Should We Condition on the Number of Points When Modelling Spatial Point Patterns?

Møller, Jesper; Vihrs, Ninna

Published in:
International Statistical Review

DOI (link to publication from Publisher):
10.1111/insr.12501

Publication date:
2022

Document Version
Publisher's PDF, also known as Version of record

International Statistical Review

International Statistical Review (2022), **90**, 3, 551–562 doi: 10.1111/insr.12501

Should We Condition on the Number of Points When Modelling Spatial Point Patterns?

Jesper Møller and Ninna Vihrs

Department of Mathematical Sciences, Aalborg University, Skjernsvej 4A, Aalborg, DK-9220, Denmark
Jesper Møller, Department of Mathematical Sciences, Aalborg University, Denmark.
Email: jm@math.aau.dk

Summary

We discuss the practice of directly or indirectly assuming a model for the number of points when modelling spatial point patterns even though it is rarely possible to validate such a model in practice because most point pattern data consist of only one pattern. We therefore explore the possibility to condition on the number of points instead when fitting and validating spatial point process models. In a simulation study with different popular spatial point process models, we consider model validation using global envelope tests based on functional summary statistics. We find that conditioning on the number of points will for some functional summary statistics lead to more narrow envelopes and thus stronger tests and that it can also be useful for correcting for some conservativeness in the tests when testing composite hypothesis. However, for other functional summary statistics, it makes little or no difference to condition on the number of points. When estimating parameters in popular spatial point process models, we conclude that for mathematical and computational reasons, it is convenient to assume a distribution for the number of points.

Key words: conditional inference; global envelope test; maximum likelihood; maximum pseudo-likelihood; spatial point process models.

1 Introduction

Consider a spatial point process defined on the d-dimensional Euclidean space $\mathbb{R}^d$, and let $W \subset \mathbb{R}^d$ be a bounded region $W \subset \mathbb{R}^d$ within which realisations of the process are observed. In the literature on statistical analysis of spatial point processes, the data usually consist of a single realisation $\{x_1, \ldots, x_n\} \subset W$, that is, a (finite) point pattern where both the number of points n and the locations $x_1, \ldots, x_n$ are considered to be random. For instance, in the `R`-package `spatstat` (Baddeley *et al*., 2015), which is widely used for analysing spatial point patterns, there are more than 50 data examples of point patterns, but only seven of these consist of more than one point pattern. Although it is impossible from a single point pattern to validate the plausibility of any claimed model for the number of points, the practice is nonetheless to fit, validate and use spatial point process models, which directly or indirectly assume a model for the number of points.

For the example of a stationary Poisson process, n is a realisation of a Poisson-distributed random variable, and conditioned on n, the points are realisations of n independent uniformly distributed random variables on W. Even for this simple example of a point process model, it is not possible to validate the assumed Poisson distribution for the number of points based on just one realization. A common procedure for testing whether a Poisson process model fits a given point pattern is to divide W into disjoint subsets of equal size and count the number of points falling within each subset. Conditioning on n, these observed counts constitute a realization from a multinomial model with equal probabilities, and the validity of this model can be checked using a goodness-of-fit test, for example, as implemented in the function **quadrat.test** from `spatstat`. However, even if we accept this multinomial model, to claim that n is a realisation from a Poisson distribution amounts to assume independence for the counts (Moran, 1952), and indeed, this assumption is hard to justify by a statistical test. Therefore, it will be impossible from a single point pattern to validate that n is a realisation from a Poisson distribution.

We still believe that it is usually reasonable to think of the number of points as a single realisation of a random variable. We merely point out that it would be inappropriate to make claims about the distribution of the number of points because we have no chance to validate these as illustrated for a stationary Poisson process above. If we are indeed willing to make some model assumptions and can make simulated point patterns under these assumptions, such simulations can of course be used to estimate the distribution of the number of points under the model if wanted, but why should we believe in such a distribution when we do not have the means to validate this? It may thus be more fair to accept that we do not know anything about the distribution of the number of points and therefore use a conditional approach instead. However, we have not found much places in the spatial point process literature where the role of conditioning on n is studied. Ripley (1977) provided a short discussion of conditional inference saying 'For Poisson models, we can justify this conditioning by an appeal to a conditionality principle. In general, all we can say is that conditional inference seems reasonable and that our revised procedures yield valid conditional tests'. Further, Ripley (1988) wrote 'Where we are interested in interactions, n may be approximately ancilliary'. He illustrated this for a Strauss process which on the unit square when ignoring boundary effects has density

$$f(\{x_1, \ldots, x_n\}) \propto \beta^n \gamma^{s(\{x_1, \ldots, x_n\})}$$

where $\beta > 0,\ 0 \le \gamma \le 1,\ R > 0$ and $s(\{x_1, \ldots, x_n\})$ is the number of R-close pairs of points in the point pattern; we return to this process in Sections 2.3 and 4.2 and Section 2 in the supporting information. Ripley (1988) derived an approximation of the probability density function for the number of points:

$$p(n) \propto \frac{\beta^n}{n!} \exp\{(\gamma - 1)n(n - 1)\pi R^2/2\}. \tag{1}$$

He noticed that this density depends on γ but wrote 'the dependence is quite weak in typical cases' where he referred to a plot of the cumulative distribution functions when $\beta = 50,\ R = 0.05$ and $\gamma = 0.0, 0.2, 0.4, 0.8, 1.0$. Mean values of n in this case and when $\beta = 200$ are shown in Table 1 where we see a good agreement between those means calculated from the approximation in (1) and those obtained by simulations. In our opinion, the mean values in Table 1 depend much on γ, but we notice in Section 4.2 that maximum pseudo-likelihood estimates do not depend much on whether or not we condition on n.

