



# Minimum contrast for the first-order intensity estimation of spatial and spatio-temporal point processes

Nicoletta D'Angelo<sup>1</sup> · Giada Adelfio<sup>1</sup>

Received: 20 March 2023 / Revised: 10 November 2023  
© The Author(s) 2024

## Abstract

In this paper, we harness a result in point process theory, specifically the expectation of the weighted  $K$ -function, where the weighting is done by the true first-order intensity function. This theoretical result can be employed as an estimation method to derive parameter estimates for a particular model assumed for the data. The underlying motivation is to avoid the difficulties associated with dealing with complex likelihoods in point process models and their maximization. The exploited result makes our method theoretically applicable to any model specification. In this paper, we restrict our study to Poisson models, whose likelihood represents the base for many more complex point process models. In this context, our proposed method can estimate the vector of local parameters that correspond to the points within the analyzed point pattern without introducing any additional complexity compared to the global estimation. We illustrate the method through simulation studies for both purely spatial and spatio-temporal point processes and show complex scenarios based on the Poisson model through the analysis of two real datasets concerning environmental problems.

**Keywords** Second-order characteristics · Spatio-temporal point processes · Local models · Minimum contrast estimation

## 1 Introduction

Modeling the complex dynamics of point processes is a fundamental challenge in various fields, ranging from ecology and epidemiology to urban planning. Researchers often deal with the complexities of estimating parameters for specific models due to

---

✉ Nicoletta D'Angelo  
nicoletta.dangelo@unipa.it

Giada Adelfio  
giada.adelfio@unipa.it

<sup>1</sup> Department of Economics, Business and Statistics, University of Palermo, Palermo, Italy

the nature of their likelihood functions. In this paper, we present a novel approach, based on point process result, that simplifies this task.

Typically, a realization of a spatio-temporal point process is characterized via its intensity function, and its parameters are usually fit via the maximum likelihood estimation (MLE) method.

Unfortunately, for many point processes, the integral term on the likelihood is often extremely difficult to compute. Even considering the benchmark Poisson process, many choices have to be made for approximating such integral.

Approximation methods proposed for certain processes, such as Hawkes processes, suggest a computationally intensive numerical integration method, but in general, the problem of computation or estimation of the integral term in the log-likelihood can be burdensome. Furthermore, Poisson likelihood estimation is proved to be consistent, with its estimates being asymptotically normal, asymptotically unbiased, and efficient, under quite general conditions. However, in real-life applications, we often do not have access to extensive datasets, which can make complex to assess rates of convergence. Despite the computational constraints, Maximum Likelihood (ML) remains the most widely used method for estimating the parameters of point process intensities. The significant work by Baddeley et al. (2015) has established the Berman–Turner technique (Berman and Turner 1992) as the predominant approach for fitting parametric Poisson spatial point process models. Furthermore, its spatio-temporal extension continues to be a convenient way to estimate parameters for spatio-temporal point process models, directly building upon purely spatial methodologies.

The scientific literature now recognizes spatial point pattern statistics as a mature discipline, while the spatio-temporal context requires further advancements. Nevertheless, parametric models that depend on external variables present specific challenges. One such challenge is defining the locations of dummy points within the quadrature scheme while respecting the locations of the covariates or, conversely, ensuring knowledge of the covariate values at the selected dummy point locations.

Due to these considerations, while our method is theoretically applicable to various model specifications, this paper confines its scope to Poisson models as a starting proposal for future development. It is important to note that Poisson models cover a diverse range of models, all of which are based on maximizing the Poisson likelihood. The examples covered in this manuscript include homogeneous processes, inhomogeneous processes dependent on both spatial and temporal coordinates, inhomogeneous processes influenced by external covariates, models with spatial and spatio-temporal varying parameters, as well as the estimation of the first-order intensity function in a log-Gaussian Cox process.

In this paper, a novel estimator for the parameters governing spatio-temporal point processes is proposed. Unlike the ML estimator, the proposed estimator does not require the computation or approximation of the computationally expensive integral, as typically found in the point process log-likelihood, making it computationally more efficient.

The proposed parametric estimator is based on the  $K$ -function (Ripley 1976; Gabriel and Diggle 2009) and its deviation from the theoretical value. This technique, either based on the  $K$ -function of the pair correlation function, is commonly referred to as the *minimum contrast* technique. Traditionally, it serves as a convenient model-fitting

procedure for estimating second-order parameters in a class of inhomogeneous spatial point processes. However, in this context, we utilize it to estimate the parameters of a first-order intensity function.

Our method builds upon a key result in point process theory: the expectation of the weighted  $K$ -function based on the true first-order intensity function regardless of the parametric form assumed for the model. The major intuition of our idea relies upon the fact that such  $K$ -function weighted by the true first-order intensity function does not identify a specific model among competitor ones, but, conditionally on the parametric form assumed for the data, it is able to identify the best set of parameters of the specified model. In other words, the weighted  $K$ -function is now used not only to diagnose a set of competing models and consequently to select the best one but also to select the best model among competitor ones with the same parametric specification.

A further notable advantage of our method lies in its ability to exploit local second-order characteristics (Adelfio et al. 2020). Indeed, a model with constant parameters may not adequately represent detailed local variations in the data, since the pattern may present spatial and temporal variations due to the influence of covariates, the scale or spacing between points, and also perhaps due to the abundance of points (D'Angelo et al. 2022). Indeed, a different way of analysing a point pattern can be based on local techniques identifying specific and undiscovered local structure, for instance, sub-regions characterized by different interactions among points, intensity and influence of covariates (D'Angelo et al. 2023a). By considering the local version of the weighted  $K$ -function (D'Angelo et al. 2023b), our approach accurately estimates the vector of local parameters corresponding to specific points within the analyzed point pattern. This level of detail in estimation is crucial for understanding the different variations within spatial and spatio-temporal point processes.

Throughout this paper, we will demonstrate the methodology within the spatio-temporal context. It's important to note that every aspect presented can be straightforwardly reduced to the purely spatial context, as illustrated in both the numerical experiments and the applications to real data.

All the analyses are carried out through the statistical software R Core Team (2023). Section 2 sets the preliminaries of spatio-temporal point processes, their first- and second-order characteristics, and recalls the most used method employed in literature for fitting *global* and *local* Poisson processes. Section 2 introduces the new idea for fitting a general Poisson process model through the minimum contrast procedure (MC) and formalizes the estimation procedure. Section 4 presents numerical experiments to study and assess the performance of the proposed fitting procedure, and Sect. 5 shows more complex challenges through applications to real datasets. The paper ends with some conclusions in Sect. 6.

## 2 Preliminaries

Let us give a simple point process defined in space and time, which is a random countable subset  $X$  of  $\mathbb{R}^2 \times \mathbb{R}$ . Every point  $(u, t) \in X$  corresponds to an event that occurred at a spatial location  $u \in \mathbb{R}^2$  and at time  $t \in \mathbb{R}$ . Its realization is a finite set  $\{(u_i, t_i)\}_{i=1}^n$  of distinct points,  $n \geq 0$  not fixed in advance. This spatio-temporal

realization is assumed to occur within a bounded region  $W \times T \subset \mathbb{R}^2 \times \mathbb{R}$ , with area and length  $|W| > 0, |T| > 0$ .

Any event close in both space and time to a given one  $(u, t)$  can be defined by a spatio-temporal cylindrical neighbourhood of the event for each spatial distance  $r$  and time lag  $h$ . This can be expressed by the Cartesian product  $b((u, t), r, h) = \{(v, s) : ||u - v|| \leq r, |t - s| \leq h\}, (u, t), (v, s) \in W \times T$ , with  $|| \cdot ||$  denoting the Euclidean distance in  $\mathbb{R}^2$ . This will be a cylinder with centre  $(u, t)$ , radius  $r$ , and height  $2h$ .

The Campbell Theorem states that, for any non-negative function  $f$  on  $(\mathbb{R}^2 \times \mathbb{R})^k$ , the following holds

$$\mathbb{E} \left[ \sum_{\zeta_1, \dots, \zeta_k \in X}^{\neq} f(\zeta_1, \dots, \zeta_k) \right] = \int_{\mathbb{R}^2 \times \mathbb{R}} \cdots \int_{\mathbb{R}^2 \times \mathbb{R}} f(\zeta_1, \dots, \zeta_k) \lambda^{(k)}(\zeta_1, \dots, \zeta_k) \prod_{i=1}^k d\zeta_i.$$

This essential result defines one of the main tools of point process theory, i.e. the *product densities*  $\lambda^{(k)}, k \in \mathbb{N}$  and  $k \geq 1$ .

The arguably most important product densities are obtained for  $k = 1$  and  $k = 2$ , called the *intensity function*  $\lambda$  and the *(second-order) product density*  $\lambda^{(2)}$ , respectively.

In short, the intensity function gives the rate of occurrence of events in the given region, and the second-order product density describes the correlation between pairs of points of the pattern.

The pair correlation function

$$g((u, t), (v, s)) = \frac{\lambda^{(2)}((u, t), (v, s))}{\lambda(u, t)\lambda(v, s)}$$

is linked to  $\lambda^{(2)}$ , formally interpretable as the standardized probability density of an event occurring in two small spatio-temporal volumes. This constitutes an important second-order tool, knowing that a Poisson process has  $g((u, t), (v, s)) = 1$ .

### 2.1 Likelihood-based inference for spatio-temporal Poisson point processes

Assuming that the template model is a Poisson process, with a parametric intensity or rate function  $\lambda(u, t; \theta), u \in W, t \in T$ , with parameters  $\theta \in \Theta$ , the log-likelihood is

$$\log L(\theta) = \sum_i \lambda(u_i, t_i; \theta) - \int_W \int_T \lambda(u, t; \theta) dt du. \tag{1}$$

In practice, intensity models of log-linear form  $\lambda(u, t; \theta) = \exp(\theta Z(u, t))$ , are often considered, with  $Z(u, t)$  a spatio-temporal covariate function, including the space or time coordinates themselves.

