Computational Statistics and Data Analysis 167 (2022) 107349

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

www.elsevier.com/locate/csda

Hierarchical Bayesian modeling of spatio-temporal area-interaction processes ☆

Jiaxun Chen [a,*], Athanasios C. Micheas [b], Scott H. Holan [b]

[a] Eli Lilly and Company, Lilly Corporate Center, Indianapolis, IN 46285, United States of America
[b] Department of Statistics, University of Missouri, 146 Middlebush Hall, Columbia, MO 65211-6100, United States of America

ARTICLE INFO

Article history:
Received 6 October 2020
Received in revised form 19 May 2021
Accepted 13 September 2021
Available online 20 September 2021

Keywords:
Autoregressive prior
Bayesian analysis
Double Metropolis-Hastings within Gibbs sampler
Hierarchical model
Spatio-temporal area-interaction process

ABSTRACT

To model spatial point patterns with discrete time stamps a flexible spatio-temporal area-interaction point process is proposed. In particular, this model is suitable for describing the dependency between point patterns over time, when the new point pattern arises from the previous point pattern. A hierarchical model is also implemented in order to incorporate the underlying evolution process of the model parameters. For parameter estimation, a double Metropolis-Hastings within Gibbs sampler is used. The performance of the estimation algorithm is evaluated through a simulation study. Finally, the point pattern forecasting procedure is demonstrated through a simulation study and an application to United States natural caused wildfire data from 2002 to 2019.

1. Introduction

Studying point processes over time has become increasingly important in many disciplines. For example, in criminology or in seismology we wish to predict future occurrences of criminal activity or earthquakes, respectively. The time an event occurs can be crucial to the understanding of several characteristics of the phenomenon, including the evolution of the point process, connections with other events in the same domain, or the time of possible observations of future events. As such, to fully understand the stochastic mechanism of the process of locations as it evolves over time, it is natural to include the time of occurrence as another coordinate in the event.

Spatio-temporal point process models can either be viewed in continuous or discrete time. The continuous time point process model assumes that events can be observed within a continuous spatial and temporal domain, such as the self-exciting point process proposed by Hawkes (1971), space-time Epidemic-Type Aftershock Sequence (ETAS) model introduced by Zhuang and Ogata (2006), and the multi-scale spatio-temporal area-interaction model developed by Iftimi et al. (2018). For discrete time models, points are aggregated over time intervals and formed as point patterns at each time stamp. Recent applications on discrete time models include the log-Gaussian Cox process with dynamical intensity function proposed by Brix and Diggle (2001) and the spatio-temporal Poisson point process model with Gaussian mixture components developed by Zhou et al. (2015).

☆ R and C++ source code of the simulation study is included in the Supplementary Material.

* Corresponding author.
E-mail address: jczz7@mail.missouri.edu (J. Chen).

https://doi.org/10.1016/j.csda.2021.107349
0167-9473/

One of the advantages of the continuous time model is that the model preserves the accuracy of the time stamp for each event and it can predict the precise time of future events. However, the predictive temporal domain needs to include the observed time interval, since these models consider the time stamp of occurrence as a random variable with support consisting of the whole temporal domain. In other words, the continuous time model generates events in the past and future simultaneously. On the other hand, discrete time models can directly predict future events without generating past events, which is computationally more efficient and suitable to the directional feature of time. Based on past events, events in the next time interval can be easily predicted. However, the exact time of the future event within the time interval cannot be provided by discrete time models.

Apart from the temporal component, the spatial structure is also crucial for the analysis. A flexible model should be able to capture the spatial inhomogeneity and interactions between points. An important class of point processes are Gibbs point processes, such as the Strauss process (Strauss, 1975), the hardcore process (Kelly and Ripley, 1976), the Geyer's saturation process (Geyer, 1998) and the area-interaction process (Baddeley and Van Lieshout, 1995). The Strauss process and its special case, the hardcore process, can be used to model inhibition between points. The saturation process and area-interaction process can be applied to both repulsive and attractive point patterns. Properties of the Poisson and Markov point process models have been extensively studied over the past fifty years. Theoretical and practical treatments can be found in the texts by Ripley (1987), Karr (1991), Cressie (1993), Barndorff-Nielsen et al. (1999), Van Lieshout (2000), Lantuejoul (2001), Lawson and Denison (2002), Møller (2013), Møller and Waagepetersen (2003), Daley and Vere-Jones (2003, 2007), Illian et al. (2008), Gelfand et al. (2010), Chiu et al. (2013), Spodarev (2013), Diggle (2013) and Baddeley et al. (2015).

In particular, Gibbs processes offer a large class of models which allow any type of interaction (attraction or repulsion) between the point process events, across space and time and, as such, have received more attention over the past few years (see Dereudre (2019), for a recent review). Moreover, space-time Gibbs processes can describe phenomena with interactions at different spatial or spatio-temporal scales. For example, in the case of seismic data, different sources of earthquakes (faults, active tectonic plates and volcanoes) produce events with different displacements (e.g., Siino et al., 2018) and can be seen as the superposition of background earthquakes and clustered earthquakes (e.g., Pei et al. 2012). Such multi-structure phenomena have motivated researchers to construct new spatial point process models, e.g., Matern-type point processes with applications to tree spatial locations, nerve fiber cells, Greyhound bus stations (Rao et al., 2017), in ecology (Wiegand et al., 2007; Picard et al., 2009), in epidemiology (Iftimi et al., 2017), and in seismology (Siino et al., 2017; Micheas, 2019; Sørbye et al., 2019), mainly based on Gibbs and Cox point processes, but not exclusively (e.g., Lavancier and Møller, 2016).

In contrast, there are very few spatio-temporal models in the literature; Gabriel et al. (2017) and Raeisi et al. (2019) modeled the multi-scale spatio-temporal structure of forest fires occurrences using the log-Gaussian Cox processes (LGCP) and the multi-scale Geyer saturation process, respectively. Additionally, Iftimi et al. (2018) developed a multi-scale area-interaction model for varicella cases and Illian et al. (2012) modeled the locations of muskoxen herds using the LGCP with a constructed covariate measuring local interactions.

In this paper, we propose a discrete-time spatio-temporal area-interaction process, where we implement a Markovian structure for the spatial interaction between the current and the previous patterns. Specifically, the main effect of the Gibbs point process at the current time depends on the point pattern of the previous time stamp. This dependence structure can also incorporate the spatial inhomogeneity of the process. For the interaction between points within a specific time stamp, an area-interaction function is used.

We propose a hierarchical model structure to describe the underlying evolution process. A time indexed parameter structure and autoregressive structure are implemented in the next level of the hierarchy. The time indexed structure allows us to evaluate the point process at each time period, while the autoregressive structure focuses more on the evolution of the underlying process.

For parameter estimation, we implement the double Metropolis-Hastings (DMH) (Liang, 2010) within Gibbs sampler for the hierarchical model. Thus, we are able to sample posterior realizations of the parameters of the area-interaction process at each time using the DMH algorithm and sample the realizations of other parameters by a standard Gibbs step. The DHM replaces the perfect sampler (Møller et al., 2006; Murray et al., 2012) with a standard MCMC sampler. According to (Park and Haran, 2018), the DMH is the most practical approach for high-dimensional data. Based on the estimates of the parameters at each time period and the underlying evolution model, we can effectively predict the parameter values. Consequently, future point patterns can be predicted based on the evolution parameter estimates and the point pattern in the previous time period.

This paper proceeds as follows. Section 2 begins by defining the proposed spatio-temporal area-interaction point process. Section 3 provides details about the hierarchical Bayesian formulation including the choices of prior distribution, joint posterior distribution, and the resulting full conditional distributions. Section 4 provides a sensitivity analysis of the model parameters. Section 5 presents the results of a simulation study that illustrates the effectiveness of the estimation method and forecast procedure. Section 6 illustrates an application to the United States natural caused wildfire data from 2002 to 2019. Finally, in Section 7, we provide concluding remarks.

2. Spatio-temporal area-interaction point process

We introduce a finite spatio-temporal point pattern $\mathbf{X}$ on $\mathcal{W}_{s,t} \subset \mathbb{R}^2 \times \mathbb{Z}^+$, where $\mathcal{W}_s$ is a finite subset of $\mathbb{R}^2$ and $\mathbb{Z}^+$ denotes the collection of positive integers. Thus, this point process is defined on a finite continuous spatial and discrete temporal domain.

Definition 1. The density function of the spatio-temporal area-interaction (STAI) process $\mathbf{X}$ on $\mathbb{R}^2 \times \mathbb{Z}^+$ is

$$f(\mathbf{X}|\boldsymbol{\beta}, \boldsymbol{\gamma}) = f(\mathbf{X}_0) \prod_{t=1}^{T} \frac{1}{C(\beta_t, \gamma_t)} \beta_t^{n_t} \prod_{i=1}^{n_t} g(x_{i,t}|\mathbf{X}_{t-1}) \gamma_t^{n_t - m\{\cup_{i=1}^{n_t} B(x_{i,t}, r_h)\}/(\pi r_h^2)}, \tag{1}$$

where

$$g(x_{i,t}|\mathbf{X}_{t-1}) = \frac{1}{2\pi|\Sigma_g|} \exp\left\{-\frac{1}{2}(x_{i,t} - \mu_{i,t-1})^T \Sigma_g^{-1} (x_{i,t} - \mu_{i,t-1})\right\},$$

$x_{i,t}$ represents the ith point in point pattern $\mathbf{X}_t$, $\mu_{i,t-1} = \mathbf{w}_i(x_{1,t-1}, \ldots, x_{n_{t-1},t-1})$, $t = 1, 2, \ldots, T$, $\mathbf{w}_i$ is the vector of indicators that identify the nearest previous point to $x_{i,t}$, i.e., $\mathbf{w}_i = [I\{||x_{i,t} - x_{1,t-1}|| = \min(||x_{i,t} - x_{j,t-1}||, j = 1, \ldots, n_{t-1})\}, \ldots, I\{||x_{i,t} - x_{n_{t-1},t-1}|| = \min(||x_{i,t} - x_{j,t-1}||, j = 1, \ldots, n_{t-1})\}]$. n_t is the number of points in point pattern at time t. m represents Lebesgue measure and $B(x_{i,t}, r_h)$ is the disc centered at $x_{i,t}$ with radius r_h. Σ_g is a positive definite matrix. The function $C(\beta_t, \gamma_t)$ is the normalizing constant for the model at time t. Notice that $\mathbf{X}_0$ is the initial point pattern (i.e., the observed point pattern at time 0) and any point process model can be used to model $f(\mathbf{X}_0)$.

