Inhomogeneous interacting marked point processes for studying morphostructures in paleobiological data

Diego Astaburuaga, Radu S. Stoica, Didier Gemmerlé, Francisco Cuevas-Pacheco

Inhomogeneous interacting marked point processes for studying morphostructures in paleobiological data.

D. Astaburuaga[1], R. S. Stoica[2], D. Gemmerlé[3], and F. Cuevas-Pacheco[1]

[1]Universidad Técnica Federico Santa María, Departamento de Matemática, Av. España 1680, Valparaíso Chile

[2]Université de Lorraine, CNRS, Inria, IECL, F-54500 Nancy France

[3]Université de Lorraine, CNRS, IECL, F-54500 Nancy France

July 18, 2024

Abstract

This paper presents inhomogeneous marked point processes with interactions that are applied to the analysis of morphostructures exhibited by a paleo-biological dataset presented in Kolesnikov (2018). Specifically, due to the nature of the dataset, we model the probability density function describing the models by considering three effects: the distance to the nearest edge, the distance to the lower right corner, and the distance to a reference circle. Furthermore, interactions between the points through the observed marks are introduced. This is done using the Strauss and Area-Interaction processes. Such models have four parameters that must be estimated. The proposed procedure is as follows. First, the sufficient statistics of the proposed model are computed from the data. Then, posterior sampling of the parameters is performed using the ABC Shadow algorithm. Next, the quality of the estimation is assessed by calculating the estimation errors and evaluating the significance of the model parameters. Finally, the model is verified using a Global Envelopes test from Myllymäki et al. (2017); Myllymäki M (2023).

The C++ library DRLib is the main programming tool used to perform model simulations, and parametric statistical inference based on the ABC Shadow algorithm the `R` package *spatstat* is used for the exploratory analysis, for the model verification analysis by global envelope tests and for the graphical presentation of the results.

1 Introduction

The data set analysed in this paper is issued from a study of specific morphostructures observed in a salina pond at Guérande (France). From a simple geometrical perspective, these morphostructures are discs with different radii located in rectangle as shown in Figure 1. The spatial distribution of these structures together with the evolution of their respective sizes is important for the specialists, since they may reveal relevant connections with the morphology of Ediacaran fossils Kolesnikov (2018).

Several aspects should be taken into account. First, the distribution of these structures is influenced by the manner the salina workers they proceed. These induce some edge and circular efects. The second element to be taken into account is the water arrival in the pond. And finally, the possible interaction between these structures should be also considered.

Exploratory analysis of these data done by Kolesnikov (2018) rejected the Poissonian distribution of the discs centres. In addition, inhomogeneous Poisson point process and area-interaction models using a fixed interaction radii were fitted to these data by Rhimi and Stoica (2019).

This work propose to make a step forward towards the modelling of this data set by fitting models that take into account the variability of the size of the observed discs structures. This is achieved through Bayesian inference Gillot et al. (2023); Stoica et al. (2017). This allows to discuss the quality of the estimation and the significativity of the model components. The performance of the fitted models is evaluated using Global envelope tests procedures developed in Myllymäki et al. (2017), and available in the R package GET (Myllymäki et al., 2017).

The structure of the paper is as follows Section 2 provides the theoretical background and framework where we describe the type of data, the used models, simulation algorithms and the Bayesian inference method via the ABC Shadow algorithm. Section 3 uses the aforementioned methodology to the salina pond data, presenting and discussing the results. Section 4 discuss the main obtained results and outlines futures perspectives.

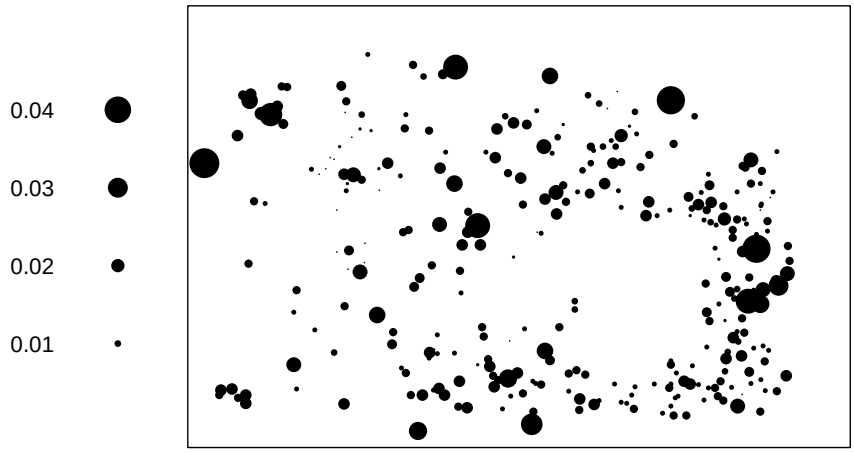

Figure 1: Observed data set: discoidal morphostructures of different sizes distributed in a rectangular pond.

2 Framework and Methods

2.1 Marked Point Pattern

Point processes are stochastic models for random configurations of points in a given region W. These processes can be generalised to marked point processes, if characteristics or marks, characterised by a probability distribution, can be attached to the points. The reader should refer to Møller and Waagepetersen (2004); Van Lieshout (2000) for a rigorous and detailed introduction to marked point processes.

In order to model the observed data, we consider a marked point process with realisations given by $\mathbf{x} = \{(w_1, r_1), \ldots, (w_n, r_n)\}$, where $\{w_1, \ldots, w_n\} \subset W \subset \mathbb{R}^2$ correspond to the point positions in a compact observation window W, and $\{r_1, \ldots, r_n\} \subset \mathbb{R}_0^+$ are the non-negative radii associated with each point.

2.2 Examples of Point Process Models

The models presented belong to the exponential family. That is, for a parameter vector θ, their corresponding probability density function with respect to the standard Poisson process, is given by

$$f(\mathbf{x}|\theta) = \frac{\exp\langle t(\mathbf{x}), \theta\rangle}{c(\theta)}, \tag{1}$$

where $t(\mathbf{x})$ corresponds to the vector of sufficient statistics for θ (Møller and Waagepetersen (2004); Van Lieshout (2000)). The difficulty in straightforward use of these models is that the normalising constant $c(\theta)$ is not available in a closed analytical form. Therefore, adapted strategies for sampling and statistical inference are required. The Metropolis-Hastings dynamics for model simulation and the ABC Shadow algorithm for parametric inference use appropriate techniques to avoid the computation of the involved normalising constants. These algorithms will be presented later in the paper. Due to these reasons, it is common to specify a model just by its un-normalised probability density

$$f(\mathbf{x}|\theta) \propto \exp\langle t(\mathbf{x}), \theta\rangle.$$

2.2.1 Poisson Point Process

This process is maybe the most known point process. Its probability density distributes points independently in W and it can be expressed as

$$f(\mathbf{x} \mid \theta) \propto \exp\left[\sum_{i=1}^{n(\mathbf{x})} \log(\rho(w_i, r_i))\right], \tag{2}$$

where ρ represents the intensity measure and $n(\mathbf{x})$ is the number of point in the configuration. If $\rho(w_i, r_i) = \rho$ for all i, the process is said to be homogeneous; otherwise, it is inhomogeneous. In this case, the intensity function models the spatial distribution of points in W.