The most direct approach to fitting this model is to adopt the method described by Berman and Turner (1992), which involves employing a finite quadrature approximation for the log-likelihood. It is actually the default implemented in the `spatstat`

package (Baddeley and Turner 2005), and for this reason, this approach is widely recognized as the standard method for fitting Poisson spatial models.

Renaming the data points as  $x_1, \dots, x_n$  with  $(u_i, t_i) = x_i$  for  $i = 1, \dots, n$ , then  $m$  additional *dummy points*  $(u_{n+1}, t_{n+1}) \dots, (u_{m+n}, t_{m+n})$  are generated, to form a set of  $n + m$  quadrature points, where  $m$  is commonly taken larger than  $n$ . Then, some *quadrature weights*  $a_1, \dots, a_m$  are determined so that integrals in Eq. (1) can be approximated by a Riemann sum  $\int_W \int_T \lambda(u, t; \theta) dt du \approx \sum_{k=1}^{n+m} a_k \lambda(u_k, t_k; \theta)$ . The quadrature weights  $a_k$  are taken such that  $\sum_{k=1}^{n+m} a_k = l(W \times T)$ , with  $l$  the Lebesgue measure. Then the log-likelihood in Eq. (1) of the template model can be approximated as

$$\log L(\theta) \approx \sum_k a_k (y_k \log \lambda(u_k, t_k; \theta) - \lambda(u_k, t_k; \theta)) + \sum_k a_k, \tag{2}$$

taking  $y_k = e_k/a_k$ , with the indicator  $e_k$  equaling 1 if  $u_k$  is a data point and 0 otherwise. Apart from the constant  $\sum_k a_k$ , this expression is formally equivalent to the weighted log-likelihood of a Poisson regression model. This connection to the log-likelihood of a Poisson regression model makes the use of standard Generalized Linear Models (GLM) software possible to maximize it, which significantly contributes to its widespread popularity. However, many choices have to be made in order to define the spatio-temporal quadrature scheme. The first one regards the spatio-temporal partition of  $W \times T$  into cubes  $C_k$  of equal volume  $v$ , and assigning the weight  $a_k = v/n_k$  to each quadrature point (dummy or data) where  $n_k$  is the number of points that lie in the same cube as the point  $u_k$ .

The number of dummy points should be sufficient for an accurate estimate of the likelihood, but at the moment of writing, there aren't guidelines on this aspect. Only Raeisi et al. (2021) and D'Angelo et al. (2023a), studying more complex models than the Poisson one, start with a number of dummy points  $m \approx 4n$ , increasing it until  $\sum_k a_k = l(W \times T)$ .

Sometimes, however, a model with constant parameters may not adequately represent detailed local variations in the data (D'Angelo et al. 2022). A local estimation approach should accurately estimate the vector of local parameters corresponding to specific points within the analyzed point pattern (D'Angelo et al. 2023a). This level of detail in estimation can reveal to be crucial for understanding the observed variations within space and time.

Assume now that the template model is a Poisson process, with a parametric intensity or rate function  $\lambda(u, t; \theta_i)$  with space and time locations  $u \in W, t \in T$  and parameters  $\theta_i \in \Theta$ , with  $i$  the data point index. Estimation can still be performed through the fitting of a GLM using a localized version of the quadrature scheme just introduced.

In order to obtain the local estimates  $\hat{\theta}_i$ , the local log-likelihood associated with the spatio-temporal location  $(v, s)$  can be written as

$$\begin{aligned} \log L((v, s); \theta) &= \sum_i w_i (u_i - v, t_i - s) \lambda(u_i, t_i; \theta) \\ &\quad - \int_W \int_T \lambda(u, t; \theta) w_i (u_i - v, t_i - s) dt du \end{aligned}$$

where for instance  $w_i(v, s) = w_{\sigma_s}(v - u_i)w_{\sigma_t}(s - t_i)$ . In this case,  $w_{\sigma_s}$  and  $w_{\sigma_t}$  are weight functions, and  $\sigma_s, \sigma_t > 0$  are their smoothing bandwidths. It is not necessary to assume that  $w_{\sigma_s}$  and  $w_{\sigma_t}$  are probability densities. For simplicity, one might consider only kernels of fixed bandwidth, even though spatially adaptive kernels could also be used. Note that if the template model is the homogeneous Poisson process with intensity  $\lambda$ , then the local likelihood estimate of  $\hat{\lambda}(v, s)$  reduces to the kernel estimator of the point process intensity (Diggle 2013) with kernel proportional to  $w_{\sigma_s}w_{\sigma_t}$ . A similar approximation of that used in Eq. (2) for the local log-likelihood associated with each desired location  $(v, s) \in W \times T$  can therefore be used as follows

$$\log L((v, s); \theta) \approx \sum_k w_k(v, s)a_k(y_k \log \lambda(u_k, t_k; \theta) - \lambda(u_k, t_k; \theta)) + \sum_k w_k(v, s)a_k. \tag{3}$$

We refer to D'Angelo et al. (2023a) for further details, but basically, for each desired location  $(v, s)$ , one replaces the vector of quadrature weights  $a_k$  by  $a_k(v, s) = w_k(v, s)a_k$ , and consequently can still use the GLM software to fit the Poisson local regression.

### 2.2 The spatio-temporal K-function and its estimator

Gabriel and Diggle (2009) define the spatio-temporal inhomogeneous  $K$ -function and propose a non-parametric estimator.

**Definition 1** A point process defined in space and time is second-order intensity reweighted stationary and isotropic if its intensity function is bounded away from zero and its pair correlation function depends only on the spatio-temporal difference vector  $(r, h)$ , where  $r = \|u - v\|$  and  $h = |t - s|$ .

**Definition 2** For a second-order intensity reweighted stationary, isotropic spatio-temporal point process, the spatio-temporal inhomogeneous  $K$ -function is

$$K(r, h) = 2\pi \int_0^r \int_0^h g(r', h')r'dr'dh'$$

where  $g(r, h) = \lambda^{(2)}(r, h)/(\lambda(u, t)\lambda(v, s))$ ,  $r = \|u - v\|$ ,  $h = |t - s|$ .

The most widely used and simplest estimator of the spatio-temporal  $K$ -function is:

$$\hat{K}(r, h) = \frac{1}{|W||T|} \sum_{i=1}^n \sum_{j>i}^n \mathbf{1}(\|u_i - u_j\| \leq r, |t_i - t_j| \leq h). \tag{4}$$

A homogeneous Poisson process has  $\mathbb{E}[\hat{K}(r, h)] = \pi r^2 h$ , regardless the first-order intensity  $\lambda$ . The spatio-temporal  $K$ -function represents a useful tool to measure

interaction and clustering in space and time. The estimator  $\hat{K}(r, h)$  is commonly compared to its theoretical counterpart  $\mathbb{E}[\hat{K}(r, h)] = \pi r^2 h$ . Values  $\hat{K}(r, h) > \pi r^2 h$  suggest spatio-temporal clustering of points, while  $\hat{K}(r, h) < \pi r^2 h$  suggests a regular pattern.

The inhomogeneous version of the  $K$ -function in Eq. (4) (Gabriel and Diggle 2009) is

$$\hat{K}_I(r, h) = \frac{|W||T|}{n(n-1)} \sum_{i=1}^n \sum_{j>i} \frac{\mathbf{1}(|u_i - u_j| \leq r, |t_i - t_j| \leq h)}{\hat{\lambda}(u_i, t_i)\hat{\lambda}(u_j, t_j)}. \tag{5}$$

Also, for the inhomogeneous case  $\mathbb{E}[\hat{K}_I(r, h)] = \pi r^2 h$ , when the weighting intensity is the true one. This represents a very important result in the spatio-temporal point process theory since it allows the usage of the weighted estimator  $\hat{K}_I(r, h)$  as a diagnostic tool for a general fitted first-order intensity function  $\lambda(\cdot, \cdot)$ . In other words, it can be used for assessing the goodness-of-fit of spatio-temporal point processes with any fitted first-order intensity function. In practice, if the fitted intensity is close enough to the true one, its expectation should be close to the theoretical Poisson one  $\mathbb{E}[\hat{K}(r, h)] = \pi r^2 h$ . Therefore, values  $\hat{K}_I(r, h)$  greater than  $\pi r^2 h$  indicate that the model is not a good fit since the distances among points exceed those of the theoretical Poisson.

In Adelfio et al. (2020), local versions of both the homogeneous and inhomogeneous spatio-temporal  $K$ -functions are provided, as diagnostic tools accounting also for local characteristics. They define an estimator of the intensity by  $\hat{\lambda} = n/(|W||T|)$ , and then propose expressing the localized version of Eq. (4) for the  $i$ -th event  $(u_i, t_i)$  as

$$\hat{K}^i(r, h) = \frac{1}{\hat{\lambda}^2 |W||T|} \sum_{(\mathbf{u}_i, t_i) \neq (v, s)} \mathbf{1}(|u_i - v| \leq r, |t_i - s| \leq h) \tag{6}$$