Table 1. *Mean values for the number of points in a Strauss process on a two-dimensional unit square with $R = 0.05$ and different values of γ and β*

	$\gamma = 0$	$\gamma = 0.2$	$\gamma = 0.4$	$\gamma = 0.6$	$\gamma = 0.8$	$\gamma = 1$
$\beta = 50$	37.33	39.13	41.19	43.61	46.48	50.00
	(36.91)	(38.91)	(41.24)	(43.78)	(46.56)	(49.87)
$\beta = 200$	95.00	104.11	115.92	132.12	156.45	200.00
	(88.37)	(100.72)	(114.53)	(132.35)	(157.36)	(200.32)

The means are calculated from the approximate distribution in (1), and the numbers in parenthesis are the means obtained from 5 000 simulations of the process.

Apart from Ripley's study of the Strauss process considered above, we are not aware of any thorough study of the effect of conditioning on n. Moreover, the practice is still to work with spatial point process models without conditioning on n. Therefore, the objective of this paper is to investigate the consequences of this practice and to explore the possible benefits of conditioning on n when considering various popular classes of spatial point process models. We investigate this through a comprehensive simulation study of model validation based on the widely used method of global envelopes and corresponding tests (Myllymäki *et al*., 2017), and by discussing the effect of conditioning on n when making frequentistic parameter estimation.

Our paper is organised such that Section 2 contains some preliminaries needed for our main contributions in Sections 3–5, where Section 5 summarises our findings. Technical details related to these sections are found in the supporting information that is available in the online version of the article. All statistical analyses were made with `R` (R. Core Team, 2019). We used the packages `spatstat` (Baddeley *et al*., 2015) for handling spatial point patterns, `GET` (Myllymäki & Mrkvička, 2019) to make global envelope tests and `ggplot2` (Wickham, 2016) for visualisation. Furthermore, we used our own implementations of simulation and estimation procedures when conditioning on the number of points, which can be found in the `R`-scripts in the supporting information.

2 Preliminaries

2.1 Setting and Notation

Throughout this paper, we use the following point process setting and notation.

For a subset $x \subset \mathbb{R}^d$, let $n(x)$ denote its cardinality (setting $n(x) = \infty$ if x is not a finite set) and $x_B := x \cap B$ its restriction to any set $B \subseteq \mathbb{R}^d$. Let $\mathbf{\Omega}$ denote the set of all locally finite subsets $x \subset \mathbb{R}^d$, that is, $n(x_B) < \infty$ whenever $B \subset \mathbb{R}^d$ is bounded. A simple locally finite point process on $\mathbb{R}^d$ is a random variable taking values in $\mathbf{\Omega}$. Here and elsewhere, we omit measure theoretical details and refer instead to Møller and Waagepetersen (2004) and the references therein. The point process can also be specified in terms of the counting measure $N(B) := n(X_B)$ for bounded sets $B \subseteq \mathbb{R}^d$ (more precisely, B should also be a Borel set, but as mentioned, we omit such details).

We assume that X is a stationary simple locally finite point process on $\mathbb{R}^d$ with intensity $\rho \in (0, \infty)$. This means that $X \in \mathbf{\Omega}$, the distribution of X is invariant under translations in $\mathbb{R}^d$ and $\mathrm{E}N(B) = \rho|B|$ where $|B|$ denotes the d-dimensional volume (Lebesgue measure) of $B \subseteq \mathbb{R}^d$. Stationarity is a common assumption for spatial point processes, and it allows us to deal with frequently used functional summary statistics, refer to Section 2.2. We also assume

that a single realisation $x = \{x_1, \ldots, x_n\}$ of X_W has been observed where the observation window $W \subset \mathbb{R}^d$ is compact and $|W| > 0$.

Let $u, v \in \mathbb{R}^d$, $r \geq 0$ and $B \subset \mathbb{R}^d$. Then, $\mathbb{1}(\cdot)$ is the indicator function; $\|u - v\|$ is the usual distance between u and v; $b(u, r)$ is the closed d-dimensional ball with centre u and radius r; $\operatorname{dist}(B, u) := \inf\{r > 0 \,|\, b(u, r) \cap B \neq \emptyset\}$ is the distance from u to B; $B_{\ominus r} := \{u \in \mathbb{R}^d \,|\, b(u, r) \subset B\}$ is B eroded by a ball of radius r; $B_{\oplus} := \cup_{u \in B} b(u, r)$ is B dilated by a ball of radius r; and $(B_{\ominus r})_{\oplus r} \subseteq B$ is the opening of B by a ball of radius r. Finally, we use the convention $0/0 := 0$.

