Statistics and Applications {ISSN 2452-7395 (online)}
Volume 18, No. 2, 2020 (New Series), pp 223-248

A Comparative Study of Approximate Bayesian Computation Methods for Gibbs Point Processes

Jiaxun Chen[1], **Athanasios C. Micheas**[1] **and Scott H. Holan**[1,2]
[1] *Department of Statistics, University of Missouri, 146 Middlebush Hall, Columbia, MO 65211-6100*
[2] *U.S. Census Bureau, 4600 Silver Hill Road, Washington, D.C. 20233-9100*

Received 11 April 2020; Revised 28 July 2020; Accepted 30 July 2020

Abstract

Crucial aspects of applying approximate Bayesian computation (ABC) for Gibbs point processes are the choice of summary statistic and method of constructing the discrepancy measure. In this paper, we present a comparative study of ABC for Gibbs point processes based on various summary statistics and different approaches of constructing the discrepancy measure. We also demonstrate the issue of identifiability of the parameter values for Gibbs point processes and provide a solution for parameter estimation. We further propose robust choices for the discrepancy measures for different point processes through an intensive simulation study. The ABC algorithm, with all of the tested discrepancy measures, is also applied to the Swedish pines data and Chicago crime data to illustrate the feasibility of the proposed approaches.

Key words: Approximate Bayesian computation; Discrepancy measure; Functional characteristic; Gibbs point processes; Informative prior.

1. Introduction

As an important class of spatial point processes, Gibbs point processes have been intensively studied over the past few decades. Since Gibbs point processes can take into account interactions between event locations, they have become reasonable choices for describing this kind of phenomena, *i.e.*, the inhibition and clustering behaviors in point processes. Here, we refer to point processes with inhibition as repulsive spatial point processes and to point processes with clustering as attractive spatial point processes. In the literature, the most well known models for repulsive point processes are the Strauss process (Strauss, 1975) and hardcore process (Ripley, 1981). For attractive point processes, a model that can be used to describe this behavior is the area-interaction process, proposed by Baddeley and Van Lieshout (1995).

Although Gibbs point processes meet the need for describing the underlying process in different fields of science, conducting inference directly using the likelihood of the point process is challenging, and standard Bayesian analysis based on Markov chain Monte Carlo (MCMC) is not feasible due to the intractable normalizing constant. Maximum pseudo-likelihood estimation (MPLE) (Besag, 1975) can be applied to spatial point processes. However, Huang and Ogata (1999) show that the performance of MPLE is

Corresponding Author: Scott H. Holan
E-mail: holans@missouri.edu

poor for repulsive point processes with strong interaction. There are also several estimation methods that have been developed in a Bayesian framework. Geyer (1991) proposed the Markov chain Monte Carlo maximum likelihood estimate (MCMCMLE) for stochastic processes with an intractable normalizing constant. One of the issues with MCMCMLE is that the required computational effort within each MCMC iteration is intensive for Gibbs point processes. Moreover, the estimation result is sensitive to the auxiliary parameters. By implementing an auxiliary variable, Møller *et al.* (2006) proposed an auxiliary variable Markov chain Monte Carlo (AVM) and an exchange algorithm approach, which can cancel out the intractable normalizing constant in the acceptance probability ratio of the Metropolis-Hastings algorithm. Extending the exchange algorithm through "bridging" was proposed by Murray *et al.* (2012).

In addition, the double Metropolis-Hastings (DMH) and an adaptive exchange (AEX) algorithm were proposed by Liang (2010) and Liang *et al.* (2016). Alternatively, approaches based on likelihood approximation were proposed by Beaumont (2003), Andrieu *et al.* (2009), Alquier *et al.* (2016), and Atchadé *et al.* (2013). A comparison of these methods was provided by Park and Haran (2018). The overall conclusion from this comparison recommends that researchers start with DMH since it is computationally efficient and does not require perfect sampling. However, the convergence of DMH cannot be guaranteed, and the auxiliary variable approaches become computationally expensive when the inner sampler for the auxiliary variable is expensive. Recently, Park and Haran (2018) focused on comparing likelihood based sampling methods and provided practical recommendations. In contrast, our study provides evaluations of a likelihood free algorithm and its applications (see details in Section 3).

As an alternative to the approaches based on classic MCMC, Stoica *et al.* (2017) and Shirota and Gelfand (2017) proposed an approximate Bayesian computation (ABC) algorithm for Gibbs point processes. The ABC algorithm provides a likelihood free approach for approximating the posterior distribution and it is straightforward to implement. In Shirota and Gelfand (2017), the simulation results for the Strauss model and the determinantal point process show that the true models can be recovered. Also, the authors compare the ABC with the exchange algorithm of Murray *et al.* (2012) and point out that the ABC is more efficient in terms of the parameter inefficiency factors (IF).

In this study, we propose a more robust approach to constructing discrepancy measures for the ABC and compare the performance of available summary statistics for both repulsive and attractive point processes. Implementing the ABC algorithm for fitting Gibbs point processes has been proven feasible and efficient. However, we still need to thoroughly investigate how to construct the best discrepancy measure for different situations. Inspired by the discussion surrounding the choices of various distance metrics between two point process realizations (Mateu *et al.*, 2010), we use the integrated distance of the available functional summary characteristics to measure the similarity of the interactive structure for two point patterns. This measure takes into account the information of the interactions among all the reasonable scales, which can be more informative compared with the measures only calculated at a given interaction distance.

An important component of the ABC algorithms for Gibbs point processes is the choice of summary statistic. In Stoica *et al.* (2017), the authors discuss the issue, but did

not conduct a comparative study, especially when the sufficient statistics are not available. In our study, we implement and compare the performance of Ripley's K-function, nearest neighbor distance distribution function (D-function), empty-space function (F-function), and J-function (Van Lieshout and Baddeley, 1996) when used in the ABC algorithm, in terms of recovering the parameters of a Gibbs point process. The final recommendation on the most robust summary statistic among the ones tested is formed based on an extensive simulation study. Note that Geyer (1998) proposed to use stochastic approximation as rough estimates of the parameter values. Thus, we implement this estimate as an informative prior and compare it with a vague prior for the ABC algorithms presented herein.

The paper proceeds as follows. Section 2 begins by reviewing the repulsive and attractive Gibbs point processes that we investigate. Section 3 provides the functional summary statistics considered for Gibbs point processes, the proposed structure of the discrepancy measure for the ABC algorithm, and the choices of prior distributions. Sections 4 and 5 include the results of the simulation study, and Section 6 presents applications to the Swedish pines data and Chicago crime data. Finally, in Section 7, we provide concluding remarks.

2. Review of Classic Gibbs Point Processes

The general density function of a Gibbs point process for a finite point pattern $\varphi_n = \{x_1, \ldots, x_n\}$ over a bounded domain $B \in \mathbb{R}^2$ (Cressie, 1993) is given by

$$f(\varphi_n) = c^{-1} \exp\left\{ \sum_{i-1}^{n} g_1(x_i) + \sum_{1 \le i < j \le n} g_{1,2}(x_i, x_j) + \cdots + g_{1,\ldots,n}(x_1, \ldots, x_n) \right\} \tag{1}$$

with respect to the Poisson process with unit intensity, where the function $g_{1,\ldots,k}(\cdot)$ describes the k-level interaction. Hence, c is a normalizing constant and the form of c cannot be provided analytically. In other words, the normalizing constant for the Gibbs point process is typically intractable. Another fundamental functional is the Papangelou conditional intensity,

$$\lambda(u|\varphi_n, \boldsymbol{\theta}) = \frac{f(\varphi_n \cup u|\boldsymbol{\theta})}{f(\varphi_n|\boldsymbol{\theta})},$$

which is the intensity at location u given the point pattern φ_n. Note that the normalizing constant in the likelihood function cancels out when calculating the Papangelou conditional intensity. Thus, it is used to construct the log pseudo-likelihood as follows

$$\log PL(\varphi_n|\boldsymbol{\theta}) = -\int_B \lambda(u|\varphi_n, \boldsymbol{\theta}) du + \sum_{i}^{n} \log\lambda(x_i|\varphi_n, \boldsymbol{\theta}), \tag{2}$$

and the MPLE is obtained by maximizing the latter equation.

In this paper, we consider point processes that are suitable for describing inhibition and attraction. Thus, two classes of Gibbs point processes are investigated: the homogeneous Strauss process and the area-interaction point process.

2.1. Strauss point process

The Strauss process was introduced by Strauss (1975) to model point processes with pairwise interactions; however, Kelly and Ripley (1976) showed that the Strauss process density is only integrable for repulsive interactions. Hence, it is extensively used to model point processes under the regularity assumption. The density function of the homogeneous Strauss model for $\varphi_n = \{x_1, \ldots, x_n\}$ over a bounded domain $B \in \mathbb{R}^2$ is given by

$$f(\varphi_n) = c(\beta, \gamma)^{-1} \beta^n \gamma^{S_r(\varphi_n)}, \varphi_n \in B^n, \tag{3}$$
$$S_r(\varphi_n) = \sum_{1 \le i < j \le n} \mathbb{I}(||x_i - x_j|| \le r),$$

where $\beta > 0$, $\gamma \in [0, 1]$, $c(\beta, \gamma)$ is the normalizing constant, r is the interaction distance and $S_r(\varphi_n)$ is the number of pairs of points that are closer than a distance r. In the Strauss process, β and γ represent the main effect and the interaction effect, respectively. For $\gamma \in (0, 1)$, the point process shows inhibition between points and smaller γ leads to stronger interaction. For $\gamma = 0$, the process is known as the hardcore process which does not allow points to be closer than distance r. For $\gamma = 1$, the process is equivalent to the homogeneous Poisson process with intensity β.