Lemma 1. *The STAI process defined in Definition 1 is a valid process, i.e., $f(\mathbf{X}|\boldsymbol{\beta}, \boldsymbol{\gamma})$ is a density.*

Proof. First we note that $f(\mathbf{X}_0)$ is assumed to be the density of any point process model and, therefore, we only need consider propriety of the evolutions $f(\mathbf{X}_t|\mathbf{X}_{t-1}, \beta_t, \gamma_t)$, so that propriety of the STAI model easily follows using the evolutions likelihood of (1), i.e.,

$$\int f(\mathbf{X}|\boldsymbol{\beta}, \boldsymbol{\gamma}) d\mathbf{X} = \left[\int f(\mathbf{X}_0) d\mathbf{X}_0\right] \left[\prod_{t=1}^{T} \int f(\mathbf{X}_t|\mathbf{X}_{t-1}, \beta_t, \gamma_t) d\mathbf{X}_t\right] = 1.$$

Now in order for the evolution $f(\mathbf{X}_t|\mathbf{X}_{t-1}, \beta_t, \gamma_t)$ to be a proper point process model, we require that

$$C(\beta_t, \gamma_t) = \sum_{n_t=0}^{+\infty} \int \beta_t^{n_t} \left[\prod_{i=1}^{n_t} g(x_{i,t}|\mathbf{X}_{t-1})\right] \gamma_t^{n_t - m\{\cup_{i=1}^{n_t} B(x_{i,t}, r_h)\}/(\pi r_h^2)} < +\infty.$$

Owing to the form of g, we have

$$\prod_{i=1}^{n_t} g(x_{i,t}|\mathbf{X}_{t-1}) \leq 1,$$

and, therefore, we can write

$$C(\beta_t, \gamma_t) \leq \sum_{n_t=0}^{+\infty} \int \beta_t^{n_t} \gamma_t^{n_t - m\{\cup_{i=1}^{n_t} B(x_{i,t}, r_h)\}/(\pi r_h^2)}.$$

Since the latter is a standard area-interaction model, following Baddeley and Van Lieshout (1995), we have that the right hand side is finite and, therefore, $C(\beta_t, \gamma_t) < +\infty$, as required. □

The function $g(x_{i,t}|\mathbf{X}_{t-1})$ in the main effect indicates that the points of the current point pattern will arise from their neighboring points at the previous time. Specifically, the density of a given location $x_{i,t}$ follows a normal kernel which is centered at the nearest point in the previous point pattern. The parameter β affects the number of points at time t, and parameter γ is the interaction parameter at time t. For $\gamma \in (0, 1)$, the point process shows inhibition between points and smaller γ values lead to stronger interactions. The STAI process generates clustered point patterns if $\gamma > 1$. For $\gamma = 1$, this process is equivalent to the Poisson point process, where the intensity function is determined by the parameter β, function $g(x_{i,t}|\mathbf{X}_{t-1})$, and the point pattern, $\mathbf{X}_{t-1}$, at the previous time point. Thus, this model implies that the points of the current pattern can interact with each other and also interact with the previous point pattern. However, these two types of interaction functions can be different. The initial point pattern $\mathbf{X}_0$ can be generated from any point process.

Algorithm 1 describes the birth-and-death process that can be used to generate point patterns from this process.

Algorithm 1 Birth-and-death process.

for $t = 1$ to T **do**
  for $i = 1$ to L **do**
    With probability 0.5 to propose a birth and generate a point $x^*_{i,t}$ uniformly within $\mathcal{W}$.
    Accept the generated point with probability

$$b(\mathbf{X}_t, x^*_{i,t}) = \frac{p(\mathbf{X}_t \cup x^*_{i,t})}{p(\mathbf{X}_t)}$$

    Or propose a death and randomly remove a point $x_{i,t}$ from the current point pattern $\mathbf{X}_t$.
    Accept the removal of the point with probability

$$d(\mathbf{X}_t \setminus \{x_{i,t}\}, x_{i,t}) = \frac{p(\mathbf{X}_t \setminus x_{i,t})}{p(\mathbf{X}_t)}$$

  end for
end for

3. Hierarchical Bayesian modeling

We define a Bayesian hierarchical model structure which is suitable for describing the underlying process evolution and parameter estimation. Additionally, we implement the double Metropolis-Hastings (DMH) algorithm for sampling the parameters of the area-interaction process and consider two types of structures for the prior distribution of β_t and γ_t, the time indexed structure and the autoregressive structure. Based on the different structures, we obtain the corresponding hierarchical models as follows.

3.1. Prior specification

We first consider a time indexed parameter structure, which assumes that the parameters at each time period are independent of the ones at the other time periods. In other words, the parameters of the area-interaction process will not evolve over time. Thus, we assume independent time indexed priors for β_t and γ_t at each time t and implement the standard DMH for parameter estimation. The vague prior distributions of $\log\beta_t$ and $\log\gamma_t$ are both set to $N(0, 1000)$, for $t = 1, \ldots, T$.

We also propose an autoregressive model for the temporal structure of the parameters of the area-interaction process. Specifically, we use autoregressive model of order 1 (AR(1)) with nonzero mean as a prior distribution for the logarithms of β_t and γ_t. That is,

$$\log\beta_1 = b_1 | \mu_{b_1}, \sigma^2_{b_1} \sim N\left(\mu_{b_1}, \sigma^2_{b_1}\right),$$
$$\log\beta_t = b_t | b_{t-1}, \phi_1, \mu_1, \sigma^2_b \sim N\left(\phi_1(b_{t-1} - \mu_1) + \mu_1, \sigma^2_b\right),$$
$$\log\gamma_1 = \xi_1 | \mu_{\xi_1}, \sigma^2_{\xi_1} \sim N\left(\mu_{\xi_1}, \sigma^2_{\xi_1}\right),$$

and

$$\log\gamma_t = \xi_t | \xi_{t-1}, \phi_2, \mu_2, \sigma^2_y \sim N\left(\phi_2(\xi_{t-1} - \mu_2) + \mu_2, \sigma^2_y\right),$$

where ϕ_1 and ϕ_2 are autoregressive coefficients of the underlying AR(1) processes, μ_1 and μ_2 are the means of the AR(1) processes, and σ^2_b and σ^2_y are the variances of random errors. A positive autoregressive coefficient means that the main effect or the interaction parameters at two consecutive time points are positively correlated, indicating that the parameter values will indeed change gradually over time. This means that the point patterns over time will be similar in terms of their interactive relationships between points. In contrast, a negative autoregressive coefficient indicates negative correlation between parameters at two consecutive time points. That is, the parameters will change more dramatically over time and the generated point patterns will be significantly different from each other at consecutive time points. If the autoregressive coefficient equals zero, the parameters are independent over time and, therefore, we do not expect any trend in terms of the change of interaction structures. Parameters μ_{b_1} and $\sigma^2_{b_1}$ are the mean and variance for the distribution of b_1, and μ_{ξ_1} and $\sigma^2_{\xi_1}$ are the mean and variance for the distribution of ξ_1. The prior distributions for these parameters are defined as follows,

$$\phi_1 \sim N_{(-1,1)}(\hat{\phi}_1, \sigma^2_{\phi_1}), \quad \phi_2 \sim \mathcal{U}(-1, 1),$$
$$\sigma^2_b \sim IG(u_b, v_b), \quad \sigma^2_y \sim IG(u_y, v_y),$$
$$\sigma^2_{b_1} \sim IG(u_{b_1}, v_{b_1}), \quad \sigma^2_{\xi_1} \sim IG(u_{\xi_1}, v_{\xi_1}),$$

$$\mu_1 \sim N(0, \tau_1), \quad \mu_2 \sim N(0, \tau_2),$$
$$\mu_{b_1} \sim N(0, \tau_{b_1}), \quad \mu_{\xi_1} \sim N(0, \tau_{\xi_1}),$$

where u_b, v_b, u_y, v_y, u_{b_1}, v_{b_1}, u_{ξ_1}, v_{ξ_1}, τ_1, τ_2 τ_{b_1} and τ_{ξ_1} are fixed hyperparameters chosen to form vague hyperprior distributions.

On average, we note that the estimates of b_t have larger biases than the biases of ξ_t. In order to improve the estimate of ϕ_1 for small T, the mean of the truncated normal prior is specified as the method of moment estimator $\hat{\phi}_1$, which is calculated based on the estimated value of $b_t, t = 1, \ldots, T$, obtained by stochastic approximation (Geyer, 1998). The variance of this prior is then provided by the asymptotic distribution of $\hat{\phi}_1$, which is $\sigma^2_{\phi_1} = (1 - \hat{\phi}_1^2)/T$. An informative prior for ϕ_2 does not significantly improve the estimate. Thus, we assume a uniform distribution with support (-1,1) as the prior distribution of ϕ_2. Simulation analyses show that the informative prior can have a significant effect on the posterior distribution when T is small. Thus, we recommend selection of an informative prior be done carefully. For large T, we use the same uniform prior for both ϕ_1 and ϕ_2. Moreover, we assume that the parameters r_g and Σ_g are fixed and known since their effects on the configuration of the point process can be confounded. Ideally, we can keep one of these parameters fixed and estimate the other. This issue is the subject of future research.