2.2 Functional Summary Statistics

Functional summary statistics $\hat{K}$, $\hat{F}$, $\hat{G}$ and $\hat{J}$ which are non-parametric (empirical) estimates of the theoretical functions below are widely used for exploratory purposes, model fitting and model checking, refer to Baddeley *et al.* (2015) and the references therein.

For every $r > 0$ and an arbitrary point $u \in \mathbb{R}^d$, Ripley's K-function is defined by

$$\rho K(r) := \mathrm{E}(N(b(u, r) \backslash \{u\}) | u \in X), \tag{2}$$

the empty space function F (or spherical contact function) and the nearest-neighbour function G are

$$F(r) := \mathrm{P}(\operatorname{dist}(X, u) \leq r), \qquad G(r) := \mathrm{P}(\operatorname{dist}(X \backslash \{u\}, u) \leq r | u \in X) \tag{3}$$

and for $F(r) < 1$ the J-function is

$$J(r) := (1 - G(r))/(1 - F(r)). \tag{4}$$

These definitions do not depend on the choice of u since X is stationary, and as indicated by the notation, we have conditioned on $u \in X$ in the definitions of K and G, meaning that $X \backslash \{u\}$ then follows the reduced Palm distribution of X at u.

When estimating the K, F, G and J-functions by non-parametric methods, different types of edge correction methods have been suggested in order to adjust for the fact that $n(x_{W \cap b(u, r)})$ tends to be smaller for points $u \in W$ which are close to the boundary of W compared with points $u \in W$ which are far from the boundary of W. Baddeley *et al.* (2015) noticed that the choice of edge correction method is usually not very important. We choose to use a particular border (or minus-sampling) correction method that is available in `spatstat` and refer to Baddeley *et al.* (2015) for the concrete estimators.

When conditioning on n, we are not aware of how to modify the definitions of K, F, G and J. If we just condition in (2)–(4), the expressions will depend on u because stationarity no longer holds. Still, when only a single point pattern is observed and hence it is impossible to validate any claimed model of $N(W)$, it seems appropriate to condition on n when calculating global envelopes and tests as in Section 3.

2.3 Models

For the simulation studies in Sections 3 and 4.2, we consider four concrete examples of point process models on $\mathbb{R}^2$ as specified in M1–M4 below. Section 1 in the supporting information

provides further details including how to make simulations both with and without conditioning on the number of points.

M1: X is a stationary Poisson process. This is the model of no spatial interaction or complete spatial randomness. When making simulations in Section 3, we let the intensity be $\rho = 100$.

M2: X is a stationary log-Gaussian Cox process. Then, X is driven by a stochastic intensity $Z = \exp(Y)$ where Y is a stationary Gaussian random field on $\mathbb{R}^2$, meaning that X conditioned on Y is a Poisson process with intensity function Z (Møller *et al*., 1998). We use an exponential covariance function $c(u, v) := \sigma^2 \exp(-\|u - v\|/\delta)$ for $u, v \in \mathbb{R}^2$ where $\sigma^2 > 0$ is a variance parameter and $\delta > 0$ is a scale parameter, and so the mean of Y is $\mu = \log(\rho) - \sigma^2/2$. Because of the positive correlation in Y, realisations of X exhibit clustered behaviour. When making simulations in Section 3, we let $\rho = 100$, $\sigma^2 = 1$ and $\delta = 0.1$.

M3: X is a stationary Strauss process. This process has parameters $\beta > 0$, $0 \leq \gamma \leq 1$ and $R > 0$. It is defined by a so-called local specification: let $B \subset \mathbb{R}^2$ be a bounded set of positive area and define the R-close neighbourhood to B by $\partial B := B \oplus R \backslash B$. Then, for every such B, X_B and $X_{\mathbb{R}^2 \backslash B \oplus R}$ are conditionally independent given $X_{\partial B}$. Furthermore, for every finite $x_{\partial B} \subseteq \partial B$, X_B conditioned on $X_{\partial B} = x_{\partial B}$ has a density with respect to a Poisson process of intensity 1 and restricted to B. This conditional density is

$$f_B(x_B \,|\, x_{\partial B}) \propto \beta^{n(x_B)} \gamma^{s(x)} \tag{5}$$

for finite $x_B \subseteq B$, $x = x_B \cup x_{\partial B}$ and $s(x) = \sum_{i<j} \mathbb{1}(\|x_i - x_j\| \leq R)$ if $x = \{x_1, \ldots, x_n\}$. The normalising constant that is omitted in (5) depends on (β, γ, R) and $x_{\partial B}$, and it is intractable unless $\gamma = 1$. When $\gamma = 1$, X is just a stationary Poisson process with intensity β. As γ decreases, X becomes more and more inhibitory, and it is a Gibbs hard core model if $\gamma = 0$. When making simulations in Section 3, we let $\beta = 200$, $\gamma = 0.3$ and $R = 0.05$. Then, the intensity is approximately 100.