2.2. Area-interaction point process

The area-interaction process (Baddeley and Van Lieshout, 1995) can be used to model point processes with attraction or repulsion. The density function of the homogeneous area-interaction process for $\varphi_n = \{x_1, \ldots, x_n\}$ in B is defined as follows

$$f(\varphi_n) = c(\beta, \gamma)^{-1} \beta^n \gamma^{-m(U_r(\varphi_n))},$$

with respect to the Poisson process with unit intensity, where $\beta > 0$, $\gamma > 0$, $c(\beta, \gamma)$ is the normalizing constant, m denotes the Lebesgue measure and

$$U_r(\varphi_n) = \cup_{i=1}^n B(x_i, r),$$

is the union of discs with radius r centered at the points contained in the point pattern. Similar to the Strauss process, γ controls interaction between points. The area-interaction process generates repulsive point patterns if $\gamma \in (0, 1)$ and clustered point patterns if $\gamma > 1$. For $\gamma = 1$, this process is also equivalent to the Poisson point process with intensity β.

In our study, we implement a canonical scale-free version of the area-interaction process proposed by Baddeley and Turner (2014), since the interpretation is easier. The density function is

$$f(\varphi_n) = c(\kappa, \eta)^{-1} \kappa^n \eta^{-A(\varphi_n)},$$

where κ is the new main effect, η is the new interaction effect and

$$A(\varphi_n) = m(U_r(\varphi_n))/(\pi r^2) - n.$$

In this way, each isolated disc has unit area and contributes a factor κ to the density. Parameter η has the same interpretation as the original parameterization. Notice that the parameter r denotes the disc radius at each point but not the interaction distance. Thus, for the area-interaction process, the interaction distance is $2r$.

3. Approximate Bayesian Computation for Gibbs Point Processes

The ABC algorithm has been widely used for estimating model parameters with an intractable normalizing constant. The fundamental algorithm is known as ABC rejection sampling that was implemented by Pritchard *et al.* (1999) and Sunnåker *et al.* (2013). The specific pseudo-code associated with ABC rejection sampling is shown in Algorithm 1, where φ_{obs} is the observed data, φ_{sim} is the simulated data, $\mathbf{T}(\varphi_n)$ is the summary statistic and $d(\cdot,\cdot)$ is the discrepancy measure that is used to evaluate the similarity between the data. Moreover, $\epsilon > 0$ is the acceptance threshold.

Algorithm 1 ABC rejection sampling

```
for i = 1 to L do
  repeat
    Sample θ* from its prior, θ* ~ π(θ).
    Generate simulated data, φ_sim ~ f(φ|θ*).
  until d(T(φ_obs), T(φ_sim)) ≤ ε
end for
```

Based on Algorithm 1, the proposed θ will be accepted if the discrepancy measure is smaller than the pre-specified threshold, which implies that the data simulated based on θ is similar to the observed one. In this way, the accepted θ is generated from the approximated posterior distribution, since the simulated data approximate the realizations from the observed likelihood. More importantly, the choices of discrepancy measures and summary statistics can determine the efficiency and unbiasedness of the approximation of the posterior distribution.

In order to improve the efficiency of the ABC rejection sampling algorithm, Marjoram *et al.* (2003) proposed the ABC-MCMC, which combines the MCMC with the ABC. This algorithm was applied to repulsive spatial point processes by Shirota and Gelfand (2017). However, the choice of proposal kernel can significantly affect the efficiency of the ABC algorithm in terms of the level of mixing. In addition, choosing the most appropriate proposal kernel is a challenging issue. Thus, we implement the ABC rejection sampling algorithm for our study in order to investigate the performance of different summary statistics, prior distributions, and discrepancy measures.

3.1. Summary Statistics

For homogeneous Gibbs point processes, it is reasonable to evaluate the similarity of point patterns by comparing summary statistics for the main and interaction effects. We use the number of points as a summary statistic of the main effect, since it is also a function of the sufficient statistic. In the literature, several useful functional summary characteristics can be used for describing the interaction effect, *e.g.*, Ripley's K-function, D-function (the nearest neighbor distance distribution function), F-function (the empty space function), and the J-function (Van Lieshout and Baddeley, 1996). The definition of each function for a stationary point process with intensity λ is as follows:

K-function

$$K(r) = \mathbf{E}_o(N(b(o,r) \setminus \{o\}))/\lambda,$$

with $N(b(o,r) \setminus \{o\})$ representing the number of points found within the distance r from the typical point o.

D-function

$$D(r) = \mathbf{P}_o(N(b(o,r) \setminus \{o\}) > 0),$$

where $\mathbf{P}_o$ represents the Palm probability, which is the conditional probability that the point process has the property $N(b(o,r) \setminus \{o\}) > 0$ given that a point of the process is in o (see Illian *et al.* (2008), Page 178).

F-function

$$F(r) = 1 - \mathbf{P}(N(b(o,r)) = 0),$$

where $\mathbf{P}$ represents the distribution for a given point process.

J-function

$$J(r) = \frac{1 - D(r)}{1 - F(r)}.$$

Illian *et al.* (2008) point out that these summary characteristics play an important role in the analysis of point processes, and that each of them has specific advantages that can reveal unique information about the point pattern. Hence, we implement the functions mentioned above and compare their performance within the ABC algorithm. In order to obtain estimates of the various functions, we use the empirical estimators with edge correction using the R package "spatstat" (Baddeley and Turner, 2014).

3.2. Prior distributions

The prior distribution in the ABC is typically used as the proposal kernel or part of the Metropolis-Hastings acceptance probability. Here, we discuss various choices for the prior distribution for the ABC algorithm. Simulation studies detailing their performance are included in Sections 4 and 5.

A common choice of prior distribution for the ABC is a vague prior over a reasonable parameter space, such as a uniform prior with a pre-specified range. However, the choice of parameter space is subjective without strong prior knowledge. Additionally, a uniform prior with a wide parameter space can significantly reduce the efficiency of the ABC algorithm.

Alternatively, a more informative distribution can be applied as a prior for the ABC to improve estimation and efficiency. Geyer (1998) proposed using the stochastic approximation to obtain the starting point for more sophisticated methods. In this context, the estimates provided by the stochastic approximation form a consistent estimator for the model parameters. Thus, we use the estimate from the stochastic approximation as the mean in an informative prior that has an associated large variance. Specifically, we implement the Robbins-Monro (R-M) algorithm (Robbins and Monro, 1951) for Gibbs

point processes. Given the canonical form of the homogeneous Gibbs point process, the model is as follows

$$f(\varphi_n) = \exp\{\boldsymbol{\theta}'\mathbf{T}(\varphi_n) - \log c(\boldsymbol{\theta})\},$$

and the R-M algorithm generates a sequence of estimates $\boldsymbol{\theta}_k, k = 1, 2, \ldots,$ given by

$$\boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}_k + \frac{A}{k}\{\mathbf{T}(\varphi_{obs}) - \mathbf{T}(\varphi_k)\},$$

where φ_k is a realization generated from the specified Gibbs process given parameters $\boldsymbol{\theta}_k$. We also implement an objective stopping rule proposed by Gu and Zhu (2001),

$$K_1 = \inf\left\{K \geq K_0 : \left\|\sum_{k=K-K_0+1}^{K} \text{sgn}(\theta_{j,k} - \theta_{j,k-1})/K_0\right\| \leq \delta, \forall j = 1, \ldots, m\right\},$$

where K_0 and δ are the pre-specified length of the subsequence of $\boldsymbol{\theta}_k$ and required precision level in order to calculate the stopping rule and $\text{sgn}(x)$ is the sign function, which equals to 1,0 or -1 depending on the sign of x. Also, A is preassigned to ensure that the step function is slowly decreasing. It is difficult to tune these parameters, and iterative adjustment can introduce a computational burden. However, since we only need the sequence to converge to an area that is close to the true value, any reasonable choice of tuning parameters should yield a satisfactory result.

3.3. Construction of the discrepancy measure

We briefly review the details of constructing discrepancy measures and propose several measures that emerge as more robust in ABC estimation. Shirota and Gelfand (2017) implement the approach of constructing discrepancy measures based on linear regression for repulsive point processes (Fearnhead and Prangle, 2012). First, $h(\varphi_u, \varphi_v)$ is defined as a vector of functions of summary statistics for two point patterns, and typically, the length of the vector $h(\varphi_u, \varphi_v)$ is the same as the number of parameters in the specified model. Shirota and Gelfand (2017) used the difference of the number of points and the squared difference of the estimated variance stabilized K-function at a given interaction distance to construct the vector of functions. In contrast, Mateu *et al.* (2010) proposed to construct these functions based on functional summaries over a certain range of interaction $[0, r_{max}]$, *e.g.*, the integrated squared difference between estimates of Ripley's K-function for two point patterns. Inspired by both approaches, we use the difference of the number of points and the integrated absolute difference between estimated functional summaries of the two point patterns to propose and construct our general discrepancy measure, *i.e.*,

$$\begin{aligned} h_1(\varphi_{obs}, \varphi_{sim}) &= n(\varphi_{sim}) - n(\varphi_{obs}), \\ h_2(\varphi_{obs}, \varphi_{sim}) &= \int_0^{r_{max}} |\mathcal{F}_{sim}(r) - \mathcal{F}_{obs}(r)| dr, \end{aligned}$$

where $\mathcal{F}$ represents the estimated summarizing functional, such as those mentioned in Section 3.1 and $h_2(\varphi_{obs}, \varphi_{sim})$ is approximated numerically.

For comparison purposes, following the approach of Shirota and Gelfand (2017), one of the discrepancy measures implemented in our study for a parameter vector of length m is,

$$d_{reg}(\mathbf{T}(\varphi_{obs}), \mathbf{T}(\varphi_{sim})) = \sum_{j}^{m} (\hat{\xi}_{l,j} - \hat{\xi}_{obs,j})^2 / \widehat{\text{var}}(\hat{\xi}_j), \tag{4}$$

and

$$\hat{\boldsymbol{\xi}}_l = \hat{\mathbf{a}} + \hat{\mathbf{b}} h(\varphi_{sim}, \varphi_{obs}),$$
$$\hat{\boldsymbol{\xi}}_{obs} = \hat{\mathbf{a}} + \hat{\mathbf{b}} h(\varphi_{obs}, \varphi_{obs}) = \hat{\mathbf{a}},$$

where $\hat{\mathbf{a}}$ and $\hat{\mathbf{b}}$ are obtained by fitting linear regression for the pilot run. Moreover, $\widehat{\text{var}}(\hat{\xi}_j)$ is the sample variance of the j-th component of $\hat{\boldsymbol{\xi}}$ and it is also provided by the pilot run. Notice that for Gibbs point processes model, $\boldsymbol{\xi}$ is the log transformed parameter vector. Specifically, $\boldsymbol{\xi}$ is $(\log\beta, \log\gamma)$ for the Strauss process and $(\log\kappa, \log\eta)$ for the area-interaction process.