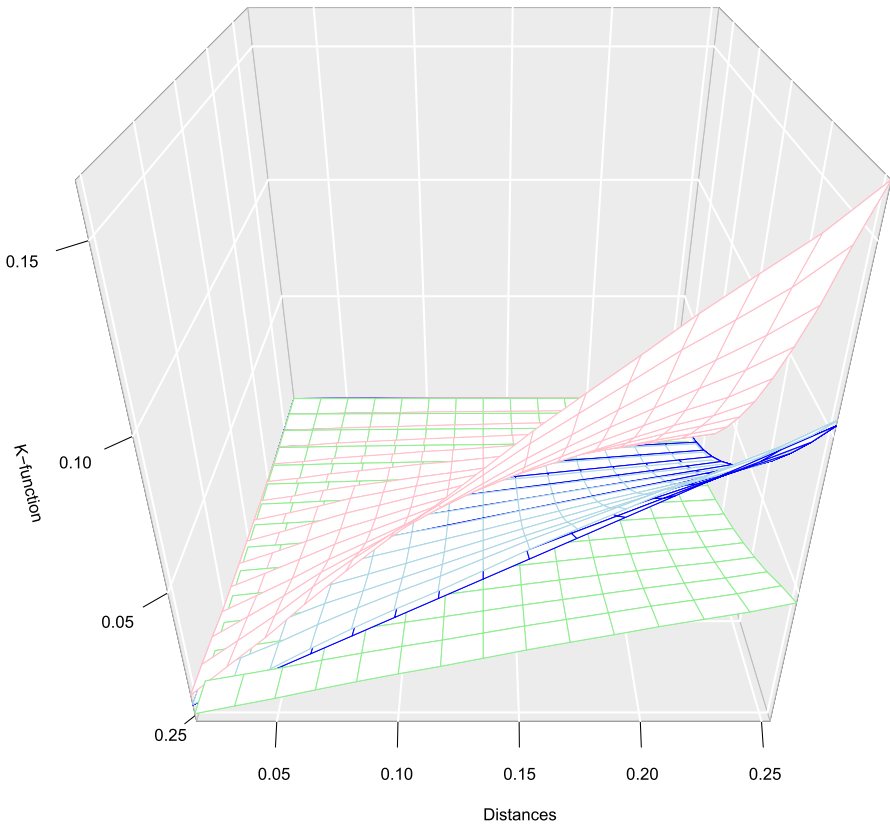
and the local version of Eq. (5) as

$$\hat{K}_I^i(r, h) = \frac{1}{|W||T|} \sum_{(\mathbf{u}_i, t_i) \neq (v, s)} \frac{\mathbf{1}(|u_i - v| \leq r, |t_i - s| \leq h)}{\hat{\lambda}(u_i, t_i)\hat{\lambda}(v, s)}, \tag{7}$$

with  $(v, s)$  being any other point's spatial and temporal coordinates. They further proved that also the local inhomogeneous estimators behave as the corresponding theoretical Poisson ones, i.e. the expectation of Eqs. (6) and (7) is  $\pi r^2 h$  as well.

### 3 Minimum contrast for first-order intensity estimation

In this section, we present the rationale behind our intuition for employing the  $K$ -function as an inferential tool for estimating model parameters in a minimum contrast approach. In particular, we refer to a graphical representation of a straightforward example, the homogeneous Poisson process model, characterized by a constant intensity.



**Fig. 1** In blue: the theoretical  $K$ -function of a simulated Poisson process with 500 points. In light blue: the estimated  $K$ -function, weighted by the true intensity function. In pink and green: the estimated  $K$ -functions, weighted by some wrong intensities ( $\lambda = \{400, 750\}$ ). (Color figure online)

Our primary insight arises from the realization that the  $K$ -function, when weighted by the true first-order intensity function, can serve as a tool for identifying the optimal set of parameters for the assumed parametric model, as opposed to the traditional approach of selecting the best model among competing alternatives.

This idea is shown graphically in Fig. 1.

The blue surface of Fig. 1 represents the theoretical  $K$ -function of a simulated spatio-temporal Poisson process with 500 points. As an example, we estimate three  $K$ -functions, respectively weighted by: the true intensity (in light blue) and two wrong constant intensities (400 in pink and 750 in green). From these plots, we can observe that the overall behaviour of the  $K$ -function is the same (i.e. increasing with the space and time lags), with the only difference in the magnitude values of the  $K$ -functions. In particular, the  $K$ -function weighted by the true intensity function reports the values closer to the theoretical one.

Indeed, the theory suggests that the squared difference between the observed  $K$ -function and the theoretical one should approach zero, as the intensity used for



weighting the observed  $K$ -function approaches the true one (Adelfio et al. 2020). The value of the sum of the squared differences are 0.07, 0.0002 and 0.06 for the  $K$ -functions weighted by 400, 500 and 750, respectively. As expected, the lowest value is obtained when weighting for the true intensity function.

We now formalize the proposed method of parameter estimation using  $K_I(r, h)$  in Eq. (4) and its estimator  $\hat{K}_I(r, h)$  of Eq. (5).

Let be given a point process model with the intensity  $\lambda(u, t; \theta)$  (in brief:  $\lambda_\theta$ ) with a vector of parameters  $\theta \in \Theta$ . The proposed *minimum contrast for first-order intensity estimation* (MC) procedure is defined by the minimization of the following objective function

$$\mathcal{M}(\theta) = \int_{h_0}^{h_{max}} \int_{r_0}^{r_{max}} \phi(r, h) \{(\hat{K}_I(r, h; \lambda_\theta)) - \pi r^2 h\}^2 dr dh \tag{8}$$

with respect to  $\theta$ , providing a vector of estimates  $\hat{\theta}$ :

$$\hat{\theta} = \arg \min_{\theta \in \Theta} \mathcal{M}(\theta).$$

Here  $r_0, h_0, r_{max}$  and  $h_{max}$  are the lower and upper space and time lag limits of the contrast criterion, and  $\phi(r, h)$  is a weight that depends on the spatio-temporal distance. Note that our MC proposal poses the basis on the assumption that  $\pi r^2 h$  is the expected value of the  $K$ -function when this is weighted by the true intensity function. Given the model assumed for the data, the combination of parameters leading to a good intensity (that is, a small discrepancy in Eq. (8)) may not be unique. This might result in biased (and unreliable) estimates. We refer to Baddeley et al. (2022) for a detailed classification of causes for these practical difficulties, with particular reference to the purely spatial case.

For this reason, we employ a further step in our proposal, by a radial penalization presented in Kreutz (2018). That work aimed at addressing the feasibility of unique parameter estimation in dynamic systems. Numerical optimization is used to test the uniqueness of parameters by means of a penalty in the radial direction which has the goal to enforce the displacement of the parameters. The suggested method is based on a comparison of the objective function of common fitting with a penalized fit pulling the parameter vector away from the first estimate. A major characteristic of their method is that it allows identifiability analysis of all parameters in a joint manner, as well as the possibility to investigate the identifiability of each parameter individually. Their proposed approach enables quick testing for parameter identifiability, whereas the several other approaches proposed in the literature are typically computationally demanding, difficult to perform and/or not applicable in many application settings. Indeed, their method is suited for any model with an optimization-based parameter estimation such as *maximum likelihood* and *least-squares*. This is the reason why it appears applicable to our purposes, particularly addressing the joint identifiability investigation of all model parameters.

**Definition 3** Let  $\theta \in \Theta$  be the parameter vector containing all  $p$  unknown constants in the model assumed for the data. A parameter  $\theta_j$  ( $j = 1, \dots, p$ ) is said to be structurally

locally identifiable, if, for almost any  $\theta_j$ , there exists a neighbourhood  $P$  such that if

$$\theta \in P \quad \wedge \quad g(\theta_j^{(1)}) = g(\theta_j^{(2)}) \quad \Rightarrow \quad \theta_j^{(1)} = \theta_j^{(2)}$$

“Almost any” means for all parameters except for isolated points. If this property holds not only within a neighbourhood but also for the whole parameter space, the parameter is termed structurally globally identifiable (Chis et al. 2011).

Therefore, alternatively to estimating  $\theta$  according to Eq. (8), we suggest the use of a penalized objective function

$$\mathcal{M}_{tot}^R(\theta) = \mathcal{M}(\theta) + \mathcal{M}_{pen}^R(\theta) \tag{9}$$

with

$$\mathcal{M}_{pen}^R(\theta) = \frac{1}{R^2} \left( \sqrt{\sum_j (\theta_j - \hat{\theta}_j)^2} - R \right)^2, \tag{10}$$

where  $1/R^2$  is the tuning parameter representing the penalization strength. In essence, the penalty acts as an extra data point that is utilized to pull the parameter in the direction where the data is least informative.

The penalty term  $\mathcal{M}_{pen}^R(\theta)$  has its minimum at a sphere with radius  $R$  centered around  $\hat{\theta}$ . The final estimates are found as

$$\theta^* = \arg \min_{\theta \in \Theta} \mathcal{M}_{tot}^R(\theta).$$

### 3.1 Selection of the radius R

Any penalized approach requires a tuning parameter which controls the degree of shrinking of the model coefficients. Traditional selection criteria for regression problems involve selecting the lowest goodness-of-fit criterion among competitors. However, the main problem when dealing with residual analysis for point processes is to find a correct definition of residuals since the one used in dependence models can not be used for point processes (Adelfio et al. 2020).

An alternative approach to defining a weighted second-order statistic is the usage of the smoothed raw residuals (Baddeley et al. 2005).

As our whole fitting method is based on a weighted second-order statistic, we choose to employ the smoothed raw residuals for the selection of the radius  $R$ .

The predicted number of points occurring in a given spatial region  $W$  is equal to  $\int_W \hat{\lambda}_\theta(u) du$ , with  $\hat{\lambda}_\theta(u)$  the intensity of a model fitted to an inhomogeneous Poisson process. Consequently, the *raw residual process* (Baddeley et al. 2005) in each region  $W \subset \mathbb{R}^2$  can be defined as the number of points which fall in  $W$ , denoted as

$$r(W) = n(\mathbf{x} \cap W) - \int_W \hat{\lambda}_\theta(u) du. \tag{11}$$

Here  $\mathbf{x}$  denotes the observed realization of a purely spatial point pattern, and  $n(\mathbf{x} \cap W)$  is its number of points falling in  $W$ . Increments of  $r(W)$  are analogous to the raw residuals (observed minus fitted values) in a linear model. The adequacy of the fitted model can be checked by inspecting whether  $r(W) \approx 0$  and various plots and transformations of  $r(W)$  can be useful diagnostics for a fitted point process model. The resulting residuals can be displayed easily by smoothing them. Hence, we can define the *smoothed residual fields* as

$$s(u) = \tilde{\lambda}_\theta(u) - \lambda^\dagger(u) \quad (12)$$

with  $\tilde{\lambda}_\theta(u)$  a non-parametric estimate of the fitted intensity, usually estimated through kernel procedures. A common practice is to select the smoothing bandwidth for the kernel estimation of the raw residuals by cross-validation as the value that minimizes the Mean Squared Error (MSE) criterion defined by Diggle (1985), by the method of Berman and Diggle (1989). This is because, among the possible alternatives proposed in the literature, the cross-validation method is typically the most adaptive, and, therefore, the one which should resemble the true intensity function the most. This is particularly relevant when assessing the goodness-of-fit of a parametric fitted intensity function. See Diggle (2013) for further details. On the other hand,  $\lambda^\dagger(u)$  is a smoothed version of the estimated intensity of the fitted model. Note that the smoothed residual fields procedure is intended for parametric specifications of the fitted intensity function. Given these, of course, smaller differences in Eq. (12) indicate that the fitted model is close to the real one. For this reason, we choose the best model among competitors as the one with the lowest values of the smoothed raw residuals.