M4: X is a stationary determinantal point process. In brief, determinantal point processes (Macchi, 1975; Lavancier *et al*., 2015) are specified by a function $C: \mathbb{R}^d \times \mathbb{R}^d \mapsto \mathbb{C}$ called the kernel, and they are repulsive at all scales, compare Section 1.4 in the supporting information. We use a Gaussian kernel $C(u, v) = \rho \exp(-\|(u - v)/\kappa\|^2)$ where $\kappa > 0$ is a scale parameter and $\rho > 0$ is the intensity. It should be satisfied that $\kappa \leq 1/\sqrt{\rho\pi}$, and this upper limit corresponds to the most repulsive case when ρ is fixed. When making simulations in Section 3, we let $\rho = 100$ and $\kappa = 0.03 \leq 1/\sqrt{100\pi} \approx 0.056$.

3 Global Envelopes and the Effect of Conditioning

In this section, we investigate the effect of conditioning on the number of points when using global envelopes for model validation. Section 3.1 first describes the set-up of the simulation study, and Section 3.2 describes and interprets the results.

3.1 Set-up

We investigate the effect of conditioning on the number of points when considering global envelopes for functional summary statistics and corresponding tests based on the extreme rank length as described in Myllymäki *et al*. (2017), Mrkvička *et al*. (2018) and Myllymäki & Mrkvička (2019). Briefly, a $(100 - \alpha)\%$ global envelope consists of a lower and an upper curve defining a region such that the observed functional summary statistic for data falls completely

between these bounding curves if and only if the global envelope test cannot be rejected at level α%. There of course exist other tests that can be used in connection with spatial point processes, refer to, for example Baddeley *et al*. (2015, chapter 10); however, because the use of a global envelope and its corresponding test statistic is by far the most popular method for performing model validation of a fitted spatial point process model, we restrict attention to this test procedure. There are ways to make a combined global envelope test based on several functional summary statistics, but for our purpose, we prefer to investigate the effect of conditioning on the number of points for each functional summary statistic.

We made the simulation study as follows. Under each of the four models M1–M4 we simulated 1 000 independent point patterns within a two-dimensional unit square (the observation window W). For each of these point patterns, we fitted the parameters of the models as described in the last paragraph of this section. Under each fitted model and each true model, we made further 2 500 simulations with and 2 500 simulations without conditioning on the number of points. From each of these four cases, or three in the case of the Poisson process because the fitted and true model is the same when conditioning on the number of points, we used the 2 500 simulations to calculate 95% global envelopes based on each of the functional summary statistics $\hat{F}$, $\hat{G}$, $\hat{J}$ and $\hat{K}$. Some further technical and practical details related to the set-up of the simulation study are deferred in Appendix A.

Clearly, more narrow envelopes are preferable when comparing envelopes for the same type of functional summary statistic. For simplicity, in order to spot a general tendency in the width of envelopes, we considered for each envelope a numerical approximation of its area $\int_0^R (c_u(r) - c_l(r))\,\mathrm{d}r$ where c_u and c_l are the upper and lower curves of the envelope, respectively, and R is the highest r value for which the considered functional summary statistic was estimated.

3.2 Results

Figure 1 shows boxplots of the approximated area of the envelopes. We see that it generally makes little difference in the areas of the envelopes whether the parameters are fitted from data or not, except for $\hat{K}$ where there is less variation in the width of the envelopes when using the true parameters especially in the unconditional case. We also see that for $\hat{K}$ and $\hat{J}$, it makes no real difference in the area of the envelopes to condition on the number of points either, but for $\hat{G}$ and especially $\hat{F}$ the envelopes are in general more narrow in the conditional case.

Figure 2 shows quantile–quantile plots comparing the distributions of the p values of the global envelope tests for each fitted model to a uniform distribution on $[0, 1]$. We see that some of the tests are too conservative, which may be because the null hypothesis is composite except in the conditional case of the Poisson process. Using $\hat{F}$ overall gives very conservative tests in the unconditional case, especially for the Poisson, Cox and determinantal point process, and this behaviour is corrected very well by conditioning on the number of points. For $\hat{G}$, it makes little difference to condition on the number of points except in the case of the Cox process where it corrects the conservativeness in the unconditional case very well. For both $\hat{J}$ and $\hat{K}$, it makes little to no difference whether we condition on the number of points or not, and in all cases, the distributions of the p values are in good agreement with the uniform distribution even though we also see some slight conservativeness in some tests. Since $\hat{K}$ was used to fit the parameters of the Cox and determinantal point processes, we usually do not want to use $\hat{K}$ for model validation as well.