3.2. Posterior distribution

In order to implement the posterior sampler, we derive the full posterior distribution of the STAI process associated with the full conditional distributions as follows:

$$\begin{aligned}
f(b, y, \phi_1, \phi_2, \sigma_b^2, \sigma_y^2, \sigma_{b_1}^2, \sigma_{\xi_1}^2, \mu_1, \mu_2, \mu_{b_1}, \mu_{\xi_1} | \mathbf{X}) &\propto \prod_{t=1}^{T} \frac{1}{C(b_t, \xi_t)} \exp(b_t n_t) \prod_{i=1}^{n_t} g(x_{ti} | \mathbf{X}_{t-1}) \\
&\times \exp(-\xi_t [n_t - m\{\cup_{i=1}^{n_t} B(x_{i,t}, r_h)\}/(\pi r_h^2)]) \left(\frac{1}{\sqrt{2\pi\sigma_b^2}} \right)^{T-1} \\
&\times \exp\left\{ -\frac{\sum_{t=2}^{T} ((b_t - \mu_1) - \phi_1 (b_{t-1} - \mu_1))^2}{2\sigma_b^2} \right\} \frac{1}{\sqrt{2\pi\sigma_{b_1}^2}} \exp\left\{ -\frac{(b_1 - \mu_{b_1})^2}{2\sigma_{b_1}^2} \right\} \left(\frac{1}{\sqrt{2\pi\sigma_\xi^2}} \right)^{T-1} \\
&\times \exp\left\{ -\frac{\sum_{t=2}^{T} ((\xi_t - \mu_2) - \phi_2 (\xi_{t-1} - \mu_2))^2}{2\sigma_\xi^2} \right\} \frac{1}{\sqrt{2\pi\sigma_{\xi_1}^2}} \exp\left\{ -\frac{(\xi_1 - \mu_{\xi_1})^2}{2\sigma_\xi^2} \right\} \\
&\times p(v_b) p(v_\xi) p(v_{b_1}) p(v_{\xi_1}) p(\phi_1) p(\phi_2) p(\mu_1) p(\mu_2) p(\mu_{b_1}) p(\mu_{\xi_1}),
\end{aligned}$$

where $p(\ldots)$ represents the prior distribution for a given parameter. The full conditional distributions are provided in the Appendix.

Note that, for b_t and ξ_t, $t = 1, \ldots, T$, the full conditional distributions contain an intractable normalizing constant. Thus, we use the DMH to obtain posterior samples from the target distribution. The rest of the parameters can be sampled by a standard Gibbs step. Overall, we implement a DMH within Gibbs sampling algorithm for the STAI process with autoregressive structure and the algorithm is given by

Algorithm 2 Double Metropolis-Hastings within Gibbs sampling algorithm.

Given $\mathbf{b}^{(k)} = b_t^{(k)}, \boldsymbol{\xi}^{(k)} = \xi_t^{(k)}, t = 1, \ldots, T$ at the k-th iteration.
Step 1. Propose $\phi_1^{prop} \sim \pi(\cdot | \mathbf{b}^{(k)}, \mu_1^{(k)}, \sigma_b^{2(k)})$ and $\phi_2^{prop} \sim \pi(\cdot | \boldsymbol{\xi}^{(k)}, \mu_2^{(k)}, \sigma_\xi^{2(k)})$.
Step 2. Propose $\mu_1^{prop} \sim \pi(\cdot | \phi_1^{prop}, \mathbf{b}^{(k)}, \sigma_b^{2(k)})$ and $\mu_2^{prop} \sim \pi(\cdot | \phi_2^{prop}, \boldsymbol{\xi}^{(k)}, \sigma_\xi^{2(k)})$.
Step 3. Propose $\sigma_b^{2(prop)} \sim \pi(\cdot | \phi_1^{prop}, \mu_1^{prop})$ and $\sigma_\xi^{2(prop)} \sim \pi(\cdot | \phi_2^{prop}, \mu_2^{prop})$.
Step 4. Propose $\mathbf{b}^{prop} \sim \pi(\cdot | \mathbf{b}^{(k)})$ and $\boldsymbol{\xi}^{prop} \sim \pi(\cdot | \boldsymbol{\xi}^{(k)})$.
Step 5. Generate auxiliary patterns $\mathbf{X}'$ using the point pattern generating algorithm.
Step 6. Accept the proposed parameters at t with probability r_{DMH}.
else reject and set $b_t^{(k+1)} = b_t^{(k)}, \xi_t^{(k+1)} = \xi_t^{(k)}$. If all the proposed b_t and ξ_t are rejected for $t = 1, \ldots, T$, set $\phi_1^{(k+1)} = \phi_1^{(k)}, \phi_2^{(k+1)} = \phi_2^{(k)}, \mu_1^{(k+1)} = \mu_1^{(k)}, \mu_2^{(k+1)} = \mu_2^{(k)}, \sigma_b^{2(k+1)} = \sigma_b^{2(k)}$ and $\sigma_\xi^{2(k+1)} = \sigma_\xi^{2(k)}$.

where the form of the acceptance ratio is

$$r_{DMH} = \frac{\pi(b_t^{prop}|\phi_1^{(k)}, \sigma_b^{2(k)})\pi(\xi_t^{prop}|\phi_2^{(k)}, \sigma_\xi^{2(k)})q(b_t^{(k)}|b_t^{prop})q(\xi_t^{(k)}|\xi_t^{prop})}{\pi(b_t^{(k)}|\phi_1^{(k)}, \sigma_b^{2(k)})\pi(\xi_t^{(k)}|\phi_2^{(k)}, \sigma_\xi^{2(k)})q(b_t^{prop}|b_t^{(k)})q(\xi_t^{prop}|\xi_t^{(k)})} \times \frac{f(\mathbf{X}_t'|b_t^{(k)}, \xi_t^{(k)})f(\mathbf{X}_t|b_t^{prop}, \xi_t^{prop})}{f(\mathbf{X}_t'|b_t^{prop}, \xi_t^{prop})f(\mathbf{X}_t|b_t^{(k)}, \xi_t^{(k)})}.$$

4. Sensitivity analysis

We provide a sensitivity analysis for several parameters of the STAI to show the effects of different parameter values on the configuration of the point patterns. This study can suggest reasonable parameter spaces for each parameter that can then be used when fitting the models to real data where we do not know the truth.

4.1. Sensitivity analysis of Σ_g

In the main effect, the function $g(x|\mathbf{X}_{t-1})$ is a normal kernel centered at the point in the previous point pattern that is closest to location x. The covariance matrix Σ_g controls the smoothness of the kernel. By assuming a new event is equally likely to be observed from all the directions of a previous event, we use a diagonal matrix for Σ_g. Figure 1 (Supplementary Material) shows the initial point pattern from a non-homogeneous Poisson point process, the realizations from the STAI at $t = 1$ with $\Sigma_g = 0.01I_2$, and with $\Sigma_g = 0.05I_2$ (I_2 is the identity matrix), respectively, given all the other parameters are fixed. From this figure, we see that the pattern is more spread out as the diagonal elements for Σ_g increase, which implies that the new event can be generated further away from a previous event. Consequently, a small value should be used for the diagonal elements in order to preserve the non-homogeneous structure over time.

4.2. Sensitivity analysis of ξ

Next, we present realizations of the STAI process with different interaction levels. To do this, we generate realizations from the processes with $\xi = \log(0.01)$, $\log(1)$, $\log(10)$, and $\log(50)$ for strong repulsion, independence, moderate clustering, and strong clustering, respectively. The interaction distance is fixed at 0.05 for all of these processes. The matrix Σ_g for the main effect function is $0.01I_2$. The realizations from the corresponding STAI process at $t = 1$ are shown in Figure 2 (Supplementary Material). For these processes, smaller values of b are used for larger y to ensure that the generated realizations have approximately 200 points. In this simulation, the values of b are 7, 3, 1, and -0.35, respectively. As we can see, the realizations from the processes with larger ξ are more clustered than the Poisson process with mixture intensity and the realization from the process with $\xi = \log(0.01)$ shows significant regularity. The pattern in Figure 2b (Supplementary Material) can be considered as a realization from a non-homogeneous Poisson process with high intensity in the neighborhood area of the previous pattern. Here, we see that the point patterns at the current time period show completely different structure, even though they are generated based on the same previous pattern.

4.3. Sensitivity analysis of b

To check how the parameter b affects the number of points while the other parameters are fixed, we also conducted a sensitivity analysis. We chose $b = \log(0.6)$, $\log(0.7)$, $\log(0.8)$, and $\log(0.9)$. The other parameters are held fixed at $\Sigma_g = 0.01I_2$ and $\xi = \log(50)$. Figure 3 (Supplementary Material) shows realizations from these processes at $t = 1$, which indicates that the number of points is very sensitive to the value of b. In particular, given the current values of Σ_g and y, the number of points will drop drastically when b is smaller than $\log(0.6)$. Overall, Figure 4 (Supplementary Material) shows that the number of points increases as the value of b increases from $\log(0.1)$ to $\log(5)$.

4.4. Analysis of interactions between b, ξ, and r_h

In this section, we conduct a robustness study in order to illustrate the effect that different values of the model parameters b, ξ, and r_h, have on the generated point patterns. In particular, we consider the following parameter values to simulate point patterns from the process: $b = \log(1)$, $\log(2)$, $\log(4)$, $\log(8)$, $\xi = \log(0.5)$, $\log(1)$, $\log(10)$, $\log(20)$, and $r_h = 0.01$, 0.05, 0.1, 0.5. The plots of the realizations are included in the Supplementary Material (Figures 5–8), with each figure corresponding to a different value of the interaction parameter. For each row of a figure, the value of parameter b increases and we notice that the number of points for each point pattern increases accordingly. For each column, the point patterns are more clustered as the value of ξ increases. The different combinations of values of b and ξ show that both parameters can affect the number of points and can generate similar point patterns from different parameter values in a finite window. This is an indication of the well documented identifiability issues that can occur with such point process (see Chen et al., 2020).

Table 1
Parameter estimation results for 50 realizations from the STAI process with 1 time period based on DMH.

Parameter	True value	Average mean	Average median	Average SD
b	-0.2231	-0.2357	-0.2247	0.3818
ξ	3.9120	3.9316	3.9212	0.4278

Moreover, by comparing the four figures across different interaction scales, we can see that larger r_h values exemplify the effect that the parameter ξ has on the point pattern. That is, for large r_h and $\xi = \log(0.5)$, we can see that there are less points and the points are more spread out in the point pattern. In addition, the different combinations of r_h and ξ can generate similar point patterns which can cause identifiability issues. For example, small r_h or ξ close to 0 can both generate point patterns with weak interactions between points. A common approach to avoid this issue is to estimate the interaction radius r_h separately and include it as a fixed parameter in the model, where the other parameters are assumed unknown and require estimation.

5. Simulation study

This simulation study demonstrates the estimation procedure and results for different model structures. The STAI process with time indexed parameters for one time period is shown in Section 5.1. The STAI process with time indexed parameters for five time periods is shown in Section 5.2, and the STAI process with autoregressive parameter structures is included in Section 5.3.