We further propose and implement another way of constructing the discrepancy measure by rescaling the function $h(\varphi_u, \varphi_v)$ based on the pilot run. The form of the discrepancy measure is

$$d_{rescale}(\mathbf{T}(\varphi_{obs}), \mathbf{T}(\varphi_{sim})) = w_1 |h_1(\varphi_{obs}, \varphi_{sim})| + w_2 |h_2(\varphi_{obs}, \varphi_{sim})|, \tag{5}$$

where w_1, w_2 are the scales provided by the inverse of the maximum absolute value of the elements of $h(\varphi_u, \varphi_v)$ for the pilot run, *i.e.*, $w_1 = 1/\max(|h_1(\varphi_{obs}, \varphi_{pilot})|)$ and $w_2 = 1/\max(\ |h_2(\varphi_{obs}, \varphi_{pilot})|)$. The goal of this approach is to rescale the effect of each element in the discrepancy measure to a common level.

The surface of the log discrepancy measure from Shirota and Gelfand (2017) and the proposed surfaces based on the pilot run are shown in Figures A1, A2, and A3 (in the Appendix). The value of each pixel in the surfaces is the log discrepancy between the observed point pattern and the simulated point pattern that is generated using the corresponding parameter values on the axes. Smaller discrepancy values indicate that the simulated and observed point pattern are similar. For the simulation study, we expect that the pixel associated with the true parameter value will minimize the discrepancy. However, we can see that the area having minimum discrepancy value is significantly larger than one pixel and that the value of neighboring pixels are also small. This behavior illustrates the identifiability issue associated with the Gibbs point process, *i.e.*, different parameter value combinations can generate similar point patterns in terms of the discrepancy measures discussed in this study. We mitigate the identifiability effect by using an informative prior; simulation results are shown in Section 5. Also, the value of the surfaces from Shirota and Gelfand (2017) are smaller than the values of the surfaces for the proposed method because of the lack of information. The flatness of these surfaces can also cause inefficiencies in the ABC algorithm.

4. Simulation Study: Vague Prior

For this study, we implement a uniform prior for the ABC rejection algorithm. In other words, we generate proposed parameter values from a uniform distribution with

reasonable lower and upper bounds for both the Strauss process and the area-interaction process. The crucial components of the discrepancy measures are the number of points in the point pattern and the estimated summary characteristics over interaction range $[0, r_{max}]$. For the K, D and F-function, we choose a moderately large $r_{max} = 0.18$ based on a sensitivity study. However, r_{max} is 0.08 for the J-function since the value of J-function is either highly unstable or undefined for large r. For the pilot run, we generate 100 parameter values equally spaced in the pre-specified parameter space, *e.g.*, a sequence of β from 50 to 400 and a sequence of γ from 0 to 1. Thus, there are 10,000 combinations of parameter values in total, and the corresponding realizations are generated to calibrate the coefficients $\hat{\mathbf{a}}$, $\hat{\mathbf{b}}$, w_1 and w_2 for the discrepancy measure. The acceptance threshold ϵ is set to be the 1% percentile of $d_{rescale}(\mathbf{T}(\varphi_{obs}), \mathbf{T}(\varphi_{sim}))$ for the pilot run. To simulate point patterns from the specified process in the ABC and pilot run, we assume that the interaction distance is known, however it can also be estimated using profile maximum pseudo-likelihood in practice. Finally, 1000 samples are accepted as the posterior realizations. The simulation study is conducted based on 50 realizations from each process, and the average estimate of the model parameters are compared in the following sections.

4.1. Strauss point process

We first consider the homogeneous Strauss process with strong interaction on the unit window, $W = [0, 1] \times [0, 1]$. Following the same parameterization of (3), we set $\beta = 200$, $\gamma = 0.1$ and $r = 0.05$. The simulation algorithm used to generate the point patterns is the *dominated coupling from the past algorithm* (Berthelsen and Møller, 2003). The prior distributions that are used as proposal kernels for β and γ are $\mathcal{U}(50, 400)$ and $\mathcal{U}(0, 1)$, and the estimation results are shown in Table 1. Each metric in this table corresponds to the discrepancy measures based on different summary statistics and structures (Formulas (4) and (5)). Specifically, $K(r)$ represents the discrepancy measure proposed by Shirota and Gelfand (2017). For the remaining metrics, the name of the metric indicates the functional characteristic that is used in the ABC and the method used to construct the discrepancy measure is shown in the parentheses, *e.g.*, D-function (regression) means that the discrepancy measure is constructed by linear regression with summary statistics based on the D-function and number of points.

Considering the point estimate and standard deviation of the posterior distribution for each parameter, the algorithm proposed by Shirota and Gelfand (2017) has the best performance, since the average posterior median provides the smallest bias for the main effect and second smallest bias for the interaction effect. Additionally, the posterior standard deviations are the smallest for both parameters among all of the algorithms. ABC with the discrepancy measure constructed based on the D-function and rescaling method also has a competitive performance, and therefore, both algorithms can be considered in order to fit the Strauss process. However, the large posterior standard deviation indicates that the estimates are affected by the apparent identifiability issue inherited from the underlying mechanism of the Gibbs processes.

4.2. Area-interaction point process

We also conducted a simulation study for the homogeneous area-interaction process with strong interaction on unit window W. By implementing the canonical scale-free

Table 1: The average posterior mean, median and standard deviation for Strauss process with vague prior: $K(r)$ represents the algorithm proposed by Shirota and Gelfand (2017), "regression" and "rescaling" indicate that the discrepancy measure is constructed based on Formulas (4) and (5). The bold numbers represent the best and second best value in each column.

Metric	Parameter	True value	Average mean	Average median	Average SD
$K(r)$	β	200	**203.5**	**200.4**	**35.65**
	γ	0.1	0.125	**0.113**	**0.070**
D-function	β	200	251.9	248.2	51.37
(regression)	γ	0.1	**0.097**	0.084	**0.063**
D-function	β	200	206.5	**202.9**	41.81
(rescaling)	γ	0.1	**0.120**	**0.103**	0.080
K-function	β	200	226.2	220.7	57.87
(regression)	γ	0.1	0.147	0.123	0.109
K-function	β	200	195.9	192.5	42.18
(rescaling)	γ	0.1	0.173	0.141	0.135
F-function	β	200	182.5	179.4	40.94
(regression)	γ	0.1	0.265	0.228	0.187
F-function	β	200	182.9	180.1	40.14
(rescaling)	γ	0.1	0.248	0.210	0.184
J-function	β	200	**197.0**	194.0	39.50
(regression)	γ	0.1	0.182	0.156	0.128
J-function	β	200	194.9	192.0	**37.70**
(rescaling)	γ	0.1	0.167	0.142	0.119

version, we set $\kappa = 50$, $\eta = 7$ and $r = 0.05$. Here, setting the disc radius of the area-interaction process to be 0.05 implies that the interaction distance is 0.1. For this process, the prior distributions of κ and η are $\mathcal{U}(10, 400)$ and $\mathcal{U}(1, 30)$. The estimation results of 50 realizations from the area-interaction process are shown in Table 2. The average mean and median show that only using the J-function as the summary statistic can provide reasonable estimates for both parameters. Although, the discrepancy measure based on the K-function and D-function can provide a reasonable estimate of one parameter, the posterior standard deviations are significantly larger than using J-function. Moreover, the discrepancy measure based on the J-function outperforms the one proposed by Shirota and Gelfand (2017).

Overall, the most robust estimator in this case is the median of the posterior sample from the ABC algorithm based on the discrepancy measure constructed by the summary statistic J-function and rescaling method. Similar to the Strauss process, the large posterior standard deviations indicate that the identifiability issue still exists for the area-interaction process.

5. Simulation Study: Informative Prior

In order to reduce the effect of the identifiability issue that was illustrated in Figures A1, A2, and A3 (in the Appendix), we implemented a class of informative prior distributions for the ABC. For the Strauss process, the prior of the main effect is a shifted gamma distribution and for the interaction effect a truncated normal distribution. Simi-

Table 2: The average posterior mean, median and standard deviation for area-interaction process with vague prior: $K(r)$ represents the algorithm proposed by Shirota and Gelfand (2017), "regression" and "rescaling" indicate that the discrepancy measure is constructed based on Formulas (4) and (5). The bold numbers represent the best and second best value in each column.

Metric	Parameter	True value	Average mean	Average median	Average SD
$K(r)$	κ	50	99.95	80.62	69.83
	η	7	**6.760**	3.912	6.647
D-function	κ	50	141.4	122.1	88.04
(regression)	η	7	4.457	2.933	**4.112**
D-function	κ	50	100.3	77.65	72.66
(rescaling)	η	7	**6.737**	4.833	5.703
K-function	κ	50	65.46	**54.12**	42.05
(regression)	η	7	9.811	7.893	6.902
K-function	κ	50	73.15	63.53	41.36
(rescaling)	η	7	7.820	**6.193**	5.646
F-function	κ	50	89.12	75.14	56.49
(regression)	η	7	6.661	4.293	6.146
F-function	κ	50	93.53	80.33	57.59
(rescaling)	η	7	6.155	3.980	5.682
J-function	κ	50	**46.53**	38.53	**29.30**
(regression)	η	7	9.604	8.450	5.387
J-function	κ	50	**58.78**	**49.16**	**35.06**
(rescaling)	η	7	9.000	**7.880**	**5.221**

larly, the shifted gamma distribution is used for both parameters of the area-interaction process. The summary statistics and discrepancy measures considered are the same as in Section 4.

