Score, Pseudo-Score and Residual Diagnostics for Spatial Point Process Models

Adrian Baddeley, Ege Rubak and Jesper Møller

Abstract. We develop new tools for formal inference and informal model validation in the analysis of spatial point pattern data. The score test is generalized to a "pseudo-score" test derived from Besag's pseudo-likelihood, and to a class of diagnostics based on point process residuals. The results lend theoretical support to the established practice of using functional summary statistics, such as Ripley's K-function, when testing for complete spatial randomness; and they provide new tools such as the *compensator* of the K-function for testing other fitted models. The results also support localization methods such as the scan statistic and smoothed residual plots. Software for computing the diagnostics is provided.

Key words and phrases: Compensator, functional summary statistics, model validation, point process residuals, pseudo-likelihood.

1. INTRODUCTION

This paper develops new tools for formal inference and informal model validation in the analysis of spatial point pattern data. The score test statistic, based on the point process likelihood, is generalized to a "pseudo-score" test statistic derived from Besag's pseudo-likelihood. The score and pseudo-score can be viewed as residuals, and further generalized to a class of residual diagnostics.

The likelihood score and the score test [61, 75], [22], pages 315 and 324, are used frequently in applied statistics to provide diagnostics for model selection and model validation [2, 15, 19, 60, 77]. In spatial statistics, the score test has been used mainly to support formal

Adrian Baddeley is Research Scientist, CSIRO Mathematics, Informatics and Statistics, Private Bag 5, Wembley WA 6913, Australia and Adjunct Professor, University of Western Australia (e-mail: Adrian.Baddeley@ csiro.au). Ege Rubak is Postdoctoral Scholar, Department of Mathematical Sciences, Aalborg University, Fredrik Bajers Vej 7G, DK-9220 Aalborg Ø, Denmark (e-mail: rubak@math.aau.dk). Jesper Møller is Professor of Statistics, Department of Mathematical Sciences, Aalborg University, Fredrik Bajers Vej 7G, DK-9220 Aalborg Ø, Denmark (e-mail: jm@math.aau.dk).
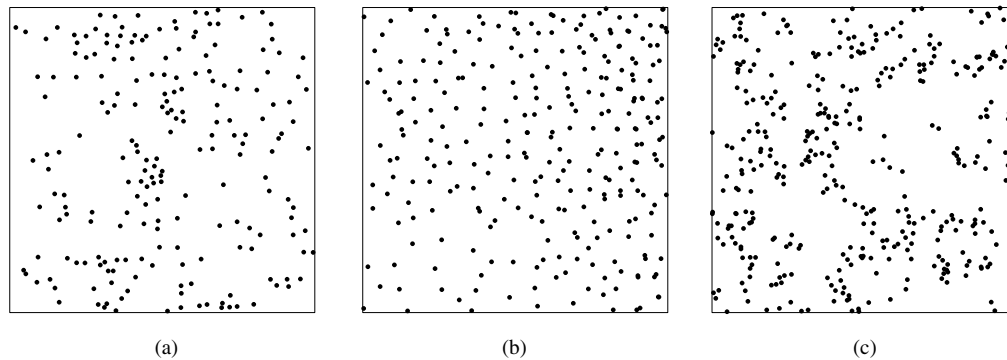
inference about covariate effects [13, 47, 76] assuming the underlying point process is Poisson under both the null and alternative hypotheses. Our approach extends this to a much wider class of point processes, making it possible (for example) to check for covariate effects or localized hot-spots in a clustered point pattern.

Figure 1 shows three example data sets studied in the paper. Our techniques make it possible to check separately for "inhomogeneity" (spatial variation in abundance of points) and "interaction" (localized dependence between points) in these data.

Our approach also provides theoretical support for the established practice of using functional summary statistics such as Ripley's K-function [63, 64] to study clustering and inhibition between points. In one class of models, the score test statistic is equivalent to the empirical K-function, and the score test procedure is closely related to the customary goodness-of-fit procedure based on comparing the empirical K-function with its null expected value. Similar statements apply to the nearest neighbor distance distribution function G and the empty space function F.

For computational efficiency, especially in large data sets, the point process likelihood is often replaced by Besag's [14] pseudo-likelihood. The resulting "pseudo-score" is a possible surrogate for the

FIG. 1. *Point pattern data sets.* (a) *Japanese black pine seedlings and saplings in a* 10×10 *metre quadrat* [53, 54]. *Reprinted by kind permission of Professors M. Numata and Y. Ogata.* (b) *Simulated realization of inhomogeneous Strauss process showing strong inhibition and spatial trend* [7], *Figure* 4b. (c) *Simulated realization of homogeneous Geyer saturation process showing moderately strong clustering without spatial trend* [7], *Figure* 4c.

likelihood score in the score test. In one model, this pseudo-score test statistic is equivalent to a *residual* version of the empirical K-function, yielding a new, efficient diagnostic for model fit. However, in general, the interpretation of the pseudo-score test statistic is conceptually more complicated than that of the likelihood score test statistic, and hence difficult to employ as a diagnostic.

In classical settings the score test statistic is a weighted sum of residuals. For point processes the pseudo-score test statistic is a weighted point process residual in the sense of [4, 7]. This suggests a simplification, in which the pseudo-score test statistic is replaced by another residual diagnostic that is easier to interpret and to compute.