5.1. Time indexed STAI process with $T = 1$

We first simulate 50 realizations from a STAI process with only one time period and evaluate the average point estimate and average standard deviation. The true parameter values are $b = \log(0.8) = -0.2231$, $\xi = \log(50) = 3.9120$, $r_h = 0.05$, and $\Sigma_g = 0.01 I_2$. All 50 realizations are generated based on the same initial point pattern $\mathbf{X}_0$, which is generated from a non-homogeneous Poisson point process and is shown in Figure 1a (Supplementary Material). For Gibbs point processes, letting the interaction range be dynamic will cause identifiability issues. For instance, different interaction parameters ξ and interaction ranges r_h can generate similar point patterns. In order to simplify the process, we assume that r_h and Σ_g are fixed and known. For each realization, 10,000 posterior realizations are generated using the DMH sampler with 2000 iterations discarded for burn-in. The average posterior means, medians, and standard deviations are provided in Table 1. The results show that the average mean and median are good estimators of the parameters for the STAI process, although the estimate of the parameters based on a single realization may deviate from the true parameter value. Notice that the average of the standard deviations of b is large, which indicates that the estimate of b might be affected by the identifiability issue inherited in Gibbs point process, i.e., different parameter values can generate similar point patterns. In this case, a given point pattern can be generated from processes with several different values of b. The similarity between point patterns can be evaluated based on several discrepancy measures, such as the integrated difference of K-functions or J-functions (see Chen et al., 2020). Also, we show that our approach can recover the parameter values of the current point pattern even if the previous pattern was generated from a different process. Next, we present parameter estimation for the time indexed STAI with more than one time period.

5.2. Time indexed STAI process with $T = 5$

In this section, we show the parameter estimation of 50 realizations from a STAI process of 5 time periods under the time indexed parameter structure. Similar to Section 5.1, we use the same initial point pattern for all 50 realizations and use the DMH algorithm to generate 10,000 posterior realizations, discarding the first 2000 realizations for burn-in. The interaction distance $r_h = 0.05$ and $\Sigma_g = 0.01 I_2$ are fixed for each time period. Table 2 includes the average and standard deviation for the posterior mean, median, and standard deviation of b and ξ at each time period. The results indicate that the DMH can provide good estimates of the parameters for all time periods. However, the average standard deviations of b_t, $t = 1, \ldots, 5$, still indicate possible identifiability issues.

5.3. Evolution STAI process

For the evolution STAI process, we implement an AR(1) prior as described in Section 3.1. We first use 20 time periods as our observed temporal window. The parameters b_t and ξ_t, $t = 1, \ldots, 20$, are considered latent variables and are generated by the stationary AR(1) model, where $b_1 = 0.7$ and $\xi_1 = 2.5$ are fixed at $t = 1$ for this simulation study. The point pattern in Figure 1a (Supplementary Material) is used as the initial point pattern for all the realizations of this process. In this simulation, we generate 10 realizations from the AR(1) model with parameters $\phi_1 = 0.25$, $\phi_2 = 0.5$, $\mu_1 = 0.5$, $\mu_2 = 3$, $\sigma_b^2 = 0.05$, and $\sigma_\xi^2 = 0.1$. The hyperparameters are $u_b = 0.05$, $v_b = 0.05$, $u_\xi = 0.05$, $v_\xi = 0.05$, $u_{b_1} = 0.05$, $v_{b_1} = 0.05$,

Table 2
The parameter estimation results for 50 realizations from the STAI process with 5 time periods.

Parameter	True value	Average (SD) mean	Average (SD) median	Average (SD) SD
b_1	$\log(0.8) = -0.2231$	-0.2234(0.38)	-0.2142(0.37)	0.3917(0.03)
ξ_1	$\log(50) = 3.912$	3.9174(0.42)	3.9083(0.42)	0.4401(0.03)
b_2	$\log(1.2) = 0.1823$	0.2141(0.38)	0.2144(0.37)	0.3705(0.04)
ξ_2	$\log(30) = 3.4012$	3.3663(0.41)	3.3638(0.41)	0.4147(0.04)
b_3	$\log(2) = 0.6931$	0.7293(0.51)	0.7242(0.51)	0.4884(0.06)
ξ_3	$\log(20) = 2.9957$	2.9655(0.54)	2.9709(0.54)	0.5140(0.06)
b_4	$\log(0.9) = -0.1054$	-0.1255(0.39)	-0.1196(0.39)	0.4101(0.06)
ξ_4	$\log(40) = 3.6889$	3.7093(0.44)	3.7036(0.44)	0.4485(0.06)
b_5	$\log(3) = 1.0986$	1.2675(0.39)	1.2574(0.38)	0.4202(0.06)
ξ_5	$\log(10) = 2.3026$	2.1147(0.43)	2.1237(0.42)	0.4608(0.06)

Table 3
The average estimates of the model parameters for 10 realizations from the STAI process with 20 time period based on DMH.

Parameter	True value	Average (SD) mean	Average (SD) median	Average (SD) SD
ϕ_1	0.25	0.2214(0.10)	0.2221(0.10)	0.1122(0.02)
ϕ_2	0.5	0.2389(0.15)	0.2470(0.15)	0.2568(0.03)
μ_1	0.5	0.5393(0.14)	0.5317(0.13)	0.1570(0.04)
μ_2	3	2.9161(0.18)	2.9434(0.16)	0.4026(0.20)
σ_b^2	0.05	0.0950(0.03)	0.0808(0.03)	0.0593(0.02)
σ_ξ^2	0.1	0.1329(0.05)	0.1137(0.05)	0.0850(0.04)

Table 4
The parameter estimation results for patterns from the STAI process with 100 time period based on DMH.

Parameter	True value	Mean	Median	95% CI
ϕ_1	0.25	0.1465	0.1497	(-0.0998, 0.3702)
ϕ_2	0.5	0.3558	0.3596	(0.1824, 0.5049)
μ_1	0.5	0.4590	0.4578	(0.3591, 0.5557)
μ_2	3	3.0820	3.0806	(2.9472, 3.2267)
σ_b^2	0.05	0.0465	0.0444	(0.0230, 0.0802)
σ_ξ^2	0.1	0.1095	0.1078	(0.0665, 0.1612)

$u_{\xi_1} = 0.05$, $v_{\xi_1} = 0.05$, $\tau_1 = 1$, $\tau_2 = 5$ $\tau_{b_1} = 1$, and $\tau_{\xi_1} = 5$. In Table 3, we show the average and standard deviation of the posterior mean, median, and standard deviation of the parameters in the AR(1) model mentioned above. As we can see the DMH within Gibbs sampler can recover the true values of the parameters except for ϕ_2 and σ_b^2, due to the lack of information for the temporal process. In order to improve the estimate of the autoregressive coefficient, we extend the time period, which increases the sample size for the time series model.

We increase the number of time periods T to 100 and repeat the process introduced above. Since calculating the informative prior for the autoregressive coefficient is computationally inefficient for large T, we use a uniform prior for both ϕ_1 and ϕ_2. We use the same parameter values for the AR(1) process and same initial point pattern as previously mentioned in this section. The range of the number of points over 100 time periods is between 24 to 1340. The estimation results for one realization of the STAI process are shown in Table 4 and the generated latent variable along with the posterior mean and 95% point-wise credible interval is shown in Fig. 1. The point estimates indicate that the autoregressive coefficients are still underestimated, although the estimates of σ_b^2 and σ_ξ^2 are significantly improved. Since the observed point patterns are used for each time point to generate the auxiliary point pattern in the DHM, we do not expect that the estimation accuracy of b_t and ξ_t will decrease when T increases. However, estimating the STAI process with large T may require significant computational resources, such as computation time. From Fig. 1, we see that the estimates of b_t and ξ_t over 100 time periods are performing well. The estimates also indicate that the proposed method can provide reasonable results for point patterns with different number of points.

In order to assess the goodness-of-fit of the proposed model, we calculate the Bayesian p-value for the point pattern at each time period and the average p-value over 100 time periods is 0.55. The average p-value indicates that the proposed model is appropriate to describe the observed point pattern. To calculate the Bayesian p-value, we use the integrated difference between estimates of Ripley's K-function for the auxiliary point pattern and the observed point pattern as the discrepancy measure and 8000 posterior realizations are used to generate the auxiliary point patterns. The current goodness-of-fit assessment focuses on the comparison at each time period and provides limited information. A more ef-

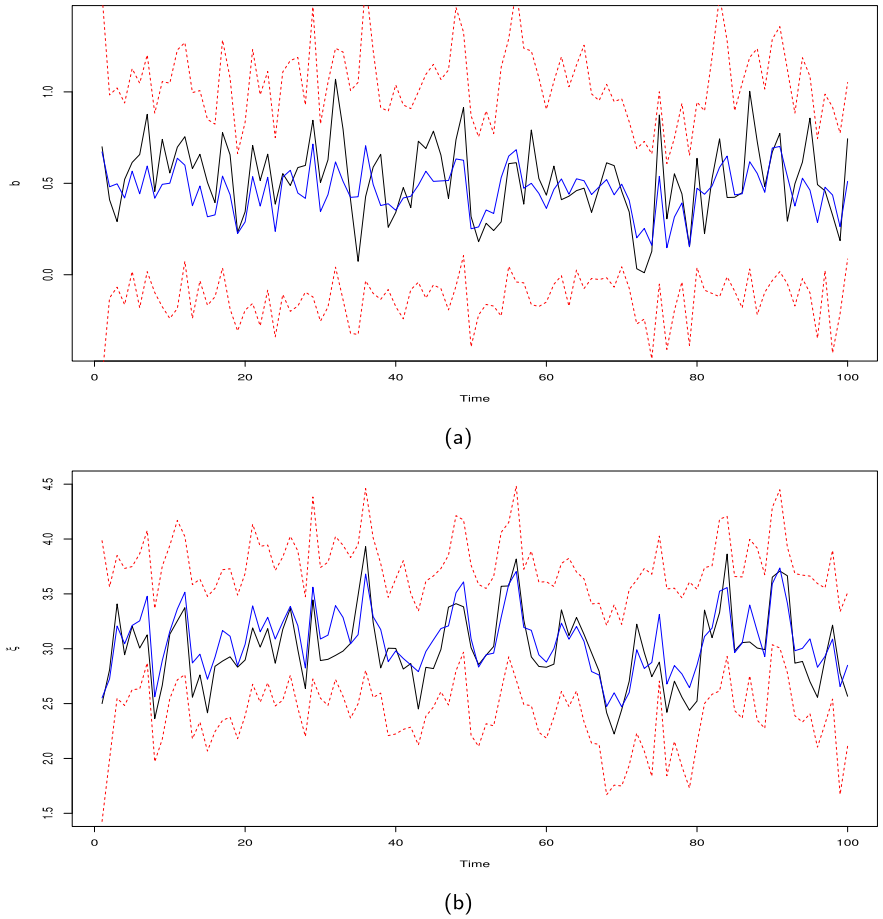

Fig. 1. True value, posterior mean and 95% CI of the latent variable b_t and ξ_t, $t = 1, \ldots, 100$. True value: black solid line; Posterior mean: blue solid line; 95% CI: red dashed lines. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

fective goodness-of-fit test based on evaluating similarities between spatio-temporal point processes is subject to future study.