5.1. Strauss point process

We consider an informative prior as the proposal kernel for the ABC. In order to generate enough points to estimate the summary characteristics, we set a lower bound for the prior of the main effect by shifting it to the right. Specifically, the prior distributions for β and γ are

$$\beta - l_\beta \sim G((\hat{\beta} - l_\beta)^2/\sigma_\beta, \sigma_\beta/(\hat{\beta} - l_\beta)), \tag{6}$$

$$\gamma \sim N_{[0,1]}(\hat{\gamma}, \sigma_\gamma^2), \tag{7}$$

where $l_\beta = 50$, $\sigma_\beta = 300$ and $\sigma_\gamma = 0.1$. $\hat{\beta}$ and $\hat{\gamma}$ are the estimated parameter values using stochastic approximation. Thus, the mean and variance of the shifted gamma distribution for β are $\hat{\beta}$ and σ_β. For γ, we choose a standard deviation that is not too large over the parameter space $[0, 1]$ to preserve the information from the stochastic approximation. In order to demonstrate the consistency of the stochastic approximation, we show that the average of the estimated parameter values for β and γ over 50 realizations are 200.7 and 0.098, respectively.

The simulation results of the same realizations of Section 4.1 are shown in Table 3. From this table, we can see the improvement in terms of the point estimate when compared to the estimate from stochastic approximation, is limited and that the estimated

Table 3: The average posterior mean, median and standard deviation for Strauss process with informative prior: $K(r)$ represents the algorithm proposed by Shirota and Gelfand (2017), "regression" and "rescaling" indicate that the discrepancy measure is constructed based on Formulas (4) and (5). The bold numbers represent the best and second best value in each column.

Metric	Parameter	True value	Average mean	Average median	Average SD
$K(r)$	β	200	200.2	202.0	16.51
	γ	0.1	0.111	0.116	0.063
D-function	β	200	206.2	**199.8**	16.40
(regression)	γ	0.1	**0.110**	**0.104**	**0.051**
D-function	β	200	200.5	**199.9**	15.87
(rescaling)	γ	0.1	**0.104**	**0.098**	**0.051**
K-function	β	200	202.6	201.9	16.51
(regression)	γ	0.1	0.122	0.116	0.063
K-function	β	200	**200.1**	199.5	15.68
(rescaling)	γ	0.1	0.112	0.105	0.062
F-function	β	200	**199.9**	199.4	**15.36**
(regression)	γ	0.1	0.123	0.118	0.067
F-function	β	200	199.3	198.8	**15.35**
(rescaling)	γ	0.1	0.118	0.111	0.066
J-function	β	200	200.4	**199.8**	15.55
(regression)	γ	0.1	0.117	0.112	0.062
J-function	β	200	199.6	199.0	15.38
(rescaling)	γ	0.1	0.113	0.107	0.060

values are close for different summary statistics and discrepancy measures. However, ABC provides significant reduction of the standard deviation when compared with the standard deviation of the prior distribution, which is an indication of Bayesian learning. The maximum standard deviation reductions for the distribution of β and γ are 11.4% and 49%, respectively. Notice that the discrepancy measure proposed by Shirota and Gelfand (2017) shows good point estimates of the parameters. However, combining the results from the point estimates and the posterior standard deviation, using the D-function as the summary statistic, and constructing the discrepancy measure by rescaling, provides the most robust performance. More importantly, implementing the informative prior can significantly improve point estimation and reduce the posterior standard deviation, in contrast to the ABC with vague prior.

5.2. Area-interaction point process

We also implement the informative prior for the area-interaction process. The right shifted gamma distribution is used as the prior distribution for κ and η,

$$\kappa - l_\kappa \sim G((\hat{\kappa} - l_\kappa)^2/\sigma_\kappa, \sigma_\kappa/(\hat{\kappa} - l_\kappa)),$$
$$\eta - l_\eta \sim G((\hat{\eta} - l_\eta)^2/\sigma_\eta, \sigma_\eta/(\hat{\eta} - l_\eta)),$$

where $l_\kappa = 10$, $l_\eta = 1$, $\sigma_\kappa = 300$, $\sigma_\eta = 100$, $\hat{\kappa}$ and $\hat{\eta}$ are the estimates from stochastic approximation. That is, the mean and variance of the shifted gamma distribution for κ and η are $(\hat{\kappa}, \sigma_\kappa)$ and $(\hat{\eta}, \sigma_\eta)$. We choose to use large prior variances to ensure that the choice of prior will not dominate the results of the analysis. Alternatively, the

Table 4: The average posterior mean, median and standard deviation for area-interaction process with informative prior: $K(r)$ represents the algorithm proposed by Shirota and Gelfand (2017), "regression" and "rescaling" indicate that the discrepancy measure is constructed based on Formulas (4) and (5). The bold numbers represent the best and second best value in each column.

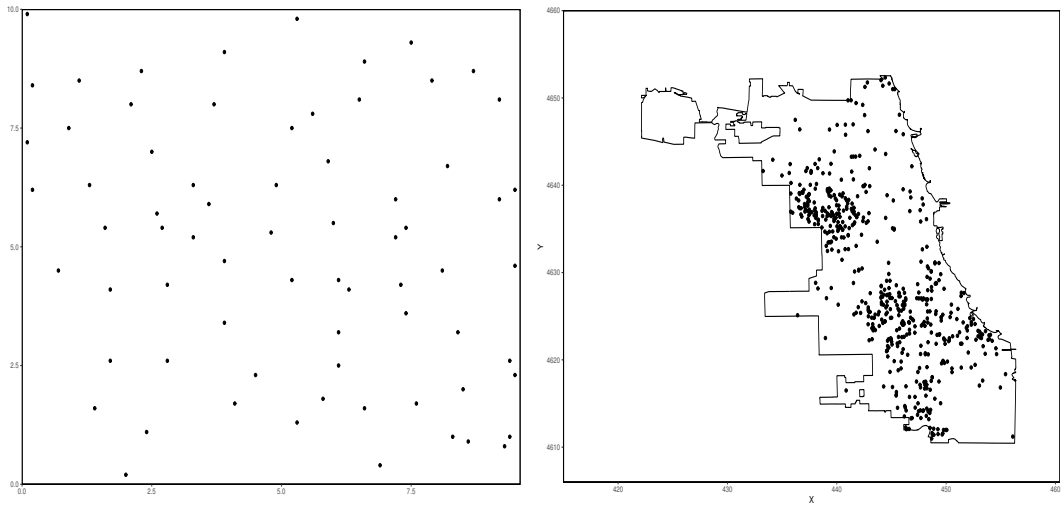
Metric	Parameter	True value	Average mean	Average median	Average SD
$K(r)$	κ	50	58.11	55.64	17.67
	η	7	7.521	6.834	3.310
D-function	κ	50	60.72	58.42	17.52
(regression)	η	7	7.360	6.873	**2.753**
D-function	κ	50	57.18	54.79	16.75
(rescaling)	η	7	7.726	7.158	3.051
K-function	κ	50	56.33	54.60	15.46
(regression)	η	7	7.843	**7.061**	3.924
K-function	κ	50	58.02	56.17	15.63
(rescaling)	η	7	**7.154**	6.504	3.335
F-function	κ	50	59.02	56.64	17.63
(regression)	η	7	7.311	6.642	3.311
F-function	κ	50	59.34	56.92	17.66
(rescaling)	η	7	7.221	6.566	3.241
J-function	κ	50	**51.10**	**49.33**	**13.83**
(regression)	η	7	**7.324**	6.808	**2.967**
J-function	κ	50	**53.74**	**51.90**	**14.32**
(rescaling)	η	7	7.623	**7.098**	3.030

truncated gamma distribution can be used as the proposal kernel to simulate parameters with a specific mean, variance, and positive lower bound. The average estimates from the stochastic approximation of κ and η over the 50 realizations used in Section 4.2 are 56.57 and 7.151, and the simulation results are shown in Table 4. By comparing the average posterior mean, median, and standard deviation with the prior distribution, we see significant improvements in terms of the point estimate and reduction in standard deviation, *i.e.*, the maximum reduction for κ and η are 20.2% and 70.3%, respectively. Among all the possible combinations, using the J-function as the summary statistic to construct the discrepancy measure provides the most robust results for both parameters and using regression or rescaling methods do not drastically affect the performance of the J-function. Similar to the results in Section 4.2, the proposed discrepancy measures based on the J-function outperform the one proposed by Shirota and Gelfand (2017). Also, implementing the informative prior significantly improves point estimation and reduces the posterior standard deviation for all the approaches.

6. Real Data Application

6.1. Swedish pines data

The Swedish pines data, shown in Figure 1(a), contains the locations of 71 pine saplings in a 9.6 by 10 meter window provided (Strand, 1972). Previous analyses on this data include Ripley (1981), Venables and Ripley (1997) and Baddeley and Turner (2000). All the results indicate that the interaction distance r is 0.7. To be proportional with the



(a) (b)

Figure 1: (a) The Swedish pines data: locations of 71 pine saplings in a 9.6 m $\times$ 10 m window first studied by Strand (1972). The data is available in the R package "spatstat". (b) Chicago crime data: locations based on UTM projection of 564 Chicago homicide incidents in 2018. The data is published by the City of Chicago Data Portal.

window size, the maximum interaction distance r_{max} is 0.8 for the J-function and 1.8 for the other summary characteristics. By assuming the Strauss process, the point estimates for β are between 1.49 and 3.29 and for γ are between 0.20 and 0.29 in the previous analyses mentioned above. In our study, we fit the Strauss process and area-interaction process to this dataset by using the ABC algorithm with different discrepancy measures and prior distributions

For the Strauss process, informative and vague priors are implemented in the ABC algorithm. Similar to Section 5, the prior distributions are gamma and truncated normal distribution with the mean provided by the estimate of the parameter values from stochastic approximation, *i.e.*, $\hat{\beta}$ and $\hat{\gamma}$ are 2.262 and 0.197, respectively. The standard deviations of the prior distributions are $\sqrt{30}$ and 0.1, respectively. The results from the ABC are shown in Table 5. For all the ABC algorithms, the acceptance threshold is the 1% percentile of the discrepancy measure in the pilot run. The overall results indicate strong inhibition between points in the point pattern and the values are consistent with the previous studies. Moreover, the standard deviations of the posterior distributions show significant reduction when compared with the prior standard deviations. For the vague priors, $\mathcal{U}(0.5, 40)$ and $\mathcal{U}(0, 1)$ are used for β and γ. The results are included in Table A1 in the Appendix and indicate larger posterior standard deviation. However, the overall results based on the vague prior are still consistent with the ones from the informative prior and previous studies.