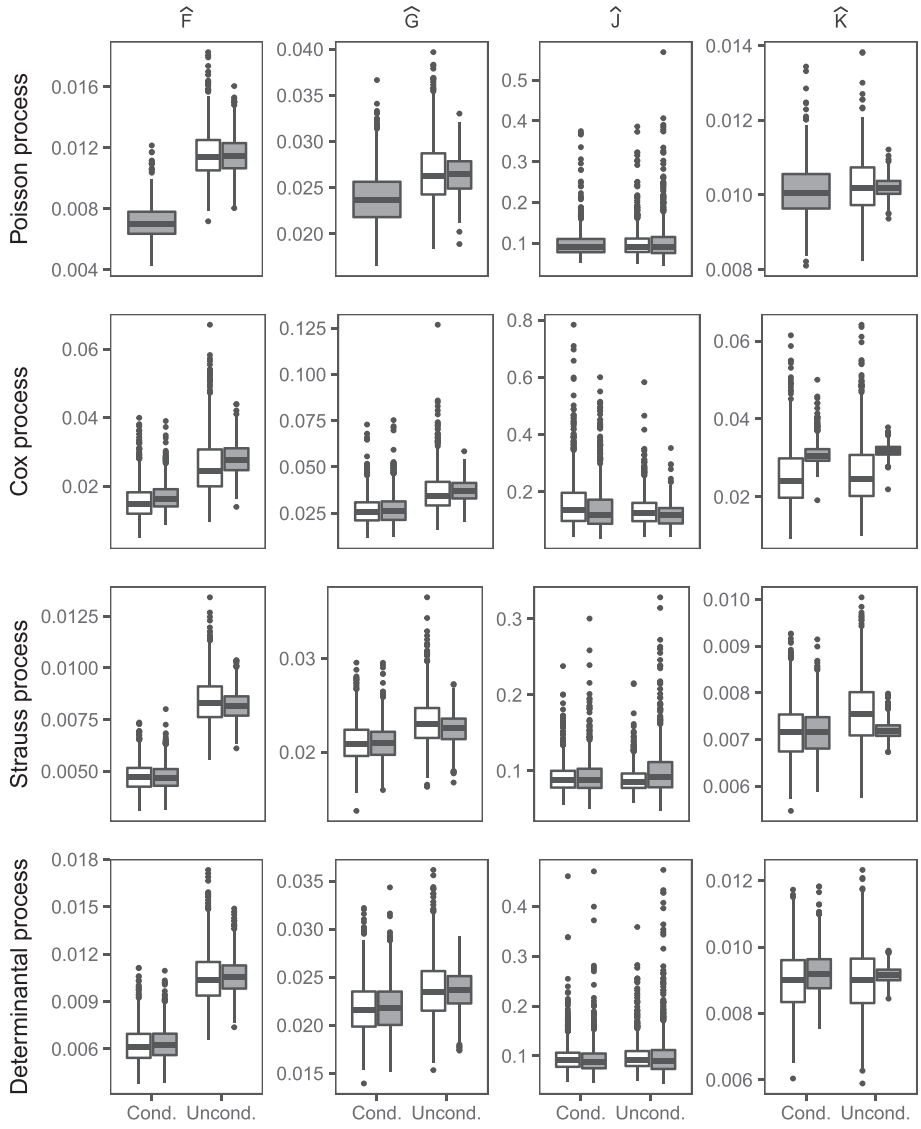

FIGURE 1. *Boxplots for the area between 95% global envelope curves for the functional summary statistic stated at the top of each column. The results were obtained from the simulation study described in the text. The grey level indicates whether the true parameters (dark grey) or fitted parameters (white) were used in the simulations. Whether simulations were made conditional on the number of points is stated at the bottom of each column. The type of model under consideration is stated to the left of each row*

Moreover, to see the effect of increasing the intensity, we made a simulation study for the stationary Poisson process with $\rho = 200$. Our conclusions remained the same, and the only real difference was that all envelopes were in general more narrow, which was to be expected since the summary statistics varies less when we observe more points. However, the simulation studies for the remaining three models are time consuming, and it will be even worse if we increase the intensity, but we do not believe the conclusions will be much different as we have indeed established for the example of a Poisson process.