5.4. Convergence diagnostics

To perform convergence diagnostics, we randomly selected several posterior sample chains from the previous sections and calculated the Monte Carlo standard error (MCSE) proposed by Flegal et al. (2008). Compared with the corresponding posterior means, the MCSE for the selected chains are sufficiently small, i.e., all MCSE are less than 0.1. Thus, the comparison results do not suggest any lack of convergence in the sample chains.

5.5. STAI process forecasting

We also evaluate the forecasting efficiency of the proposed STAI process by comparing the true point pattern and simulated point patterns of future times. Based on the true evolution process over 100 time periods that was considered in Section 5.3, we generate the true parameters $b_{101} = 0.421$ and $\xi_{101} = 2.842$ and the corresponding point pattern at $t = 101$. Then, we calculate the predicted parameters $\tilde{b}_{101}$ and $\tilde{\xi}_{101}$ using the posterior mean estimates of the parameters in the AR(1) processes (see Table 4). Finally, 100 corresponding forecast realizations are generated from the STAI process using the predicted parameter values. In order to compare the observed (true) point pattern with the forecasts, we calculate the approximated Papangelou conditional intensity surface for the observed point pattern in Fig. 2a and the average Papangelou surface for the forecast point patterns in Fig. 2b. The surface of the Papangelou conditional intensity can uniquely determine a point process and it can be interpreted as the intensity surface for the next event given the current events. However, the forecast point patterns are usually generated within a certain area that are affected by the previous point pattern and are highly different than each other. The Papangelou surfaces, calculated based on forecast realizations, will also show significant variability. Thus, the averaged Papangelou surface will be smoother and have a significantly larger high-intensity area than individual Papangelou surfaces.

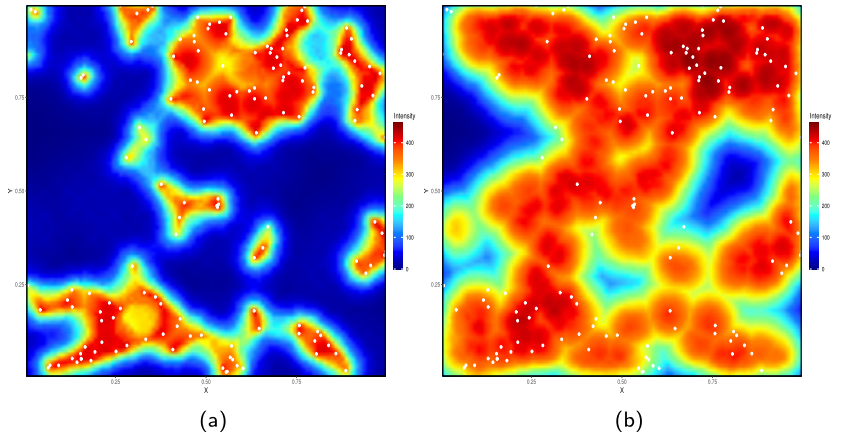

Fig. 2. a: The approximated Papangelou conditional intensity surface for the observed point pattern. b: The average Papangelou surface based on 100 forecast point patterns. The white dots represent the observed point pattern at $t = 101$.

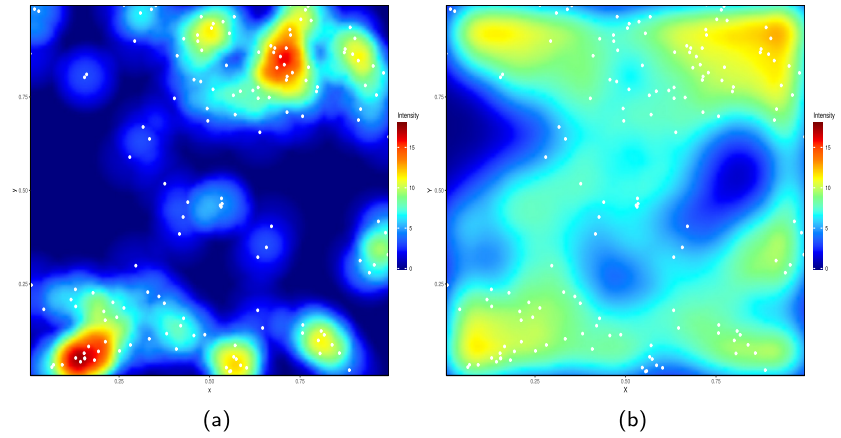

Fig. 3. a: The nonparametric estimation of the intensity surface for the observed pattern based on the Epanechnikov kernel with bandwidth 0.1. b: The average nonparametric estimate of the intensity surfaces based on the Epanechnikov kernel with bandwidth 0.1 for 100 forecast point patterns. The white dots represent the observed point pattern at $t = 101$.

We also compare the nonparametric estimate of the intensity surface for the observed pattern based on the Epanechnikov kernel with bandwidth 0.1, shown in Fig. 3a, and the average of the same nonparametric estimate surface for the forecast patterns, shown in Fig. 3b. Notice that this nonparametric estimate of the surface does not contain information about the specified model. It is only used to evaluate the similarity between point patterns. The average Papangelou surface of the predicted point patterns exhibits some deviation from the surface of the observed point pattern. However, most of the observed points are covered by the high intensity areas in the average Papangelou surface. This result indicates that, on average, the forecast point patterns can provide a reasonable prediction of the true point pattern.

The nonparametric estimates of the intensity surfaces show similar features as the Papangelou surfaces. The average surface of the predicted point patterns shows the same hotspots on the bottom left and top right corners, along with some extra high intensity spots at the other corners of the surface. In addition, the average surface is smoother, since it is the average of 100 predicted surfaces. Overall, the comparisons based on the Papangelou surfaces and intensity surfaces show that the proposed model can provide a reasonable forecast for future events from this the process.

For the uncertainty measure of the predictive surfaces, we present the 2.5 and 97.5 pixel-wise percentile surfaces of the Papangelou intensity and the nonparametric estimate intensity for the forecast realizations in Figs. 4 and 5. The 2.5 and 97.5 pixel-wise percentile surfaces of the Papangelou show that the points of the observed pattern are located in the high intensity area, which is similar to the average Papangelou surface shown in Fig. 2b. This result confirms that the posterior predicted realizations can provide reasonable forecasts of the future events. However, due to the dynamic mechanism of this process, the variation of the predictive surfaces is large.

Based on our simulations, we observed a high variability of this point process in terms of the number and locations of points and strong correlation between the previous point pattern and the current point pattern. Owing to these features,

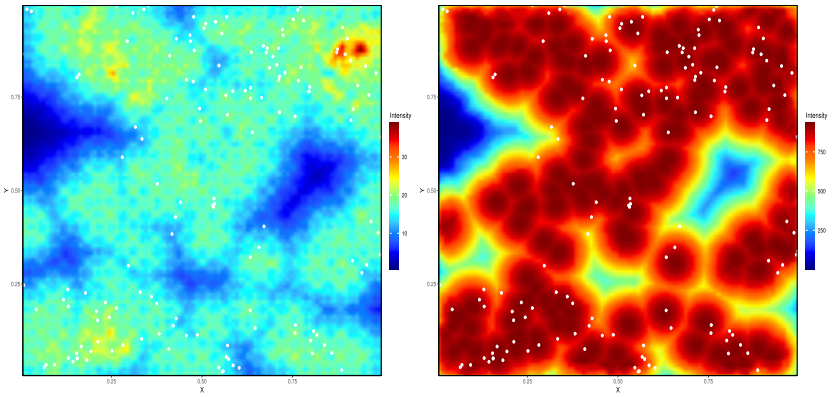

Fig. 4. For the simulated example with 100 time periods: a: 2.5 pixel-wise percentile of the Papangelou surface based on 100 forecast point patterns; b: 97.5 pixel-wise percentile of the Papangelou surface based on 100 forecast point patterns. The white dots represent the observed point pattern at $t = 101$.

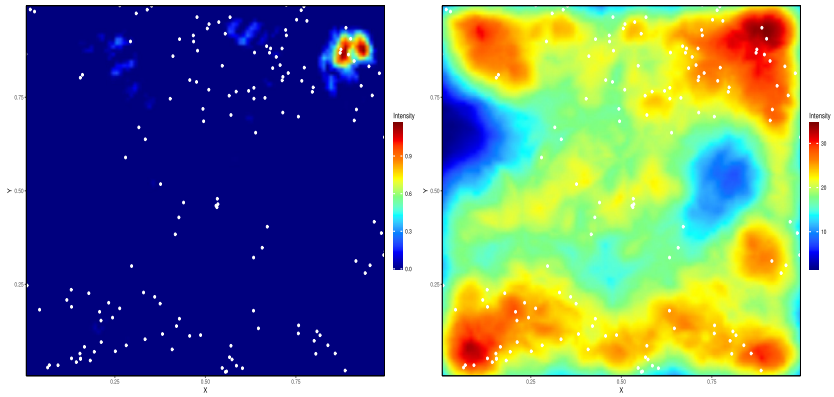

Fig. 5. For the simulated example with 100 time periods: a: 2.5 pixel-wise percentile of the nonparametric estimate surface based on 100 forecast point patterns; b: 97.5 pixel-wise percentile of the nonparametric estimate surface based on 100 forecast point patterns. The white dots represent the observed point pattern at $t = 101$.

the predicted point patterns rapidly deviate from the true point patterns as the predicted time period increases. Thus, it is very difficult to make multi-step ahead predictions and, consequently, we recommend making predictions for only one time point ahead.