We also fit an area-interaction process to the Swedish pines data, and the estimates show strong inhibition as well. However, a goodness-of-fit test shows lack of fit for the area-interaction process compared with the Strauss process. We included the estimates of the area-interaction process in the Appendix.

Table 5: The posterior mean, median, standard deviation and 95% credible interval of Strauss process for the Swedish pines data with informative prior: $K(r)$ represents the algorithm proposed by Shirota and Gelfand (2017), "regression" and "rescaling" indicate that the discrepancy measure is constructed based on Formulas (4) and (5).

Metric	Parameter	Mean	Median	SD	95% CI
$K(r)$	β	2.877	2.494	1.766	(1.471, 7.175)
	γ	0.203	0.196	0.089	(0.047, 0.381)
D-function	β	1.960	1.732	1.024	(0.694, 4.479)
(regression)	γ	0.211	0.204	0.078	(0.078, 0.379)
D-function	β	2.047	1.878	0.945	(0.782, 4.401)
(rescaling)	γ	0.207	0.201	0.077	(0.077, 0.265)
K-function	β	5.134	4.006	4.190	(1.070, 16.06)
(regression)	γ	0.201	0.196	0.069	(0.080, 0.343)
K-function	β	2.783	2.449	1.462	(0.935, 6.691)
(rescaling)	γ	0.201	0.200	0.077	(0.058, 0.361)
F-function	β	2.894	2.755	0.846	(1.659, 5.029)
(regression)	γ	0.168	0.163	0.089	(0.016, 0.360)
F-function	β	2.881	2.748	0.850	(1.658, 4.951)
(rescaling)	γ	0.167	0.163	0.087	(0.162, 0.356)
J-function	β	2.406	2.234	0.761	(1.311, 4.250)
(regression)	γ	0.156	0.150	0.077	(0.021, 0.311)
J-function	β	2.616	2.479	0.801	(1.465, 4.605)
(rescaling)	γ	0.164	0.158	0.082	(0.022, 0.343)

6.2. Chicago crime data

The Chicago crime data contains locations of reported homicide incidents in Chicago during 2018. The source of the raw data is the City of Chicago Data Portal which is an online resource summarizing incidents of crime that occurred in the city of Chicago (`https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-present/ijzp-q8t2`). As demonstrated in Figure 1(b), we use the Universal Transverse Mercator (UTM) coordinates for incident locations and the city boundary as the irregular domain of this point pattern. It is obvious that the point pattern shows significant clustering. Thus, we have reason to assume the data is generated from an attractive point process.

In order to be consistent with the simulation study, we assume that the process is homogeneous and proceed to fit the area-interaction process to this point pattern using the ABC algorithm. The disc radius of the area-interaction process is 0.35, which is calculated based on maximizing the profile pseudo-likelihood. Following the same choices in the simulation study, the maximum interaction distance r_{max} is 2 for the J-function and 7 for the other summary characteristics. Also, we implemented the gamma distribution as the informative prior along with the prior mean provided by the estimates from stochastic approximation. The prior mean and standard deviation for κ are 0.3575 and 1 and for η are 24.2648 and 30.

For all the ABC algorithms, the acceptance threshold is the 1% sample percentile

Table 6: The posterior mean, median, standard deviation and 95% credible interval of area-interaction process for the Chicago crime data with informative prior: $K(r)$ represents the algorithm proposed by Shirota and Gelfand (2017), "regression" and "rescaling" indicate that the discrepancy measure is constructed based on Formulas (4) and (5).

Metric	Parameter	Mean	Median	SD	95% CI
$K(r)$	κ	0.346	0.265	0.114	(0.175, 0.586)
	η	35.74	27.18	26.69	(9.035, 102.4)
D-function	κ	0.495	0.448	0.216	(0.180, 0.994)
(regression)	η	21.68	14.65	22.54	(1.605, 88.19)
D-function	κ	0.542	0.468	0.292	(0.197, 1.300)
(rescaling)	η	20.46	13.40	22.69	(0.605, 77.62)
K-function	κ	0.078	0.012	0.162	(0.001, 0.561)
(regression)	η	18.40	8.497	25.86	(0.046, 90.69)
K-function	κ	0.590	0.447	0.469	(0.159, 1.841)
(rescaling)	η	27.03	15.30	34.82	(0.094, 126.0)
F-function	κ	0.372	0.353	0.131	(0.172, 0.642)
(regression)	η	32.60	24.29	27.39	(5.625, 105.9)
F-function	κ	0.488	0.392	0.316	(0.175, 1.383)
(rescaling)	η	28.57	20.32	28.35	(0.484, 100.6)
J-function	κ	0.381	0.363	0.119	(0.191, 0.661)
(regression)	η	29.32	25.31	17.75	(8.026, 75.16)
J-function	κ	0.417	0.381	0.189	(0.165, 0.912)
(rescaling)	η	31.31	21.59	30.50	(2.258, 114.1)

of the pilot run. The results are shown in Table 6 and indicate significant attraction between points. Similar to the results in Section 5, the ABC algorithm provides significant standard deviation reduction for both parameters. We also use $\mathcal{U}(0,5)$ and $\mathcal{U}(0,150)$ as the vague prior for κ and η. The results are included in Table A4 in the Appendix, which show larger posterior variances for the parameters compared with the ones from the ABC with informative prior. This indicates that the informative prior helps alleviate the lack of identifiability.

7. Concluding Remarks

In this study, we compared the performances of several discrepancy measures for the ABC algorithm and proposed new measures. The simulation results confirm several aspects of our assumptions: the effect of different summary statistics on the estimation results, the improved performance of comparing the functional characteristics over a range of interaction distances and the effect of implementing the informative prior for the ABC.

First of all, the candidate summary characteristics show different performances for different point processes. This indicates that the choices of the most robust summary statistic for the ABC algorithm is point process specific. Although the literature points out that the choice should be a function of the sufficient statistic based on the sufficiency principle, we observed that the nearest neighbor distance distribution function can also provide competitive results for the Strauss process. Also, the F-function and J-function are both sufficient statistics for the area-interaction process, but the simulation shows that the J-function is more robust. Secondly, the discrepancy measure constructed by

the K-function given a specific interaction distance $K(r)$ out-performs the one that is constructed by the K-function over a range of interaction distances for the Strauss process, but not for the area-interaction process. This result indicates that evaluating attractive point processes requires more information than repulsive point processes. Finally, we demonstrated that applying an informative prior can significantly improve the point estimates of the parameters and reduce the posterior standard deviation. However, the identifiability issue of the Gibbs process cannot be completely eliminated by the current approaches. The applications to the Swedish pines data and Chicago crime data show that the ABC algorithm can be easily applied to point processes with regular and irregular windows.

In practice, we recommend that the analysis of the Gibbs point processes using the ABC algorithm should proceed as follows. First, an appropriate model for the point pattern should be specified based on an initial exploratory analysis. The Strauss process can be considered for patterns with strong inhibition and the area-interaction process can be considered for general clustered patterns. After choosing a specific model, one should decide the best discrepancy measure for the chosen model. For the models included in this study, the most robust discrepancy measures are provided. For other models, a similar simulation study, to the ones conducted in this paper, can provide a good indication for the best discrepancy measure. Finally, assuming that an informative prior is implemented, the median of the posterior realizations from the ABC is usually the most robust estimator of the parameters.

Since the Gibbs point process can be uniquely determined by the Papangelou conditional intensity, our future work will be investigating possible choices of the nonparametric estimate of the Papangelou and its implementation as the summary statistic for the ABC algorithm in order to resolve the identifiability issue. Moreover, a non-homogeneous version of the ABC algorithm will be developed. However, examining the non-homogeneous Gibbs point processes can be challenging since the non-homogeneity of the main effect is always confounded with the interaction effect.

Acknowledgements

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. This report is released to inform interested parties of ongoing research and to encourage discussion of work in progress. The views expressed are those of the authors, and not those of the U. S. Census Bureau.

References

Alquier, P., Friel, N., Everitt, R., and Boland, A. (2016). Noisy Monte Carlo: Convergence of Markov chains with approximate transition kernels. *Statistics and Computing*, **26(1-2)**, 29–47.

Andrieu, C., Roberts, G. O., *et al.* (2009). The pseudo-marginal approach for efficient Monte Carlo computations. *The Annals of Statistics*, **37(2)**, 697–725.

Atchadé, Y. F., Lartillot, N., Robert, C., *et al.* (2013). Bayesian computation for statistical models with intractable normalizing constants. *Brazilian Journal of Probability*

and Statistics, **27(4)**, 416–436.

Baddeley, A. and Turner, R. (2000). Practical Maximum Pseudolikelihood for Spatial Point Patterns: (with Discussion). *Australian and New Zealand Journal of Statistics*, **42(3)**, 283–322.

Baddeley, A. and Turner, R. (2014). Package 'spatstat'. *The Comprehensive R Archive Network (CRAN)*.

Baddeley, A. J. and Van Lieshout, M. (1995). Area-interaction point processes. *Annals of the Institute of Statistical Mathematics*, **47(4)**, 601–619.

Beaumont, M. A. (2003). Estimation of population growth or decline in genetically monitored populations. *Genetics*, **164(3)**, 1139–1160.

Berthelsen, K. K. and Møller, J. (2003). Likelihood and non-parametric bayesian mcmc inference for spatial point processes based on perfect simulation and path sampling. *Scandinavian Journal of Statistics*, **30(3)**, 549–564.

Besag, J. (1975). Statistical analysis of non-lattice data. *Journal of the Royal Statistical Society: Series D (The Statistician)*, **24(3)**, 179–195.

