


Figure 1. Study area and location of rainfall stations.

- $\{t_{*,k}\}_{k=1}^{n_*}$ represents the set of inferred past event times at s^* .
- $\mu_*(t - t_{*,k} | \boldsymbol{\varrho}_*) = \alpha_* e^{-\beta_*(t - t_{*,k})}$ is the interpolated excitation function, incorporating the influence of past events at s^* .

Since no direct event observations exist at s^* , we must infer the event history $\{t_{*,k}\}_{k=1}^{n_*}$. A natural approach is to simulate event times from the interpolated intensity function itself. At each iteration of the MCMC algorithm, we follow these steps:

1. *Draw samples of the background and excitation parameters:* Using the previously computed interpolations, obtain $\boldsymbol{\varphi}_*$ and $\boldsymbol{\varrho}_*$ for the new location s^* .
2. *Simulate event times at s^* .* Generate a set of candidate event times $\{t_{*,k}\}_{k=1}^{n_*}$ from a Poisson process with intensity $\lambda_*(t | \boldsymbol{\varphi}_*, \boldsymbol{\varrho}_*)$.
3. *Compute $\lambda_*(t)$ at each sampled time.* Using Eq. (22), evaluate the interpolated conditional intensity function at each time step.

By iterating over this procedure during the MCMC sampling process, we obtain a probabilistic estimate of $\lambda_*(t)$ at the unobserved location, allowing us to characterize the expected occurrence rate of extreme events in regions without direct observations.

This framework ensures that $\lambda_*(t)$ maintains consistency with the spatiotemporal structure of the observed data. The background intensity function preserves large-scale spatial patterns, while the excitation function captures local clustering behavior, ensuring that interpolated event dynamics remain realistic.

5 Modeling Extreme Precipitation Events (R20mm) in the Northern Region of Maranhão State

Understanding the occurrence and intensity of extreme precipitation events is crucial for hydrological planning and disaster mitigation. In this study, we model the frequency and spatial distribution of extreme daily precipitation events, defined as days with rainfall exceeding 20 mm (R20mm), in the northern region of Maranhão State, Brazil.

To achieve this, we analyze daily accumulated precipitation data (mm) over a 10-year period, from 1 January 2013 to 31 December 2022. These data were obtained from 20 rain gauges located at national meteorological stations, managed by the National Institute of Meteorology (INMET) and the National Water and Basic Sanitation Agency (ANA). The datasets are publicly available through the INMET meteorological database (<https://bdmep.inmet.gov.br>, last access: 25 June 2026) and the ANA open data portal (<https://dadosabertos.ana.gov.br>, last access: 25 June 2026).

Table 1. MAD and MSE values obtained in the cross-validation study for each excluded station (s_j , with $j = 1, \dots, 20$), along with the station-wide median of each metric, generated by Model A with radius R_1 and different weight vectors (P_1, P_2, P_3).

Station	R_1					
	P_1		P_2		P_3	
	MAD	MSE	MAD	MSE	MAD	MSE
s_1	17	316	21	487	21	498
s_2	2	7	3	12	3	10
s_3	5	51	4	22	4	29
s_4	5	34	6	53	6	53
s_5	5	31	10	120	14	227
s_6	18	493	12	212	16	304
s_7	22	624	13	207	9	107
s_8	7	60	10	114	10	114
s_9	18	346	17	319	17	317
s_{10}	20	535	16	339	14	251
s_{11}	6	58	7	64	12	251
s_{12}	19	461	7	64	11	149
s_{13}	11	164	21	473	23	599
s_{14}	28	949	18	374	13	203
s_{15}	11	154	4	21	4	21
s_{16}	3	12	2	9	2	9
s_{17}	6	59	2	8	2	9
s_{18}	6	41	5	39	5	38
s_{19}	15	252	9	90	8	79
s_{20}	43	2370	49	3176	54	3843
Median	11	159	9.5	102	10.5	131.5

Figure 1 presents the spatial distribution of the rain gauges (P) used in this study. The study area spans parts of the states of Maranhão and Piauí, in northeastern Brazil, a region marked by significant climatic and geographical diversity (Moura and Shukla, 1981; Vale et al., 2024) and persistent socio-economic challenges (IBGE, 2024). Covering approximately 145 611 km², it extends between latitudes 3.2 and 6.5° S and longitudes 41.9 and 45.5° W. The region's topography varies considerably, with elevations ranging from below 25 m to over 275 m, influencing local precipitation patterns. According to the National Institute of Meteorology, annual rainfall in this area ranges from 1000 to 1800 mm (INMET, 2025), with extreme daily events occasionally exceeding 100 mm (Rodrigues et al., 2020).

Precipitation in the region is governed by multiple atmospheric systems throughout the year. The primary driver of rainfall is the Intertropical Convergence Zone (ITCZ), which migrates towards northern and northeastern Brazil during summer and autumn, generating intense precipitation between February and May (Uvo, 1989; Utida et al., 2019). Additionally, the Upper Tropospheric Cyclonic Vortex (UTCV) plays a significant role, particularly during the summer months from December to February, further shaping

Table 2. MAD and MSE values obtained in the cross-validation study for each excluded station (s_j , with $j = 1, \dots, 20$), along with the station-wide median of each metric, generated by Model A with radius R_2 and different weight vectors (P_1, P_2, P_3).

Station	R_2					
	P_1		P_2		P_3	
	MAD	MSE	MAD	MSE	MAD	MSE
s_1	18	343	21	487	21	498
s_2	3	11	3	12	3	10
s_3	4	32	4	21	4	29
s_4	4	31	6	54	6	53
s_5	5	32	10	119	14	227
s_6	13	258	12	193	16	304
s_7	18	405	12	191	9	107
s_8	7	60	10	114	10	114
s_9	18	346	17	319	17	317
s_{10}	20	535	16	339	14	251
s_{11}	6	61	7	65	12	252
s_{12}	19	469	13	223	11	149
s_{13}	11	181	21	482	23	599
s_{14}	26	813	17	359	13	204
s_{15}	12	164	4	21	4	21
s_{16}	2	9	2	9	2	9
s_{17}	3	16	2	8	2	9
s_{18}	5	37	5	39	5	38
s_{19}	12	171	8	87	8	79
s_{20}	52	3531	53	3695	55	3938
Median	11.50	167.50	10.00	116.50	10.50	131.50

the regional precipitation regime (Kousky and Alonso Gan, 1981; Lyra et al., 2020).

5.1 Application and Predictive Performance Analysis

The proposed model is applied to analyze the spatiotemporal dynamics of extreme precipitation events in the northern region of Maranhão. By estimating its parameters, we aim to characterize the frequency, intensity, and clustering patterns of these events, providing insights into their underlying drivers. This analysis allows us to better understand the precipitation regime in the region and assess how extreme rainfall events are distributed over time and space. To ensure the robustness of our approach, we compare its performance with alternative models commonly used in the literature, evaluating their ability to capture the observed precipitation dynamics.

The following models are considered for comparison:

- *Model A*: The proposed Hawkes process model.
- *Model B*: A Poisson model with a seasonal component.
- *Model C*: A standard Poisson model without seasonality.