6. United States natural caused wildfire data

We apply the proposed model to natural caused wildfire data in the western United States from 2002 to 2019 available the GEOMAC wildland fire support (https://www.geomac.gov/viewer/viewer.shtml). The data contain the reported locations and time stamps of the wildfires. We aggregate the events by years and rescale the data to the unit window. The observed point patterns from 2002 to 2019 are illustrated in the supplementary material (Figures 9-12). Additionally, we treat the wildfire locations in 2002 as the initial point pattern $\mathbf{X}_0$. Next, we fit the wildfire data from 2003 to 2018 to a STAI process over 16 time periods. The data in 2019, shown in Fig. 6, is used as the "truth" to compare with the forecasts from the fitted model. In Fig. 6, we show the polygon that is used as a plausible window for this point process, which approximates the shape of west coast.

We first implement the Albers equal area projection with standard parallels 30° N and 50° N to the wildfire data. Based on a visual inspection, the point patterns over 17 year show similar shapes. Thus, the parameter Σ_g in the main effect function is set to be $0.005I_2$ in order to preserve the similar features over time. The interaction distance of the area-interaction process is set to $r_h = 0.025$ based on sensitivity analysis. Similar to the simulation study, we also assume that the parameters of the area-interaction process are generated by some underlying AR(1) processes. Since the data are aggregated by year, we do not expect significant direction effects. However, an anisotropic kernel can be used for data aggregated based on shorter time intervals in order to incorporate the dominant directions effect. We implement the DMH within Gibbs sampler algorithm to obtain the 10,000 posterior samples of the model parameters with 2000 iterations discarded for burn-in. The estimation results of the AR(1) processes are shown in Table 5. The posterior means and 95%

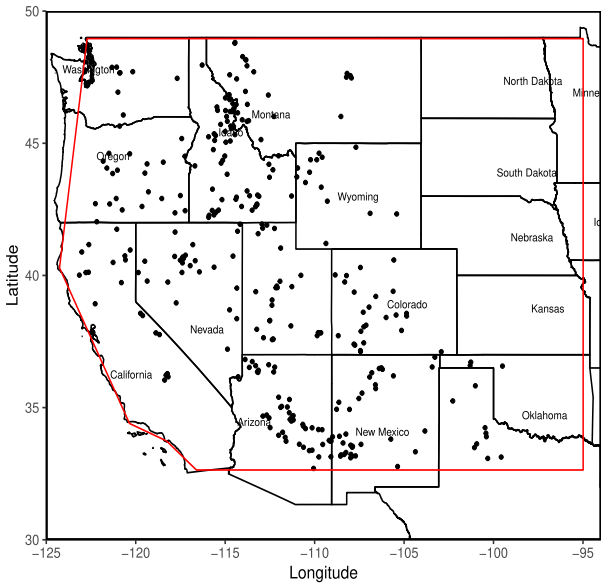

Fig. 6. The observed natural caused wildfire locations in 2019. The red lines are the boundaries of the polygon that is used as the observation window, which covers the west United States.

Table 5
The parameter estimation results, using the Albers equal area projection, for the United States natural caused wildfire data from 2003 to 2018 based on DMH with $\Sigma = 0.005 I_2$ and $r_h = 0.025$.

Parameter	Mean	Median	95% CI
ϕ_1	0.2284	0.2356	(-0.4061, 0.8924)
ϕ_2	0.3358	0.3561	(-0.3574, 0.9995)
μ_1	0.3438	0.3586	(0.0287, 0.5794)
μ_2	3.6755	3.9098	(-0.3503, 4.2629)
σ_b^2	0.0338	0.0281	(0.0101, 0.0932)
σ_y^2	0.0721	0.0569	(0.0153, 0.2257)

point-wise credible intervals of the parameters for the area-interaction process are shown in Fig. 7. Notice that the posterior means of ξ_t are all larger than 0, which indicates that the locations of wildfire are significantly clustered for every year.

Using the estimated parameters, we generate 100 predicted parameters $\tilde{b}_{2019}$ and $\tilde{y}_{2019}$ for the point process in 2019. Then, we generate 100 forecast point patterns based on the predicted parameters and calculate the average of the Papangelou surfaces and nonparametric estimate intensity surfaces for these realizations. These surfaces were introduced in Section 5.5. The average surface of the Papangelou conditional intensity along with the observed point pattern in 2019 is shown in Fig. 8a and the average nonparametric estimates of the intensity surfaces with the true pattern are shown in Fig. 9a. The Papangelou and intensity surfaces show that the true point pattern can be covered by high intensity area based on the fitted model, which indicates the proposed model can provide reasonable forecasts for possible wildfire locations. More importantly, this information can help wildland fire management officers to efficiently allocate resources to monitor the natural environment in order to prevent wildfires.

Moreover, the 2.5 and 97.5 pixel-wise percentile surfaces of the Papangelou and nonparametric estimate intensity for the forecast point patterns are shown in the Figs. 8b, 8c, 9b, and 9c. The 2.5 and 97.5 pixel-wise percentile surfaces of the Papangelou show that most of the wildfires in 2019 are located in the high intensity area, which confirm that the posterior predicted realizations can provide reasonable forecasts of the future wildfire locations. Due to the dynamic mechanism of this model, the variation of the forecasts is large. More importantly, this information can help wildland fire management officers to efficiently allocate resources to monitor the natural environment in order to prevent wildfires.

7. Concluding remarks

In this paper, we propose a flexible spatio-temporal point process model that can be used to describe spatial and temporal interactions between points. Specifically, we use a Markovian structure to model the dependence of point patterns

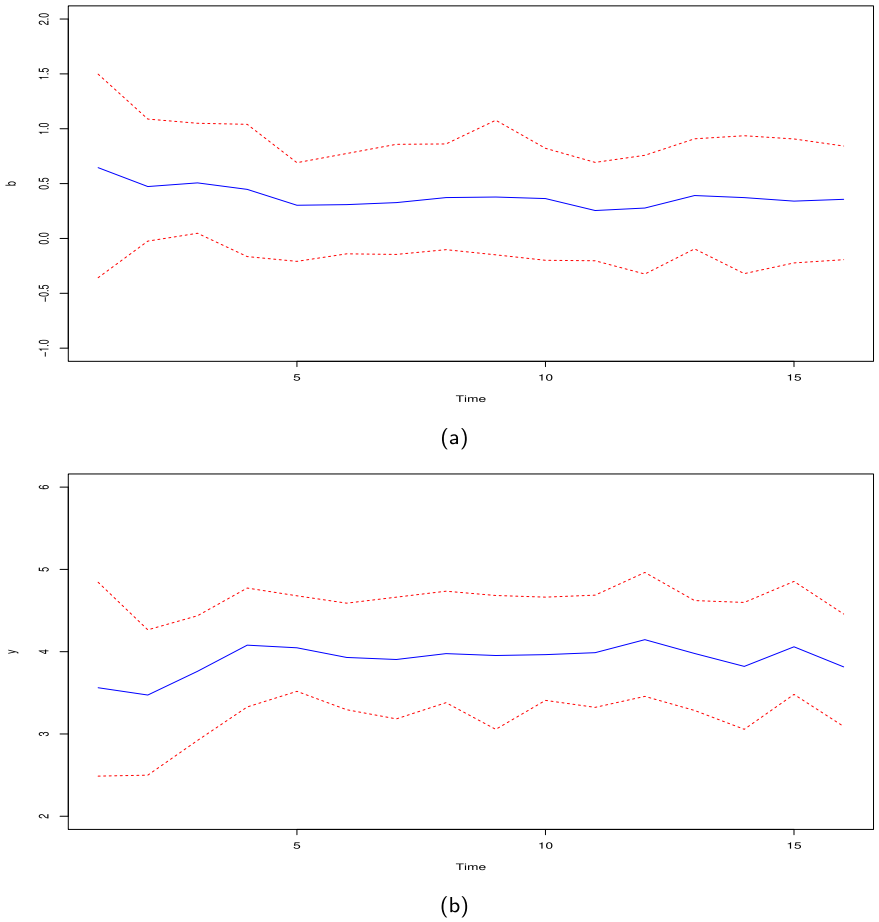

Fig. 7. Posterior mean and 95% CI of the latent variables b_t and ξ_t, $t = 1, \ldots, 16$ for the United States natural caused wildfire data from 2003 to 2018 using the Albers equal area projection. Posterior mean: blue solid line; 95% CI: red dashed lines.

over consecutive time periods, which assumes that the current main effect function is only affected by the point pattern in the previous time period. This structure provides the opportunity to predict the point pattern in the following time period without the use of covariate information. Unlike continuous time processes, the prediction from our model does not require generating point patterns across the whole temporal window, and is, thus, more computationally efficient.

For the parameter structure, a time indexed parameter model and a hierarchical model are evaluated in our study. Based on simulations, we show that the DMH sampler can provide good estimates for the time indexed parameters and the latent parameters in the hierarchical model. More importantly, we also demonstrate that the parameters of the autoregressive process model can be recovered by the Bayesian estimation method. These results indicate that it is possible to discover the underlying structure of a spatio-temporal point process and predict future point patterns by evolving the latent parameters. However, our results show that it is difficult to recover the autoregressive structures due to the deviations between the true latent parameter values and their estimates, which is introduced by the sampling algorithm for the parameters of the data model. Based on the evolution process, we also provide an example of predicting a future point pattern (via a simulation study). By comparing the surfaces of Papangelou conditional intensity and nonparametric estimate of the intensity for the true point pattern and predicted patterns, we evaluate the predictions made by the proposed model. The results indicate that our model can provide reasonable predictions for future events.

For the interaction structure, we use the area-interaction function to describe the interaction level of the points at the current time. Since the area-interaction process is suitable for repulsive, independent, and clustered events, our model can be used as a general approach for modeling any point pattern under consideration. Additionally, the interaction function can be easily replaced by other functions; e.g., if there is strong information to indicate a different interaction function, such as the Strauss process, say, for forestry data.