A note is in order. Raw residuals in Eq. (11) are proposed by Baddeley et al. (2000). Whereas previous works (Lawson 1993; Stoyan and Grabarnik 1991) have defined diagnostic values for the data points only, these residuals are also ascribed to locations which are not points of the pattern. This is related to an important methodological issue for point processes. In a point pattern dataset, the observed information isn't limited to the locations of the observed points within the pattern. The absence of points in other locations also provides valuable information. With the additional advantage of graphical presentation, smoothed raw residuals become a straightforward and effective diagnostic tool.

We, therefore, select the radius  $R$  as the one minimizing the integration of the smoothed raw residuals over the analysed area:

$$\hat{R} = \arg \min_R \int_W (\tilde{\lambda}(u) - \lambda_R^\dagger(u)) du, \quad (13)$$

where  $\lambda_R^\dagger(u) = \lambda^\dagger(u; \hat{\theta}_R)$ , which is the intensity obtained by imputing the estimated parameter vector  $\hat{\theta}$  with the  $R$  radius value in the penalization procedure.

Moving to the spatio-temporal context, fixed bandwidths for spatial or spatio-temporal data based on the maximal smoothing (over-smoothing) principle of Terrell (1990) can be employed. The optimal values minimize the asymptotic mean integrated squared error assuming normally distributed data (Silverman 1986). Two separate values can be obtained for the purely spatial and temporal components by independently

applying the normal scale rule to the spatial and temporal margins of the supplied data. Alternatively, bandwidth selection for standalone spatio-temporal density/intensity can be based on either unbiased least squares cross-validation (LSCV), likelihood (LIK) cross-validation, or bootstrap estimation of the MISE, providing an isotropic scalar spatial bandwidth and a scalar temporal bandwidth.

### 3.2 Local extension

Our proposal can also be extended in a local context, representing an alternative to the most standard local likelihood procedure, which consists of many choices, including the quadrature scheme but also the weighting functions for the spatio-temporal kernels (as in Eq. (3)). Suppose then that the model incorporates a vector of parameters  $\theta_i$  for each point  $i$ . Let  $\hat{K}_I^i(r, h; \lambda_\theta)$  denote the local estimators calculated from the data.

For each point indexed by  $i$  we consider

$$\mathcal{M}_{local}(\theta_i) = \int_{h_0}^{h_{max}} \int_{r_0}^{r_{max}} \phi(r, h) \{(\hat{K}_I^i(r, h; \lambda_\theta) - \pi r^2 h)\}^2 dr dh.$$

Then, we can obtain a vector of estimates  $\hat{\theta}_i$ , one for each point  $i$ , as

$$\hat{\theta}_i = \arg \min_{\theta \in \Theta} \mathcal{M}_{local}(\theta_i).$$

In practice, one could also wish to specify a different weight function  $\phi(r, h)$  for each point  $i$ , making the objective function as follows

$$\mathcal{M}_{local}(\theta_i) = \int_{h_0}^{h_{max}} \int_{r_0}^{r_{max}} \phi_i(r, h) \{(\hat{K}_I^i(r, h; \lambda_\theta) - \pi r^2 h)\}^2 dr dh.$$

Finally, the penalized optimization could also be implemented locally, giving rise to the estimated parameters

$$\theta_i^* = \arg \min_{\theta \in \Theta} \mathcal{M}_{local;tor}^R(\theta_i),$$

where the component  $\mathcal{M}_{local;pen}^R(\theta_i)$  could depend on a fixed  $R$  value, or either on an individual one  $R_i$  for each point.

### 3.3 Cox processes

Cox processes are point process models typically used when observing clustering among points of the analysed pattern. In general, any Cox model can be estimated by a two-step procedure, involving first the first-order intensity and then the cluster or correlation parameters.

First, a Poisson process with a particular model for the log-intensity is fitted to the point pattern data, providing the estimates of the coefficients of all the terms that characterize the intensity.

Then, the estimated intensity is taken as the true one and the cluster or correlation parameters are estimated using either the *method of minimum contrast* (Pfanzagl 1969; Eguchi 1983; Diggle 1979; Diggle and Gratton 1984; Møller et al. 1998; Davies and Hazelton 2013; Siino et al. 2018), *Palm likelihood* (Ogata and Katsura 1991; Tanaka et al. 2008), or *composite likelihood* (Guan 2006).

The most common technique for this second stage is the *minimum contrast*, and it is the method which we shall refer to here.

Log-Gaussian Cox processes are one of the most prominent clustering models. By specifying the intensity of the process and the moments of the underlying Gaussian Random Field (GRF), it is possible to estimate both the first and second-order characteristics of the process. Following the inhomogeneous specification in Diggle et al. (2013), a LGCP for a generic point in space and time has the intensity  $\Lambda(u, t) = \lambda(u, t) \exp(S(u, t))$  where  $S$  is a Gaussian process with  $\mathbb{E}(S(u, t)) = \mu = -0.5\sigma^2$  and so  $\mathbb{E}(\exp S(u, t)) = 1$  and with variance and covariance matrix  $\mathbb{C}(S(u_i, t_i), S(u_j, t_j)) = \sigma^2\gamma(r, h)$  under the stationary assumption, with  $\gamma(\cdot)$  the correlation function of the GRF, and  $r$  and  $h$  some spatial and temporal distances. Following Møller et al. (1998), the first-order product density and the pair correlation function of an LGCP are  $\mathbb{E}(\Lambda(u, t)) = \lambda(u, t)$  and  $g(r, h) = \exp(\sigma^2\gamma(r, h))$ , respectively.

Driven by a GRF, controlled in turn by a specified covariance structure, the implementation of the LGCP framework in practice requires a proper estimate of the intensity function. Usually, this is achieved through the maximization of the Poisson likelihood in Eq. (1), with all the challenges already introduced. We substitute the fitting of the first-order intensity function with the proposed minimum contrast procedure, making the *two-step minimum contrast estimation procedure* as follows:

- The intensity parameters  $\theta$  are estimated by minimizing either  $\mathcal{M}(\theta)$  in Eq. (8) or  $\mathcal{M}_{tot}^R(\theta)$  in Eq. (9), depending on the necessity (instead of the maximising likelihood of the Poisson process with intensity  $\lambda(\cdot; \theta)$ )
- The interaction/covariance parameters  $\psi$  are estimated by minimizing the discrepancy between a second-order summary statistics (either the pcf or  $K$ -function) and its theoretical value under the assumed covariance function.

In particular, we propose to perform the second step relying on the *joint minimum contrast* (Siino et al. 2018) procedure for obtaining the interaction parameters  $\psi$

$$\mathcal{M}_J(\psi) = \int_{h_0}^{h_{max}} \int_{r_0}^{r_{max}} \phi(r, h) \{ \hat{J}(r, h) - J(r, h; \psi) \}^2 dr dh,$$

with  $\hat{J}(r, h)$  the estimate of the second-order summary statistics and  $J(r, h; \psi)$  its theoretical value depending on the functional form assumed for the covariance structure.

**Table 1** Means, MSE, and SE values of the estimates obtained over 1000 simulations for the purely spatial scenarios

Process	$\mathbb{E}[n]$	True par	Mean	MSE	SE
Homogeneous Poisson	500	$\theta_0 = 6.21$	6.21	0.002	0.03
Inhomogeneous Poisson	500	$\theta_1 = 8.34$	8.64	2.81	0.69
Inhomogeneous Poisson	500	$\theta_0 = 2$	3.37	2.12	20.06
		$\theta_1 = 6$	4.03	2.06	23.71
Inhomogeneous Poisson $R \in [0.25, 10]$ Eq. (14)	500	$\theta_0 = 2$	2.01	2.83	0.79
		$\theta_1 = 6$	5.96	2.81	1.52
Inhomogeneous Poisson $R = 2.5$	500	$\theta_0 = 2$	1.94	2.89	0.87
		$\theta_1 = 6$	6.02	2.89	1.52
Inhomogeneous Poisson $R \in [0.25, 10]$ Eq. (13)	500	$\theta_0 = 2$	1.81	3.01	0.91
		$\theta_1 = 6$	6.22	3.04	1.69

We would like to emphasize that in this paper, particularly in the simulation studies, we execute the first step using the proposed procedure based on the  $K$ -function, and the second step through the pair-correlation function.

## 4 Simulation results

In this section, we report some experimental results, for illustrating the proposed estimation procedure, both for the space and spatio-temporal case. Note indeed that the whole theory introduced in Sect. 3 is easily implementable for the purely spatial case, as the theory regarding the properties of global and local weighted second-order statistics holds the same. We refer to Adelfio et al. (2020) and references therein.

### 4.1 Space

We simulate 1000 spatial point patterns from three different purely spatial point processes in the unit square with 500 points on average following these scenarios:

1. homogeneous Poisson process with constant intensity  $\lambda = \exp(\theta_0)$ ;
2. inhomogeneous Poisson process with intensity  $\lambda(x, y) = \exp(\theta_1 x)$ ;
3. inhomogeneous Poisson process with intensity  $\lambda(x, y) = \exp(\theta_0 + \theta_1 x)$ .