2.2.2 Strauss Point Process

The Strauss point process was introduced by Kelly and Ripley (1976); Strauss (1975) and it is appropriate for modelling repulsive point patterns. Its probability density is given by

$$f(\mathbf{x} \mid \beta, \gamma_s) \propto \exp\left[n(\mathbf{x}) \log(\beta) + s_R(\mathbf{x}) \log(\gamma_s)\right]. \tag{3}$$

The parameter $\beta > 0$ controls the number of points in a configuration. The parameter $\gamma_s \in]0, 1[$ controls the number of R-close pairs of points in $\mathbf{x}$, via the statistic

$$s_R(\mathbf{x}) = \sum_{\{\xi,\eta\}\subseteq\mathbf{x}} \mathbf{1}[\|w_\xi - w_\eta\| \leq R],$$

with the parameter R controlling the interaction range.

For our purpose, the marks of the observed point pattern are integrated within the interaction function as range parameters. Specifically, for the marked point pattern $\mathbf{x} = \{(w_1, r_1), \ldots, (w_n, r_n)\}$, we consider the following interaction statistic:

$$s(\mathbf{x}) = \sum_{\{(w_i,r_i),(w_j,r_j)\}\subseteq\mathbf{x}} \mathbf{1}\left[\|w_i - w_j\| \leq r_i + r_j\right].$$

2.2.3 Area-Interaction Point Process

The Strauss process presented previously exhibits distance based interactions and it does not produce clustered patterns. The Area-Interaction process introduced by A. J. Baddeley and van Lieshout (1995) exhibits territorial interactions and it is able to produce both clustered and regular point patterns.

Similar to the Strauss process, set $R \geq 0$ as the interaction range. With $\beta > 0$ and $\gamma_A > 0$, the probability density function of an Area-Interaction point process takes the form:

$$f(\mathbf{x}) \propto \exp\left(n(\mathbf{x})\log(\beta) + a_R(\mathbf{x})\log(\gamma_A)\right), \tag{4}$$

where the sufficient statistic $a_R(\mathbf{x}) = -\left|\bigcup_{\xi\in\mathbf{x}} b(w_\xi, R)\right|$, with $b(w_\xi, R)$ being the ball of radius R and centre w_ξ, denotes the area covered by balls with centres $\{w_\xi : \xi \in \mathbf{x}\}$ and fixed radius R. Therefore this model controls the area covered by the point pattern. The process tends to favour configurations that occupy a minimised area whenever $\gamma_A \geq 1$ hence producing clustering. The other way around the process encourage configurations that occupy a more territory if $0 < \gamma_A < 1$, hence producing regular or repulsive point patterns.

Analogous to the Strauss point process, we introduce the marks into the sufficient statistics statistic as:

$$a(\mathbf{x}) = -\frac{1}{\pi\bar{r}^2}\left|\bigcup_{(w_i,r_i)\in\mathbf{x}} b(w_i, r_i)\right|,$$

such a modification yields to a general occupation of the territory depending on the contribution to it of each marked point. Here, the term $\bar{r}$ is given by the mean value of the observed the radii.

2.3 Bayesian inference: ABC Shadow

The ABC Shadow algorithm, introduced in Stoica et al. (2017), operates within a Bayesian framework. Its primary goal is to estimate parameters by effectively sampling from the posterior distribution. This process is mathematically represented as:

$$f(\theta \mid \mathbf{x}) \propto f(\mathbf{x} \mid \theta)p(\theta),$$

where $p(\cdot)$ represents the prior distribution of the parameters.

The algorithm employs simulation techniques to mimic the behaviour of an ideal chain, which in theory allows for drawing samples directly from the posterior distribution. Nonetheless, this approach is prone to an approximation error. Importantly, the magnitude of this error can be controlled and minimised (Stoica et al., 2021). For a given vector of perturbation parameters δ of the same dimension as θ, initial condition θ_0, and number of iterations m, the ABC Shadow algorithm can be described by the following steps:

Step 1: Assume the pattern $\mathbf{x}$ is observed.

Step 2: Generate $\mathbf{y}$ according to $f(\mathbf{y}|\theta)$.

Step 3: For $k = 1$ to m:

(a) Generate a new parameter ψ according to the density $U_\delta(\theta_{k+1} \to \psi)$ defined by $U_\delta(\theta \to \psi) = \frac{1}{|b(\theta,\delta/2)|}\mathbf{1}_{b(\theta,\delta/2)}\{\psi\}$.

(b) Accept the new state $\theta_k = \psi$ with probability $\alpha_s(\theta_{k-1} \to \psi) = \min\left\{1, \frac{f(\mathbf{x},\theta_k)p(\theta_k)}{f(\mathbf{x}|\theta_{k-1})p(\theta_{k-1})}\frac{f(\mathbf{y}|\theta_{k-1})}{f(\mathbf{y}|\theta_k)}\right\}$.

Step 4: Return θ_m.

Step 5: If more samples are needed, repeat from step 1 with $\theta_0 = \theta_m$.

The output are samples from the approximated posterior density $f(\theta|\mathbf{x})$. This approximation is well controlled under rather smooth conditions (Stoica et al., 2017). The probability density f should be continuously differentiable with respect the parameters θ. Comparing with Møller et al. (2006); Murray et al. (2006) no exact simulation is required for the simulation of the auxiliary pattern $\mathbf{y}$. In our case, the needed samples are obtained via the Algorithm 1 which is a Metropolis-Hastings algorithm that is $\phi-$irreducible, Harris recurrent and geometric ergodic (C. Geyer, 1999; C. J. Geyer & Møller, 1994; Møller

& Waagepetersen, 2004; Van Lieshout, 2019). The implemented ABC Shadow algorithm is detailled in Algorithm 2.

Estimation errors can be approximated. The data is assumed to be a sample of the considered model depending on a "true" parameter which is unknown. The asymptotic normality of the Maximum Likelihood Estimator (MLE) is used to compute estimation errors. Precisely, there are two types of error to consider: the error related to the discrepancy between the maximum likelihood estimator (MLE) and the true parameter, denoted as $\widehat{\theta} - \theta$; and the error associated with the difference between the Monte Carlo maximum likelihood estimation and the unknown exact MLE, denoted as $\widehat{\theta}_n - \widehat{\theta}$. The methods to compute these asymptotic errors have been presented and utilised in several studies, including C. Geyer (1994, 1999); Gillot et al. (2023); Van Lieshout and Stoica (2003).

Model validation procedure has been done using Global Envelope Tests Myllymäki et al. (2017) based on chosen summary statistics such as the empty space function, the nearest neighbour distribution, the pair correlation function and the J function.

Most of the work was developed using the **C++ library DRlib** (Gemmerlé et al., 2022). This library was specially build to perform simulation and to produce inference for point processes with interactions. The current available models are Poisson, Strauss and Area-Interaction with and without marks. The previously mentioned Metropolis-Hastings and ABC Shadow algorithms are also implemented. However, at its current state of development, the use of **spatstat R library** (A. Baddeley et al., 2015) may be required in order to perform calculations of point patterns summary statistics. These GET envelope tests were performed using the R package GET (Myllymäki M, 2023).

3 Application: morphostructure spatial distribution and interactions

In the following, the models from the previous section are fitted to the data. The summary statistics of the point pattern data are depicted in Figure 2 and they will be used to verify the quality of the modelling. The very first model that we fit is the inhomogeneous Poisson point processes. Since this process exhibits no interaction between points, it takes into account only the impact of the workers on the salina pond together with the influence of the position of the water source. In order to explore the possible interactions between the structures, to the Poisson process, an Area-Interaction and a Strauss point processes are superposed, alternatively.