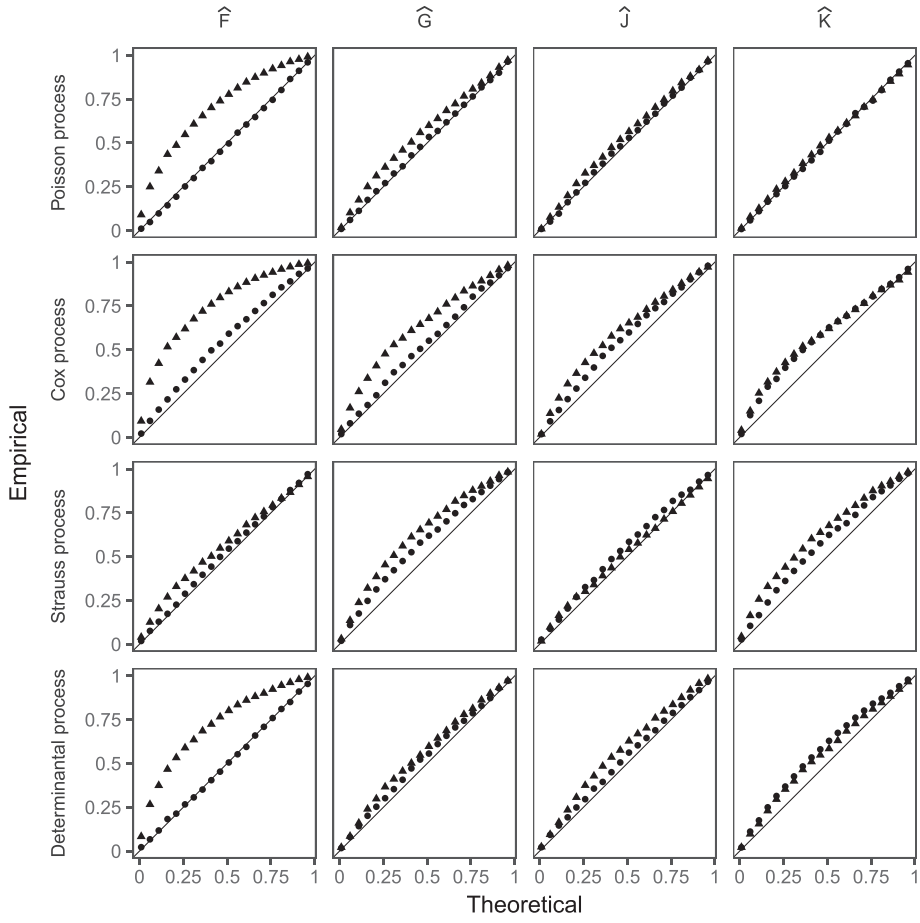

FIGURE 2. *Quantile–quantile plots comparing the empirical distributions of the p values of global envelope tests to a uniform distribution on* $[0, 1]$. *The considered p values are the results of the simulation study described in the text. The type of point indicates whether the global envelope tests were based on conditional (dots) or unconditional (triangles) simulations under the fitted model. The functional summary statistic used is stated at the top of each column, the type of model is stated to the left of each row and the straight line is the identity line*

4 Conditional Estimation

In this section, we investigate the possibility to estimate parameters in spatial point process models conditional on the number of points. We consider Cox, Gibbs and determinantal point processes and in each case discuss whether conditional estimation offers any advantages over the unconditional approach. There is of course no reason to consider conditional estimation in the case of a stationary Poisson process because the conditional case is a binomial point process with no unknown parameters.

4.1 Cox Processes

Parametric models for a stationary Cox process X driven by a random field Z on $\mathbb{R}^d$ are usually of the form $Z(u) = \rho R(u)$ where $R = \{R(u)\}_{u \in \mathbb{R}^d}$ is a non-negative unit-mean stochastic process following a parametric model with a parameter ψ so that (ρ, ψ) has range $(0, \infty) \times \Psi$ for some set $\Psi \subseteq \mathbb{R}^p$. The process X_W then has a density

$$f(x) = \rho^n \mathrm{E}\left[R(x_1)\ldots R(x_n)\exp\left\{|W| - \rho\int_W R(u)\,\mathrm{d}u\right\}\right] \quad (6)$$

with respect to a Poisson process of intensity 1 and restricted to W. In general, this density is not expressible on closed form except for rather trivial cases, and so likelihood based inference is difficult although a missing-data Markov chain Monte Carlo approach can be used for approximate maximum likelihood estimation, refer to Møller and Waagepetersen (2004). Since second-order moments of the counts $N(B)$ are often expressible on closed form, moment-based and composite likelihood estimation procedures are usually preferred, refer to Møller & Waagepetersen (2017) and the references therein.

When conditioning on $N(W) = n$, any ordering $(X_1, \ldots, X_n)$ of the n points in X_W has a density with respect to Lebesgue measure on W^n which is proportional to the right-hand side in (6). Hence, it is also not expressible on closed form and depends on both ρ and ψ. Furthermore, moment-based estimation is no longer possible. Consequently, we do not see any advantage in conditioning on the number of points.

4.2 Gibbs Processes

Suppose that X is a stationary Gibbs point process with known interaction radius $R > 0$ (details for this general setting are provided in Section 1.3 in the supporting information). Further, suppose that $X_W = x$ has been observed, and let $x_{W\ominus R} = \{x_1, \ldots, x_m\}$. We define the R-close neighbourhood to $W_{\ominus R}$ by $\partial W_{\ominus R} := (W_{\ominus R})_{\oplus R} \backslash W_{\ominus R}$ and base inference on the conditional distribution of $X_{W\ominus R}$ given $X_{\partial W\ominus R} = x_{\partial W\ominus R}$. Thereby, we account for edge effects due to the unobserved points in $X_{\mathbb{R}^d\backslash W}$ because $X_{W\ominus R}$ and $X_{\mathbb{R}^d\backslash W}$ are independent conditioned on $X_{\partial W\ominus R} = x_{\partial W\ominus R}$. Below, we discuss reasons and consequences of further conditioning on $N(W_{\ominus R}) = m$.

For specificity and simplicity, let us think of X as the stationary Strauss process given in Section 2.3. Then, the likelihood function is of exponential family form with canonical parameter $(\log\beta, \log\gamma)$ and minimal sufficient statistic $(n(x_{W\ominus R}), s(x_{W\ominus R}\cup x_{\partial W\ominus R}))$. However, estimation of (β, γ) is complicated by the fact that the normalising constant is not expressible on closed form for $\gamma < 1$. Often, the interaction parameter γ is of main interest; if we also condition on $N(W_{\ominus R}) = m$, we obtain a likelihood function which only depends on γ. The normalising constant of that likelihood function is also not expressible on closed form for $\gamma < 1$, but it is simpler to approximate, compare Section 2 in the supporting information. In particular, as noticed at the end of Section 2 in the supporting information, simulations and computations will be faster when conditioning on $N(W_{\ominus R}) = m$. However, it is still slower and more difficult than using maximum pseudo-likelihood estimation, which will be described next.