Table 1 shows the results of the proposed minimum contrast estimation procedure, reporting the means, MSE, and standard errors (SE) of the estimates obtained for the 1000 replications. The spatial lags in the  $K$ -function are 153 values ranging between 0 and 1/4 of the maximum distance.

The procedure performs effectively when the parameter to be estimated is unique. However, challenges arise in the third scenario, likely stemming from issues related to parameter identifiability. Indeed, as anticipated in Sect. 3, the minimum of the objective function in Eq. (8) is not unique, but the combination of the estimated parameters

represents the best fit in terms of intensity. Therefore, we proceed by adding a penalty to the already employed objective function only for the third scenario.

Having simulated from known parameters, we choose the radius  $R$  to be used in the penalization procedure as the one minimizing the discrepancy between the true parameters and the estimated ones. Formally:

$$\hat{R} = \arg \min_R \sum_j \frac{(\theta_j - \hat{\theta}_{j,R})^2}{\hat{\theta}_{j,R}}. \quad (14)$$

We explore different ranges for  $R$ , omitted for brevity in Table 1, and we note that the choice of the  $R$  range, in the minimization of Eq. (14), does not seem to be relevant. Indeed, the penalization procedure clearly overcomes the identifiability problem previously encountered. Moreover, we can spot smaller standard errors than the unpenalized version. Another relevant result is that the mean of all the selected  $R$  values over the 1000 simulations, i.e.  $R = 2.5$ , leads to comparable results. We interpret this numerical result as an indication that an optimal value of  $R$  should exist, which is related to the combination of the true parameters.

Naturally, in real data applications, one does not have the knowledge of the true parameter values, making it impossible to carry out the minimization as described in Eq. (14). Therefore, we proceed by employing the tuning parameter selection criterion proposed in Sect. 3.1. It is important to note that further investigations could be conducted to assess the performance of the  $R$  selection procedure as outlined in Sect. 3.1. However, at the time of writing, the testing of the  $R$  selection procedure falls outside the scope of this research.

## 4.2 Space–time

Moving to the spatio-temporal context, we simulate 1000 spatio-temporal point patterns in the unit cube with 500 points on average from the following point processes:

1. homogeneous Poisson process with constant intensity  $\lambda = \exp(\theta_0)$ ;
2. inhomogeneous Poisson process with intensity  $\lambda(x, y, t) = \exp(\theta_0 + \theta_1 x)$ .

The spatial and temporal distances used in the observed weighted  $K$ -functions are 15 values ranging from 0 to  $r_{max}$ , equal to 1/4 of the maximum (spatial or temporal) distances.

Table 2 reports the means, MSE, and SE of the estimates of the two considered processes, obtained over 1000 simulations.

We notice that the mean of the intensity function for the homogeneous scenario appears systematically overestimated. For the inhomogeneous point processes, we employ the penalized procedure with radius  $R = 2.5$  since, as noticed in the purely spatial case, this tends to make the estimated parameters more similar to the true values than in the unpenalized case.

**Table 2** Means, MSE, and SE values of the estimates obtained over 1000 simulations for the spatio-temporal scenarios

Process	$\mathbb{E}[n]$	True par	Mean	MSE	SE
Homogeneous Poisson	500	$\theta_0 = 6.21$	6.27	0.006	2.62
Inhomogeneous Poisson	500	$\theta_0 = 2$	2.85	2.38	1.19
$R = 2.5$		$\theta_1 = 6$	4.89	2.34	1.52

**Table 3** Mean and quartiles of the distributions of the estimated local parameters, averaged over 100 simulated point patterns

Process	$\mathbb{E}[n]$	True par	25%	50% (MSE)	Mean (MSE)	75%
Inhomogeneous Poisson	500	$\theta_0 = 2$	- 1.28	3.55 (2.10)	2.59 (2.49)	6.08
		$\theta_1 = 6$	0.95	4.34 (2.10)	5.09 (2.34)	9.50
Inhomogeneous Poisson	500	$\theta_0 = 2$	0.95	2.49 (2.51)	3.09 (2.22)	4.44
		$\theta_1 = 6$	2.87	5.39 (2.45)	6.12 (2.96)	9.08

MSE values come in parentheses

#### 4.2.1 Local space–time

Section 3.2 introduced the further advantage of our proposal, showing the possibility of fitting local parameters. Preliminary analyses not reported for brevity showed a systematic overestimation of the local parameters in the homogeneous case. Moving to the inhomogeneous scenario, we show results averaged over 100 simulated patterns in Table 3.

The means and medians are similar to the true values of the parameters, and the variability of the local estimates is smaller when adding the penalty. For the considered simulation scenarios we highlight that the fitted intensity, obtained imputing the average of the local parameters, greatly resembles the true one.

#### 4.3 Convergence study

Having assessed the performance of the proposed procedure, we now proceed to investigate the impact of varying sample sizes in the simulated patterns to gain a deeper understanding of the convergence behavior.

At this scope, Tables 4 and 5 contain the means, MSE, and SE of the estimates obtained over 100 simulations of both homogeneous and inhomogeneous processes, for three sample sizes:  $\mathbb{E}[n] = \{250, 500, 750\}$ .

Table 4 reports results for the purely spatial scenarios, while Table 5 contains those of the spatio-temporal one. In both cases, we employed the  $R = 2.5$  penalization for the inhomogeneous processes, as it proved to be the optimal value in the previous simulations.

We have established that the most favorable scenario for our proposal is the purely spatial homogeneous one, characterized by minimal bias and limited variability in the



**Table 4** Means, MSE, and SE of the estimates obtained over 100 simulations for the purely spatial scenarios

Process	$\mathbb{E}[n]$	True par	MC			MLE		
			Mean	MSE	SE	Mean	MSE	SE
Homogeneous Poisson	250	$\theta_0 = 5.52$	5.51	0.004	0.03	5.51	0.004	0.06
	500	$\theta_0 = 6.21$	6.21	0.002	0.03	6.22	0.002	0.04
	750	$\theta_0 = 6.62$	6.62	0.001	0.03	6.62	0.001	0.04
Inhomogeneous Poisson $R = 2.5$	250	$\theta_0 = 1.3215$	1.49	0.26	0.95	1.31	0.10	0.34
		$\theta_1 = 6$	5.59	0.54	1.44	6.00	0.14	0.40
	500	$\theta_0 = 2$	1.89	0.13	0.87	1.95	0.05	0.24
		$\theta_1 = 6$	6.09	0.21	1.52	6.05	0.07	0.29
	750	$\theta_0 = 2.42$	2.27	0.11	0.79	2.41	0.04	0.19
		$\theta_1 = 6$	6.29	0.21	1.59	6.02	0.06	0.23

**Table 5** Means, MSE, and SE of the estimates obtained over 100 simulations for the spatio-temporal scenarios

Process	$\mathbb{E}[n]$	True par	MC			MLE		
			Mean	MSE	SE	Mean	MSE	SE
Homogeneous Poisson	250	$\theta_0 = 5.52$	5.59	0.01	2.33	5.59	1.06	0.06
	500	$\theta_0 = 6.21$	6.27	0.006	2.62	6.26	0.13	0.04
	750	$\theta_0 = 6.62$	6.68	0.005	2.79	6.69	0.006	0.04
Inhomogeneous Poisson $R = 2.5$	250	$\theta_0 = 1.3215$	2.43	2.21	0.78	1.84	0.33	0.33
		$\theta_1 = 6$	4.56	3.46	0.78	5.37	0.49	0.39
	500	$\theta_0 = 2$	2.76	0.92	0.72	2.53	0.33	0.23
		$\theta_1 = 6$	5.01	1.58	0.74	5.33	0.50	0.27
	750	$\theta_0 = 2.42$	3.18	0.94	0.71	2.97	0.33	0.19
		$\theta_1 = 6$	5.01	1.65	0.74	5.32	0.50	0.22

estimates. The sample size primarily impacts the MSE, which naturally decreases with the increasing number of points in the simulated pattern, as expected.

In contrast, the spatio-temporal estimates of the homogeneous intensity exhibit a higher level of bias, providing a slight overestimation of the single parameter.

Focusing to the inhomogeneous case, we observe that both the spatial and spatio-temporal scenarios display some degree of bias. Among these scenarios, the one with 500 points stands out as the most favorable. This is likely due to the value of  $R$  set to 2.5, which was actually obtained from simulations with that sample size.

However, the MSE value shows minimal improvement as the number of points increases, moving from 500 to 750 points, whether in spatial or spatio-temporal contexts.

Moving to the comparison with MLE results, we can note there are differences in the standard errors and performance of these methods in different scenarios. In the spatial homogeneous case, MLE and MC results are pretty similar and MLE standard

errors are slightly higher than the MC ones. In the homogeneous spatio-temporal case, MLE standard errors are quite smaller than MC standard errors, that is, MLE provides more precise and reliable parameter estimates compared to MC in this particular scenario. In inhomogeneous cases (both spatial and spatio-temporal), we can still note a difference in standard errors between MLE and MC but of lower magnitude, suggesting that MLE still outperforms MC in terms of parameter estimation, but the difference is not as pronounced as in the homogeneous spatio-temporal case. Finally, the better performance in spatial contexts is attributed to the additional complexity of the temporal component in the spatio-temporal scenarios. In summary, as expected, it seems that MLE performs slightly better, mostly for scenarios involving spatio-temporal complexity, as it provides more precise parameter estimates with smaller standard errors compared to MC. However, our proposal may represent a promising and valid alternative, especially when considering the computational complexity of the specific statistical methods and observed data.