In special cases this diagnostic is a residual version of one of the classical functional summary statistics K, G or F obtained by subtracting a "*compensator*" from the functional summary statistic. The compensator depends on the fitted model, and may also depend on the observed data. For example, suppose the fitted model is the homogeneous Poisson process. Then (ignoring some details) the compensator of the empirical K-function $\hat{K}(r)$ is its expectation $K_0(r) = \pi r^2$ under the model, while the compensator of the empirical nearest neighbor function $\hat{G}(r)$ is the empirical empty space function $\hat{F}(r)$ for the same data. This approach provides a new class of residual summary statistics that can be used as informal diagnostics for model fit, for a wide range of point process models, in close analogy with current practice. The diagnostics apply under very general conditions, including the case of inhomogeneous point process models, where exploratory methods are underdeveloped or inapplicable. For instance,

Figure 2 shows the compensator of $K(r)$ for an inhomogeneous Strauss process.

Section 2 introduces basic definitions and assumptions. Section 3 describes the score test for a general point process model, and Section 4 develops the important case of Poisson point process models. Section 5 gives examples and technical tools for non-Poisson point process models. Section 6 develops the general theory for our diagnostic tools. Section 7 applies these tools to tests for first order trend and hotspots. Sections 8–11 develop diagnostics for interaction between
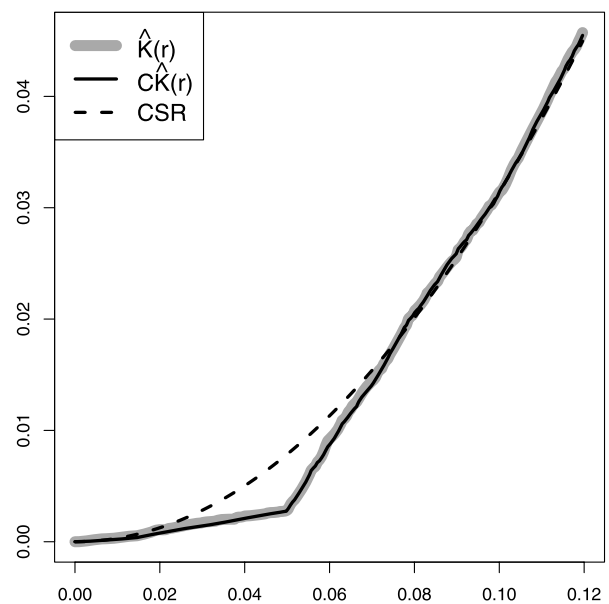

FIG. 2. *Empirical K-function* (*thick grey line*) *for the point pattern data in Figure* 1(b), *compensator of the K-function* (*solid black line*) *for a model of the correct form, and expected K-function for a homogeneous Poisson process* (*dashed line*).

points, based on pairwise distances, nearest neighbor distances and empty space distances, respectively. The tools are demonstrated on data in Sections 12–15. Further examples of diagnostics are given in Appendix A. Appendices B–E provide technical details.

2. ASSUMPTIONS

2.1 Fundamentals

A spatial point pattern data set is a finite set $\mathbf{x} = \{x_1, \ldots, x_n\}$ of points $x_i \in W$, where the number of points $n(\mathbf{x}) = n \geq 0$ is not fixed in advance, and the domain of observation $W \subset \mathbb{R}^d$ is a fixed, known region of d-dimensional space with finite positive volume $|W|$. We take $d = 2$, but the results generalize easily to all dimensions.

A point process model assumes that $\mathbf{x}$ is a realization of a finite point process $\mathbf{X}$ in W without multiple points. We can equivalently view $\mathbf{X}$ as a random finite subset of W. Much of the literature on spatial statistics assumes that $\mathbf{X}$ is the restriction $\mathbf{X} = \mathbf{Y} \cap W$ of a stationary point process $\mathbf{Y}$ on the entire space $\mathbb{R}^2$. We do not assume this; there is no assumption of stationarity, and some of the models considered here are intrinsically confined to the domain W. For further background material including measure theoretical details, see, for example, [50], Appendix B.

Write $\mathbf{X} \sim \text{Poisson}(W, \rho)$ if $\mathbf{X}$ follows the Poisson process on W with intensity function ρ, where we assume $\nu = \int_W \rho(u) \, du$ is finite. Then $n(\mathbf{X})$ is Poisson distributed with mean ν, and, conditional on $n(\mathbf{X})$, the points in $\mathbf{X}$ are i.i.d. with density $\rho(u)/\nu$.

Every point process model considered here is assumed to have a probability density with respect to $\text{Poisson}(W, 1)$, the unit rate Poisson process, under one of the following scenarios.

2.2 Unconditional Case

In the *unconditional case* we assume $\mathbf{X}$ has a density f with respect to $\text{Poisson}(W, 1)$. Then the density is characterized by the property

$$(1) \qquad \mathbb{E}[h(\mathbf{X})] = \mathbb{E}[h(\mathbf{Y}) f(\mathbf{Y})]$$

for all nonnegative measurable functionals h, where $\mathbf{Y} \sim \text{Poisson}(W, 1)$. In particular, the density of $\text{Poisson}(W, \rho)$ is

$$(2) \qquad f(\mathbf{x}) = \exp\left(\int_W (1 - \rho(u)) \, du \right) \prod_i \rho(x_i).$$

We assume that f is hereditary, that is, $f(\mathbf{x}) > 0$ implies $f(\mathbf{y}) > 0$ for all finite $\mathbf{y} \subset \mathbf{x} \subset W$. Processes

satisfying these assumptions include (under integrability conditions) inhomogeneous Poisson processes with an intensity function, finite Gibbs processes contained in W, and Cox processes driven by random fields. See [38], Chapter 3, for an overview of finite point processes including these examples. In practice, our methods require the density to have a tractable form, and are only developed for Poisson and Gibbs processes.