Cressie, N. (1993). *Statistics for Spatial Data. 2nd Edition*. John Wiley & Sons.

Fearnhead, P. and Prangle, D. (2012). Constructing summary statistics for approximate Bayesian computation: semi-automatic approximate Bayesian computation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **74(3)**, 419–474.

Geyer, C. (1998). Likelihood inference for spatial point. *Stochastic Geometry: Likelihood and Computation*, **80**, 79.

Geyer, C. J. (1991). Markov chain Monte Carlo maximum likelihood. Interface Foundation of North America. Retrieved from the University of Minnesota Digital Conservancy, http://hdl.handle.net/11299/58440.

Gu, M. G. and Zhu, H.-T. (2001). Maximum likelihood estimation for spatial models by Markov chain Monte Carlo stochastic approximation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **63(2)**, 339–355.

Huang, F. and Ogata, Y. (1999). Improvements of the maximum pseudo-likelihood estimators in various spatial statistical models. *Journal of Computational and Graphical Statistics*, **8(3)**, 510–530.

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

Kelly, F. P. and Ripley, B. D. (1976). A note on strauss's model for clustering. *Biometrika*, **63**, 357–360.

Liang, F. (2010). A double Metropolis–Hastings sampler for spatial models with intractable normalizing constants. *Journal of Statistical Computation and Simulation*, **80(9)**, 1007–1022.

Liang, F., Jin, I. H., Song, Q. and Liu, J. S. (2016). An adaptive exchange algorithm for sampling from distributions with intractable normalizing constants. *Journal of the American Statistical Association*, **111(513)**, 377–393.

Marjoram, P., Molitor, J., Plagnol, V. and Tavaré, S. (2003). Markov chain monte carlo without likelihoods. *Proceedings of the National Academy of Sciences*, **100(26)**, 15324–15328.

Mateu, J., Schoenberg, F. P. and Diez, D. (2010). On distances between point patterns and their applications. *American Statistician, in Review*, **9**.

Møller, J., Pettitt, A. N., Reeves, R. and Berthelsen, K. K. (2006). An efficient Markov chain Monte Carlo method for distributions with intractable normalising constants.

Biometrika, **93(2)**, 451–458.

Murray, I., Ghahramani, Z., and MacKay, D. (2012). MCMC for doubly-intractable distributions. *arXiv preprint arXiv:1206.6848*.

Park, J. and Haran, M. (2018). Bayesian inference in the presence of intractable normalizing functions. *Journal of the American Statistical Association*, **113(523)**, 1372–1390.

Pritchard, J. K., Seielstad, M. T., Perez-Lezaun, A. and Feldman, M. W. (1999). Population growth of human Y chromosomes: a study of Y chromosome microsatellites. *Molecular biology and evolution*, **16(12)**, 1791–1798.

Ripley, B. D. (1981). *Spatial Statistics*. vol. **24**. Wiley Online Library.

Robbins, H. and Monro, S. (1951). A stochastic approximation method. *The Annals of Mathematical Statistics*, 400–407.

Shirota, S. and Gelfand, A. E. (2017). Approximate Bayesian computation and model assessment for repulsive spatial point processes. *Journal of Computational and Graphical Statistics*, **26(3)**, 646–657.

Stoica, R. S., Philippe, A., Gregori, P. and Mateu, J. (2017). ABC Shadow algorithm: a tool for statistical analysis of spatial patterns. *Statistics and Computing*, **27(5)**, 1225–1238.

Strand, L. (1972). A model for stand growth. In *IUFRO Third Conference Advisory Group of Forest Statisticians*, vol. **72**. INRA, Institut National de la Recherche Agronomique Paris, p. 3.

Strauss, D. J. (1975). A model for clustering. *Biometrika*, **62(2)**, 467–475.

Sunnåker, M., Busetto, A. G., Numminen, E., Corander, J., Foll, M. and Dessimoz, C. (2013). Approximate Bayesian computation. *PLoS Computational Biology*, **9(1)**, e1002803.

Van Lieshout, M. and Baddeley, A. (1996). A nonparametric measure of spatial interaction in point patterns. *Statistica Neerlandica*, **50(3)**, 344–361.

Venables, W. and Ripley, B. D. (1997). Modern Applied Statistic with S-plus. Second Edition. Springer Verlag.

Appendix

Table A1: The posterior mean, median, standard deviation and 95% credible interval of Strauss process for the Swedish pines data with vague prior: $K(r)$ represents the algorithm proposed by Shirota and Gelfand (2017), "regression" and "rescaling" indicate that the discrepancy measure is constructed based on Formulas (4) and (5).

Metric	Parameter	Mean	Median	SD	95% CI
$K(r)$	β	5.379	2.812	5.909	(1.367, 12.38)
	γ	0.242	0.190	0.201	(0.018, 0.674)
D-function	β	2.176	1.727	1.514	(0.617, 6.450)
(regression)	γ	0.283	0.237	0.177	(0.057, 0.712)
D-function	β	2.226	1.873	1.372	(0.660, 5.633)
(rescaling)	γ	0.272	0.227	0.171	(0.053, 0.674)
K-function	β	11.90	9.761	8.942	(1.345, 34.63)
(regression)	γ	0.181	0.166	0.093	(0.052, 0.409)
K-function	β	3.665	3.118	2.398	(0.900, 9.879)
(rescaling)	γ	0.213	0.189	0.129	(0.029, 0.528)
F-function	β	2.894	2.755	0.846	(1.659, 5.029)
(regression)	γ	0.168	0.163	0.089	(0.016, 0.360)
F-function	β	3.227	3.065	1.194	(1.393, 5.981)
(rescaling)	γ	0.169	0.117	0.158	(0.004, 0.561)
J-function	β	2.810	2.650	0.993	(1.307, 5.263)
(regression)	γ	0.123	0.096	0.103	(0.004, 0.338)
J-function	β	3.011	2.919	1.037	(1.381, 5.417)
(rescaling)	γ	0.137	0.105	0.120	(0.003, 0.437)

A.1. Swedish pines data: fitted by the area-interaction process

The gamma distribution is used as the informative prior, with prior means 1.128 and 0.109 for κ and η. The standard deviations for the priors are $\sqrt{30}$ and $\sqrt{0.1}$, respectively. The estimation results are shown in Table A2. The results also show strong inhibition in the point pattern. We can see that the ABC algorithms with the J-function yield smaller standard deviations for the main function and closer values to the estimate from the Strauss process (Table 5). For the ABC with the vague prior, the prior distributions are $\mathcal{U}(0.1, 40)$ and $\mathcal{U}(0, 1)$. The corresponding results are included in the Appendix, see Table A3.

Table A2: The posterior mean, median, standard deviation and 95% credible interval of area-interaction process for the Swedish pines data with informative prior: $K(r)$ represents the algorithm proposed by Shirota and Gelfand (2017), "regression" and "rescaling" indicate that the discrepancy measure is constructed based on Formulas (4) and (5).

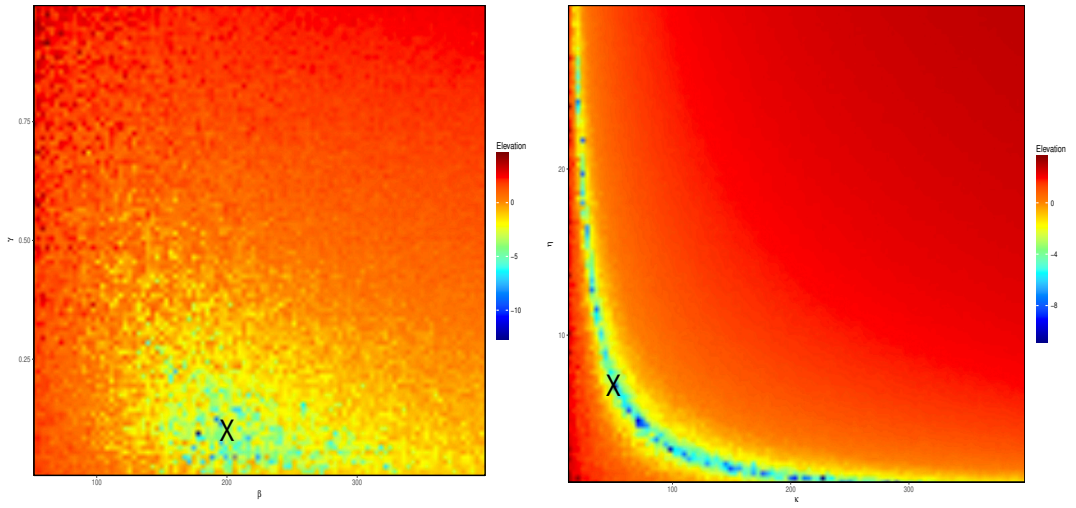
Metric	Parameter	Mean	Median	SD	95% CI
$K(r)$	κ	4.557	1.821	7.405	(0.344, 24.80)
	η	0.055	2.410×10^{-4}	0.210	(0.000, 0.450)
D-function	κ	5.343	2.057	8.555	(0.330, 28.88)
(regression)	η	0.046	2.083×10^{-4}	0.181	(0.000, 0.417)
D-function	κ	4.541	1.752	7.609	(0.265, 25.49)
(rescaling)	η	0.057	2.630×10^{-4}	0.236	(0.000, 0.461)
K-function	κ	5.012	1.939	8.166	(0.272, 27.87)
(regression)	η	0.036	1.839×10^{-4}	0.137	(0.000, 0.307)
K-function	κ	4.169	1.605	6.583	(0.252, 23.54)
(rescaling)	η	0.058	2.470×10^{-4}	0.223	(0.000, 0.495)
F-function	κ	3.892	1.018	7.607	(0.071, 25.30)
(regression)	η	0.083	15.08×10^{-4}	0.291	(0.000, 0.704)
F-function	κ	4.689	2.408	6.296	(0.618, 21.54)
(rescaling)	η	0.065	2.640×10^{-4}	0.245	(0.000, 0.610)
J-function	κ	1.785	0.843	2.421	(0.127, 8.619)
(regression)	η	0.090	9.084×10^{-4}	0.247	(0.000, 0.867)
J-function	κ	1.758	0.764	2.345	(0.005, 8.618)
(rescaling)	η	0.075	4.930×10^{-4}	0.263	(0.000, 0.664)

Table A3: The posterior mean, median, standard deviation and 95% credible interval of area-interaction process for the Swedish pines data with vague prior: $K(r)$ represents the algorithm proposed by Shirota and Gelfand (2017), "regression" and "rescaling" indicate that the discrepancy measure is constructed based on Formulas (4) and (5).