3.1 Inhomogeneity Effects

In view of modelling the possible interactions in the point pattern, it is necessary first to consider the inherent inhomogeneity of the phenomenon, which can be addressed by proposing a function, ρ, in a suitable form. To this end, and motivated by Rhimi and Stoica (2019), we consider three proposals that are depicted in Figure 3, where the first two are related to the effect of salina workers and the third one is considering the direction of water arrival:

- Edge effect: it is considered that points appear more likely as they move away from the edge

$$\varphi_1(w_i) = d(w_i, \partial W).$$

- Influenced area: we assume that the affected zone by the salina workers can be modelled as a region where the probability to observer a point is low. In this regard, we consider the function

$$\varphi_2(w_i) = 1 - g\left(\frac{r_2 - \|w_i - z_0\|}{r_2 - r_1}\right),$$

  where the function $g(x)$ is a mollifier function given by $g(x) = f(x)/(f(x) + f(1-x))$ and $f(x) = \exp\{-1/x\}$ for $x > 0$ and 0 if $x \leq 0$. The point $z_0 = (0.777, 0.289)$ is the center of the empty area, $r1 = 0.0799$ is the minimum distance between the z_0 and $\mathbf{x}$, whilst $r2 = 0.1667$ is obtained as the $10th$-closest point to z_0.

- Preferred direction: from the bottom right corner, say c_0, a preferred direction for the appearance of structures is assumed

$$\varphi_3(w_i) = d(w_i, c_0).$$

3.2 Proposed models

The different models considered here are based on the structure of Equations (2), (3), and (4). The proposed models are build within three steps:

Step 1: consider only the inhomogeneity effect with no interaction between points, i.e., the Poisson component $f(\mathbf{x} \mid \theta)$ given by the equation (2), where we define $\theta_\beta = (\log(\beta_1), \log(\beta_2), \log(\beta_3))$ and:

$$\rho(w_i \mid \theta_\beta) = \beta_1^{\varphi_1(w_i)} \beta_2^{\varphi_2(w_i)} \beta_3^{\varphi_3(w_i)}.$$

Step 2: include interaction between points by adding the Strauss component to model repulsion behaviour, taking into account the number of points close enough to each other, that is

$$f_s(\mathbf{x} \mid \theta_\beta, \gamma_s) \propto f(\mathbf{x} \mid \theta) \exp\left[s(\mathbf{x}) \log(\gamma_s)\right]$$

Step 3: use an Area-Interaction component instead of a Strauss component to model either repulsion or attraction by controlling the area generated by the discs:

$$f_A(\mathbf{x} \mid \theta_\beta, \gamma_A) \propto f(\mathbf{x} \mid \theta_\beta) \exp\left[a(\mathbf{x}) \log(\gamma_A)\right].$$

It can be noticed that all those models belong to the exponential family. In consequence, the parameter vector can be written as

$$\theta = \begin{cases} \theta_\beta, & \text{if Poisson} \\ (\theta_\beta, \log(\gamma_s)), & \text{if with Strauss} \\ (\theta_\beta, \log(\gamma_A)), & \text{if with Area-Interaction} \end{cases}$$

and the sufficient statistic vector is

$$t(\mathbf{x}) = \begin{cases} (t_1(\mathbf{x}), t_2(\mathbf{x}), t_3(\mathbf{x})), & \text{if Poisson} \\ (t_1(\mathbf{x}), t_2(\mathbf{x}), t_3(\mathbf{x}), s(\mathbf{x})), & \text{if with Strauss} \\ (t_1(\mathbf{x}), t_2(\mathbf{x}), t_3(\mathbf{x}), a(\mathbf{x})), & \text{if with Area-Interaction} \end{cases}, \text{ with } \quad t_j(\mathbf{x}) = \sum_{i=1}^{n(\mathbf{x})} \varphi_j(w_i),$$

where those values for the described data set are calculated using DRlib and showed in the table 1.

Table 1: Sufficient statistics calculated for the Sel de Guerande point pattern

$t_1(\mathbf{x})$	$t_2(\mathbf{x})$	$t_3(\mathbf{x})$	$s(\mathbf{x})$	$a(\mathbf{x})$
59.051	300.383	200.006	214	-88.2704

3.3 ABC Shadow set-up and results analysis

3.4 Set-up of the algorithm

The use of the ABC Shadow algorithm requires to set hyper-parameters such as: delta step to update the proposal, the initial values of the parameters, define the interval where to look the real parameter, the number of iterations and the number of samples. There is no a exact rule to select them and these parameter have a direct influence on the convergence of the algorithm. Here, these parameters were chosen after several trials and error on simulated data. In fact, model with known parameters were simulated, hence the estimation result from these simulation is perfectly known, this allowing the set-up of the hyper-parameters . The choices adopted here are presented in Table 2.

Given those hyper-parameters, only the sufficient statistics of the data set regarding of the respective exponential family model and the mark distribution are required. Here for the marks, the uniform distribution is chosen.

The sampling from the posterior of the different models are presented in the Figure 7, for the models Poisson, with Strauss and with Area-Interaction respectively. It is noted that the posterior effectively look as a Gaussian, then the respective estimation of each parameter is selected by the mean of the posterior.

Parameter	Value
General setting	
Observation window	1.2×0.8
Probability of birth (p_b)	0.5
Probability of death (p_d)	0.5
Number of loop theta (N)	1×10^5
Number of iterations theta (m)	100
MH time for generating an x sample	250
Observed statistics	
Poisson Edge effect ($t_1(\mathbf{x})$)	59.051
Poisson Influenced area ($t_2(\mathbf{x})$)	300.383
Poisson Preferred direction ($t_3(\mathbf{x})$)	200.006
Strauss ($s(\mathbf{x})$)	214
Area-Interaction ($a(\mathbf{x})$)	-88.2704
Poisson components parameters	
Interval of $\log(\beta_i)$ range	$[-30.0, 30.0]$
Perturbation parameter (δ)	0.01
Initial value $\log(\beta_i^{(0)})$	0.0
Strauss model parameters	
Interval of $\log(\gamma_s)$ range	$[-10.0, 0.0]$
Perturbation parameter (δ)	0.001
Initial value $\log(\gamma_s^{(0)}))$	-0.5
Strauss random radius range	$U[0.0025, 0.0450]$
Area Interaction model parameters	
Interval of $\log(\gamma_A)$	$[-10.0, 10.0]$
Perturbation parameter (δ)	0.01
Initial value $\log(\gamma_A^{(0)}))$	0.0
Area-Interaction random radius range	$U[0.0025, 0.0450]$

Table 2: Parameter setting for the ABC Shadow algorithm.

3.5 Results analysis

The ABC Shadow algorithm was run in order to obtain $10,000$ samples of the parameter vector for each considered model. For all the models, their corresponding parameters were estimated by the Maximum Posterior Mode. For these values of the parameters, simulations of the models were drawn in order to compute asymptotics and to perform global envelope tests.

The histogram of the selected parameters via ABC Shadow for the Poisson, Strauss, and Area interaction point process are depicted in Figure 7. For all the models we conclude that the parameters are statistically significant.

The evolution of the sufficient statistics of the models with the estimated parameters are presented in Figures 4, 5, 6 indicate good behaviour of the algorithms: the sufficient statistics curves look stable suggesting the stationary regime is reached. The average values of the sufficient statistics are close the values of the observed statistics. This announces a rather good quality of the parameter estimation, which is assessed by the asymptotic errors presented in Table 3.

The adequacy of each model was explored using statistical tests based on different summary statistics. In this regard, we sample from the posterior distribution of the parameters by storing a point pattern after two hundred simulations until 999 simulations are reached. Then, under the hypothesis of independence, the stored samples are used to build 95% area envelopes (Myllymäki et al., 2017; Myllymäki M, 2023) based on the following summary statistic: the K, F, G, J- functions and the pair correlation function. The resulting envelopes under the Poisson, Strauss, and Area interaction models are depicted in Figures 8, 9, and 10 respectively. Following Figures 8, 9, the Poisson model and Strauss models are rejected for all the tests since they have a rather low p-value. Indeed, following the K, G, and the pair correlation function, both models do not reproduce the aggregation levels of the observed data. For the case of the Area-Interaction model, Figure 10 the p-values of the tests for the K, F, and pcf tend to be greater. Since, the envelope test is not rejected for theses statistics, this indicates a better performance of this model with respect to the Poisson and the Strauss process.