The United States natural caused wildfire data show strong dependence between the point patterns at two consecutive years, which is suitable for the proposed model. Based on the estimated parameters, we provide forecasts for future events and compared these forecasts with the observed data. The results indicate that the forecasts provide reasonable estimates

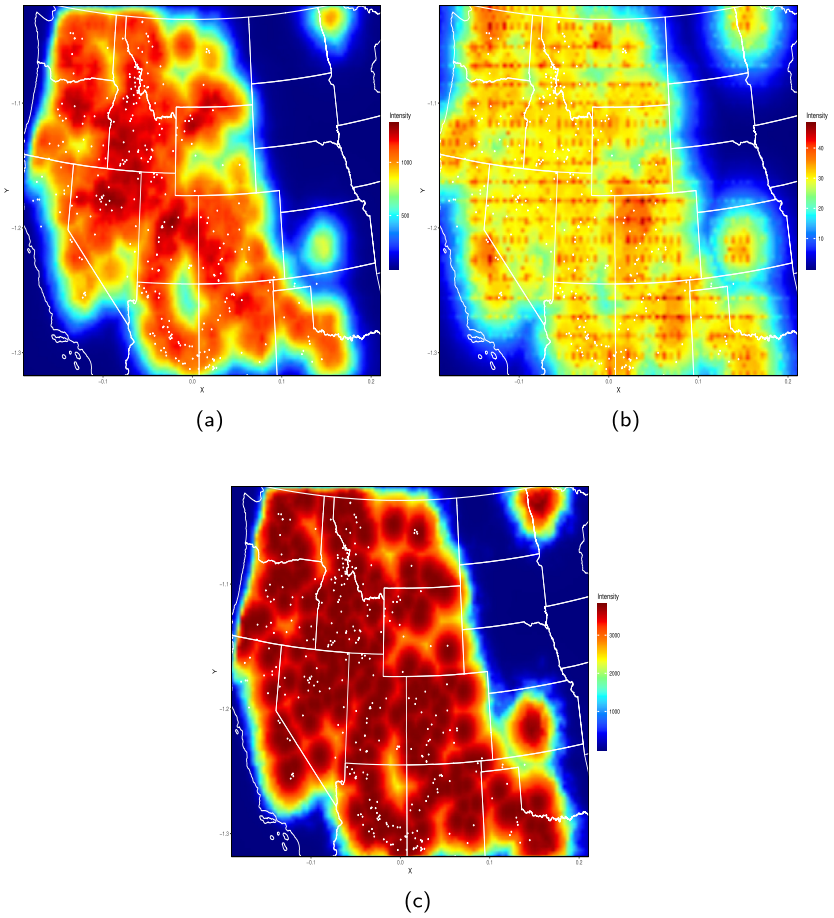

Fig. 8. The STAI model forecast for wildfire locations in 2019 using the Albers equal area projection with $\Sigma_g = 0.005I_2$ and $r_h = 0.025$. a: The average Papangelou surface for 100 forecast point patterns. b: The 2.5 pixel-wise percentile of the Papangelou surface with edge correction for 100 forecast point patterns; c: The 97.5 pixel-wise percentile of the Papangelou surface with edge correction for 100 forecast point patterns. The white dots represent the observed natural caused wildfire locations in 2019.

for future events. Thus, we believe the proposed model can be very useful in terms of describing the process and making forecasts. Moreover, we believe that the forecast intensity and Papangelou conditional intensity surfaces can provide information about high risk areas for wildfires. More importantly, the strategies of managing wildfires can be improved based on the obtained information.

In principle, a higher order autoregressive model could be implemented; e.g., an AR(2) (for the United States natural caused wildfire data). However, in practice, significant estimation challenges arise. The main issue manifests in trying to ensure that the estimated parameters result in a stationary process. For the AR(1) model this is straightforward, as there is only one constraint on the model coefficients that needs to be satisfied. In contrast, with higher-order autoregressive models the number of constraints increases. Consequently, this either leads to a less efficient sampling algorithm or explosive behavior (the case where stationarity is not met). This is a subject of future research.

Overall, the proposed model provides a general framework for evaluating spatio-temporal point processes. In future work, a more sophisticated parameter evolution structure will be investigated in order to capture more complex features of the point processes. Moreover, extra covariate information, such as temperature and precipitation level can be included in the evolution model to improve the prediction. Additionally, it is possible to organize the Bayesian estimation algorithm for each time period in a parallel structure to improve the computational efficiency.

Acknowledgement

This work was done by Jiaxun Chen as part of his Ph.D dissertation prior to joining Eli Lilly and Company. This research was partially supported by the U.S. National Science Foundation (NSF) under NSF SES-1853096 and through the Air Force Research Laboratory (AFRL) Contract No. 19C0067.

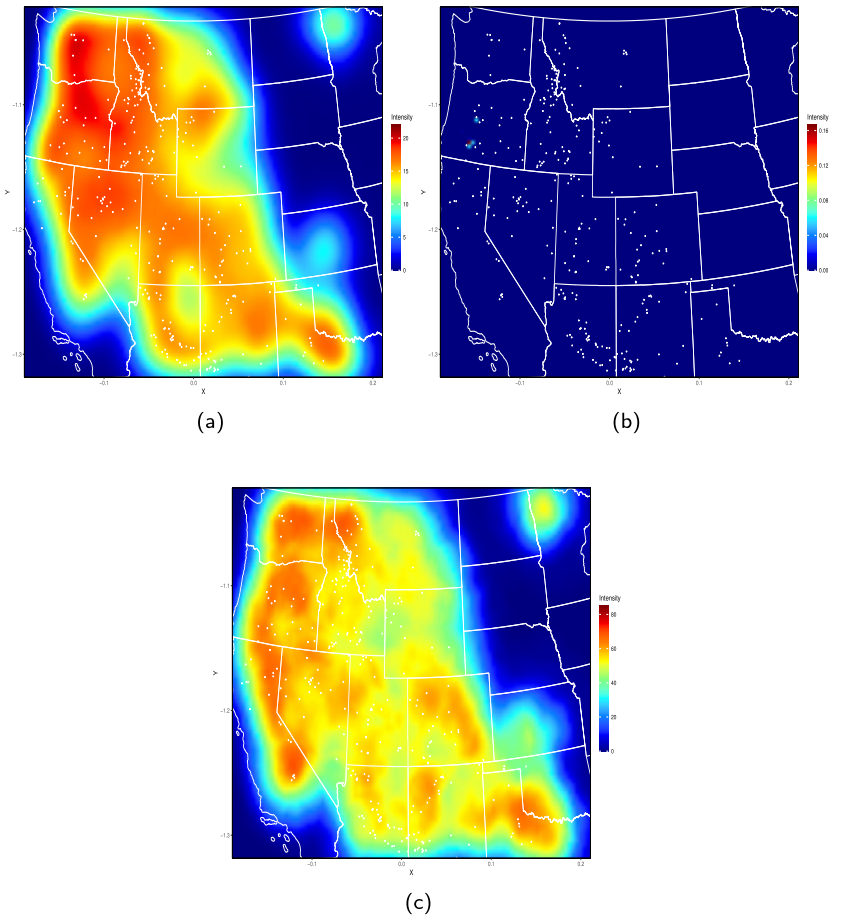

Fig. 9. The STAI model forecast for wildfire locations in 2019 using the Albers equal area projection with $\Sigma_g = 0.005 I_2$ and $r_h = 0.025$. a: The average nonparametric estimate of the intensity surfaces based on the Epanechnikov kernel with bandwidth 0.075 for 100 forecast point patterns; b: The 2.5 pixel-wise percentile of the nonparametric estimate of the intensity surface for 100 forecast point patterns; c: The 97.5 pixel-wise percentile of the nonparametric estimate of the intensity surface for 100 forecast point patterns. The white dots represent the observed natural caused wildfire locations in 2019.

Appendix

The full conditional distributions for the evolution STAI process are shown as follows

$$\pi(b_1|\cdot) \propto \frac{1}{C(b_1,\xi_1)} \exp\left\{-\frac{1}{2}\frac{\sigma_b^2+\phi_1^2\sigma_{b_1}^2}{\sigma_{b_1}^2\sigma_b^2}\left(b_1 - \frac{n_1\sigma_{b_1}^2\sigma_b^2 + \sigma_b^2\mu_{b_1} + \phi_1^2\mu_1\sigma_{b_1}^2 + \sigma_{b_1}^2\phi_1(b_2-\mu_1)}{\sigma_b^2+\phi_1^2\sigma_{b_1}^2}\right)\right\}.$$

$$\pi(b_t|\cdot) \propto \frac{1}{C(b_t,\xi_t)} \exp\left\{-\frac{1}{2}\frac{1+\phi_1^2}{\sigma_b^2}\left(b_t - \mu_1 - \frac{\phi_1(b_{t+1}+b_{t-1}-2\mu_1)+\sigma_b^2 n_t}{\phi_1^2+1}\right)\right\}, t = 2,\ldots,T.$$

$$\pi(\xi_1|\cdot) \propto \frac{1}{C(b_1,\xi_1)} \exp\left\{-\frac{1}{2}\frac{\sigma_y^2+\phi_2^2\sigma_{\xi_1}^2}{\sigma_{\xi_1}^2\sigma_y^2}\left(\xi_1 - \frac{\sigma_b^2\mu_{b_1} + \phi_1^2\mu_1\sigma_{b_1}^2 + \sigma_{b_1}^2\phi_1(b_2-\mu_1)}{\sigma_b^2+\phi_1^2\sigma_{b_1}^2}\right.\right.$$
$$\left.\left. - \frac{\{n_1 - m\{\cup_{i=1}^{n_1} B(x_{i,1},r_h)\}/(\pi r_h^2)\}\sigma_{b_1}^2\sigma_b^2 +}{\sigma_b^2+\phi_1^2\sigma_{b_1}^2}\right)\right\}.$$

$$\pi(\xi_t|\cdot) \propto \frac{1}{C(b_t,\xi_t)} \exp\left\{-\frac{1}{2}\frac{1+\phi_2^2}{\sigma_y^2}\right.$$
$$\left.\times\left(\xi_t - \mu_2 - \frac{\phi_2(\xi_{t+1}+\xi_{t-1}-2\mu_2)+\sigma_y^2\{n_t - m\{\cup_{i=1}^{n_t} B(x_{i,t},r_h)\}/(\pi r_h^2)\}}{\phi_2^2+1}\right)\right\},$$
$$t = 2,\ldots,T.$$

$$\phi_1|\cdot \sim N_{(-1,1)}\left(\frac{\sigma_{\phi_1}^2\sum_{t=2}^{T}(b_{t-1}-\mu_1)(b_t-\mu_1)+\hat{\phi}_1\sigma_b^2}{\sigma_{\phi_1}^2\sum_{t=2}^{T}(b_{t-1}-\mu_1)^2+\sigma_b^2}, \frac{\sigma_b^2\sigma_{\phi_1}^2}{\sigma_{\phi_1}^2\sum_{t=2}^{T}(b_{t-1}-\mu_1)^2+\sigma_b^2}\right).$$