2.3 Conditional Case

In the *conditional case*, we assume $\mathbf{X} = \mathbf{Y} \cap W$ where $\mathbf{Y}$ is a point process. Thus, $\mathbf{X}$ may depend on unobserved points of $\mathbf{Y}$ lying outside W. The density of $\mathbf{X}$ may be unknown or intractable. Under suitable conditions (explained in Section 5.4) modeling and inference can be based on the conditional distribution of $\mathbf{X}^\circ = \mathbf{X} \cap W^\circ$ given $\mathbf{X}^+ = \mathbf{X} \cap W^+ = \mathbf{x}^+$, where $W^+ \subset W$ is a subregion, typically a region near the boundary of W, and only the points in $W^\circ = W \setminus W^+$ are treated as random. We assume that the conditional distribution of $\mathbf{X}^\circ = \mathbf{X} \cap W^\circ$ given $\mathbf{X}^+ = \mathbf{X} \cap W^+ = \mathbf{x}^+$ has an hereditary density $f(\mathbf{x}^\circ | \mathbf{x}^+)$ with respect to $\text{Poisson}(W^\circ, 1)$. Processes satisfying these assumptions include Markov point processes [74], [50], Section 6.4, together with all processes covered by the unconditional case. Our methods are only developed for Poisson and Markov point processes.

For ease of exposition, we focus mainly on the unconditional case, with occasional comments on the conditional case. For Poisson point process models, we always take $W = W^\circ$ so that the two cases agree.

3. SCORE TEST FOR POINT PROCESSES

In principle, any technique for likelihood-based inference is applicable to point process likelihoods. In practice, many likelihood computations require extensive Monte Carlo simulation [31, 50, 51]. To minimize such difficulties, when assessing the goodness of fit of a fitted point process model, it is natural to choose the score test which only requires computations for the null hypothesis [61, 75].

Consider any parametric family of point process models for $\mathbf{X}$ with density f_θ indexed by a k-dimensional vector parameter $\theta \in \Theta \subseteq \mathbb{R}^k$. For a *simple null hypothesis* $H_0 : \theta = \theta_0$ where $\theta_0 \in \Theta$ is fixed, the score test against any alternative $H_1 : \theta \in \Theta_1$, where $\Theta_1 \subseteq \Theta \setminus \{\theta_0\}$, is based on the score test statistic ([22], page 315),

$$(3) \qquad T^2 = U(\theta_0)^\top I(\theta_0)^{-1} U(\theta_0).$$

Here $U(\theta) = \frac{\partial}{\partial \theta} \log f_\theta(\mathbf{x})$ and $I(\theta) = \mathbb{E}_\theta[U(\theta)U(\theta)^\top]$ are the score function and Fisher information, respectively, and the expectation is with respect to f_θ. Here and throughout, we assume that the order of integration and differentation *with respect to* θ can be interchanged. Under suitable conditions, the null distribution of T^2 is χ^2 with k degrees of freedom. In the case $k = 1$ it may be informative to evaluate the signed square root

$$(4) \qquad T = U(\theta_0)/\sqrt{I(\theta_0)},$$

which is asymptotically $N(0, 1)$ distributed under the same conditions.

For a *composite* null hypothesis $H_0 : \theta \in \Theta_0$ where $\Theta_0 \subset \Theta$ is an m-dimensional submanifold with $0 < m < k$, the score test statistic is defined in [22], page 324. However, we shall not use this version of the score test, as it assumes differentiability of the likelihood with respect to nuisance parameters, which is not necessarily applicable here (as exemplified in Section 4.2).

In the sequel we often consider models of the form

$$(5) \qquad f_{(\alpha,\beta)}(\mathbf{x}) = c(\alpha, \beta)h_\alpha(\mathbf{x})\exp(\beta S(\mathbf{x})),$$

where the parameter β and the statistic $S(\mathbf{x})$ are one dimensional, and the null hypothesis is $H_0 : \beta = 0$. For fixed α, this is a linear exponential family and (4) becomes

$$T(\alpha) = \left(S(\mathbf{x}) - \mathbb{E}_{(\alpha,0)}[S(\mathbf{X})]\right)/\sqrt{\mathbb{V}\mathrm{ar}_{(\alpha,0)}[S(\mathbf{X})]}.$$

In practice, when α is unknown, we replace α by its MLE under H_0 so that, with a slight abuse of notation, the signed square root of the score test statistic is approximated by

$$(6) \quad \begin{aligned} T &= T(\hat{\alpha}) \\ &= \left(S(\mathbf{x}) - \mathbb{E}_{(\hat{\alpha},0)}[S(\mathbf{X})]\right)/\sqrt{\mathbb{V}\mathrm{ar}_{(\hat{\alpha},0)}[S(\mathbf{X})]}. \end{aligned}$$

Under suitable conditions, T in (6) is asymptotically equivalent to T in (4), and so a standard Normal approximation may still apply.