#### 4.4 External covariates

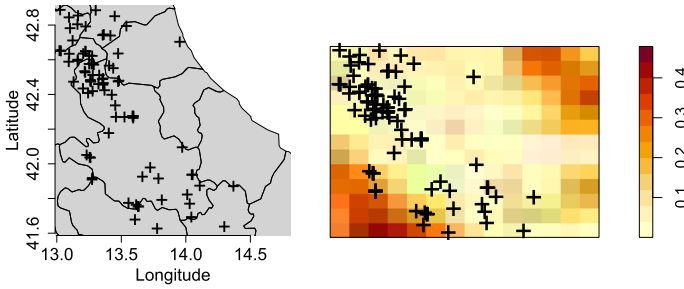
This section addresses the further challenge of estimating the first-order intensity function depending on external covariates. In spatial point process theory, these are known as spatial covariates, and they pose additional challenges with respect to the most common homogeneous or inhomogeneous Poisson processes since, for computational feasibility, their value must be known theoretically at each location.

In real data analysis, external covariates, often representing environmental phenomena, are not collected in the same detail as the observed pattern.

This implies that the standard quadrature scheme must either be customized to align with the locations of covariates or, conversely, external covariate values must be interpolated at the positions of both data and dummy points. Naturally, this requirement can complicate the implementation of the quadrature scheme method, particularly as the number of covariates increases, considering that each covariate may potentially be gathered at distinct sites.

Here we illustrate an example of a spatial point pattern whose realization is assumed to depend on an external covariate, to prove the applicability of our method. It's important to highlight that a notable advantage of our proposed method is that precise knowledge of covariate values is required only at the data point locations. The advantage of this approach is its capability to treat both marks and spatial covariates in a consistent manner within the linear predictor of the fitted first-order intensity function.

The Italian catalogue considered in this section is downloaded from the Istituto Nazionale di Geofisica e Vulcanologia (INGV) archive. As done in D'Angelo et al. (2022), we focus on the seismic sequence of the Abruzzo region. Therefore, the analysed earthquakes that occurred between May 2012 and May 2016 in Abruzzo are displayed in the left panel of Fig. 2, consisting of 85 events with 2.5 as the threshold magnitude. On the right panel of Fig. 2, we display the spatial covariate, whose effect on the intensity of earthquakes we are interested in studying: the distance from the nearest seismic station, henceforth denoted  $D_{ns}(u)$ .



**Fig. 2** Left panel: Earthquakes occurred in Abruzzo between May 2012 and May 2016, consisting of 85 events with 2.5 as the threshold magnitude. Right panel: Spatial covariate representing the distance to the nearest seismic station

**Table 6** Estimates of model for the Italian earthquake data obtained by MLE and MC procedures

	MLE		MC	
	Estimate	SE	Estimate	SE
Intercept	4.83	0.20	5.36	0.35
Distance	- 0.08	0.02	- 0.06	0.05

We therefore proceed by fitting the following model

$$\lambda(u) = \exp(\theta_0 + \theta_1 D_{ns}(u)). \tag{15}$$

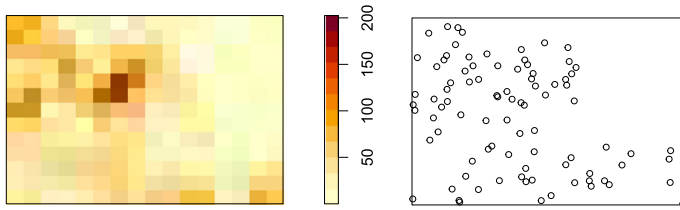
Table 6 contains the estimates of the model for the Italian earthquake data obtained by both MLE and the MC procedures.

As expected, the MC procedure tends to estimate higher values for the intercept. Furthermore, both MLE and MC estimate a negative covariate’s coefficient. A negative value of the coefficient of the spatial covariate is reasonable since usually, as the distance from the nearest station recording the event increases, the intensity decreases, as detection of earthquakes is usually more accurate if they occur not far from the network station. This result is in line with those in D’Angelo et al. (2022). To corroborate these results and reinforce the numerical experiments already shown, we run simulations in the presence of an external spatial covariate.

In particular, we simulate 100 realizations from the model in Eq. (15), with the MLE estimated parameters as the true ones, that is, with  $\theta_0 = 4.83$  and  $\theta_1 = - 0.08$ . Figure 3 shows the obtained intensity function used to simulate the patterns (left panel) and an example of a simulated pattern (right panel).

Table 7 shows the result for this scenario and also those where the number of points has increased by the double and the quadruple. In these scenarios, the true coefficient of the spatial covariate is kept fixed, and the number of points is increased by adding the logarithm of 2 and 4, respectively, to the true intercept.

The MC procedure accurately estimates values close to the true ones. Hence, again, its performance appears to be robust and relatively independent of the sample size.



**Fig. 3** Left panel: Intensity function used to simulate the patterns. Right panel: One pattern simulated according to the intensity function

**Table 7** Means, MSE, and SE values obtained over 100 simulations from the model in Eq. (15), both for MC and MLE methods, for three different sample sizes

	Mean	MC (R = 0.5)			MLE			E [n]
		Mean	MSE	SE	Mean	MSE	SE	
Intercept	4.83	4.81	0.03	0.19	4.63	0.08	0.34	87.33
Distance	- 0.08	- 0.08	0.0002	0.02	- 0.06	0.0006	0.03	
Intercept	5.52	5.5	0.02	0.14	5.45	0.35	0.35	175.81
Distance	- 0.08	- 0.08	0.0001	0.01	- 0.08	0.01	0.03	
Intercept	6.21	6.19	0.01	0.10	6.29	0.55	0.34	349.11
Distance	- 0.08	- 0.08	0.00008	0.008	- 0.09	0.01	0.03	

We note that both methods in Eqs. (14) and (13), independently select the same value of  $R = 0.5$ , providing strong evidence that our selection process is robust and reliable.

A further notice regards the comparison between MC and MLE estimates. The MC estimates exhibit lower bias, with smaller MSE and standard errors when compared to the MLE method. This difference in performance may be due to the application of the radial penalty, with the penalty radius R optimized for this specific case, enhancing the accuracy of the MC estimates compared to the MLE method in this particular scenario.

### 4.5 Log-Gaussian Cox processes

This section is devoted to the assessment of the two-step minimum contrast procedure introduced in Sect. 3.3, specifically tailored for LGCPs.

We consider a scenario used in Siino et al. (2018), used there to compare the performance with that of the then-proposed joint minimum contrast (JMC) estimation method.

The objective is to investigate the possible differences in the estimates of interaction parameters when the summary statistic used in the second stage is weighted by a first-order intensity function estimated through our proposed method.

We consider a separable structure for the covariance function of the GRF (Brix and Diggle 2001) that has exponential form for both the spatial and the temporal compo-

**Table 8** Estimates' means and MSE values of 200 simulated log-Gaussian Cox processes with 1000 expected number of points, obtained by both the MLE and MC procedure at the first step, and minimum contrast based on JMC at the second

	True values	MLE		MC	
		Mean	MSE	Mean	MSE
$\sigma^2$	5	6.54	5.03	6.57	3.88
$\alpha$	0.05	0.08	0.001	0.05	0.0002
$\beta$	2	2.43	0.92	1.79	0.40

ments,  $\mathbb{C}(r, h) = \sigma^2 \exp(-r/\alpha) \exp(-h/\beta)$ , where  $\sigma^2$  is the variance,  $\alpha$  is the scale parameter for the spatial distance and  $\beta$  is the scale parameter for the temporal one. 200 point patterns are generated with  $n = 1000$  expected number of points in the spatio-temporal window  $W \times T = [0, 1]^2 \times [0, 50]$ , with constant first-order intensity equal to  $b = \log(n/|W \times T|) = 20$ .

We consider a moderate degree of clustering in the processes with variance  $\sigma^2 = 5$  and scale parameters in space and time,  $\alpha = 0.10$  and  $\beta = 2$ . The mean of the GRF is fixed  $\mu = -0.5\sigma^2$ .

Table 8 reports the estimates' means and MSE values of the 200 simulated log-Gaussian Cox processes obtained with the MLE and MC procedure at the first step and the JMC procedure at the second one.

These results are indeed quite promising, especially when considering the second-order parameter estimates obtained using the MC method in the first step. While these estimates differ from the MLE results, they still provide comparable outcomes. Indeed, the variance parameter tends to be overestimated; however, the spatial and temporal range parameter estimates are way more precise compared to those obtained using MLE in the first stage. This improvement can be attributed to the estimated value of the first-order constant intensity function, which has an average of 29.33 when using the MC method, higher than the value of 20 estimated by MLE.

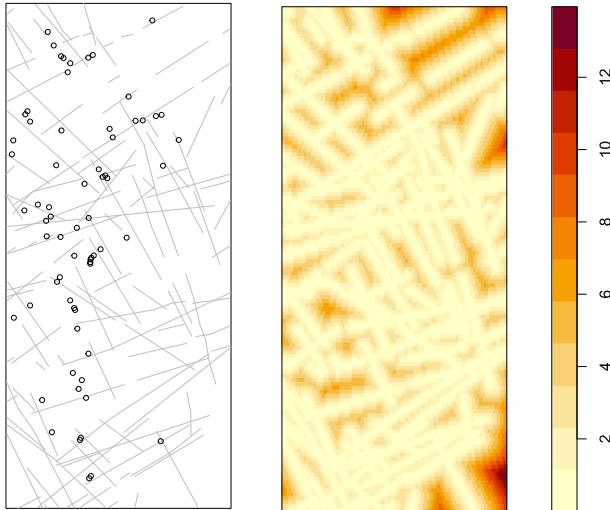
## 5 Applications to real data

This section is dedicated to the analysis of real datasets using the newly proposed inferential framework, with a specific emphasis on the local characteristics of the point patterns.

### 5.1 Analysis of copper data

We examine the Berman–Huntington points and lines dataset, also analyzed by Baddeley (2017). The origins and analysis of these data were first presented by Berman (1986) and have since been explored by Berman and Diggle (1989), Berman and Turner (1992), Baddeley and Turner (2000), Foxall and Baddeley (2002), Baddeley et al. (2005), to cite a few.