$$\phi_2|\cdot \sim N_{(-1,1)}\left(\frac{\sum_{t=2}^{T}(\xi_{t-1}-\mu_2)(\xi_t-\mu_2)}{\sum_{t=2}^{T}(\xi_{t-1}-\mu_2)^2}, \frac{\sigma_y^2}{\sum_{t=2}^{T}(\xi_{t-1}-\mu_2)^2}\right).$$

$$\mu_{b_1}|\cdot \sim N\left(\frac{\tau_{b_1}b_1}{\sigma_{b_1}^2+\tau_{b_1}}, \frac{\sigma_b^2\tau_{b_1}}{\sigma_{b_1}^2+\tau_{b_1}}\right).$$

$$\mu_1|\cdot \sim N\left(\frac{\tau_1(1-\phi_1)}{(T-1)(1-\phi_1)^2\tau_1+\sigma_b^2}\left\{\sum_{t=2}^{T}(b_t-\phi_1 b_{t-1})\right\}, \frac{\sigma_b^2\tau_1}{(T-1)(1-\phi_1)^2\tau_1+\sigma_b^2}\right).$$

$$\mu_{\xi_1}|\cdot \sim N\left(\frac{\tau_{\xi_1}\xi_1}{\sigma_{\xi_1}^2+\tau_{\xi_1}}, \frac{\sigma_y^2\tau_{\xi_1}}{\sigma_{\xi_1}^2+\tau_{\xi_1}}\right).$$

$$\mu_2|\cdot \sim N\left(\frac{\tau_2(1-\phi_2)}{(T-1)(1-\phi_2)^2\tau_2+\sigma_y^2}\left\{\sum_{t=2}^{T}(\xi_t-\phi_2\xi_{t-1})\right\}, \frac{\sigma_y^2\tau_2}{(T-1)(1-\phi_2)^2\tau_2+\sigma_y^2}\right).$$

$$\sigma_{b_1}^2|\cdot \sim IG\left(\frac{1}{2}+u_{b_1}, \frac{1}{2}(b_1-\mu_{b_1})^2+v_{b_1}\right).$$

$$\sigma_b^2|\cdot \sim IG\left(\frac{T-1}{2}+u_b.\frac{1}{2}\sum_{t=2}^{T}((b_t-\mu_1)-\phi_1(b_{t-1}-\mu_1))^2+v_b\right).$$

$$\sigma_{\xi_1}^2|\cdot \sim IG\left(\frac{1}{2}+u_{\xi_1}, \frac{1}{2}(\xi_1-\mu_{\xi_1})^2+v_{\xi_1}\right).$$

$$\sigma_y^2|\cdot \sim IG\left(\frac{T-1}{2}+u_y, \frac{1}{2}\sum_{t=2}^{T}((\xi_t-\mu_2)-\phi_2(\xi_{t-1}-\mu_2))^2+v_y\right).$$

Appendix A. Supplementary material

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.csda.2021.107349.

References

Baddeley, A., Rubak, E., Turner, R., 2015. Spatial Point Patterns: Methodology and Applications with R. CRC Press.

Baddeley, A.J., Van Lieshout, M., 1995. Area-interaction point processes. Ann. Inst. Stat. Math. 47, 601–619.

Barndorff-Nielsen, O.E., Kendall, W.S., Van Lieshout, M., 1999. Stochastic Geometry, Likelihood and Computation. Chapman and Hall.

Brix, A., Diggle, P.J., 2001. Spatiotemporal prediction for log-Gaussian Cox processes. J. R. Stat. Soc., Ser. B, Stat. Methodol. 63, 823–841.

Chen, J., Micheas, A.C., Holan, S.H., 2020. A comparative study of approximate Bayesian computation methods for Gibbs point processes. J. Stat. Appl. 18, 223–248.

Chiu, S.N., Stoyan, D., Kendall, W.S., Mecke, J., 2013. Stochastic Geometry and Its Applications. John Wiley & Sons.

Cressie, N., 1993. Statistics for Spatial Data, 2nd edition. John Wiley & Sons.

Daley, D.J., Vere-Jones, D., 2003. An Introduction to the Theory of Point Processes, Volume 1: Elementary Theory and Methods. Springer Verlag, New York Berlin Heidelberg.

Daley, D.J., Vere-Jones, D., 2007. An Introduction to the Theory of Point Processes: Volume II: General Theory and Structure. Springer Science & Business Media.

Dereudre, D., 2019. Introduction to the theory of Gibbs point processes. In: Stochastic Geometry. Springer, pp. 181–229.

Diggle, P.J., 2013. Statistical Analysis of Spatial and Spatio-Temporal Point Patterns. CRC Press.

Flegal, J.M., Haran, M., Jones, G.L., 2008. Markov chain Monte Carlo: can we trust the third significant figure? Stat. Sci., 250–260.

Gabriel, E., Opitz, T., Bonneu, F., 2017. Detecting and modeling multi-scale space-time structures: the case of wildfire occurrences. J. Soc. Fr. Stat. 158, 86–105.

Gelfand, A.E., Diggle, P., Guttorp, P., Fuentes, M., 2010. Handbook of Spatial Statistics. CRC Press.

Geyer, C., 1998. Likelihood inference for spatial point. In: Stochastic Geometry: Likelihood and Computation, vol. 80, p. 79.

Hawkes, A.G., 1971. Spectra of some self-exciting and mutually exciting point processes. Biometrika 58, 83–90.

Iftimi, A., van Lieshout, M.C., Montes, F., 2018. A multi-scale area-interaction model for spatio-temporal point patterns. Spat. Stat. 26, 38–55.

Iftimi, A., Montes, F., Mateu, J., Ayyad, C., 2017. Measuring spatial inhomogeneity at different spatial scales using hybrids of Gibbs point process models. Stoch. Environ. Res. Risk Assess. 31, 1455–1469.

Illian, J., Penttinen, A., Stoyan, H., Stoyan, D., 2008. Statistical Analysis and Modelling of Spatial Point Patterns, vol. 70. John Wiley & Sons.

Illian, J.B., Sørbye, S.H., Rue, H., Hendrichsen, D.K., 2012. Using INLA to fit a complex point process model with temporally varying effects-a case study. J. Environ. Stat. 3.
Karr, A., 1991. Point Processes and Their Statistical Inference, vol. 7. CRC Press.
Kelly, F.P., Ripley, B.D., 1976. A note on Strauss's model for clustering. Biometrika, 357–360.
Lantuejoul, C., 2001. Geostatistical Simulation: Models and Algorithms, vol. 1139. Springer Science & Business Media.
Lavancier, F., Møller, J., 2016. Modelling aggregation on the large scale and regularity on the small scale in spatial point pattern datasets. Scand. J. Stat. 43, 587–609.
Lawson, A.B., Denison, D.G., 2002. Spatial Cluster Modelling. CRC Press.
Liang, F., 2010. A double Metropolis–Hastings sampler for spatial models with intractable normalizing constants. J. Stat. Comput. Simul. 80, 1007–1022.
Micheas, A.C., 2019. Cox point processes: why one realisation is not enough. Int. Stat. Rev. 87, 306–325.
Møller, J., 2013. Spatial Statistics and Computational Methods, vol. 173. Springer Science & Business Media.
Møller, J., Pettitt, A.N., Reeves, R., Berthelsen, K.K., 2006. An efficient Markov chain Monte Carlo method for distributions with intractable normalising constants. Biometrika 93, 451–458.
Møller, J., Waagepetersen, R.P., 2003. Statistical Inference and Simulation for Spatial Point Processes. CRC Press.
Murray, I., Ghahramani, Z., MacKay, D., 2012. MCMC for doubly-intractable distributions. arXiv preprint. arXiv:1206.6848.
Park, J., Haran, M., 2018. Bayesian inference in the presence of intractable normalizing functions. J. Am. Stat. Assoc. 113, 1372–1390.
Pei, T., Gao, J., Ma, T., Zhou, C., 2012. Multi-scale decomposition of point process data. GeoInformatica 16, 625–652.
Picard, N., Bar-Hen, A., Mortier, F., Chadœuf, J., 2009. The multi-scale marked area-interaction point process: a model for the spatial pattern of trees. Scand. J. Stat. 36, 23–41.
Raeisi, M., Bonneu, F., Gabriel, E., 2019. A spatio-temporal multi-scale model for Geyer saturation point process: application to forest fire occurrences. arXiv preprint. arXiv:1911.06999.
Rao, V., Adams, R.P., Dunson, D.D., 2017. Bayesian inference for Matérn repulsive processes. J. R. Stat. Soc., Ser. B, Stat. Methodol. 79, 877–897.
Ripley, B.D., 1987. Stochastic Simulation.
Siino, M., Adelfio, G., Mateu, J., Chiodi, M., D'Alessandro, A., 2017. Spatial pattern analysis using hybrid models: an application to the Hellenic seismicity. Stoch. Environ. Res. Risk Assess. 31, 1633–1648.
Siino, M., D'Alessandro, A., Adelfio, G., Scudero, S., Chiodi, M., 2018. Multiscale processes to describe the eastern Sicily seismic sequences. Ann. Geophys.
Sørbye, S.H., Illian, J.B., Simpson, D.P., Burslem, D., Rue, H., 2019. Careful prior specification avoids incautious inference for log-Gaussian Cox point processes. J. R. Stat. Soc., Ser. C, Appl. Stat. 68, 543–564.
Spodarev, E., 2013. Stochastic Geometry, Spatial Statistics and Random Fields: Asymptotic Methods, vol. 2068. Springer.
Strauss, D.J., 1975. A model for clustering. Biometrika 62, 467–475.
Van Lieshout, M., 2000. Markov Point Processes and Their Applications. World Scientific.
Wiegand, T., Gunatilleke, S., Gunatilleke, N., Okuda, T., 2007. Analyzing the spatial structure of a Sri Lankan tree species with multiple scales of clustering. Ecology 88, 3088–3102.
Zhou, Z., Matteson, D.S., Woodard, D.B., Henderson, S.G., Micheas, A.C., 2015. A spatio-temporal point process model for ambulance demand. J. Am. Stat. Assoc. 110, 6–15.
Zhuang, J., Ogata, Y., 2006. Properties of the probability distribution associated with the largest event in an earthquake cluster and their implications to foreshocks. Phys. Rev. E 73, 046134.