4. SCORE TEST FOR POISSON PROCESSES

Application of the score test to Poisson point process models appears to originate with Cox [21]. Consider a parametric family of Poisson processes, Poisson(W, ρ_θ), where the intensity function is indexed by $\theta \in \Theta$. The score test statistic is (3), where

$$U(\theta) = \sum_i \kappa_\theta(x_i) - \int_W \kappa_\theta(u)\rho_\theta(u)\,\mathrm{d}u,$$

$$I(\theta) = \int_W \kappa_\theta(u)\kappa_\theta(u)^\top \rho_\theta(u)\,\mathrm{d}u$$

with $\kappa_\theta(u) = \frac{\partial}{\partial \theta} \log \rho_\theta(u)$. Asymptotic results are given in [45, 62].

4.1 Log-Linear Alternative

The score test is commonly used in spatial epidemiology to assess whether disease incidence depends on environmental exposure. As a particular case of (5), suppose the Poisson model has a log-linear intensity function

$$(7) \qquad \rho_{(\alpha,\beta)}(u) = \exp(\alpha + \beta Z(u)),$$

where $Z(u), u \in W$, is a known, real-valued and nonconstant covariate function, and α and β are real parameters. Cox [21] noted that the uniformly most powerful test of $H_0 : \beta = 0$ (the homogeneous Poisson process) against $H_1 : \beta > 0$ is based on the statistic

$$(8) \qquad S(\mathbf{x}) = \sum_i Z(x_i).$$

Recall that, for a point process $\mathbf{X}$ on W with intensity function ρ, we have Campbell's Formula ([24], page 163),

$$(9) \qquad \mathbb{E}\left(\sum_{x_i \in \mathbf{X}} h(x_i)\right) = \int_W h(u)\rho(u)\,\mathrm{d}u$$

for any Borel function h such that the integral on the right-hand side exists; and for the Poisson process Poisson(W, ρ),

$$(10) \qquad \mathbb{V}\mathrm{ar}\left(\sum_{x_i \in \mathbf{X}} h(x_i)\right) = \int_W h(u)^2 \rho(u)\,\mathrm{d}u$$

for any Borel function h such that the integral on the right-hand side exists. Hence, the standardized version of (8) is

$$(11) \quad T = \left(S(\mathbf{x}) - \hat{\kappa}\int_W Z(u)\,\mathrm{d}u\right)\Big/\sqrt{\hat{\kappa}\int_W Z(u)^2\,\mathrm{d}u},$$

where $\hat{\kappa} = n/|W|$ is the MLE of the intensity $\kappa = \exp(\alpha)$ under the null hypothesis. This is a direct application of the approximation (6) of the signed square root of the score test statistic.

Berman [13] proposed several tests and diagnostics for spatial association between a point process $\mathbf{X}$ and a covariate function $Z(u)$. Berman's Z_1 test is equivalent to the Cox score test described above. Waller et al. [76] and Lawson [47] proposed tests for the dependence of disease incidence on environmental exposure, based on data giving point locations of disease cases. These are also applications of the score test. Berman conditioned on the number of points when making inference. This is in accordance with the observation that the statistic $n(\mathbf{x})$ is S-ancillary for β, while $S(\mathbf{x})$ is S-sufficient for β.

4.2 Threshold Alternative and Nuisance Parameters

Consider the Poisson process with an intensity function of "threshold" form,

$$\rho_{z,\kappa,\phi}(u) = \begin{cases} \kappa \exp(\phi) & \text{if } Z(u) \leq z, \\ \kappa & \text{if } Z(u) > z, \end{cases}$$

where z is the threshold level. If z is fixed, this model is a special case of (7) with $Z(u)$ replaced by $\mathbb{I}\{Z(u) \leq z\}$, and so (8) is replaced by

$$S(\mathbf{x}) = S(\mathbf{x}, z) = \sum_i \mathbb{I}\{Z(x_i) \leq z\},$$

where $\mathbb{I}\{\cdot\}$ denotes the indicator function. By (11) the (approximate) score test of $H_0 : \phi = 0$ against $H_1 : \phi \neq 0$ is based on

$$T = T(z) = \left(S(\mathbf{x}, z) - \hat{\kappa} A(z)\right) / \sqrt{\hat{\kappa} A(z)},$$

where $A(z) = |\{u \in W : Z(u) \leq z\}|$ is the area of the corresponding level set of Z.

If z is not fixed, then it plays the role of a nuisance parameter in the score test: the value of z affects inference about the canonical parameter ϕ, which is the parameter of primary interest in the score test. Note that the likelihood is not differentiable with respect to z.

In most applications of the score test, a nuisance parameter would be replaced by its MLE under the null hypothesis. However, in this context, z is not identifiable under the null hypothesis. Several solutions have been proposed [18, 25, 26, 33, 68]. They include replacing z by its MLE under the alternative [18], maximizing $T(z)$ or $|T(z)|$ over z [25, 26], and finding the maximum p-value of $T(z)$ or $|T(z)|$ over a confidence region for z under the alternative [68].

These approaches appear to be inapplicable to the current context. While the null distribution of $T(z)$ is asymptotically $N(0, 1)$ for each fixed z as $\kappa \to \infty$, this convergence is not uniform in z. The null distribution of $S(\mathbf{x}, z)$ is Poisson with parameter $\kappa A(z)$; sample paths of $T(z)$ will be governed by Poisson behavior where $A(z)$ is small.