These data were collected during an extensive geological survey of a region measuring  $70 \times 158$  km in central Queensland, Australia. The dataset comprises 67 points rep-



**Fig. 4** Left panel: Berman–Huntington points and lines dataset. Black points are the locations of copper ore deposits, and grey lines are the geological lineaments. Right panel: The available covariate for the copper data: *Distance from the faults* ( $D_f(u)$ ), computed as the Euclidean distances from the spatial location  $u$  of events and the map of geological information

**Table 9** Estimates of model for the copper data in Eq. (16) obtained by MLE and MC procedures

	MLE		MC	
	Estimate	SE	Estimate	SE
Intercept	− 4.93	0.18	− 4.52	0.0003
Distance from the faults	− 0.10	0.08	− 0.11	0.0001

resenting copper ore deposits and 146 line segments depicting geological lineaments (left panel of Fig. 4).

Lineaments, visible on satellite imagery, are linear features primarily believed to be geological faults (Berman 1986). Typically, the focus lies on predicting copper deposits based on the pattern of these lineaments, which can be readily observed in satellite images.

For this reason, we construct the spatial covariate *Distance from the faults* ( $D_f$ ), computed as the Euclidean distances from the spatial location  $u$  of events and the map of geological information (Baddeley et al. 2015). The covariate surface is displayed in the right panel of Fig. 4.

We proceed by fitting the following model for the copper ore deposit intensity

$$\lambda(u) = \exp(\theta_0 + \theta_1 D_f(u)). \quad (16)$$

The estimates and their uncertainty are reported in Table 9.

The difference between the unpenalized and the penalized procedure is negligible, both in terms of estimates and standard errors, suggesting that the unpenalized proce-

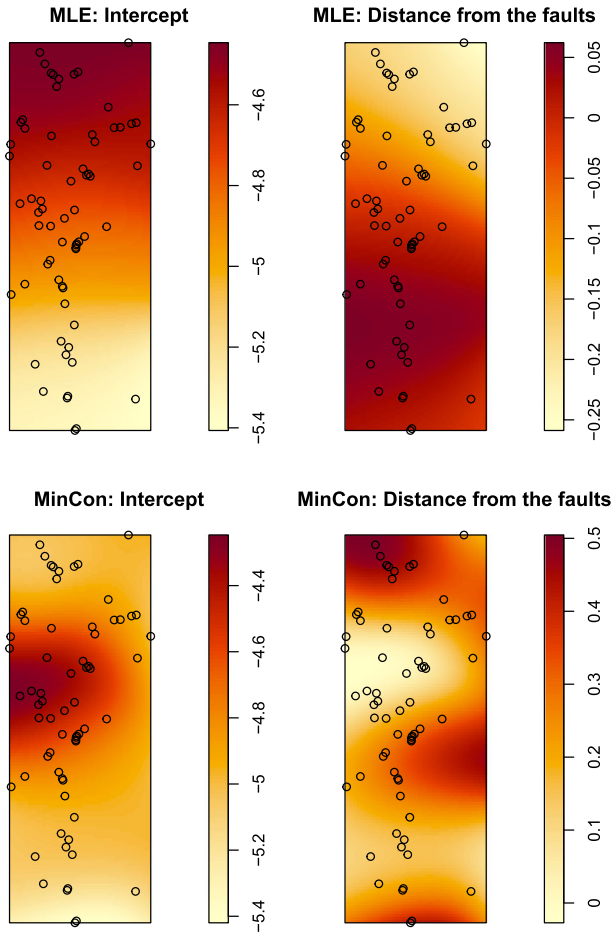


Fig. 5 Smoothed local parameters. Top panels: MLE estimates. Bottom panels: Minimum contrast estimates

ture suffices for this particular scenario. We want to reiterate the noteworthy similarity between the MC estimates and the MLE estimates.

Finally, we also compare the MC local estimation approach with that of Baddeley (2017). This means to fit the following model, with space-varying parameters

$$\lambda(u) = \exp(\theta_0(u) + \theta_1(u)D_f(u)). \tag{17}$$

Note that this model differs from the one in Eq. (16) since both the parameters are indexed by the spatial location  $u$ .

Figure 5 reports the smoothed local estimates for both parameters and both estimation procedures.

These findings highlight the higher variability of the MC estimates in comparison to the MLE estimates. Furthermore, the relatively smoother MLE estimates can be

**Table 10** Estimates' values (SE in brackets) of the log-Gaussian Cox processes applied to the Greek earthquake data with the MLE and MC procedure at the first step

	MLE Estimate (SE)	MC Estimate (SE)
$\lambda$	0.006 (0.030)	0.007 (0.000)
$\sigma^2$	8.15	10.95
$\alpha$	0.23	0.18
$\beta$	83.97	51.69

attributed to the underlying kernel technique employed in the log-likelihood function, which is optimized during the MLE process. On the other hand, the MC procedure tends to identify more distinct and separated regions in individual point spatial domains. However, it's important to note that a substantial portion of the analyzed region, such as the right-center part, exhibits different values of intensity. This variation is likely attributed to the presence of kernel smoothing artefacts, which can lead to localized inconsistencies in the estimated intensity values.

## 5.2 Analysis of Greek seismicity

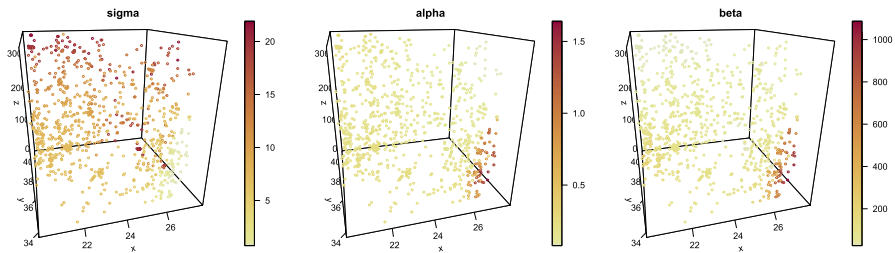
For this last scenario, we consider the same data analysed in D'Angelo et al. (2022), related to 1111 earthquakes that occurred in Greece between 2005 and 2014, coming from the Hellenic Unified Seismic Network (H.U.S.N.), with the specific aim of fitting both a global and a local version of Log-Gaussian Cox process (LGCP) model. Indeed, D'Angelo et al. (2023a) proposed a local version of spatio-temporal log-Gaussian Cox processes by using Local Indicators of Spatio-Temporal Association (LISTA) functions plugged into the minimum contrast procedure to obtain space as well as time-varying parameters of the covariance structure. The dataset and the functions to fit both global and local LGCPs are available from the R Core Team (2023) package `stopp` (D'Angelo and Adelfio 2023).

Our specification involves a constant intensity for the first-order intensity function. Therefore, our primary focus is to investigate how variations in the estimate of the first-order intensity function affects the estimation of the second-order structure, following the approach taken in our previous numerical experiments. Specifically, we utilize the same separable doubly exponential covariance function to ensure a comparable evaluation for our purpose.

Table 10 reports the estimates of the log-Gaussian Cox processes applied to the Greek earthquake data with the MLE and MC procedure at the first step.

The covariance parameters are obtained by using the pair-correlation function in the second step, weighted by the first-order intensity function previously estimated. As expected, in the first step, the MC procedure tends to estimate a higher first-order constant intensity function, and this influence is evident in the estimates of the covariance function. The MLE procedure, on the other hand, appears to attribute a greater portion of the process's variability to the clustered structure of the pattern, resulting in smaller variance and higher range parameters. Otherwise, the MC procedure yields an overall higher estimate of the first-order constant intensity function, but less dense





**Fig. 6** Local estimates of the local LGCP fitted to the Greek seismic data, with the MC method for the estimation of the first-order intensity function

clusters with higher variance and smaller range parameters. In other words, the two methods seem to interpret and model the underlying spatial structure differently, with implications for the estimated covariance function and the overall pattern characterization.

Note that the magnitude of the parameters, particularly the covariance parameters, depends on spatio-temporal units. Consequently, both MLE and MC procedures may provide overall higher values. In other words, both procedures may lead to the overall conclusion that the observed pattern exhibits clustering characteristics. Indeed, previous studies showed that this seismic catalog clearly exhibits a clustered structure which can be well described by an LGCP.

Furthermore, D'Angelo et al. (2023a) showed that a local version of the LGCP further improves the fitting to the data. For this reason, we proceed by fitting a local LGCP to the data, again with both the estimates (MLE and MC) at the first step, and compare the results.

With local LGCP, we mean that the first-order intensity function is the same (global one) previously fitted and reported in Table 10, but we fit individual interaction parameters for each of the points in the observed pattern. Of course, we also expect these local estimates to change if compared to the ones obtained using MLE at the first step. Figure 6 shows the result of the two-step minimum contrast fitting.

As expected, the local estimates slightly change between the two methods. Indeed the conclusions one could draw from this application could be the same as in D'Angelo et al. (2023a), meaning that both estimations permit to observe regions where points tend to cluster in smaller clusters (bottom right area: small variance and high range parameters) and other where clustering is more mild.

However, we still identify some differences, particularly regarding the MC selecting a smaller area with the most dense clustering.

## 6 Conclusions and future work

This paper presents a novel fitting procedure for the first-order intensity function of point processes, based on the minimum contrast procedure. Knowing the expectation of the  $K$ -function when weighted by the true intensity function, we utilize this result to construct a model estimation procedure that is adaptable to any model specification.

The only prerequisite is the knowledge of the expression of the first-order intensity function, completely circumventing the need to face the likelihood, which can often be complex to maximize, in point process models.

The motivation behind our research stems from the desire to simplify the estimation process and broaden its applicability. By employing the expectation of the weighted  $K$ -function, our method offers an intuitive solution to the complexities associated with point process models. In detail, compared to MLE and standard point process estimation methods, our method completely avoids dealing with the likelihood and, therefore, avoids the complexities of its two- or three-dimensional integrals. Turning to the local estimation context, our method is able to provide local estimates without introducing any additional complexity, than the global estimation.