All three models employ a Weibull intensity function. The processes W , M , and U incorporate the covariates $X_W =$

Table 3. MAD and MSE values obtained in the cross-validation study for each excluded station (s_j , with $j = 1, \dots, 20$), along with the station-wide median of each metric, generated by Model A with radius R_3 and different weight vectors (P_1, P_2, P_3).

Station	R_3					
	P_1		P_2		P_3	
	MAD	MSE	MAD	MSE	MAD	MSE
s_1	19	391	21	487	21	498
s_2	4	22	3	12	3	10
s_3	4	26	4	21	4	29
s_4	5	36	6	56	6	54
s_5	6	50	10	123	14	227
s_6	11	205	12	197	16	305
s_7	16	338	12	185	9	106
s_8	6	49	10	114	10	114
s_9	19	384	17	318	17	317
s_{10}	17	394	16	330	14	251
s_{11}	7	72	7	63	12	252
s_{12}	16	338	13	218	11	146
s_{13}	15	282	21	498	23	600
s_{14}	17	357	15	269	13	193
s_{15}	5	34	3	19	4	21
s_{16}	3	12	2	9	2	9
s_{17}	2	8	2	8	2	9
s_{18}	5	38	5	39	5	38
s_{19}	9	103	8	85	8	79
s_{20}	60	4748	60	4748	60	4748
Median	8.00	87.50	10.00	118.50	10.50	130.00

(1, Latitude, Longitude), with $X_M = X_U = X_W$. The variance parameters σ_W^2, σ_M^2 , and σ_U^2 follow an Inverse-Gamma prior distribution, $IG(0.001, 0.001)$, while the scale parameters ϕ_W, ϕ_M , and ϕ_U follow a Gamma prior distribution, $Gamma(0.001, 0.001)$.

In the case of the Hawkes process model, the excitation decay parameter β is defined as $\beta = aZ$, where $a = 2$ and $Z \sim \text{Beta}(2\tau, (1 - \tau) \cdot 2)$, allowing for flexibility in capturing self-excitation effects.

Finally, for models incorporating seasonal effects, the prior distributions for the seasonal component parameters δ and f are given by:

$$p(\delta) = \frac{1}{\sqrt{\delta(100 - \delta)}},$$

$$p(f) = \frac{1}{\sqrt{\left(f - \frac{1}{365+10}\right)\left(\frac{1}{365-10} - f\right)}}.$$

This comparative analysis enables us to evaluate how well the proposed model represents precipitation extremes in Maranhão and how it performs relative to other modeling strategies. By contrasting its predictive capabilities with existing methodologies, we assess its effectiveness in capturing the temporal and spatial structure of extreme rainfall events,

ensuring a comprehensive understanding of the region’s precipitation patterns.

5.1.1 Sensitivity Analysis of the Interpolation Method

Before comparing the predictive performance of the models, we conduct a sensitivity analysis to evaluate the impact of key parameters in the interpolation process for the function Λ . Specifically, we assess the influence of the radius R used to define neighboring stations and the weight vector P assigned to each neighbor.

To analyze the effect of R , we consider three different definitions:

- R_1 : The maximum distance between the observed stations.
- R_2 : The midpoint between the maximum and average distances.
- R_3 : The average distance between stations.

Similarly, we test three different weight functions for the neighboring stations:

$$- P_{1i} = \frac{|s_i - s_k|^{-1}}{\sum_k |s_i - s_k|^{-1}}$$

$$- P_{2i} = \frac{|s_i - s_k|^{-3}}{\sum_k |s_i - s_k|^{-3}}$$

$$- P_{3i} = \frac{|s_i - s_k|^{-6}}{\sum_k |s_i - s_k|^{-6}}$$

where s_i represents the location where the function $\Lambda(s_i, t)$ is being predicted, and s_k are the neighboring stations within the selected radius R_j , for $j = 1, 2, 3$.

To evaluate interpolation accuracy, we compute the mean absolute deviation (MAD) and mean squared error (MSE) at each excluded station s_j in a cross-validation procedure. These metrics are defined as

$$MAD_j = \frac{1}{n_j} \sum_{t \in A_j} |n_{j,t} - \hat{\Lambda}_j(t)|, \tag{23}$$

$$MSE_j = \frac{1}{n_j} \sum_{t \in A_j} (n_{j,t} - \hat{\Lambda}_j(t))^2, \tag{24}$$

where $n_{j,t}$ is the observed number of extreme precipitation events in the interval $(0, t]$ at location s_j , and $A_j = \{t_{1,j}, \dots, t_{n_j,j}\}$ is the set of observed event times at that location.

Tables 1, 2, and 3 present the MAD and MSE values obtained in the cross-validation study for each excluded station (s_j , with $j = 1, \dots, 20$), considering Model A applied to each radius (R_1, R_2 , and R_3) and the different weight vectors (P_1, P_2, P_3), as well as the median MAD and MSE across all stations for each radius–weight combination. The results indicate that the combination of R_3 with P_1 yielded the lowest MAD and MSE values in most cases, both at the station level and in terms of the reported medians, leading to its selection for use in the interpolation method of Model A.

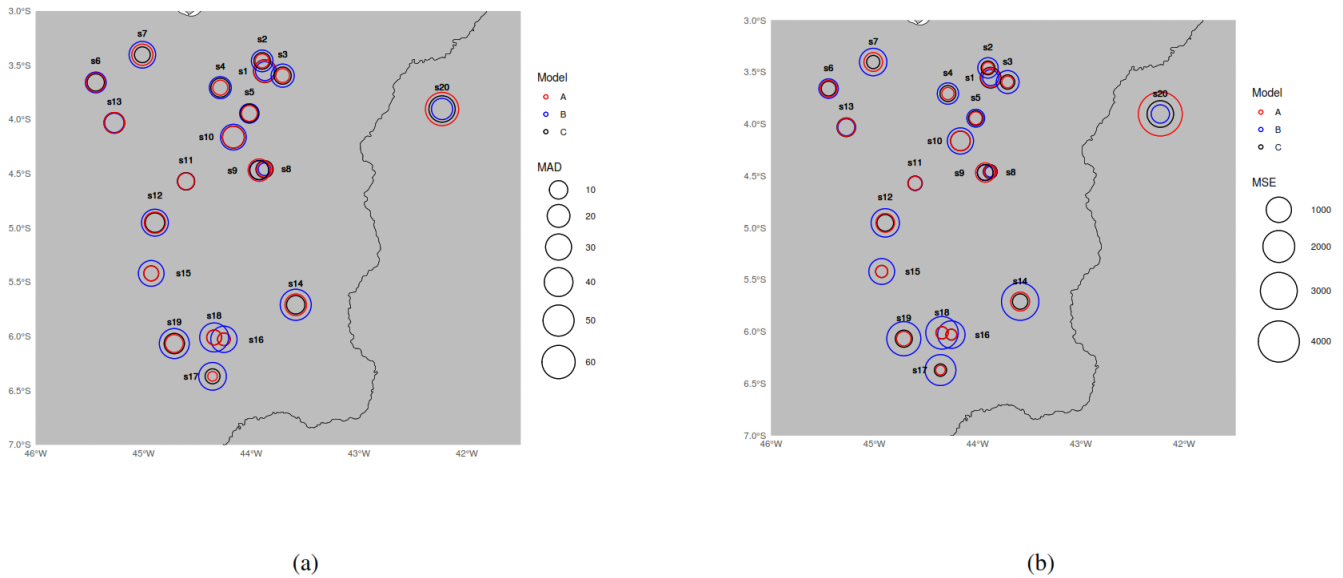


Figure 2. MDA (a) and MSE (b) results in the cross-validation study for each excluded station s_j , $j = 1, \dots, 20$, generated by the competing models. In each subplot, the colored circles represent the models on the study-area map with a radius length scale representing MAD/MSE measurement.