4 Conclusions

In this paper we provide a methodology to analyse point patterns with marks using statistical Bayesian learning with the ABC Shadow algorithm with models such allows interaction behaviour in the points related to their marks.

The results shows that the integration of interaction model as the Strauss and Area-Interaction have a significant role in order to understand the behaviour of the studied data set.

Moreover, the analysis realised in conjoint with the summary statistic functions indicates that the Area-Interaction model, with attraction parameter value, represent better the relations between the points and their marks. However, this model is not enough to understand the point pattern because of a remaining repulsion which can be modelled by including the Strauss component in this model.

Some perspectives deserve to be outlined. First, the information related to the radii distribution of the pattern should be better integrated into the model as mark distribution or into more adequate interaction functions. As an argument in favour of it, Figure 11, panels 11a and 11b shows the envelopes of the radii distribution obtained from the samples of the Strauss and Area-Interaction models simulated with the estimated parameters, respectively. It appears, that even if not perfectly, the Area-Interaction model captures important features of the radii distribution of the observed pattern. Then, a Bayesian framework can be considered for the choice of the hyper-parameters of the algorithm in an integrated form. And finally, investigate whether it is possible that the residual analysis of point processes A. Baddeley et al. (2015) may be incorporated within this framework in order to determine simultaneously the best model and its related parameters.

Considering the milestone role of DRLib in this work, it is proposed to continue to develop this library since it allows intensive simulation study of point process based inference.

5 Acknowledgements

The work of the first author was mainly performed during his stay within the Inria Pasta team and the IECL research laboratory from Université de Lorraine. His scientific journey was funded by an Inria grant together with the project *Ingenieria 2030* from CORFO-ANID.

The last author, Francisco Cuevas-Pacheco, was supported by *Proyecto ANID/FONDECYT/INICIACION* 11240330 from Agencia Nacional de Investigación y Desarrollo, Chile.

References

Baddeley, A., Rubak, E., & Turner, R. (2015). *Spatial Point Patterns: Methodology and Applications with R* (1st ed.). Boca Raton, CRC Press.

Baddeley, A. J., & van Lieshout, M. N. M. (1995). Area-interaction point processes. *Annals of the Institute of Statistical Mathematics*, *47*, 601-619.

Gemmerlé, D., Stoica, R. S., & Reype, C. (2022, August). *DRlib: a C++ library for marked Gibbs point processes simulation and inference.* 21st Annual Conference of the International Association for Mathematical Geosciences, IAMG 2022. Retrieved from `https://hal.science/hal-04047676` (Poster)

Geyer, C. (1994). On the Convergence of Monte Carlo Maximum Likelihood Calculations. *Journal of the Royal Statistical Society: Series B (Methodological)*, *56*(1), 261–274.

Geyer, C. (1999). Likelihood Inference for Spatial Point Processes. In *Chapter 1.*

Geyer, C. J., & Møller, J. (1994). Simulation Procedures and Likelihood Inference for Spatial Point Processes. *Scandinavian Journal of Statistics*, *21*(4), 359-373.

Gillot, N., Stoica, R. S., & Gemmerlé, D. (2023, September). *Study the galaxy distribution characterisation via Bayesian statistical learning of spatial marked point processes.* RING Meeting, École nationale supérieure de géologie (ENSG) Nancy. Nancy, France. Retrieved from `https://hal.archives-ouvertes.fr/hal-04163649v2/document` (Poster presented at the RING Meeting)

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

Kolesnikov, A. (2018). *Paléobiologie des biotes édiacariens durant et après la crise kotlinienne: apport des palaeopascichnidés et des morphostructures d'origine microbienne* (PhD thesis). Lille University.

Møller, J., Pettitt, A. N., Reeves, R. W., & Berthelsen, K. K. (2006). An efficient Markov chain Monte Carlo method for distributions with intractable normalizing constants. *Biometrika*, *93*, 451-458.

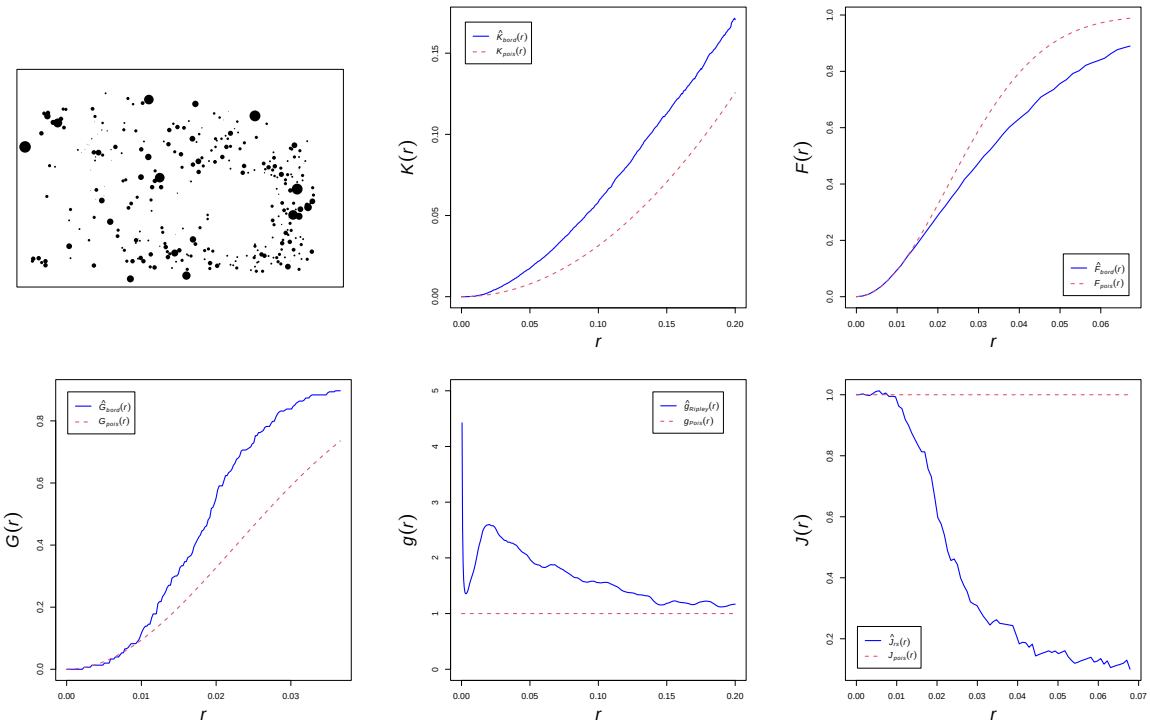

Figure 2: From left to right and up to down: Real point pattern (without the marks) with its empirical K-function, F-function, G-function, pair correlation function and J-function. All of the estimated functions where computed considering the edge correction function and without considering the marks.