Metric	Parameter	Mean	Median	SD	95% CI
$K(r)$	κ	5.379	2.812	5.901	(1.367, 23.38)
	η	0.242	0.190	0.201	(0.018, 0.674)
D-function	κ	1.982	0.966	3.611	(0.305, 11.06)
(regression)	η	0.308	0.237	0.282	(0.001, 0.942)
D-function	κ	1.726	1.117	2.151	(0.276, 7.722)
(rescaling)	η	0.342	0.290	0.281	(0.002, 0.945)
K-function	κ	2.084	1.455	1.875	(0.300, 7.506)
(regression)	η	0.342	0.290	0.258	(0.004, 0.913)
K-function	κ	1.717	1.281	1.513	(0.270, 6.373)
(rescaling)	η	0.350	0.268	0.289	(0.005, 0.956)
F-function	κ	2.102	1.701	1.573	(0.565, 6.080)
(regression)	η	0.375	0.334	0.287	(0.006, 0.958)
F-function	κ	1.988	1.624	1.394	(0.587, 6.366)
(rescaling)	η	0.373	0.302	0.294	(0.005, 0.965)
J-function	κ	1.553	1.247	1.360	(0.199, 6.024)
(regression)	η	0.390	0.335	0.297	(0.006, 0.965)
J-function	κ	1.420	1.161	1.165	(0.165, 4.511)
(rescaling)	η	0.388	0.348	0.286	(0.008, 0.958)

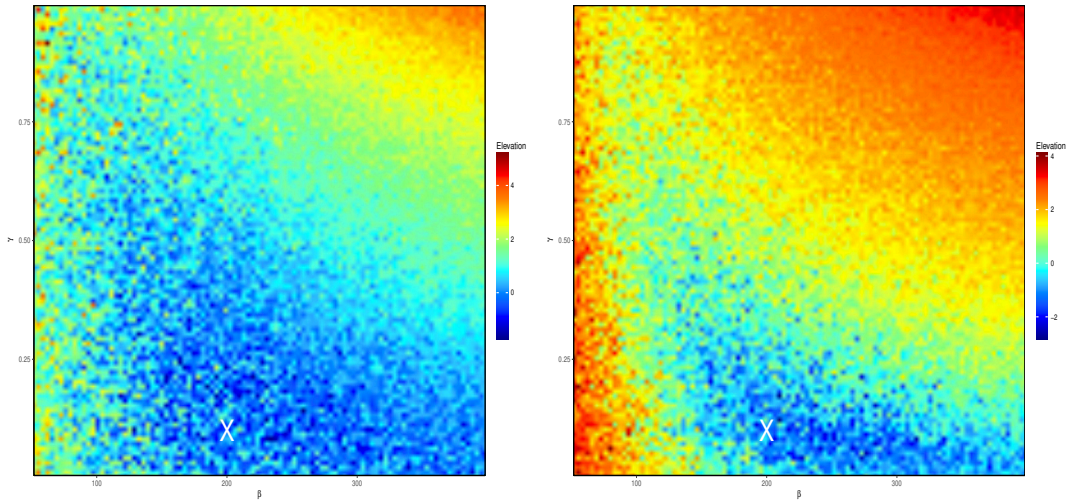
Table A4: The posterior mean, median, standard deviation and 95% credible interval of area-interaction process for the Chicago crime data with vague prior: $K(r)$ represents the algorithm proposed by Shirota and Gelfand (2017), "regression" and "rescaling" indicate that the discrepancy measure is constructed based on Formulas (4) and (5).

Metric	Parameter	Mean	Median	SD	95% CI
$K(r)$	κ	0.304	0.273	0.110	(0.173, 0.568)
	η	48.18	44.88	26.90	(9.206, 97.20)
D-function	κ	0.398	0.334	0.208	(0.174, 0.927)
(regression)	η	36.49	29.24	28.82	(1.648, 95.90)
D-function	κ	0.434	0.343	0.272	(0.181, 1.188)
(rescaling)	η	34.72	27.56	28.42	(0.867, 95.26)
K-function	κ	0.123	0.032	0.174	(0.003, 0.653)
(regression)	η	42.18	25.71	43.63	(0.138, 139.9)
K-function	κ	0.370	0.245	0.320	(0.142, 1.372)
(rescaling)	η	61.18	56.44	46.72	(0.557, 145.0)
F-function	κ	0.304	0.266	0.121	(0.171, 0.616)
(regression)	η	47.99	45.41	27.58	(6.391, 97.21)
F-function	κ	0.367	0.282	0.241	(0.177, 1.045)
(rescaling)	η	44.05	41.26	29.05	(1.182, 96.66)
J-function	κ	0.423	0.408	0.111	(0.236, 0.656)
(regression)	η	23.77	20.71	13.10	(7.700, 58.63)
J-function	κ	0.300	0.250	0.160	(0.145, 0.733)
(rescaling)	η	61.35	53.80	41.85	(4.342, 141.7)

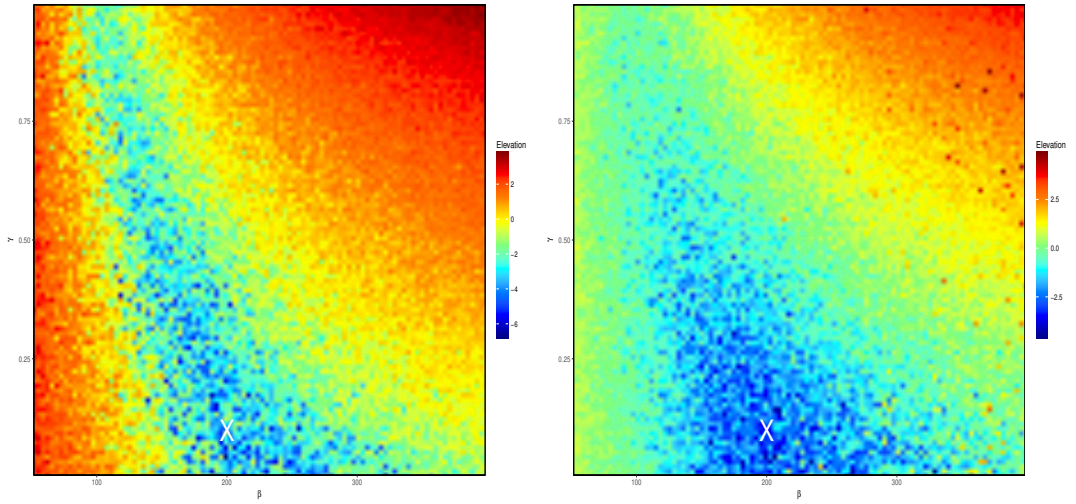


(a) (b)

Figure A1: Surfaces of the log discrepancy measure from Shirota and Gelfand (2017) for pilot run of Strauss and area-interaction process. (a) Strauss process; (b) Area-interaction process. The symbol "X" indicates the true values of the parameters.



(a) (b)



(c) (d)

Figure A2: Surfaces of the log discrepancy measure constructed by regression $d_{reg}(\mathbf{T}(\varphi_{obs}), \mathbf{T}(\varphi_{sim}))$ for the pilot runs of the Strauss process. Each discrepancy measure is constructed based on different summary characteristics: (a) K-function; (b) D-function; (c) F-function; (d) J-function. The symbol "X" indicates the true values of the parameters.