Maximum pseudo-likelihood estimation is known to be a fast and often reliable alternative to maximum likelihood estimation, refer to, for example Jensen and Møller (1991), Baddeley *et al*. (2014) and the references therein. The definition of the pseudo-likelihood function depends on the context and is specified in Section 3 in the supporting information both with and without conditioning on $N(W_{\ominus R}) = m$ and in a way which accounts for edge effects. In the conditional case, we consider Besag's original pseudo-likelihood function (Besag, 1975); in the unconditional case, we consider Besag's extension to spatial point processes (Besag, 1977; Jensen & Møller, 1991). The computational advantage of using the pseudo-likelihood functions is that they do not depend on the intractable normalising constant that appears in maximum likelihood estimation.

We wanted to investigate whether it makes a real difference in the maximum pseudo-likelihood estimate of γ to condition on the number of points. In order to do this, we

considered 1 000 simulations on the unit square of a stationary Strauss process when $\beta = 200$, $R = 0.05$, and γ was sampled uniformly in the interval $[0.01, 1]$. For each simulation, we calculated the maximum pseudo-likelihood estimate of (β, γ) with the function `exactMPLEstrauss` in `spatstat`, which computes the maximum pseudo-likelihood estimate to a high accuracy. We also found the maximum pseudo-likelihood estimate in the conditional case by implementing a function where we used the same numerical methods for optimisation and evaluation of integrals as in `exactMPLEstrauss`. In both cases, we let $R = 0.05$ be given. The average absolute difference between the estimates of γ obtained with the two methods was 0.005, and the largest absolute difference was 0.01. So it makes very little difference to condition on the number of points. Furthermore, the computations for the pseudo-likelihood function in the conditional case may be more cumbersome since more integrals have to be evaluated, compare Section 3 in the supporting information. Therefore, there is no apparent reason to use the more complicated pseudo-likelihood method of the conditional case.

4.3 Determinantal Point Processes

Parametric models for stationary determinantal point processes with intensity $\rho > 0$ are specified by a parametric class of kernel functions, which are usually of the form $C_\theta(u, v) = \rho R_\psi(u - v)$ where $\theta = (\rho, \psi)$ and $(u, v) \mapsto R_\psi(u - v)$ is a (complex) correlation function. Under weak assumptions, for example that C_θ is a continuous complex covariance function, the kernel restricted to $W \times W$ has a spectral representation

$$C_\theta(u, v) = \sum_{i=1}^{\infty} \lambda_i \phi_i(u) \overline{\phi_i(v)}, \qquad u, v \in W,$$

where $\{\phi_i\}_{i=1,2,\ldots}$ is an orthonormal basis for the L^2-space of square-integrable complex functions on W and $\lambda_1, \lambda_2, \ldots$ are corresponding eigenvalues. We have $\lambda_i = \rho \lambda_i'$ where (ϕ_i, λ_i') depends only on ψ and the condition $0 \le \lambda_i \le 1$ is needed to ensure existence of the process. Therefore, the parameter space Ψ_ρ of ψ will depend on the value of ρ; briefly speaking, there is a trade-off between intensity and repulsion, and the set Ψ_ρ decreases as ρ increases, compare Lavancier *et al*. (2015). For parameter estimation based on maximum likelihood and moment-based methods, refer to Lavancier *et al*. (2015).

Conditioned on $N(W) = n$, any ordering $(X_1, \ldots, X_n)$ of the n points in X_W has probability density function

$$\frac{\sum_{i_1 < \ldots < i_n} \left(\prod_{j=1}^n \lambda_{i_j} \right) \left(\prod_{j \notin \{i_1, \ldots, i_n\}} (1 - \lambda_j) \right) \frac{1}{n!} \det \left\{ \sum_{k=1}^n \phi_{i_k}(x_i) \overline{\phi_{i_k}(x_j)} \right\}_{i,j=1,\ldots,n}}{\sum_{i_1 < \ldots < i_n} \left(\prod_{j=1}^n \lambda_{i_j} \right) \left(\prod_{j \notin \{i_1, \ldots, i_n\}} (1 - \lambda_j) \right)}$$

for $(x_1, \ldots, x_n) \in W^n$. For parametric models as considered above, it follows that the conditional distribution of X_W given $N(W) = n$ depends on both ρ and ψ in a complicated way; in fact, it is more complicated than the likelihood in the unconditional case, compare Lavancier *et al*. (2015). Therefore, we do not see any advantage in conditioning on $N(W) = n$ when making parameter estimation.

5 Concluding Remarks

It is worth repeating that any claimed model for the number of points cannot be justified based on just one realisation. However, in case of Cox, Gibbs and determinantal point processes, it is convenient for computational reasons to assume a distribution for the number of points when estimating parameters.