Table 3: Estimation of parameters with their asymptotic standard deviations

Model	$\log \beta_1$	$\log \beta_2$	$\log \beta_3$	$\log \gamma_s$	$\log \gamma_A$
Poisson	6.0581 ± 0.5421	1.5947 ± 0.1618	6.6125 ± 0.2181	-	-
Poisson + Strauss	7.5097 ± 0.6748	1.1852 ± 0.2143	7.7208 ± 0.2669	-0.4367 ± 0.0554	-
Poisson + AreaInt	7.1454 ± 0.4233	2.7318 ± 0.1422	7.1497 ± 0.1837	-	2.4065 ± 0.1719

Table 4: Monte Carlo Errors for Parameter Estimations

Parameter	$\log \beta_1$	$\log \beta_2$	$\log \beta_3$	$\log \gamma_s$	$\log \gamma_A$
Poisson	0.00074	0.0001	7.6e-05	-	-
Poisson + Strauss	0.0018	0.0002	0.0001	3.1e-05	-
Poisson + AreaInt	0.0006	0.0001	7.8e-5	-	9.1e-5

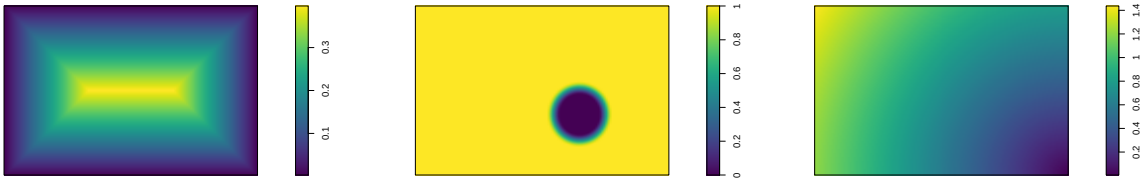

Figure 3: From left to Right: Ilustration of the Edge effect, Influenced Area and Preferred Direction functions.

Murray, I., Ghahramani, Z., & MacKay, D. J. C. (2006). MCMC for doubly-intractable distributions. *In Proceedings of the 22nd Annual Conference on Uncertainty in Artificial Intelligence*, 359–366.

Myllymäki, M., Mrkvička, T., Grabarnik, P., Seijo, H., & Hahn, U. (2017). Global envelope tests for spatial processes. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *79*(2), 381–404.

Myllymäki M, M. T. (2023). Get: Global envelopes in R. *arXiv:1911.06583*.

Møller, J., & Waagepetersen, R. (2004). Statistical Inference and Simulation for Spatial Point Processes.

Rhimi, F., & Stoica, R. S. (2019). *Spatial Point Processes with Interaction.* (Research project at Ecole des Mines Nancy ARTEM)

Stoica, R. S., Deaconu, M., Philippe, A., & Hurtado-Gil, L. (2021). Shadow Simulated Annealing: A new algorithm for approximate Bayesian inference of Gibbs point processes. *Spatial Statistics*, *43*.

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

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

Van Lieshout, M. (2000). *Markov Point Processes and Their Applications.* World Scientific.

Van Lieshout, M. (2019). Theory of Spatial Statistics: A Concise Introduction.

Van Lieshout, M., & Stoica, R. (2003). The Candy model: properties and inference. *Statistica Neerlandica*, *57*(2), 177–206.

6 Appendix A: Metropolis Hastings Algorithms

Algorithm 1 Metropolis-Hastings algorithm to simulate a Point Pattern

Require: $N \geq 0$, a initial configuration $\mathbf{x}^{(0)} = \{(w_1, r_1), \ldots, (w_n, r_n)\}$, parameter vector θ, two positive real numbers p_b, p_d such that $p_b + p_d = 1$

for $k = 1, \ldots, N$ **do**

Set $\mathbf{x}^{(k)} = \mathbf{x}^{(k-1)}$

Simulate a Bernoulli random variable B with $P(B = 1) = p_b$.

if $B = 1$ **then** ▷ Birth procedure

Generate a random point (ξ, r) within region W and set $\mathbf{x}' = \mathbf{x} \cup \{(\xi, r)\}$.

Compute $r_b = \min\left\{1, \frac{p_d}{p_b} \frac{f(\mathbf{x} \cup (\xi,r)|\theta)}{f(\mathbf{x}|\theta)} \frac{|W|}{n(\mathbf{x})+1}\right\}$.

With probability r_b, set $\mathbf{x}^{(k)} = \mathbf{x}'$; otherwise, do nothing.

else ▷ Death procedure

Choose a random point (ξ, r) from $\mathbf{x}$ and set $\mathbf{x}' = \mathbf{x} \setminus \{(\xi, r)\}$.

Calculate $r_d = \min\left\{1, \frac{p_b}{p_d} \frac{f(\mathbf{x} \setminus (\xi,r)|\theta)}{f(\mathbf{x}|\theta)} \frac{n(\mathbf{x})}{|W|}\right\}$.

With probability r_d, set $\mathbf{x}^{(k)} = \mathbf{x}'$; otherwise, do nothing.

end if

end for

7 Appendix B: ABC Shadow Algorithm

Algorithm 2 ABC Shadow algorithm to approxmate samples from the posterior distribution

Require: A integer $N > 0$ of number of samples needed, a perturbation parameter $\delta > 0$, a initial condition θ_0, m number of iterations. Assume a pattern $\mathbf{x}$ is observed.

for $k = 1, \ldots, N$ **do**

Generate a point pattern $\mathbf{y}$ according to $f(\mathbf{y}|\theta_0)$.

for $n = 1$ to m **do**

Generate a new parameter ψ according to the density $U_\delta(\theta_{n-1} \to \psi)$ defined by $U_\delta(\theta \to \psi) = \frac{1}{|b(\theta,\delta/2)|}\mathbf{1}_{b(\theta,\delta/2)\{\psi\}}$.

Compute $r = \frac{f(\mathbf{x}|\psi)p(\psi)}{f(\mathbf{x}|\theta_{n-1})p(\theta_{n-1})} \times \frac{f(\mathbf{y}|\theta_{n-1})}{f(\mathbf{y}|\psi)}$

With probability $\alpha = \min\{1, r\}$, set $\theta_n = \psi$; otherwise, $\theta_n = \theta_{n-1}$.

end for

A sample $\theta^{(k)} = \theta_n$ is obtained.

Set $\theta_0 = \theta^{(k)}$.