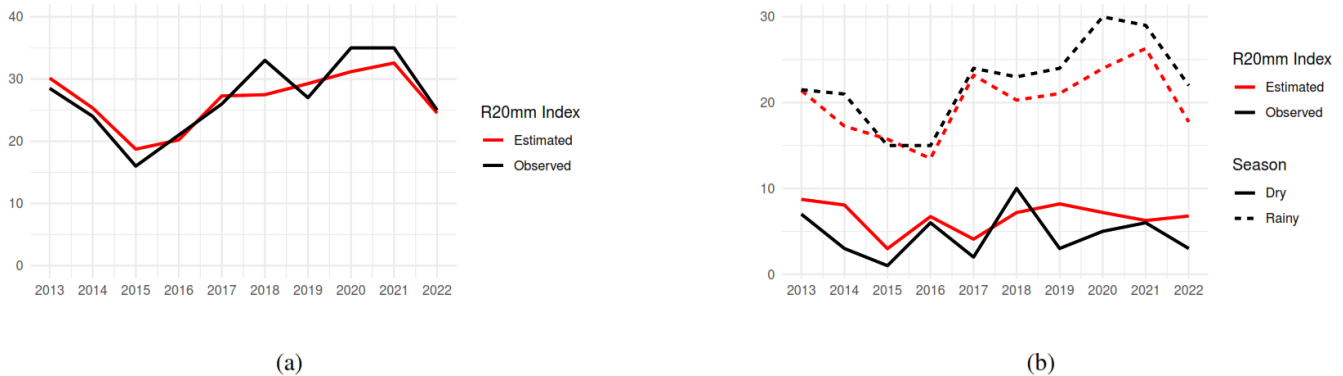


Figure 3. The estimated and the observed R20mm index by year (a) and season (b) at station s_3 .

5.1.2 Cross-Validation for Predictive Performance

To assess the predictive performance of the models in interpolating the R20mm index, we conduct a leave-one-out cross-validation study. The procedure consists of systematically removing data from one station s_i at a time and fitting the models using the remaining stations. This process is repeated for all 20 stations in the dataset.

After fitting each model under these 20 different configurations, we apply interpolation methods to estimate the integrated intensity function $\Lambda_j(t)$ at the removed station s_j . The quality of these predictions is then evaluated using the MAD and MSE metrics defined in Eqs. (23) and (24).

By combining the sensitivity analysis and the cross-validation study, we aim to ensure that the proposed model not only accurately captures the spatiotemporal structure of

extreme precipitation events but also demonstrates superior predictive performance compared to alternative approaches.

In Fig. 2, we observe that Model A generally exhibited the lowest MAD and MSE values compared to the other models, indicating its superior predictive performance. These results suggest that Model A is the most suitable for capturing the observed data patterns, providing more accurate forecasts. Figure 4 displays the predictions of the function $\Lambda_3(t)$ generated by each model, along with their respective 95% credibility intervals. It is evident from this figure that Model A yields the most consistent predictions, with lower uncertainty and a better fit to the observed data, confirming its superior predictive performance. In Fig. 3, we present time series plots comparing the estimated and observed R20mm index at station s_3 by year and season (dry/rainy); in Maranhão, the rainy season occurs from January to June, while the dry

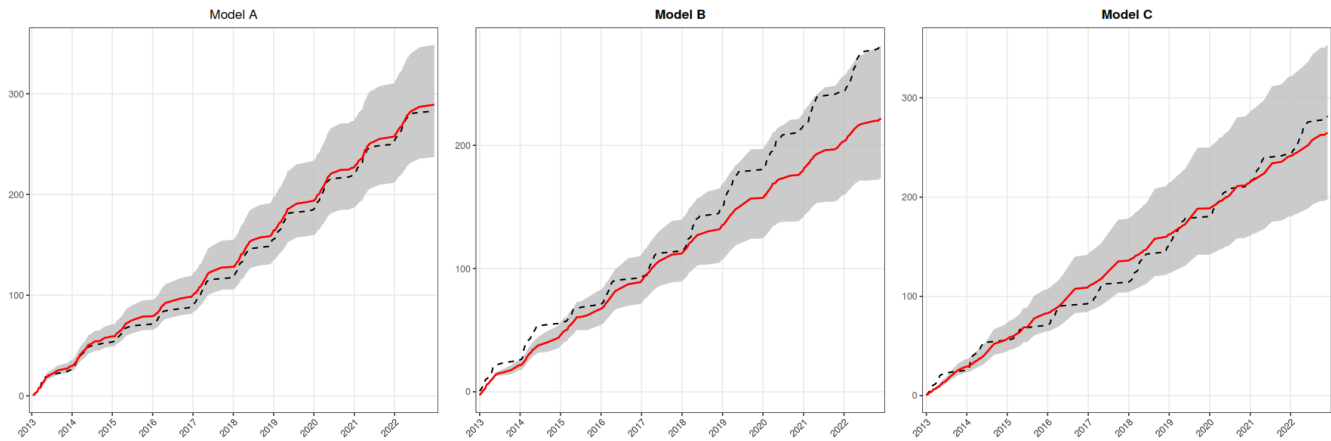


Figure 4. Estimated function $\Lambda_3(t)$ (solid red line) with a 95 % credibility interval (shaded area) and the observed number of event occurrences in the interval $(0, t]$, $n_{j,t}$ (black dashed line), for each of the analyzed models.

season occurs from July to December. These plots demonstrate the high accuracy of Model A in predicting the R20mm index and effectively capturing its temporal dynamics.

Figure 2 also shows that Model A outperforms Model B in terms of predictive accuracy. Model A achieved lower MAD and MSE values in 85 % of cases compared to Model B. Additionally, Model A also demonstrated superior performance compared to Model C, achieving lower MAD values in 50 % of cases, tying in 20 %, and showing higher values in only 30 % of instances. These results indicate that, overall, Model A provides better predictive performance than Model C.

Table 4 summarizes the estimated values of the parameters ϕ , σ^2 , Ψ_W , Ψ_M , Ψ_U , δ , θ , and f for Models A, B, and C. The results indicate notable differences in parameter estimates across models, particularly in the spatial dependence parameters (ϕ_W , ϕ_M , ϕ_U) and the variance components (σ_W^2 , σ_M^2 , σ_U^2). Table 5 presents the estimates of the shape parameter η_j of the Weibull intensity function for each station, showing that Model C generally yields higher mean estimates compared to Models A and B, with relatively narrow credibility intervals across all stations. Table 6 presents the estimates of the scale parameter γ_j of the Weibull intensity function for each station, showing that Model B generally yields higher mean estimates compared to Models A and C, with relatively narrow credibility intervals across all stations. Finally, Table 7 reports the estimates of the parameters α_j and β_j governing the excitation function μ in the Hawkes process, specifically for Model A. In our application context, the parameter α_j reflects the instantaneous increase in the intensity function following the occurrence of a new event, while the parameter β_j describes the rate at which the intensity decays back to its baseline level. That is, the occurrence of a precipitation event exceeding 20 mm in a single day increases the probability of a subsequent similar event by an increment of 0.049 in the intensity, as observed at location s_1 (Table 7). This influence decays at a rate of 0.064 (Table 7).