In this paper, our approach is simply to plot the score test statistic as a function of the nuisance parameter. This turns the score test into a graphical exploratory tool, following the approach adopted in many other areas [2, 15, 19, 60, 77]. A second style of plot based on $S(x, z) - \hat{\kappa} A(z)$ against z may be more appropriate visually. Such a plot is the lurking variable plot of [7]. Berman [13] also proposed a plot of $S(\mathbf{x}, z)$ against z, together with a plot of $\hat{\kappa} A(z)$ against z, as a diagnostic for dependence on Z. This is related

to the Kolmogorov–Smirnov test since, under H_0, the values $Y_i = Z(x_i)$ are i.i.d. with distribution function $\mathbb{P}(Y \leq y) = A(y)/|W|$.

4.3 Hot Spot Alternative

Consider the Poisson process with intensity

$$(12) \qquad \rho_{\kappa,\phi,v}(u) = \kappa \exp(\phi k(u - v)),$$

where k is a kernel (a probability density on $\mathbb{R}^2$), $\kappa > 0$ and ϕ are real parameters, and $v \in \mathbb{R}^2$ is a nuisance parameter. This process has a "hot spot" of elevated intensity in the vicinity of the location v. By (11) and (9)–(10) the score test of $H_0 : \phi = 0$ against $H_1 : \phi \neq 0$ is based on

$$T = T(v) = \left(S(\mathbf{x}, v) - \hat{\kappa} M_1(v)\right) / \sqrt{\hat{\kappa} M_2(v)},$$

where

$$S(\mathbf{x}, v) = \sum_i k(x_i - v)$$

is the usual nonparametric kernel estimate of point process intensity [28] evaluated at v without edge correction, and

$$M_i(v) = \int_W k(u - v)^i \, du, \quad i = 1, 2.$$

The numerator $S(\mathbf{x}, v) - \hat{\kappa} M_1(v)$ is the *smoothed residual field* [7] of the null model. In the special case where $k(u) \propto \mathbb{I}\{\|u\| \leq h\}$ is the uniform density on a disc of radius h, the maximum $\max_v T(v)$ is closely related to the *scan statistic* [1, 44].

5. NON-POISSON MODELS

The remainder of the paper deals with the case where the alternative (and perhaps also the null) is not a Poisson process. Key examples are stated in Section 5.1. Non-Poisson models require additional tools including the Papangelou conditional intensity (Section 5.2) and pseudo-likelihood (Section 5.3).

5.1 Point Process Models with Interaction

We shall frequently consider densities of the form

$$(13) \qquad f(\mathbf{x}) = c \left[\prod_i \lambda(x_i) \right] \exp(\phi V(\mathbf{x})),$$

where c is a normalizing constant, the first order term λ is a nonnegative function, ϕ is a real interaction parameter, and $V(\mathbf{x})$ is a real nonadditive function which specifies the interaction between the points. We refer to V as the interaction potential. In general, apart from the

Poisson density (2) corresponding to the case $\phi = 0$, the normalizing constant is not expressible in closed form.

Often the definition of V can be extended to all finite point patterns in $\mathbb{R}^2$ so as to be invariant under rigid motions (translations and rotations). Then the model for $\mathbf{X}$ is said to be homogeneous if λ is constant on W, and inhomogeneous otherwise.

Let

$$d(u, \mathbf{x}) = \min_j \|u - x_j\|$$

denote the distance from a location u to its nearest neighbor in the point configuration $\mathbf{x}$. For $n(\mathbf{x}) = n \geq 1$ and $i = 1, \ldots, n$, define

$$\mathbf{x}_{-i} = \mathbf{x} \setminus \{x_i\}.$$

In many places in this paper we consider the following three motion-invariant interaction potentials $V(\mathbf{x}) = V(\mathbf{x}, r)$ depending on a parameter $r > 0$ which specifies the range of interaction. The *Strauss process* [73] has interaction potential

$$(14) \qquad V_S(\mathbf{x}, r) = \sum_{i < j} \mathbb{I}\{\|x_i - x_j\| \leq r\},$$

the number of r-close pairs of points in $\mathbf{x}$; the *Geyer saturation model* [31] with saturation threshold 1 has interaction potential

$$(15) \qquad V_G(\mathbf{x}, r) = \sum_i \mathbb{I}\{d(x_i, \mathbf{x}_{-i}) \leq r\},$$

the number of points in $\mathbf{x}$ whose nearest neighbor is closer than r units; and the Widom–Rowlinson penetrable sphere model [78] or *area-interaction process* [11] has interaction potential

$$(16) \qquad V_A(\mathbf{x}, r) = -\left| W \cap \bigcup_i B(x_i, r) \right|,$$

the negative area of W intersected with the union of balls $B(x_i, r)$ of radius r centered at the points of $\mathbf{x}$. Each of these densities favors spatial clustering (positive association) when $\phi > 0$ and spatial inhibition (negative association) when $\phi < 0$. The Geyer and area-interaction models are well-defined point processes for any value of ϕ [11, 31], but the Strauss density is integrable only when $\phi \leq 0$ [43].