end for

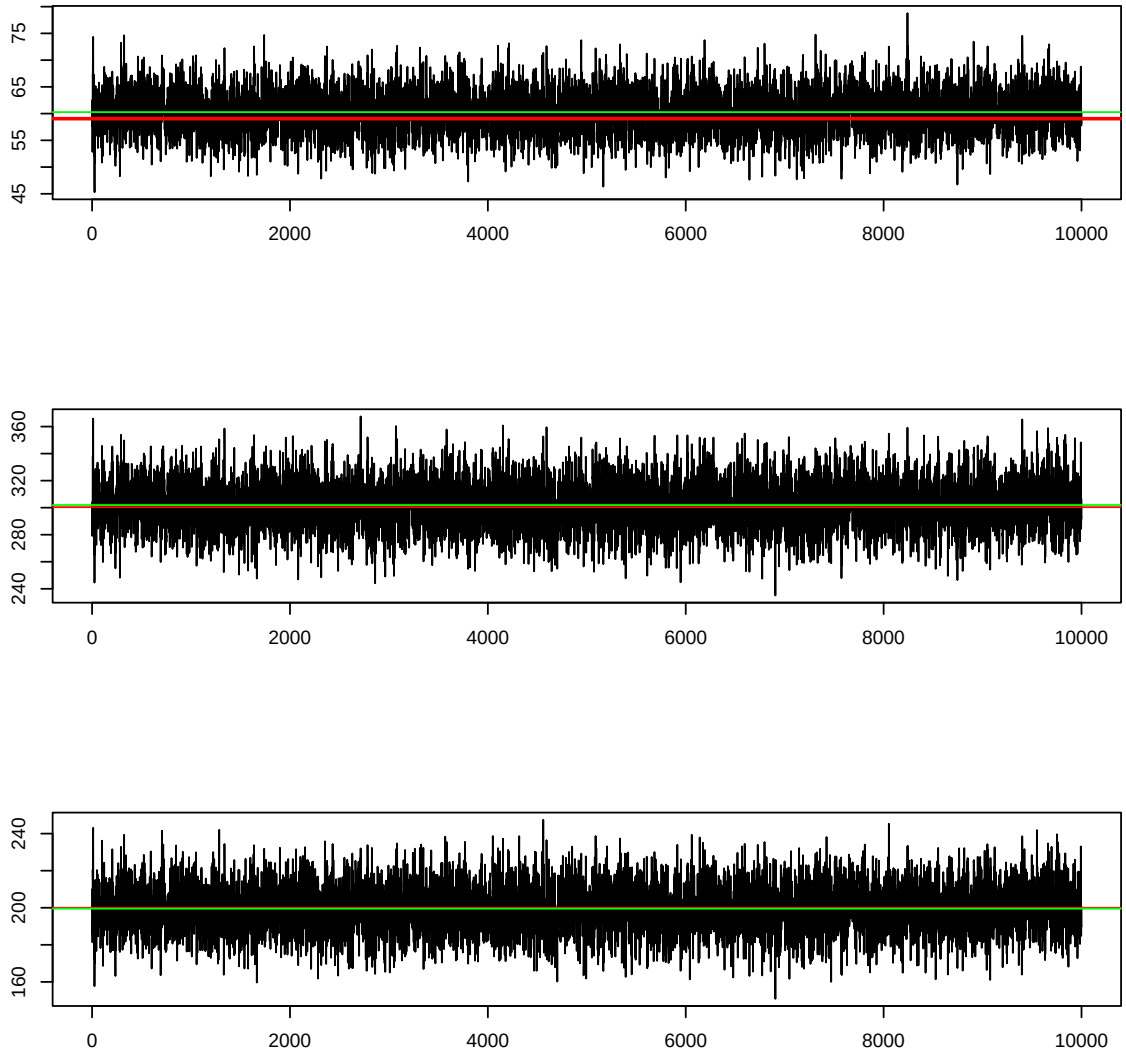

Figure 4: Plot of the computed summary statistic of the first 10,000 simulations of the Metropolis-Hastings algorithm to simulate a Poisson point process. The summary statistic of the sample is depicted with red and the average of the 10,000 simulations is depicted with green.

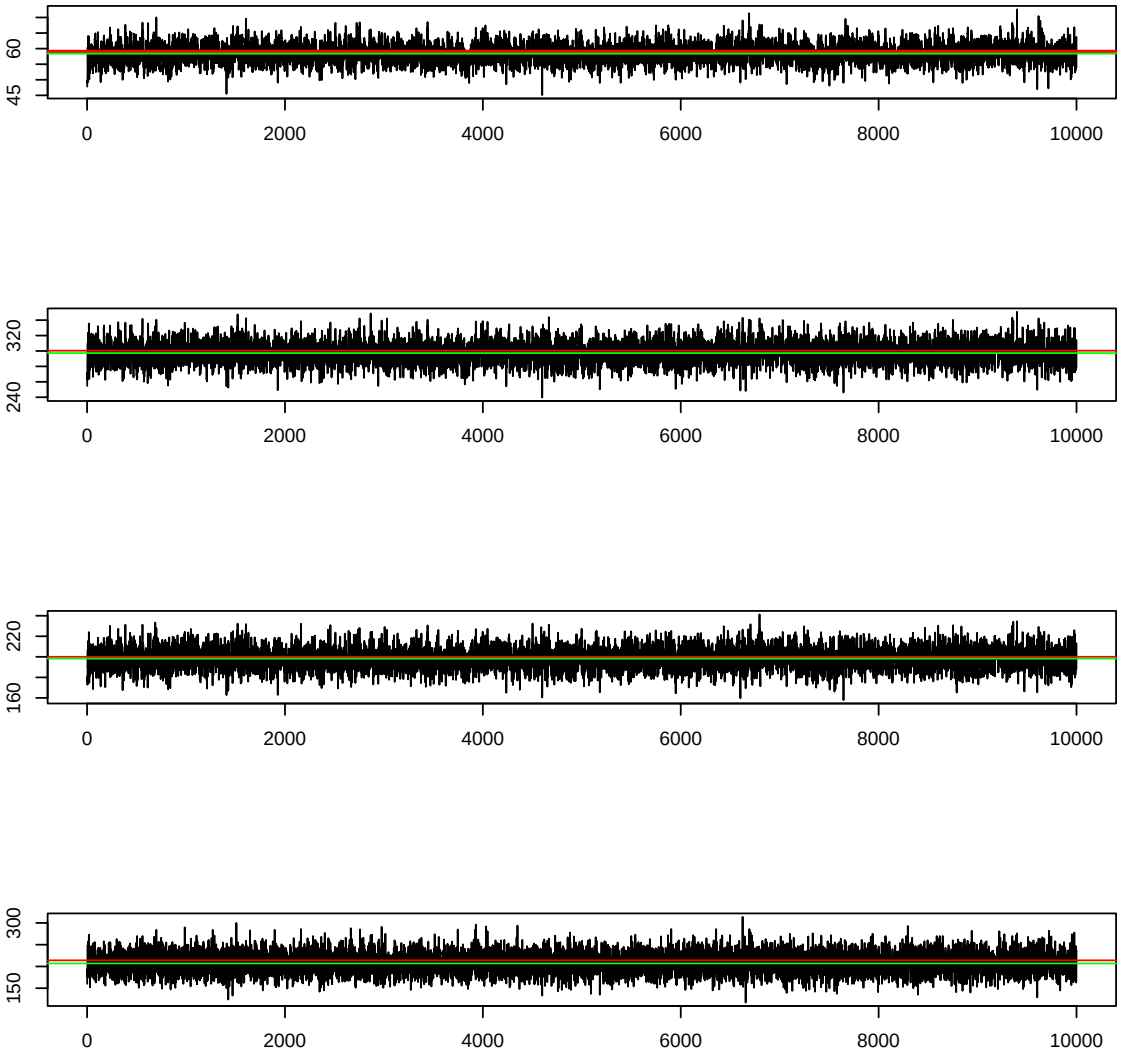

Figure 5: Plot of the computed summary statistic of the first 10,000 simulations of the Metropolis-Hastings algorithm to simulate a Strauss point process. The summary statistic of the sample is depicted with red and the average of the 10,000 simulations is depicted with green.