The results suggest some variability in the excitation intensity (α_j) and decay rate (β_j) across different stations, reflecting the spatial heterogeneity in extreme precipitation events. Higher values of α_j and β_j are concentrated in the northern part of the study area, whereas the lowest values are observed at the southernmost stations (s_{15} , s_{16} , s_{17} , s_{18} , and s_{19}), where α_j ranges from 0.035 to 0.037, and β_j from 0.053 to 0.058 (Table 7). These variations in α_j and β_j between the northern and southern regions may be partly attributed to differences in elevation and rainfall regimes influenced by proximity to the coast and the activity of the Intertropical Convergence Zone (ITCZ). Northern regions, characterized by lower elevation and higher atmospheric humidity, tend to exhibit greater intensity and persistence of extreme rainfall events, while southern, more elevated and inland areas experience lower frequency and weaker clustering of such events.

6 Conclusions

This study proposed a novel geostatistical model based on self-exciting Hawkes processes for modeling the R20mm climate index, representing an innovative extension of the class of non-homogeneous spatio-temporal Poisson models. The main motivation for developing the model stems from the empirical observation that extreme rainfall events in northern Maranhão tend to occur in clusters, especially during the rainy season, suggesting temporal dependence between events.

The proposed model incorporates temporal dependence through an excitation function and spatial dependence via hierarchical Gaussian processes, allowing for interpolation at locations with no observed data. Parameter estimation was conducted under a Bayesian framework using Markov Chain Monte Carlo (MCMC) methods. The model's predictive performance was assessed through a leave-one-out cross-

Table 4. Summary of parameter estimates for ϕ , σ^2 , Ψ_W , Ψ_M , Ψ_U , δ , θ , and f for Models A, B, and C.

Parameter	Model A			Model B			Model C		
	mean	2.5	97.5	mean	2.5	97.5	mean	2.5	97.5
ϕ_W	0.076	0.005	0.496	0.009	0.005	0.023	0.043	0.005	0.299
σ_W^2	0.773	0.038	4.716	0.304	0.117	0.733	0.557	0.150	1.641
ϕ_M	0.007	0.005	0.015	0.006	0.005	0.008	0.006	0.005	0.011
σ_M^2	0.322	0.131	0.734	0.115	0.061	0.217	0.165	0.082	0.320
ϕ_U	0.029	0.005	0.155						
σ_U^2	1.079	0.205	4.082						
Ψ_{W0}	6.436	-6.158	20.275	-13.780	-17.245	-10.183	-1.280	-11.875	9.808
Ψ_{W1}	0.223	-0.066	0.539	-0.289	-0.364	-0.211	-0.005	-0.232	0.235
Ψ_{W2}	-0.141	-0.393	0.088	0.170	0.093	0.249	0.399	0.166	0.647
Ψ_{M0}	-1.501	-5.328	2.405	1.381	-0.411	3.153	-0.478	-3.072	2.044
Ψ_{M1}	-0.033	-0.120	0.052	0.036	-0.002	0.074	-0.008	-0.063	0.046
Ψ_{M2}	0.024	-0.057	0.110	-0.020	-0.060	0.019	-0.024	-0.076	0.028
Ψ_{U0}	-1.256	-10.517	7.155						
Ψ_{U1}	0.033	-0.165	0.227						
Ψ_{U2}	0.106	-0.137	0.315						
δ				3.168	3.110	3.215			
θ				3.505	3.415	3.590			
f				0.00276	0.00275	0.00277			

Table 5. Summary of the shape parameter estimates η_j of the Weibull intensity function, for $j = 1, \dots, 20$, obtained by Models A, B, and C.

Station	Model A			Model B			Model C		
	mean	2.5	97.5	mean	2.5	97.5	mean	2.5	97.5
s_1	0.838	0.782	0.891	0.876	0.874	0.881	0.989	0.960	1.028
s_2	0.838	0.781	0.890	0.875	0.873	0.879	0.989	0.959	1.028
s_3	0.830	0.774	0.888	0.887	0.886	0.893	0.994	0.966	1.036
s_4	0.848	0.787	0.903	0.862	0.861	0.869	0.998	0.972	1.030
s_5	0.828	0.772	0.885	0.876	0.873	0.882	1.012	0.984	1.046
s_6	0.882	0.811	0.955	0.826	0.824	0.831	1.001	0.967	1.028
s_7	0.871	0.804	0.934	0.839	0.836	0.848	0.989	0.958	1.015
s_8	0.815	0.764	0.882	0.890	0.889	0.894	1.014	0.991	1.047
s_9	0.812	0.763	0.876	0.887	0.886	0.891	1.012	0.986	1.046
s_{10}	0.825	0.772	0.887	0.871	0.869	0.878	1.025	1.000	1.058
s_{11}	0.820	0.764	0.875	0.865	0.863	0.869	1.018	0.987	1.045
s_{12}	0.832	0.770	0.886	0.863	0.861	0.871	1.007	0.978	1.038
s_{13}	0.868	0.801	0.929	0.837	0.833	0.844	1.010	0.973	1.043
s_{14}	0.777	0.705	0.867	0.921	0.920	0.925	1.040	1.005	1.076
s_{15}	0.834	0.778	0.894	0.876	0.873	0.883	1.024	0.992	1.059
s_{16}	0.807	0.756	0.887	0.908	0.906	0.913	1.045	1.023	1.073
s_{17}	0.803	0.744	0.889	0.913	0.911	0.921	1.053	1.018	1.089
s_{18}	0.812	0.763	0.896	0.906	0.903	0.911	1.046	1.021	1.075
s_{19}	0.822	0.769	0.901	0.894	0.893	0.898	1.043	1.009	1.086
s_{20}	0.798	0.703	0.874	0.936	0.934	0.939	0.969	0.894	1.049

validation, comparing the results to those obtained from Poisson models with and without seasonality.

The results indicate that the Hawkes-based model outperformed the competing models in terms of predictive accuracy, particularly in regions with pronounced rainfall seasonality. Additionally, the excitation function parameters pro-

vided further insights into the intensity and persistence of extreme events, revealing spatio-temporal patterns not adequately captured by conventional models.

We conclude that the proposed model is promising for applications in climatology, especially in regions with high spatio-temporal rainfall variability. It contributes to the im-

Table 6. Summary of the scale parameter estimates γ_j of the Weibull intensity function, for $j = 1, \dots, 20$, obtained by Models A, B, and C.