This paper particularly deals with Poisson process models, whose likelihood represents the foundation of many more complex models.

We have presented simulated results for both purely spatial and spatio-temporal contexts, along with the analysis of two real datasets related to environmental issues. These real datasets serve to illustrate more complex scenarios than the Poisson one. An important finding from our simulation studies is that our approach appears to be robust to variations in sample size, whether in spatial or spatio-temporal patterns. This finding is particularly relevant as it suggests that our method may not require exceptionally large datasets to yield reliable results.

In conclusion, our approach opens the path for future research utilizing the minimum contrast procedure for first-order intensity estimation in considerably more complex models than the Poisson model, as already sketched in this study.

It is important to stress that the results presented in this paper concerning the Poisson case are highly noteworthy. Indeed, in our approach, we have entirely overcome the need for likelihood maximization and the subsequent approximation of the integral over the intensity function. Furthermore, we have not required any quadrature scheme or the selection of kernel weights for local estimation. Despite these simplifications, our method has yielded results that are comparable to those obtained through Maximum Likelihood Estimation. This highlights the effectiveness and validity of the proposed approach, which can significantly reduce the computational complexity, maintaining accuracy and reliability in parameter estimation, and producing results that do not deviate much from the more traditional MLE approach.

Many research paths could be deepened in future. First, we believe that the optimization procedure could be improved by weighting the objective function in Eq. (8) by a  $\phi(r, h)$  function. This would basically correspond to giving more importance to some specific spatial and temporal lags. For instance, Diggle (2013) suggested using  $\phi(r, h)$  to weight the discrepancy measure by the inverse (approximate) variance of the  $K$ -function (in which Guan and Sherman (2007), used their sub-sampling method to achieve). The variance of the  $K$ -function, however, is typically unknown. For a spatial Poisson process, it is suggested to use  $\phi(r) = r^{-2}$ , but no other specific recommendations for other types of process are given. Then, we want to run extended simulation studies to assess the performance of the proposed procedure in more complex settings, for instance, with Self-Exciting models such as the Epidemic-Type Aftershock-Sequence (ETAS) ones (Ogata 1988).

Furthermore, in parallel to our idea, Kresin and Schoenberg (2023) showed that parameters in spatio-temporal point process models, alternatively to MLE, can be estimated consistently, under general conditions by minimizing the Stoyan-Grabarnik statistic. Therefore, we wish to compare our proposal based on the  $K$ -function to Kresin and Schoenberg (2023)'s proposal.

**Funding** Open access funding provided by Università degli Studi di Palermo within the CRUI-CARE Agreement. This work was supported by the Targeted Research Funds 2024 (FFR 2024) of the University of Palermo (Italy), the Mobilità e Formazione Internazionali - Miur INT project “Sviluppo di metodologie per processi di punto spazio-temporali marcati funzionali per la previsione probabilistica dei terremoti”, and by the European Union—NextGenerationEU, in the framework of the GRINS—Growing Resilient, INclusive and Sustainable project (GRINS PE00000018—CUP C93C22005270001). The views and opinions expressed are solely those of the authors and do not necessarily reflect those of the European Union, nor can the European Union be held responsible for them.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Adelfio G, Siino M, Mateu J et al (2020) Some properties of local weighted second-order statistics for spatio-temporal point processes. *Stoch Env Res Risk Assess* 34(1):149–168
- Baddeley A (2017) Local composite likelihood for spatial point processes. *Spatial Stat* 22:261–295
- Baddeley A, Turner R (2000) Practical maximum pseudolikelihood for spatial point patterns: (with discussion). *Aust N Z J Stat* 42(3):283–322
- Baddeley A, Turner R (2005) spatstat: An R package for analyzing spatial point patterns. *J Stat Softw* 12(6):1–42. <http://www.jstatsoft.org/v12/i06/>
- Baddeley AJ, Møller J, Waagepetersen R (2000) Non- and semi-parametric estimation of interaction in inhomogeneous point patterns. *Stat Neerl* 54(3):329–350
- Baddeley A, Turner R, Møller J et al (2005) Residual analysis for spatial point processes (with discussion). *J R Stat Soc Ser B* 67(5):617–666
- Baddeley A, Rubak E, Turner R (2015) *Spatial point patterns: methodology and applications with R*. Chapman and Hall/CRC Press, London
- Baddeley A, Davies TM, Hazelton ML et al (2022) Fundamental problems in fitting spatial cluster process models. *Spatial Stat* 52(3):100709
- Berman M (1986) Testing for spatial association between a point process and another stochastic process. *Appl Stat* 35(1):54–62
- Berman M, Diggle P (1989) Estimating weighted integrals of the second-order intensity of a spatial point process. *J R Stat Soc Ser B* 51(1):81–92
- Berman M, Turner TR (1992) Approximating point process likelihoods with GLIM. *J R Stat Soc Ser C* 41(1):31–38
- Brix A, Diggle PJ (2001) Spatiotemporal prediction for log-Gaussian cox processes. *J R Stat Soc Ser B* 63(4):823–841
- Chis OT, Banga JR, Balsa-Canto E (2011) Structural identifiability of systems biology models: a critical comparison of methods. *PLoS ONE* 6(11):e27755
- D'Angelo N, Adelfio G (2023) stopp: Spatio-temporal point pattern methods, model fitting, diagnostics, simulation, local tests. R package version 0.1.0. <https://CRAN.R-project.org/package=stopp>

- D'Angelo N, Siino M, D'Alessandro A et al (2022) Local spatial log-Gaussian Cox processes for seismic data. *AStA Adv Stat Anal* 106(4):633–671
- D'Angelo N, Adelfio G, Mateu J (2023a) Locally weighted minimum contrast estimation for spatio-temporal log-Gaussian Cox processes. *Comput Stat Data Anal* 180:107679. <https://doi.org/10.1016/j.csda.2022.107679>
- D'Angelo N, Adelfio G, Mateu J et al (2023b) Local inhomogeneous weighted summary statistics for marked point processes. *J Comput Graph Stat*. <https://doi.org/10.1080/10618600.2023.2206441>
- Davies TM, Hazelton ML (2013) Assessing minimum contrast parameter estimation for spatial and spatiotemporal log-Gaussian Cox processes. *Stat Neerl* 67(4):355–389
- Diggle PJ (1979) On parameter estimation and goodness-of-fit testing for spatial point patterns. *Biometrics* 35:87–101
- Diggle P (1985) A kernel method for smoothing point process data. *J R Stat Soc Ser C* 34(2):138–147
- Diggle PJ (2013) *Statistical analysis of spatial and spatio-temporal point patterns*. Chapman and Hall/CRC Press, London
- Diggle PJ, Gratton RJ (1984) Monte Carlo methods of inference for implicit statistical models. *J R Stat Soc Ser B* 46(2):193–212
- Diggle PJ, Moraga P, Rowlingson B et al (2013) Spatial and spatio-temporal log-Gaussian Cox processes: extending the geostatistical paradigm. *Stat Sci* 28(4):542–563
- Eguchi S (1983) Second order efficiency of minimum contrast estimators in a curved exponential family. *Ann Stat* 11:793–803
- Foxall R, Baddeley A (2002) Nonparametric measures of association between a spatial point process and a random set, with geological applications. *J R Stat Soc Ser C* 51(2):165–182
- Gabriel E, Diggle PJ (2009) Second-order analysis of inhomogeneous spatio-temporal point process data. *Stat Neerl* 63(1):43–51
- Guan Y (2006) A composite likelihood approach in fitting spatial point process models. *J Am Stat Assoc* 101(476):1502–1512
- Guan Y, Sherman M (2007) On least squares fitting for stationary spatial point processes. *J R Stat Soc Ser B* 69(1):31–49
- Kresin C, Schoenberg F (2023) Parametric estimation of spatial–temporal point processes using the Stoyan–Grabarnik statistic. *Ann Inst Statl Math* 75(6):887–909
- Kreutz C (2018) An easy and efficient approach for testing identifiability. *Bioinformatics* 34(11):1913–1921
- Lawson AB (1993) A deviance residual for heterogeneous spatial Poisson processes. *Biometrics* 49:889–897
- Møller J, Syversveen AR, Waagepetersen RP (1998) Log Gaussian Cox processes. *Scand J Stat* 25(3):451–482
- Ogata Y (1988) Statistical models for earthquake occurrences and residual analysis for point processes. *J Am Stat Assoc* 83(401):9–27
- Ogata Y, Katsura K (1991) Maximum likelihood estimates of the fractal dimension for random spatial patterns. *Biometrika* 78:463–474
- Pfanzagl J (1969) On the measurability and consistency of minimum contrast estimates. *Metrika* 14(1):249–272
- R Core Team (2023) *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna. <https://www.R-project.org/>
- Raeisi M, Bonneau F, Gabriel E (2021) A spatio-temporal multi-scale model for Geyer saturation point process: application to forest fire occurrences. *Spatial Stat* 41:100492
- Ripley BD (1976) The second-order analysis of stationary point processes. *J Appl Probab* 13:255–266
- Siino M, Adelfio G, Mateu J (2018) Joint second-order parameter estimation for spatio-temporal log-Gaussian Cox processes. *Stoch Env Res Risk Assess* 32(12):3525–3539
- Silverman BW (1986) *Density estimation for statistics and data analysis*, vol 26. Chapman and Hall/CRC, London
- Stoyan D, Grabarnik P (1991) Second-order characteristics for stochastic structures connected with Gibbs point processes. *Math Nachr* 151(1):95–100
- Tanaka U, Ogata Y, Stoyan D (2008) Parameter estimation and model selection for Neyman-Scott point processes. *Biom J* 50(1):43–57
- Terrell GR (1990) The maximal smoothing principle in density estimation. *J Am Stat Assoc* 85(410):470–477

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.