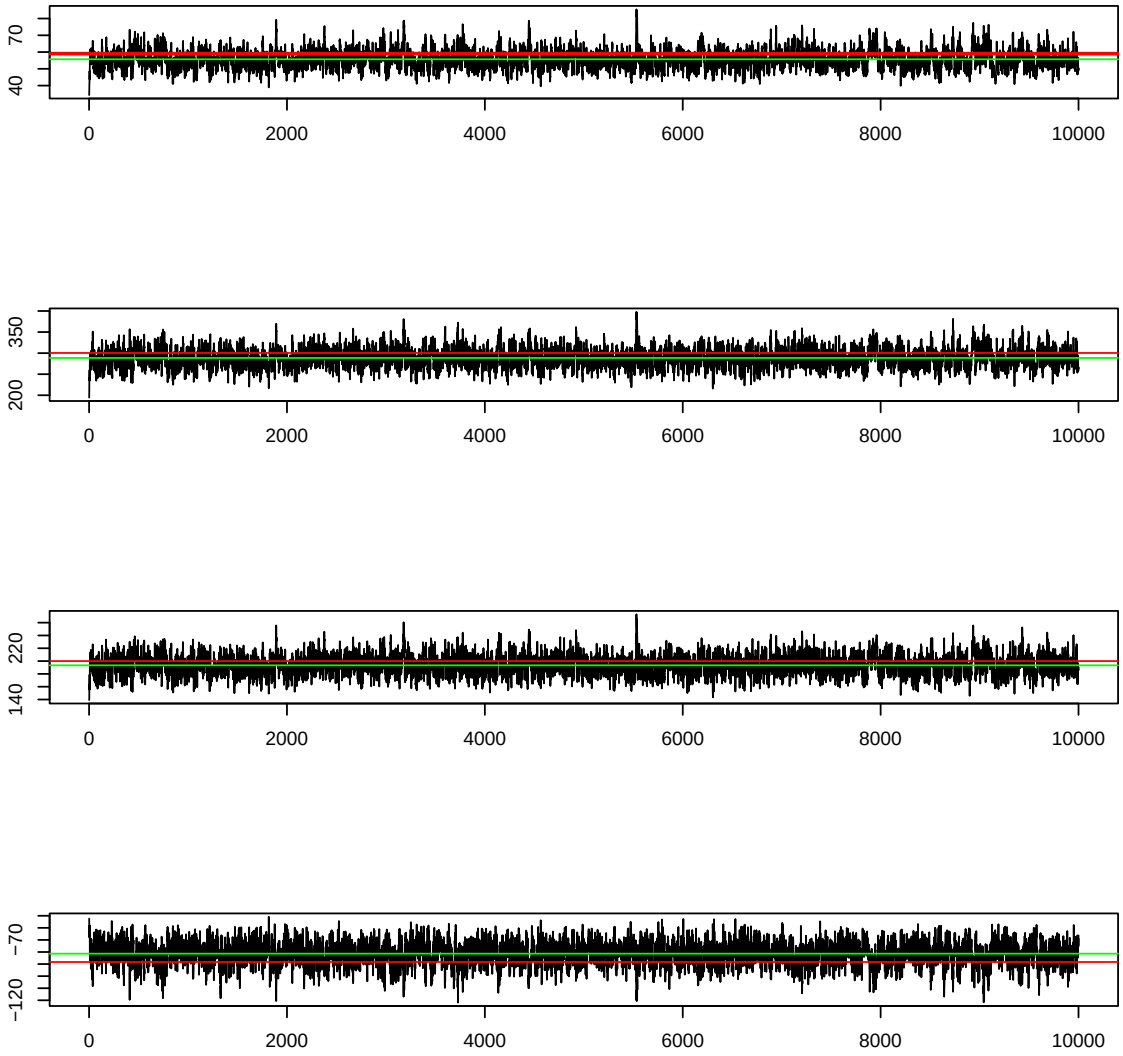

Figure 6: Plot of the computed summary statistic of the first 10,000 simulations of the Metropolis-Hastings algorithm to simulate an Area interaction point process. The summary statistic of the sample is depicted with red and the average of the 10,000 simulations is depicted with green.

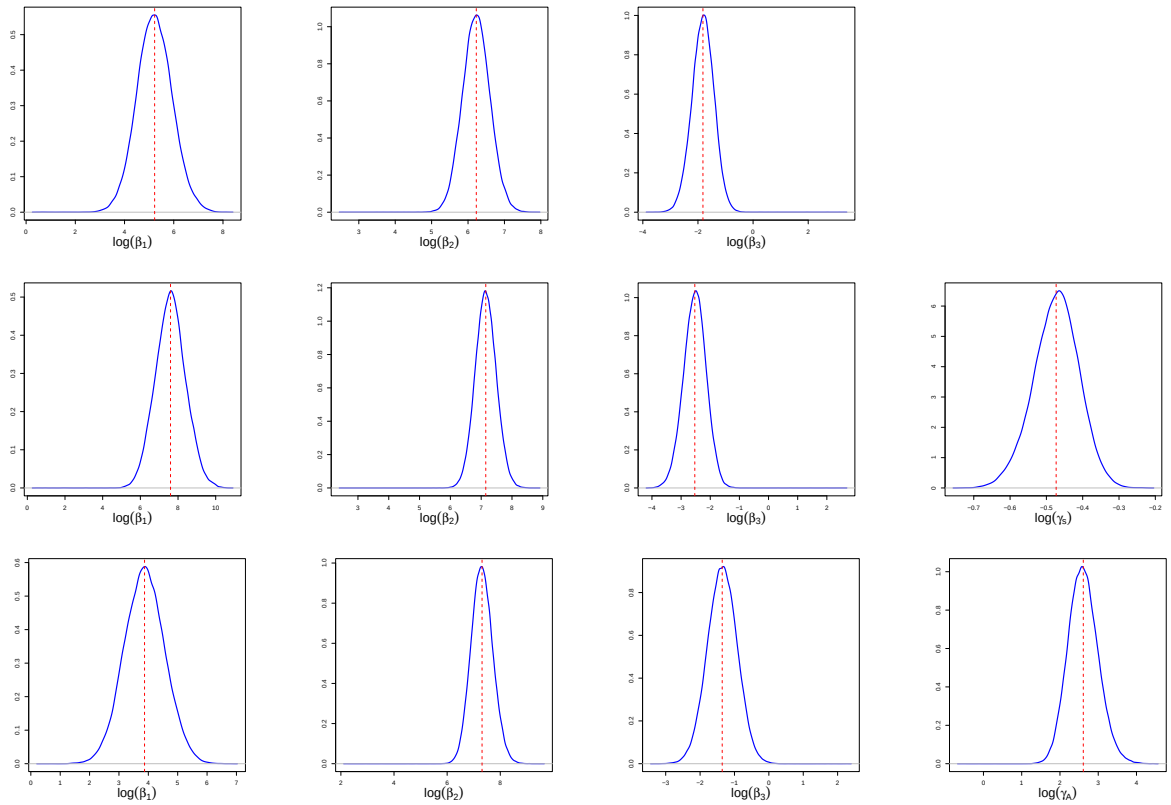

Figure 7: From up to down: estimated posterior density from the sampling of the parameters using ABC Shadow for the Poisson (upper row), Strauss (middle row) and Area-Interaction models (lower row).