Station	Model A			Model B			Model C		
	mean	2.5	97.5	mean	2.5	97.5	mean	2.5	97.5
s_1	0.062	0.042	0.085	0.178	0.173	0.182	0.081	0.061	0.098
s_2	0.061	0.041	0.084	0.181	0.175	0.187	0.085	0.062	0.108
s_3	0.065	0.043	0.091	0.168	0.163	0.175	0.082	0.059	0.100
s_4	0.056	0.040	0.076	0.201	0.190	0.209	0.076	0.061	0.091
s_5	0.061	0.042	0.082	0.178	0.172	0.184	0.068	0.055	0.081
s_6	0.042	0.026	0.065	0.283	0.270	0.298	0.078	0.063	0.103
s_7	0.044	0.028	0.062	0.254	0.242	0.264	0.089	0.073	0.112
s_8	0.069	0.044	0.092	0.156	0.150	0.161	0.055	0.043	0.066
s_9	0.068	0.044	0.090	0.159	0.154	0.163	0.055	0.043	0.065
s_{10}	0.061	0.042	0.080	0.184	0.176	0.193	0.063	0.050	0.074
s_{11}	0.060	0.042	0.090	0.195	0.187	0.202	0.054	0.044	0.068
s_{12}	0.058	0.041	0.087	0.197	0.186	0.208	0.046	0.037	0.056
s_{13}	0.045	0.032	0.066	0.257	0.242	0.273	0.065	0.050	0.085
s_{14}	0.088	0.050	0.138	0.115	0.111	0.119	0.032	0.023	0.040
s_{15}	0.062	0.042	0.089	0.175	0.162	0.185	0.039	0.030	0.048
s_{16}	0.079	0.047	0.108	0.131	0.123	0.137	0.032	0.026	0.038
s_{17}	0.080	0.047	0.114	0.125	0.117	0.132	0.028	0.021	0.036
s_{18}	0.077	0.046	0.103	0.134	0.129	0.140	0.033	0.026	0.038
s_{19}	0.071	0.043	0.098	0.148	0.143	0.154	0.032	0.023	0.042
s_{20}	0.098	0.053	0.173	0.103	0.098	0.106	0.075	0.039	0.130

Table 7. Summary of the estimates for the parameters α_j and β_j of the function μ in the Hawkes process, for $j = 1, \dots, 20$, obtained for Model A.

Station	α			β		
	mean	2.5	97.5	mean	2.5	97.5
s_1	0.049	0.042	0.057	0.064	0.053	0.078
s_2	0.049	0.042	0.057	0.062	0.051	0.074
s_3	0.049	0.042	0.057	0.063	0.052	0.075
s_4	0.049	0.042	0.057	0.064	0.053	0.078
s_5	0.048	0.041	0.056	0.060	0.050	0.073
s_6	0.046	0.038	0.054	0.057	0.047	0.069
s_7	0.048	0.039	0.057	0.059	0.048	0.071
s_8	0.044	0.037	0.053	0.057	0.047	0.072
s_9	0.045	0.037	0.053	0.058	0.047	0.073
s_{10}	0.048	0.041	0.058	0.060	0.050	0.073
s_{11}	0.049	0.040	0.064	0.067	0.054	0.085
s_{12}	0.040	0.033	0.048	0.056	0.045	0.071
s_{13}	0.045	0.037	0.053	0.057	0.046	0.070
s_{14}	0.040	0.032	0.050	0.059	0.045	0.077
s_{15}	0.037	0.028	0.045	0.053	0.040	0.069
s_{16}	0.037	0.030	0.046	0.057	0.044	0.074
s_{17}	0.035	0.028	0.044	0.055	0.041	0.074
s_{18}	0.037	0.030	0.045	0.058	0.045	0.078
s_{19}	0.035	0.028	0.044	0.053	0.040	0.071
s_{20}	0.046	0.034	0.060	0.067	0.048	0.092

provement of climate extremes analysis and forecasting, with potential applications in the planning of climate change adaptation strategies and natural disaster mitigation.

Appendix A: Simulation Study

The study area considered corresponds to the square spatial domain $[0, 1] \times [0, 1]$, while the temporal interval is defined as $(0, 1000]$. The proposed model is simulated over a regular spatial grid of 8×8 points, as illustrated in Fig. A1a.

The parameters used in the simulation process are specified as:

$$\Psi_W = (-2.78, 0, 0)^T, \quad \phi_W = 0.2, \quad \sigma_W^2 = 1, \quad (A1)$$

$$\Psi_U = (-3.02, 0, 0)^T, \quad \phi_U = 0.2, \quad \sigma_U^2 = 1, \quad (A2)$$

and we assume $\eta = 1$ throughout the spatial domain.

For the inference stage, weakly informative prior distributions (*diffuse priors*) are considered for all model parameters, given by:

$$\Psi_W \sim N(0, 100 I), \quad \Psi_U \sim N(0, 100 I), \quad (A3)$$

$$\begin{aligned} \phi_W &\sim \text{Gamma}(0.001, 0.001), \\ \phi_U &\sim \text{Gamma}(0.001, 0.001), \end{aligned} \quad (A4)$$

$$\sigma_W^2 \sim \text{IG}(0.001, 0.001), \quad \sigma_U^2 \sim \text{IG}(0.001, 0.001). \quad (A5)$$

The objective of this simulation study is to evaluate the performance of the parameter estimation procedure of the