5.2 Conditional Intensity

Consider a parametric model for a point process $\mathbf{X}$ in $\mathbb{R}^2$, with parameter $\theta \in \Theta$. Papangelou [59] defined

the *conditional intensity* of $\mathbf{X}$ as a nonnegative stochastic process $\lambda_\theta(u, \mathbf{X})$ indexed by locations $u \in \mathbb{R}^2$ and characterized by the property that

$$(17)$$
$$\mathbb{E}_\theta \left[\sum_{x_i \in \mathbf{X}} h(x_i, \mathbf{X} \setminus \{x_i\}) \right]$$
$$= \mathbb{E}_\theta \left[\int_{\mathbb{R}^2} h(u, \mathbf{X}) \lambda_\theta(u, \mathbf{X}) \, du \right]$$

for all measurable functions h such that the left or right-hand side exists. Equation (17) is known as the *Georgii–Nguyen–Zessin (GNZ) formula* [30, 41, 42, 52]; see also Section 6.4.1 in [50]. Adapting a term from stochastic process theory, we will call the random integral on the right-hand side of (17) the *(Papangelou) compensator* of the random sum on the left-hand side.

Consider a finite point process $\mathbf{X}$ in W. In the unconditional case (Section 2.2) we assume $\mathbf{X}$ has density $f_\theta(\mathbf{x})$ which is hereditary for all $\theta \in \Theta$. We may simply define

$$(18) \qquad \lambda_\theta(u, \mathbf{x}) = f_\theta(\mathbf{x} \cup \{u\}) / f_\theta(\mathbf{x})$$

for all locations $u \in W$ and point configurations $\mathbf{x} \subset W$ such that $u \notin \mathbf{x}$. Here we take $0/0 = 0$. For $x_i \in \mathbf{x}$ we set $\lambda_\theta(x_i, \mathbf{x}) = \lambda_\theta(x_i, \mathbf{x}_{-i})$, and for $u \notin W$ we set $\lambda_\theta(u, \mathbf{x}) = 0$. Then it may be verified directly from (1) that (17) holds, so that (18) is the Papangelou conditional intensity of $\mathbf{X}$. Note that the normalizing constant of f_θ cancels in (18). For a Poisson process, it follows from (2) and (18) that the Papangelou conditional intensity is equivalent to the intensity function of the process.

In the conditional case (Section 2.3) we assume that the conditional distribution of $\mathbf{X}^\circ = \mathbf{X} \cap W^\circ$ given $\mathbf{X}^+ = \mathbf{X} \cap W^+ = \mathbf{x}^+$ has a hereditary density $f_\theta(\mathbf{x}^\circ | \mathbf{x}^+)$ with respect to Poisson$(W^\circ, 1)$, for all $\theta \in \Theta$. Then define

$$(19) \qquad \lambda_\theta(u, \mathbf{x}^\circ | \mathbf{x}^+) = \frac{f_\theta(\mathbf{x}^\circ \cup \{u\} | \mathbf{x}^+)}{f_\theta(\mathbf{x}^\circ \setminus \{u\} | \mathbf{x}^+)}$$

if $u \in W^\circ$, and zero otherwise. It can similarly be verified that this is the Papangelou conditional intensity of the conditional distribution of $\mathbf{X}^\circ$ given $\mathbf{X}^+ = \mathbf{x}^+$.

It is convenient to rewrite (18) in the form

$$\lambda_\theta(u, \mathbf{x}) = \exp(\Delta_u \log f(\mathbf{x})),$$

where Δ is the one-point difference operator

$$(20) \qquad \Delta_u h(\mathbf{x}) = h(\mathbf{x} \cup \{u\}) - h(\mathbf{x} \setminus \{u\}).$$

Note the Poincaré inequality for the Poisson process $\mathbf{X}$,

$$(21) \qquad \mathbb{V}\mathrm{ar}[h(\mathbf{X})] \leq \mathbb{E} \int_W [\Delta_u h(\mathbf{X})]^2 \rho(u) \, du$$

holding for all measurable functionals h such that the right-hand side is finite; see [46, 79].

5.3 Pseudo-Likelihood and Pseudo-Score

To avoid computational problems with point process likelihoods, Besag [14] introduced the *pseudo-likelihood* function

$$\mathsf{PL}(\theta) = \left[\prod_i \lambda_\theta(x_i, \mathbf{x})\right]$$
$$(22)$$
$$\cdot \exp\left(-\int_W \lambda_\theta(u, \mathbf{x})\, du\right).$$

This is of the same functional form as the likelihood function of a Poisson process (2), but has the Papangelou conditional intensity in place of the Poisson intensity. The corresponding *pseudo-score*

$$\mathsf{PU}(\theta) = \frac{\partial}{\partial\theta} \log \mathsf{PL}(\theta)$$
$$(23)$$
$$= \sum_i \frac{\partial}{\partial\theta} \log \lambda_\theta(x_i, \mathbf{x}) - \int_W \frac{\partial}{\partial\theta} \lambda_\theta(u, \mathbf{x})\, du$$

is an unbiased estimating function, $\mathbb{E}_\theta \mathsf{PU}(\theta) = 0$, by virtue of (17). In practice, the pseudo-likelihood is applicable only if the Papangelou conditional intensity $\lambda_\theta(u, \mathbf{x})$ is tractable.