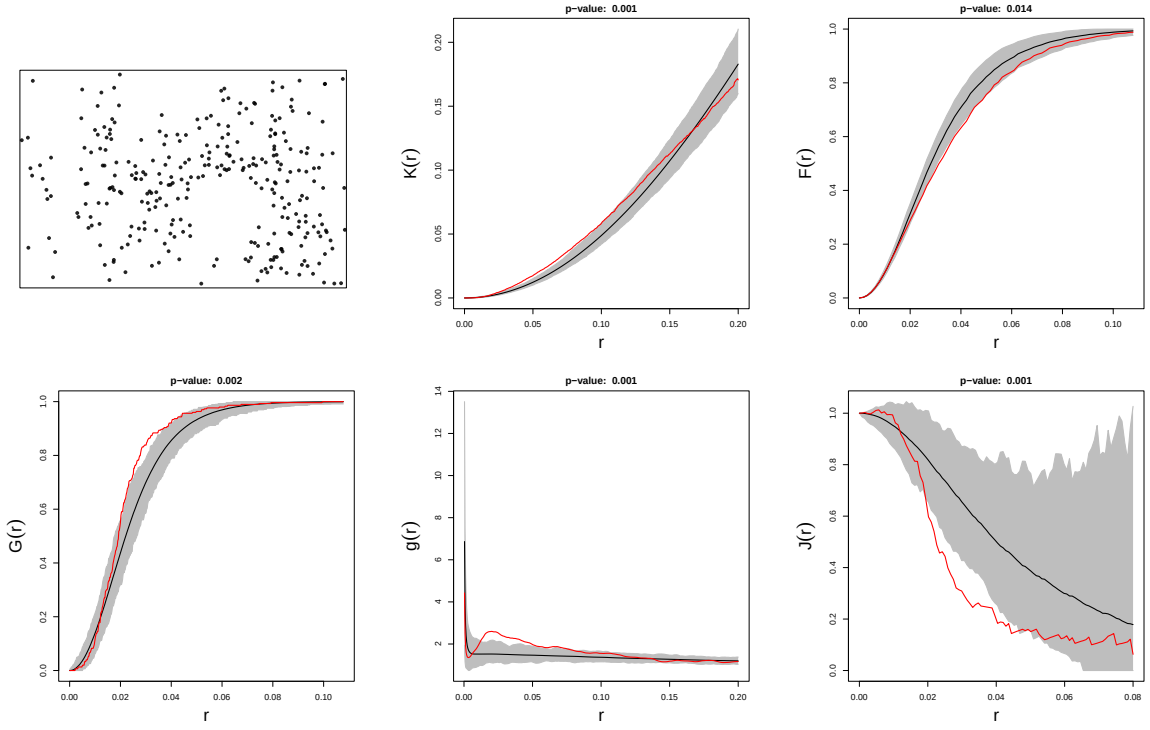

Figure 8: From left to right and then up to down: Simulation of the last point pattern and envelopes for the K, F, G, partial correlation function and J function under the Poisson model.

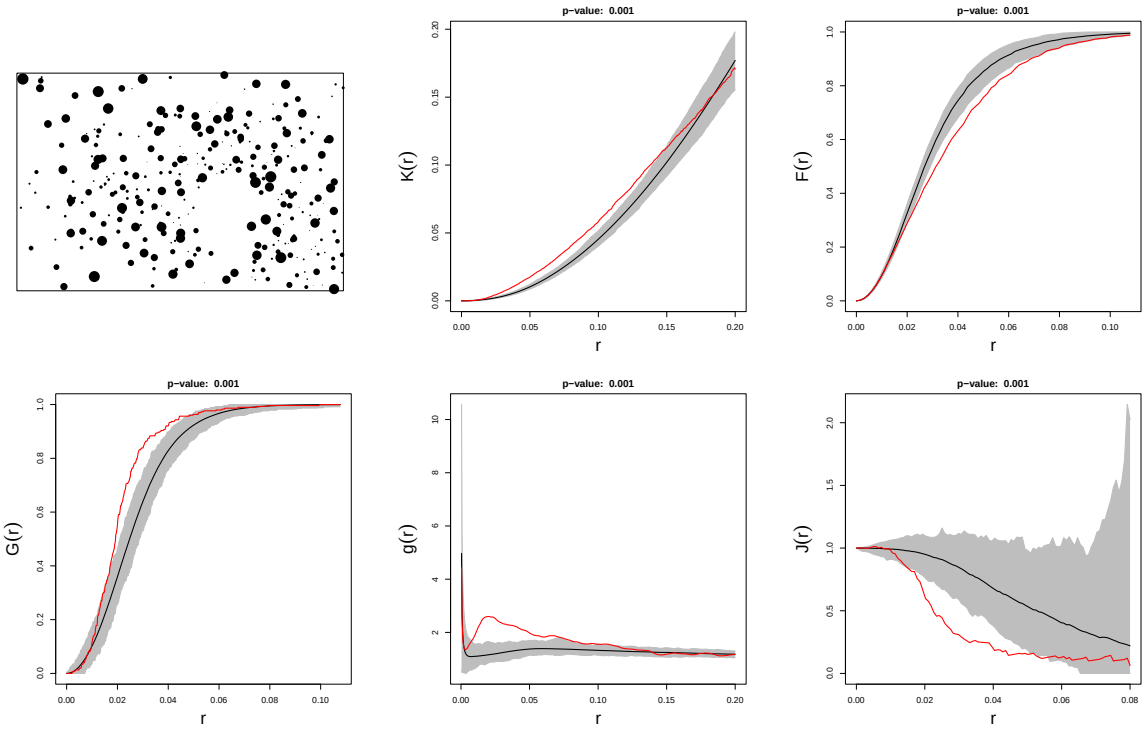

Figure 9: From left to right and then up to down: Simulation of the last point pattern and envelopes for the K, F, G, partial correlation function and J function under the Strauss model.

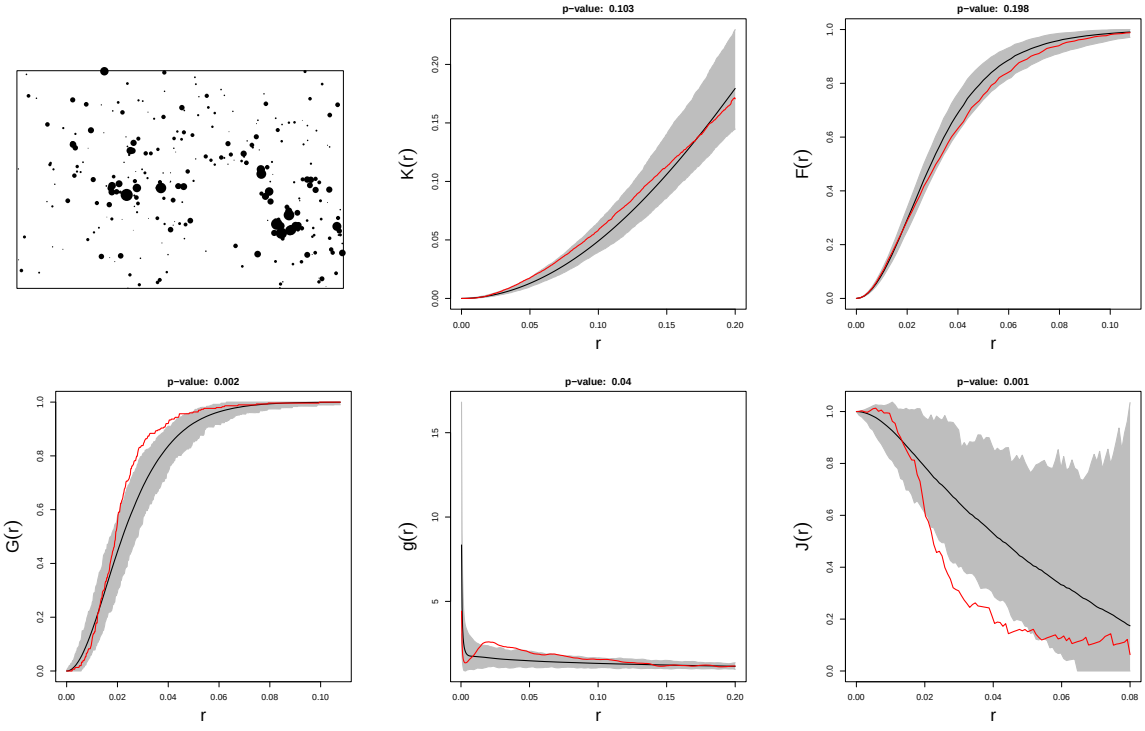

Figure 10: From left to right and then up to down: Simulation of the last point pattern and envelopes for the K, F, G, partial correlation function and J function under the Area interaction model.

(a) Strauss point process

(b) Area-Interaction point process

Figure 11: Envelopes for the radii cumulative distribution build from the simulation of two estimated models.