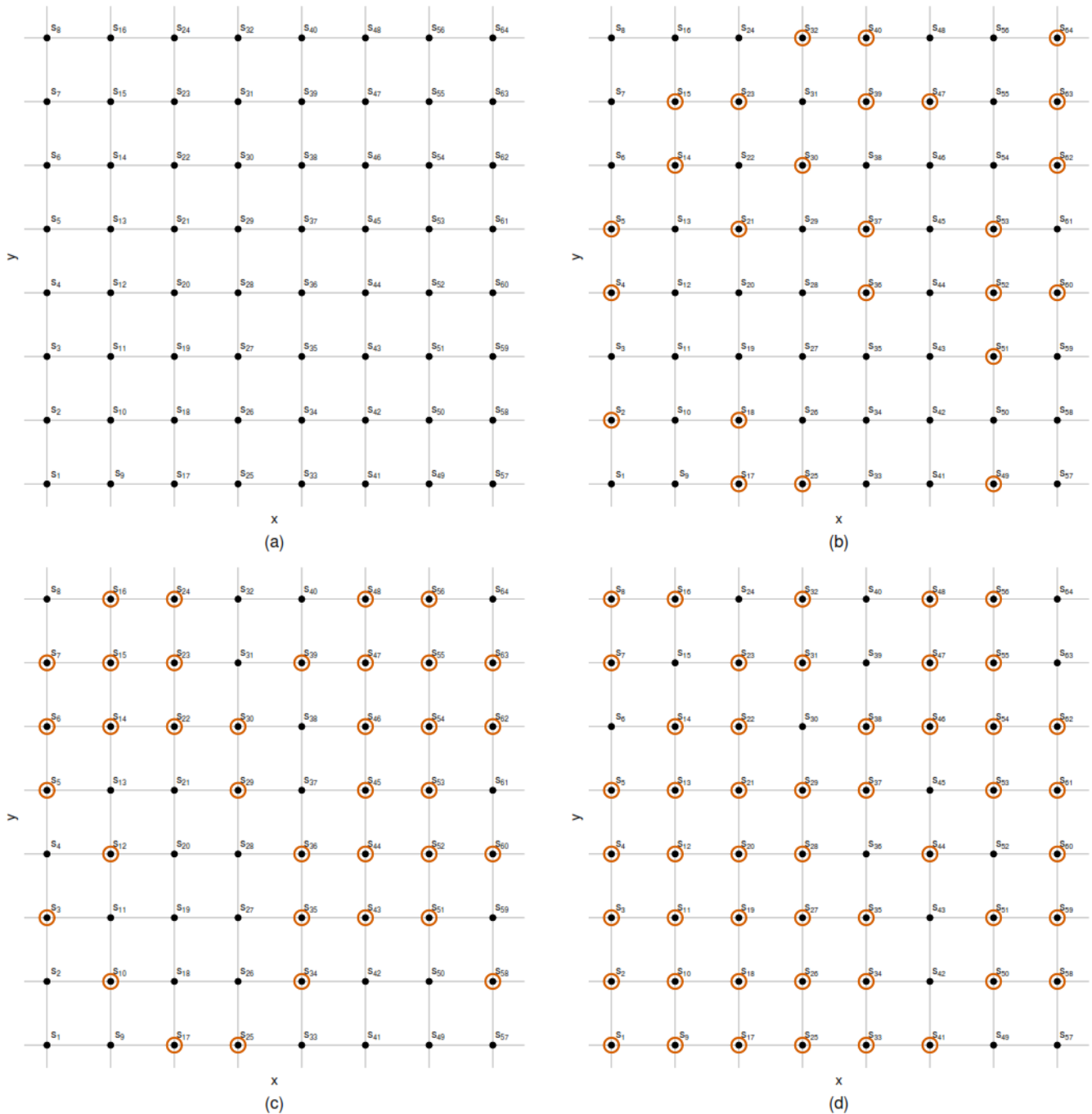


Figure A1. Regular 8×8 spatial grid over the domain $[0, 1] \times [0, 1]$, with locations denoted by s_i , for $i = 1, \dots, 64$. Panels illustrate different sampling configurations: (a) full set of 64 locations; (b) random subsample of 25 locations; (c) random subsample of 36 locations; and (d) random subsample of 49 locations. Circled points indicate the selected locations in each subsample.

spatial process, with emphasis on its stability and convergence properties as the number of observed locations varies. To this end, subsets of different sizes are randomly selected from the full set of locations, considering samples with 25, 36, and 49 spatial points. These configurations correspond to

the subsampling scenarios presented in Fig. A1b, c, and d, respectively.

For each configuration, the model is fitted and the parameter estimates are analyzed in terms of bias, variability, and convergence behavior. Additionally, the model is estimated

using the full set of locations, allowing for a direct comparison with the subsampling scenarios.

This experimental design makes it possible to systematically investigate the impact of spatial sampling density on the inferential quality of the model, providing evidence on the robustness of the proposed method under different levels of available spatial information.

For each configuration, four independent MCMC chains were run from distinct initial values in order to assess parameter convergence. The number of iterations was adjusted according to the spatial sample size to ensure stability of the estimates. Specifically, for $N = 25$, it was used 300 000 iterations, with a burn-in period of 200 000 and thinning of 100; for $N = 36$ and $N = 49$, we used 500 000 iterations, with burn-in of 300 000 and thinning of 100; and for $N = 64$, it was used 600 000 iterations, with burn-in of 400 000 and thinning of 100. This scheme was adopted to ensure adequate exploration of the parameter space and to allow for a robust assessment of convergence across chains.

The results of the simulation study indicate an overall satisfactory performance of the proposed estimation procedure across all sampling configurations considered. For the smallest sample size ($N = 25$), most parameters exhibit good convergence behavior, as indicated by the potential scale reduction factor (PSRF), with the majority of the upper bounds of the 95 % credible intervals below 1.2 (Table A1). Small deviations from full convergence are observed for the parameters ϕ_u and σ_w^2 , although their PSRF values remain close to the acceptable threshold, suggesting near convergence.

As the sample size increases to $N = 36$ and $N = 49$, an overall improvement in convergence diagnostics is observed, with most parameters showing stable estimates; however, the parameter ϕ_u still exhibits relatively slower convergence compared to the others (Tables A1 and A2). For the full dataset ($N = 64$), all parameters show clear convergence, with PSRF values essentially equal to 1.

In terms of accuracy, the 95 % credible intervals for most parameters include the true values used in the simulation, indicating the reliability of the inferential procedure. This behavior is further supported by the posterior distributions presented in Fig. A2, where an increasing concentration of the densities around the true values is observed as the number of spatial locations increases. Overall, the results highlight the positive effect of increasing spatial sampling density on both convergence and estimation accuracy, indicating that the proposed model provides robust inference even under moderate levels of spatial subsampling.

Table A1. Posterior summaries and convergence diagnostics for the simulated study with subsamples of $N = 25$ and $N = 36$ spatial locations. The posterior summaries include the median, 2.5 % and 97.5 % posterior quantiles, and posterior standard deviation (SD). Convergence is assessed using the potential scale reduction factor, reported as point estimate and upper confidence interval (upper C.I.).

Parameter	$N = 25$						$N = 36$					
	Posterior quantiles				PSRF		Posterior quantiles				PSRF	
	50.0 %	2.5 %	97.5 %	SD	Point est.	Upper C.I.	50.0 %	2.5 %	97.5 %	SD	Point est.	Upper C.I.
ϕ_W	0.23	0.01	4.38	1.30	1.02	1.03	0.22	0.01	3.65	1.04	1.01	1.01
ϕ_U	0.20	0.01	8.94	2.80	1.13	1.32	0.05	0.01	1.47	0.49	1.20	1.52
σ_W^2	2.09	0.13	80.70	25.56	1.12	1.23	2.10	0.14	70.56	20.76	1.01	1.01
σ_U^2	1.12	0.26	12.68	4.44	1.02	1.03	0.91	0.24	10.25	4.35	1.10	1.13
Ψ_{0W}	-0.82	-6.80	6.03	2.78	1.02	1.02	-1.24	-7.04	4.93	2.71	1.00	1.01
Ψ_{1W}	-1.28	-2.98	0.54	0.87	1.03	1.09	-0.62	-2.15	1.06	0.80	1.00	1.01
Ψ_{2W}	-0.04	-1.82	1.76	0.87	1.01	1.02	0.46	-1.17	2.18	0.83	1.01	1.03
Ψ_{0U}	-2.41	-5.36	0.72	1.49	1.01	1.03	-1.92	-4.56	0.79	1.37	1.02	1.05
Ψ_{1U}	0.07	-1.82	1.88	0.92	1.06	1.15	-0.22	-1.46	1.06	0.61	1.13	1.35
Ψ_{2U}	-0.66	-2.73	1.04	0.93	1.04	1.08	-1.03	-2.04	0.44	0.63	1.09	1.25

Table A2. Posterior summaries and convergence diagnostics for the simulated study with subsamples of $N = 49$ spatial locations and the full grid with $N = 64$ locations. The posterior summaries include the median, 2.5 % and 97.5 % posterior quantiles, and posterior standard deviation (SD). Convergence is assessed using the potential scale reduction factor, reported as point estimate and upper confidence interval (upper C.I.).