Regarding global envelope tests based on $\hat{K}$ or $\hat{J}$, it made little or no difference in our simulation study whether we condition on the number of points. However, when we instead used $\hat{F}$ or $\hat{G}$, conditioning on the number of points gave more narrow envelopes and hence stronger tests, and it corrected for conservativeness in the tests.

Global envelopes and tests are usually calculated from simulations of a single point process model. For a composite null hypothesis, it is possible to make an adjusted global envelope test but at the expense of many simulations (Myllymäki & Mrkvička, 2019). Conditioning on the number of points may offer an alternative which requires fewer simulations, but whether this will be faster in practice depends on the actual speed of the simulation procedures. We leave this for future research.

Regarding conditional estimation, we concluded that it is impractical and offers no clear advantage to consider conditional estimation for Cox and determinantal point process models. For Gibbs point process models, it simplifies maximum likelihood estimation but complicates the more commonly used and faster method of maximum pseudo-likelihood estimation where there is also little difference in the estimates achieved with and without conditioning. We therefore overall have found no apparent reason to use conditional estimation.

ACKNOWLEDGEMENTS

The research of the authors was supported by The Danish Council for Independent Research — Natural Sciences, grant DFF (7014-00074) ‘Statistics for point processes in space and beyond’.

REFERENCES

Baddeley, A., Coeurjolly, J.-F., Rubak, E. & Waagepetersen, R. (2014). Logistic regression for spatial Gibbs point processes. *Biometrika*, **101**, 377–392.

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

Besag, J. (1975). Statistical analysis of non-lattice data. *Statistician*, **24**, 179–195.

Besag, J. (1977). Some methods of statistical analysis for spatial data. *Bulletin of the International Statistical Institute*, **47**, 77–91.

Jensen, J.L. & Møller, J. (1991). Pseudolikelihood for exponential family models of spatial point processes. *Ann. Appl. Probab*, **1**, 445–461.

Lavancier, F., Møller, J. & Rubak, E. (2015). Determinantal point process models and statistical inference. *J. R. Stat Soc.: Ser. B (Stat Methodol.)*, **77**, 853–877.

Macchi, O. (1975). The coincidence approach to stochastic point processes. *Adv. Appli. Probab.*, **7**, 83–122.

Møller, J., Syversveen, A.R. & Waagepetersen, R.P. (1998). Log Gaussian Cox processes. *Scand. J. Stat.*, **25**, 451–482.

Møller, J. & Waagepetersen, R. (2017). Some recent developments in statistics for spatial point patterns. *Annual Review Stat. Applications*, **4**, 317–342.

Møller, J. & Waagepetersen, R.P. (2004). *Statistical Inference and Simulation for Spatial Point Processes*. Chapman and Hall/CRC: Boca Raton.

Moran, P. (1952). A characteristic property of the Poisson distribution. *Math. Proc. Camb. Philos. Soc*, **48**, 206–207.

Mrkvička, T., Myllymäki, M., Jílek, M. & Hahn, U. (2020). A one-way ANOVA test for functional data with graphical interpretation. *Kypernetika*, **56**, 432–458.

Myllymäki, M. & Mrkvička, T. (2019). GET: global envelopes in R. Available at arXiv:1911.06583.

Myllymäki, M., Mrkvička, T., Grabarnik, P., Seijo, H. & Hahn, U. (2017). Global envelope tests for spatial processes. *J. Royal Stat. Soc.: Seri. B (Stat Methodol.)*, **79**, 381–404.

R. Core Team. (2019). R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria.

Ripley, B.D. (1977). Modelling spatial patterns. *J. Royal Stat. Soc.: Ser. B (Stat. Methodol.)*, **39**, 172–192.

Ripley, B.D. (1988). *Statistical Inference for Spatial Processes*. Cambridge University Press: Cambridge.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org

APPENDIX A: Details Regarding the Simulation Study in Section 3

The following gives some practical and technical details regarding the simulation study described in Section 3.1.

When calculating functional summary statistics, we always used the default range of r values in `spatstat`. For details on simulation procedures, especially conditional simulation, refer to Section 1 in the supporting information. When fitting parameters, we always did it without conditioning on the number of points because we argue in Section 4 that there is no apparent reasons to use conditional estimates. We used the natural estimate $n(x)$ of the intensity in the case of the Poisson process, used the method of minimum contrast estimation based on Ripley's K -function (refer to Baddeley *et al*. 2015) in the cases of the log-Gaussian Cox process and the Gaussian determinantal point process and used the method of profile maximum pseudo-likelihood in the case of the Strauss process (where we considered 41 equally spaced values of R in the interval $(0.03, 0.07)$). When fitting the parameters of the log-Gaussian Cox process, we found that sometimes the scale parameter δ was seriously overestimated, which caused the conditional simulation procedure to be extremely slow. Therefore, we decided only to use realizations of the log-Gaussian Cox process M2 where the fitted scale parameter was below 0.3, which left 953 realisations in the simulation study. When fitting parameters in the case of the Gaussian determinantal point process, we found that κ was seriously underestimated for a few realizations, which either slowed down the simulation procedures considerably or caused them to fail. We therefore excluded realisations with a fitted value of κ less than 0.001 after which 996 realisations remained.

[*Received September 2021; accepted March 2022*]