The pseudo-likelihood function can also be defined in the conditional case [39]. In (22) the product is instead over points $x_i \in \mathbf{x}^\circ$ and the integral is instead over W°; in (23) the sum is instead over points $x_i \in \mathbf{x}^\circ$ and the integral is instead over W°; and in both places $\mathbf{x} = \mathbf{x}^\circ \cup \mathbf{x}^+$. The Papangelou conditional intensity $\lambda_\theta(u, \mathbf{x})$ must also be replaced by $\lambda_\theta(u, \mathbf{x}^\circ|\mathbf{x}^+)$.

5.4 Markov Point Processes

For a point process $\mathbf{X}$ constructed as $\mathbf{X} = \mathbf{Y} \cap W$ where $\mathbf{Y}$ is a point process in $\mathbb{R}^2$, the density and Papangelou conditional intensity of $\mathbf{X}$ may not be available in simple form. Progress can be made if $\mathbf{Y}$ is a *Markov point process* of interaction range $R < \infty$; see [30, 52, 66, 74] and [50], Section 6.4.1. Briefly, this means that the Papangelou conditional intensity $\lambda_\theta(u, \mathbf{Y})$ of $\mathbf{Y}$ satisfies $\lambda_\theta(u, \mathbf{Y}) = \lambda_\theta(u, \mathbf{Y} \cap B(u, R))$, where $B(u, R)$ is the ball of radius R centered at u. Define the erosion of W by distance R,

$$W_{\ominus R} = \{u \in W : B(u, R) \subset W\},$$

and assume this has nonzero area. Let $B = W \setminus W_{\ominus R}$ be the border region. The process satisfies a spatial Markov property: the processes $\mathbf{Y} \cap W_{\ominus R}$ and $\mathbf{Y} \cap W^c$ are conditionally independent given $\mathbf{Y} \cap B$.

In this situation we shall invoke the conditional case with $W^\circ = W_{\ominus R}$ and $W^+ = W \setminus W^\circ$. The conditional

distribution of $\mathbf{X} \cap W^\circ$ given $\mathbf{X} \cap W^+ = \mathbf{x}^+$ has Papangelou conditional intensity

$$(24) \quad \lambda_\theta(u, \mathbf{x}^\circ|\mathbf{x}^+) = \begin{cases} \lambda_\theta(u, \mathbf{x}^\circ \cup \mathbf{x}^+) & \text{if } u \in W^\circ, \\ 0 & \text{otherwise.} \end{cases}$$

Thus, *the unconditional and conditional versions of a Markov point process have the same Papangelou conditional intensity* at locations in W°.

For $\mathbf{x}^\circ = \{x_1, \ldots, x_{n^\circ}\}$, the conditional probability density given $\mathbf{x}^+$ becomes

$$f_\theta(\mathbf{x}^\circ|\mathbf{x}^+)$$

$$= c_\theta(\mathbf{x}^+)\lambda_\theta(x_1, \mathbf{x}^\circ) \prod_{i=2}^{n^\circ} \lambda_\theta(x_i, \{x_1, \ldots, x_{i-1}\} \cup \mathbf{x}^+)$$

if $n^\circ > 0$, and $f_\theta(\varnothing|\mathbf{x}^+) = c_\theta(\mathbf{x}^+)$, where $\varnothing$ denotes the empty configuration, and the inverse normalizing constant $c_\theta(\mathbf{x}^+)$ depends only on $\mathbf{x}^+$.

For example, instead of (13) we now consider

$$f(\mathbf{x}^\circ|\mathbf{x}^+) = c(\mathbf{x}^+)\left[\prod_{i=1}^{n^\circ} \lambda(x_i)\right] \exp(\phi V(\mathbf{x}^\circ \cup \mathbf{x}^+)),$$

assuming $V(\mathbf{y})$ is defined for all finite $\mathbf{y} \subset \mathbb{R}^2$ such that for any $u \in \mathbb{R}^2 \setminus \mathbf{y}$, $\Delta_u V(\mathbf{y})$ depends only on u and $\mathbf{y} \cap B(u, R)$. This condition is satisfied by the interaction potentials (14)–(16); note that the range of interaction is $R = r$ for the Strauss process, and $R = 2r$ for both the Geyer and the area-interaction models.

6. SCORE, PSEUDO-SCORE AND RESIDUAL DIAGNOSTICS

This section develops the general theory for our diagnostic tools.

By (6) in Section 3 it is clear that comparison of a summary statistic $S(\mathbf{x})$ to its predicted value $\mathbb{E}S(\mathbf{X})$ under a null model is effectively equivalent to the score test under an exponential family model where $S(\mathbf{x})$ is the canonical sufficient statistic. Similarly, the use of a *functional* summary statistic $S(\mathbf{x}, z)$, depending on a function argument z, is related to the score test under an exponential family model *where z is a nuisance parameter* and $S(\mathbf{x}, z)$ is the canonical sufficient statistic for fixed z. In this section we construct the corresponding exponential family models, apply the score test, and propose surrogates for the score test statistic.

6.1 Models

Let $f_\theta(\mathbf{x})$ be the density of any point process $\mathbf{X}$ on W governed by a parameter θ. Let $S(\mathbf{x}, z)$ be a functional

summary statistic of the point pattern data set **x**, with function argument z belonging to any space.