Parameter	$N = 49$						$N = 64$					
	Posterior quantiles				PSRF		Posterior quantiles				PSRF	
	50.0 %	2.5 %	97.5 %	SD	Point est.	Upper C.I.	50.0 %	2.5 %	97.5 %	SD	Point est.	Upper C.I.
ϕ_W	0.18	0.01	4.07	1.20	1.04	1.09	0.21	0.01	3.49	1.02	1.00	1.00
ϕ_U	0.28	0.01	5.31	1.87	1.22	1.56	0.14	0.01	2.59	0.85	1.05	1.06
σ_W^2	1.45	0.08	44.22	12.42	1.03	1.05	1.39	0.10	41.55	11.56	1.01	1.02
σ_U^2	1.18	0.26	16.85	6.28	1.00	1.00	1.18	0.24	14.85	6.15	1.06	1.08
Ψ_{0W}	-1.23	-6.14	3.91	2.30	1.01	1.01	-1.18	-6.34	3.71	2.29	1.01	1.02
Ψ_{1W}	-0.69	-1.84	0.62	0.62	1.01	1.03	-0.75	-1.94	0.48	0.60	1.02	1.06
Ψ_{2W}	0.27	-0.92	1.50	0.60	1.03	1.15	0.05	-1.12	1.25	0.59	1.01	1.01
Ψ_{0U}	-2.35	-5.67	0.83	1.60	1.00	1.00	-2.59	-5.58	0.75	1.57	1.01	1.02
Ψ_{1U}	-0.15	-1.86	1.41	0.81	1.01	1.04	-0.02	-1.50	1.19	0.68	1.01	1.04
Ψ_{2U}	-0.64	-2.34	0.93	0.83	1.02	1.10	-0.26	-1.55	1.16	0.66	1.03	1.08

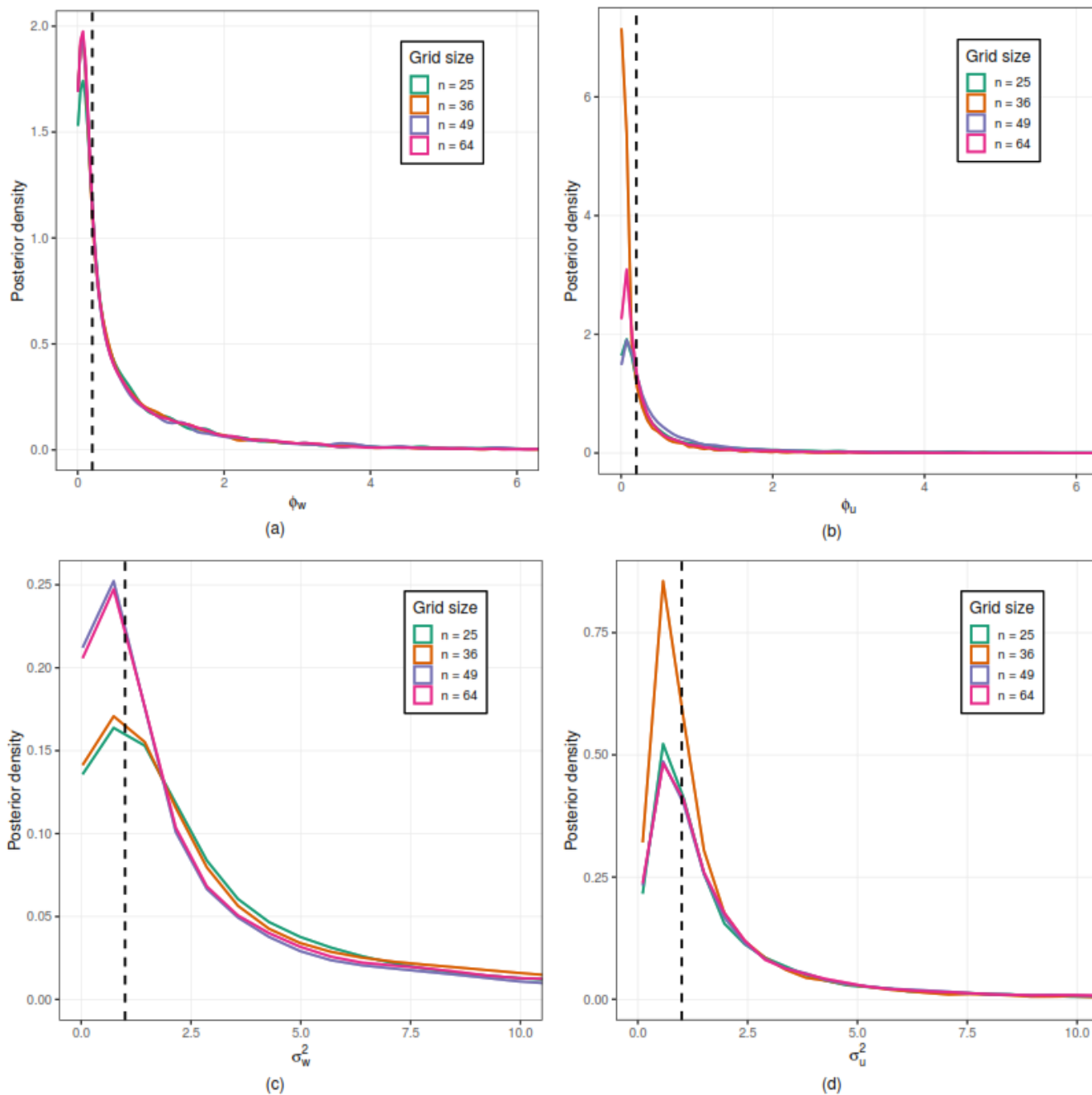


Figure A2. Posterior distributions of the model parameters under different spatial sampling configurations. Panels (a) and (b) display the posterior densities of the spatial decay parameters ϕ_w and ϕ_u , respectively, while panels (c) and (d) correspond to the variance parameters σ_w^2 and σ_u^2 . Results are shown for grid sizes $n = 25, 36, 49$, and 64 . The dashed vertical lines indicate the true parameter values used in the simulation. The figure illustrates the impact of increasing spatial sample size on the concentration and stability of the posterior distributions.

Code availability. The spatio-temporal non-homogeneous Poisson models are implemented in the `STprocPoisson` R package (<https://doi.org/10.5281/zenodo.15651335>, Projeto-CNPq-Clima, 2024b), while the proposed Hawkes-based models are implemented in the `STprocHawkes` R package (<https://doi.org/10.5281/zenodo.15652279>, Projeto-CNPq-Clima, 2024a).

Data availability. The data utilized in this article are freely accessible. Specifically, we analysed data collected from the SISDAGRO (Agricultural Decision Support System) platform, developed by INMET (the National Institute of Meteorology, Brazil) and the National Water and Basic Sanitation Agency (ANA). To ensure full reproducibility, the dataset is provided in both the `STprocHawkes` R package (<https://doi.org/10.5281/zenodo.15652279>, Projeto-CNPq-Clima, 2024a) and the `STprocPoisson` R package (<https://doi.org/10.5281/zenodo.15651335>, Projeto-CNPq-Clima, 2024b).

Author contributions. FECM conceptualized the proposed model and coordinated the team in developing and implementing it in the R software environment. AMBN and MSP developed the statistical properties of the model and assisted in constructing the interpolation method, as well as conducting the literature review on statistical models relevant to the study. DTR contributed to the preprocessing of precipitation data for the state of Maranhão, Brazil, supported the conceptual development of the model by advising on appropriate assumptions for extreme event analysis based on her expertise in climate extremes, and helped interpret the results from a climate science perspective. CMA implemented the model code in R and organized it into an R package format.