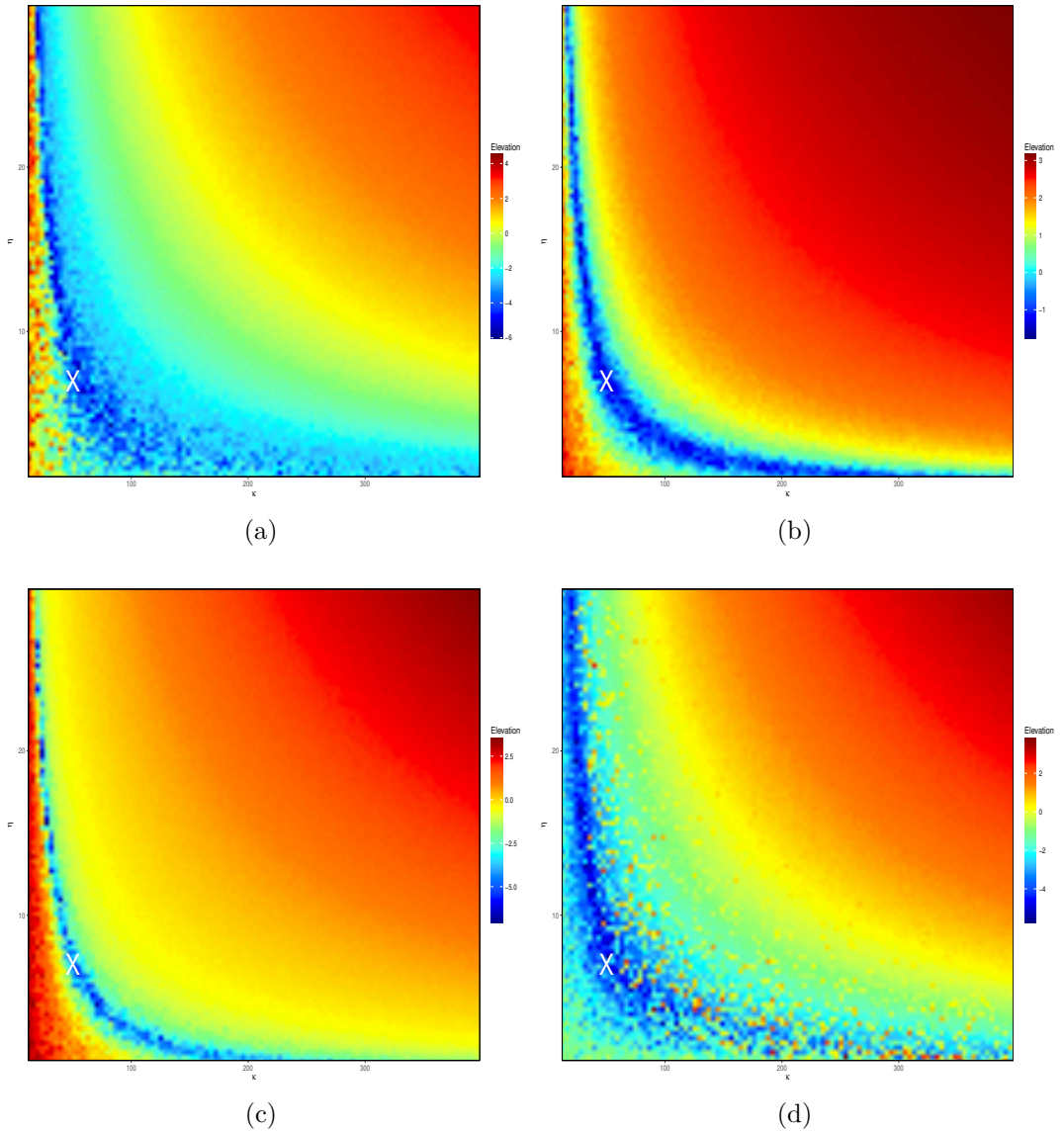
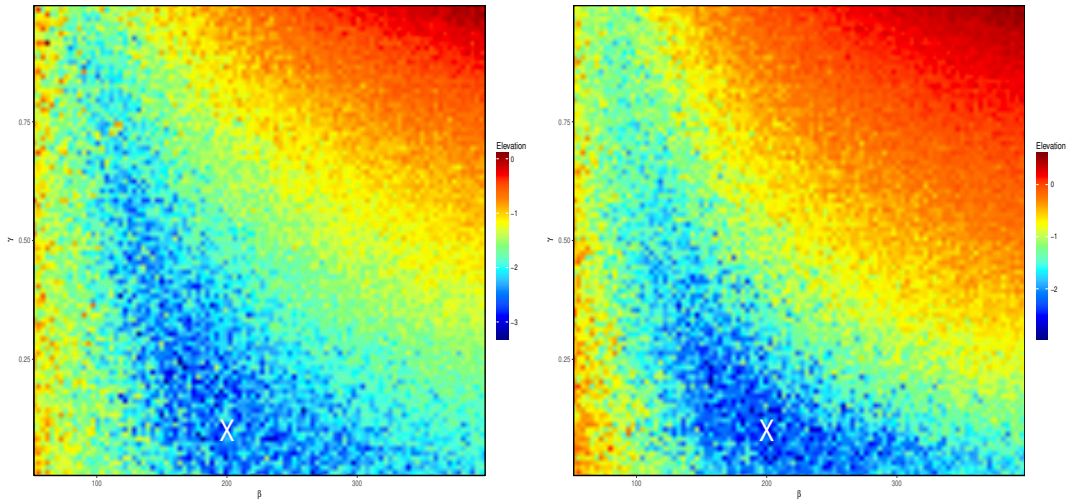
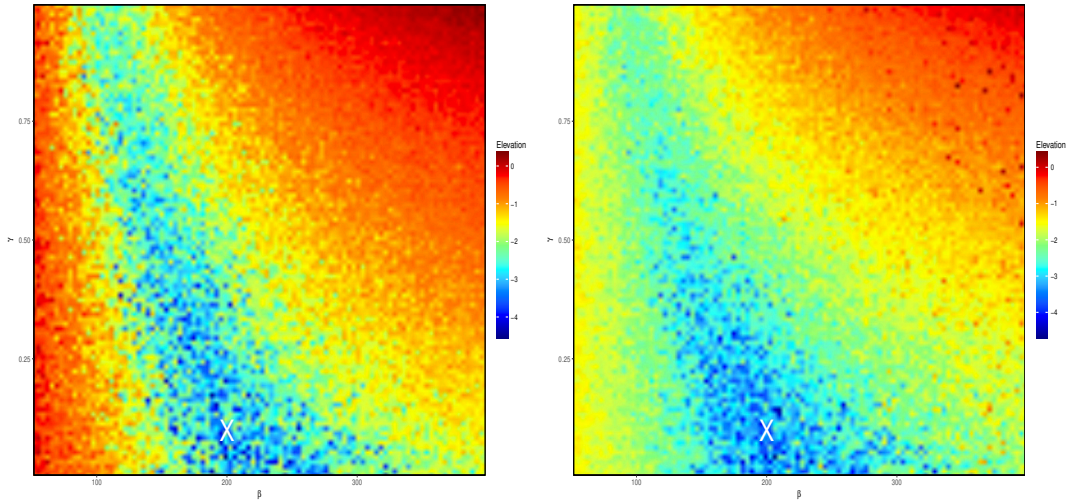


Figure A3: Surfaces of the log discrepancy measure constructed by regression $d_{reg}(\mathbf{T}(\varphi_{obs}), \mathbf{T}(\varphi_{sim}))$ for the pilot runs of the area-interaction process. Each discrepancy measure is constructed based on different summary characteristics: (a) K-function; (b) D-function; (c) F-function; (d) J-function. The symbol "X" indicates the true values of the parameters.

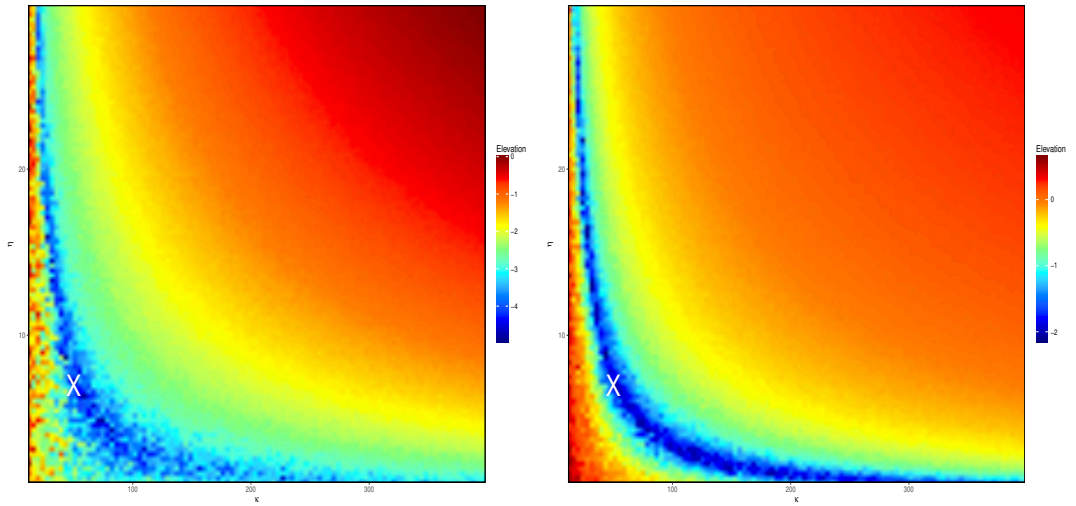


(a) (b)

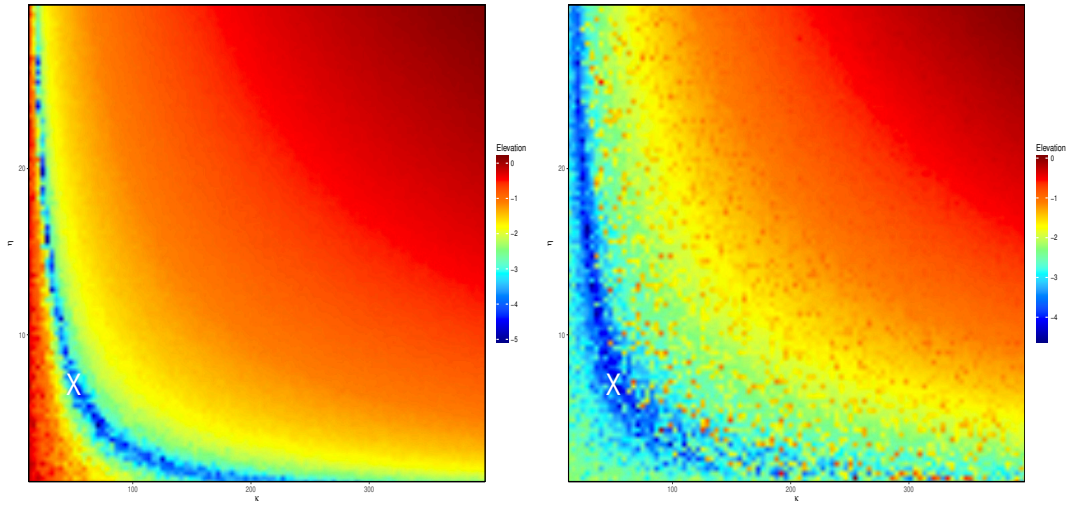


(c) (d)

Figure A4: Surfaces of the log discrepancy measure constructed by rescaling $d_{rescale}(\mathbf{T}(\varphi_{obs}), \mathbf{T}(\varphi_{sim}))$ for the pilot runs of the Strauss model. Each discrepancy measure is constructed based on different summary characteristics: (a) K-function; (b) D-function; (c) F-function; (d) J-function. The symbol "X" indicates the true values of the parameters.



(a) (b)



(c) (d)

Figure A5: Surfaces of the log discrepancy measure constructed by rescaling $d_{rescale}(\mathbf{T}(\varphi_{obs}), \mathbf{T}(\varphi_{sim}))$ for the pilot runs of the area-interaction model. Each discrepancy measure is constructed based on different summary characteristics: (a) K-function; (b) D-function; (c) F-function; (d) J-function. The symbol "X" indicates the true values of the parameters.