Competing interests. The contact author has declared that none of the authors has any competing interests.

Disclaimer. Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. The authors bear the ultimate responsibility for providing appropriate place names. Views expressed in the text are those of the authors and do not necessarily reflect the views of the publisher.

Acknowledgements. We deeply thank the National Council for Scientific and Technological Development (CNPq) and the Ministry of Science, Technology and Innovation (MCTI) for the generous funding granted to the project, identified by process number 405750/2022-6, included in the so-called Call 59/2022 – Line 1 – Modelling the Global Climate System, Impacts, Vulnerability and Adaptation to Climate Change and Monitoring and Forecasting Natural Disasters. The trust and investment of these agencies were fundamental to the success of this work.

We express our gratitude to the LABEST and BME laboratories of the Statistics Department of the Federal University of Rio

Grande do Norte. We thank you for generously granting access to and for permission to use your computer laboratories to implement the computational part of this project. The collaboration and support of these laboratories were essential for the successful completion of this research.

Financial support. This research has been supported by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (grant no. 405750/2022-6).

Review statement. This paper was edited by Rohitash Chandra and reviewed by two anonymous referees.

References

- Batista, F. F., Rodrigues, D. T., and e Silva, C. M. S.: Analysis of climatic extremes in the Parnaíba River Basin, Northeast Brazil, using GPM IMERG-V6 products, *Weather and Climate Extremes*, 43, 100646, <https://doi.org/10.1016/j.wace.2024.100646>, 2024.
- Curtarelli, M. P., Rennó, C. D., and Alcântara, E. H.: Evaluation of the Tropical Rainfall Measuring Mission 3B43 product over an inland area in Brazil and the effects of satellite boost on rainfall estimates, *J. Appl. Remote Sens.*, 8, 083589, <https://doi.org/10.1117/1.JRS.8.083589>, 2014.
- dos Santos, A. L. M., Gonçalves, W. A., Rodrigues, D. T., Andrade, L. D. M. B., and e Silva, C. M. S.: Evaluation of extreme precipitation indices in Brazil's Semiarid region from Satellite Data, *Atmosphere*, 13, 1598, <https://doi.org/10.3390/atmos13101598>, 2022.
- Hawkes, A. G.: Spectra of some self-exciting and mutually exciting point processes, *Biometrika*, 58, 83–90, 1971.
- Holbrook, A. J., Loeffler, C. E., Flaxman, S. R., and Suchard, M. A.: Scalable Bayesian inference for self-excitatory stochastic processes applied to big American wildfire data, *Stat. Comput.*, 31, 1–15, 2021.
- IBGE: Síntese de indicadores sociais: uma análise das condições de vida da população brasileira: 2024/IBGE, Coordenação de População e Indicadores Sociais, IBGE, ISBN 978-85-240-4641-4, 2024.
- INMET: Banco de dados meteorológicos para ensino e pesquisa, BDMEP, <https://bdmep.inmet.gov.br> (last access: 25 June 2026), 2025.
- Kousky, V. E. and Alonso Gan, M.: Upper tropospheric cyclonic vortices in the tropical South Atlantic, *Tellus*, 33, 538–551, 1981.
- Laub, P. J., Lee, Y., Pollett, P. K., and Taimre, T.: Hawkes models and their applications, *Annu. Rev. Stat. Appl.*, 12, 233–258, 2025.
- Lyra, M. J. A., Fedorova, N., Levit, V., and Freitas, I. G. F. D.: Características dos complexos convectivos de mesoescala no Nordeste Brasileiro, *Revista Brasileira de Meteorologia*, 35, 727–734, 2020.
- Miscouridou, X., Caron, F., and Teh, Y. W.: Modelling sparsity, heterogeneity, reciprocity and community structure in temporal interaction data, *Adv. Neur. In.*, 31, 2018.

- Morales, F. E. C.: State-space prior distribution for parameter of nonhomogeneous Poisson spatiotemporal models, *Biometrical J.*, 65, 2200125, 2023.
- Morales, F. E. C. and Rodrigues, D. T.: Spatiotemporal nonhomogeneous poisson model with a seasonal component applied to the analysis of extreme rainfall, *J. Appl. Stat.*, 50, 2108–2126, 2023.
- Morales, F. E. C. and Vicini, L.: A non-homogeneous Poisson process geostatistical model with spatial deformation, *Adv. Stat. Anal.*, 104, 503–527, 2020.
- Morales, F. E. C., Vicini, L., Hotta, L. K., and Achcar, J. A.: A non-homogeneous Poisson process geostatistical model, *Stoch. Environ. Res. Risk A.*, 31, 493–507, 2017.
- Moura, A. D. and Shukla, J.: On the dynamics of droughts in northeast Brazil: Observations, theory and numerical experiments with a general circulation model, *J. Atmos. Sci.*, 38, 2653–2675, 1981.
- Ogata, Y.: Statistical models for earthquake occurrences and residual analysis for point processes, *J. Am. Stat. Assoc.*, 83, 9–27, 1988.
- Projeto-CNPq-Clima: STprocHawkes: Repository for Spatio-Temporal Hawkes Process Models, Zenodo [code], <https://doi.org/10.5281/zenodo.15652279>, 2024a.
- Projeto-CNPq-Clima: STprocPoisson: Spatio-Temporal Non-homogeneous Poisson Process Models, Zenodo [code], <https://doi.org/10.5281/zenodo.15651335>, 2024b.
- Reinhart, A.: A review of self-exciting spatio-temporal point processes and their applications, *Stat. Sci.*, 33, 299–318, 2018.
- Rodrigues, D. T., Gonçalves, W. A., Spyrides, M. H. C., and Santos e Silva, C. M.: Spatial and temporal assessment of the extreme and daily precipitation of the Tropical Rainfall Measuring Mission satellite in Northeast Brazil, *Int. J. Remote Sens.*, 41, 549–572, 2020.
- Utida, G., Cruz, F. W., Etourneau, J., Bouloubassi, I., Schefuß, E., Vuille, M., Novello, V. F., Prado, L. F., Sifeddine, A., Klein, V., Zular, A., Viana, J. C. C., and Turcq, B.: Tropical South Atlantic influence on Northeastern Brazil precipitation and ITCZ displacement during the past 2300 years, *Sci. Rep.*, 9, 1698, <https://doi.org/10.1038/s41598-018-38003-6>, 2019.
- Uvo, C. R. B.: A Zona de Convergência Intertropical (ZCIT) e sua relação com a precipitação da região norte do Nordeste Brasileiro, M.Sc. Dissertation, Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos, Brazil, INPE-4887-TDL/378, 1989.
- Vale, T. M. C. d., Spyrides, M. H. C., Cabral Júnior, J. B., Andrade, L. d. M. B., Bezerra, B. G., Rodrigues, D. T., and Mutti, P. R.: Climate and water balance influence on agricultural productivity over the Northeast Brazil, *Theor. Appl. Climatol.*, 155, 879–900, 2024.
- West, M. and Harrison, J.: Bayesian Forecasting and Dynamic Models, Springer Series in Statistics, 2nd Edn., Springer Series in Statistics, Springer, New York, ISBN 0-387-94725-6, 1997.