Consider the *extended model* with density

$$(25) \qquad f_{\theta,\phi,z}(\mathbf{x}) = c_{\theta,\phi,z} f_\theta(\mathbf{x}) \exp(\phi S(\mathbf{x}, z)),$$

where ϕ is a real parameter, and $c_{\theta,\phi,z}$ is the normalizing constant. The density is well-defined provided

$$M(\theta, \phi, z) = \mathbb{E}[f_\theta(\mathbf{Y}) \exp(\phi S(\mathbf{Y}, z))] < \infty,$$

where $\mathbf{Y} \sim \text{Poisson}(W, 1)$. The extended model is constructed by "exponential tilting" of the original model by the statistic S. By (6), for fixed θ and z, assuming differentiability of M with respect to ϕ in a neighborhood of $\phi = 0$, the signed root of the score test statistic is approximated by

$$(26) \quad T = (S(\mathbf{x}, z) - \mathbb{E}_{\hat{\theta}}[S(\mathbf{X}, z)]) / \sqrt{\mathbb{V}\text{ar}_{\hat{\theta}}[S(\mathbf{X}, z)]},$$

where $\hat{\theta}$ is the MLE under the null model, and the expectation and variance are with respect to the null model with density $f_{\hat{\theta}}$.

Insight into the qualitative behavior of the extended model (25) can be obtained by studying the *perturbing model*

$$(27) \qquad g_{\phi,z}(\mathbf{x}) = k_{\phi,z} \exp(\phi S(\mathbf{x}, z)),$$

provided this is a well-defined density with respect to Poisson$(W, 1)$, where $k_{\phi,z}$ is the normalizing constant. When the null hypothesis is a homogeneous Poisson process, the extended model is identical to the perturbing model, up to a change in the first order term. In general, the extended model is a qualitative hybrid between the null and perturbing models.

In this context the score test is equivalent to naive comparison of the observed and null-expected values of the functional summary statistic S. The test statistic T in (26) may be difficult to evaluate; typically, apart from Poisson models, the moments (particularly the variance) of S would not be available in closed form. The null distribution of T would also typically be unknown. Hence, implementation of the score test would typically require moment approximation and simulation from the null model, which in both cases may be computationally expensive. Various approximations for the score or the score test statistic can be constructed, as discussed in the sequel.

6.2 Pseudo-Score of Extended Model

The extended model (25) is an exponential family with respect to ϕ, having Papangelou conditional intensity

$$\kappa_{\theta,\phi,z}(u, \mathbf{x}) = \lambda_\theta(u, \mathbf{x}) \exp(\phi \Delta_u S(\mathbf{x}, z)),$$

where $\lambda_\theta(u, \mathbf{x})$ is the Papangelou conditional intensity of the null model. The pseudo-score function with respect to ϕ, evaluated at $\phi = 0$, is

$$\mathsf{PU}(\theta, z) = \sum_i \Delta_{x_i} S(\mathbf{x}, z) - \int_W \Delta_u S(\mathbf{x}, z) \lambda_\theta(u, \mathbf{x}) \, du,$$

where the first term

$$(28) \qquad \Sigma \Delta S(\mathbf{x}, z) = \sum_i \Delta_{x_i} S(\mathbf{x}, z)$$

will be called the *pseudo-sum* of S. If $\hat{\theta}$ is the maximum pseudo-likelihood estimate (MPLE) under H_0, the second term with θ replaced by $\hat{\theta}$ becomes

$$(29) \qquad \mathcal{C}\Delta S(\mathbf{x}, z) = \int_W \Delta_u S(\mathbf{x}, z) \lambda_{\hat{\theta}}(u, \mathbf{x}) \, du$$

and will be called the *(estimated) pseudo-compensator* of S. We call

$$(30) \qquad \begin{aligned} \mathcal{R}\Delta S(\mathbf{x}, z) &= \mathsf{PU}(\hat{\theta}, z) \\ &= \Sigma \Delta S(\mathbf{x}, z) - \mathcal{C}\Delta S(\mathbf{x}, z) \end{aligned}$$

the *pseudo-residual* since it is a weighted residual in the sense of [7].

The pseudo-residual serves as a surrogate for the numerator in the score test statistic (26). For the denominator, we need the variance of the pseudo-residual. Appendix B gives an exact formula (66) for the variance of the pseudo-score $\mathsf{PU}(\theta, z)$, which can serve as an approximation to the variance of the pseudo-residual $\mathcal{R}\Delta S(\mathbf{x}, z)$. This is likely to be an overestimate, because the effect of parameter estimation is typically to deflate the residual variance [7].

The first term in the variance formula (66) is

$$(31) \quad \mathcal{C}^2 \Delta S(\mathbf{x}, z) = \int_W [\Delta_u S(\mathbf{x}, z)]^2 \lambda_{\hat{\theta}}(u, \mathbf{x}) \, du,$$

which we shall call the *Poincaré pseudo-variance* because of its similarity to the Poincaré upper bound in (21). It is easy to compute this quantity alongside the pseudo-residual. Rough calculations in Sections 9.4 and 10.3 suggest that the Poincaré pseudo-variance is likely to be the dominant term in the variance, except at small r values. The variance of residuals is also studied in [17].

For computational efficiency we propose to use the square root of (31) as a surrogate for the denominator in (26). This yields a *"standardized" pseudo-residual*

$$(32) \quad \mathcal{T}\Delta S(\mathbf{x}, z) = \mathcal{R}\Delta S(\mathbf{x}, z) / \sqrt{\mathcal{C}^2 \Delta S(\mathbf{x}, z)}.$$

We emphasize that this quantity is not guaranteed to have zero mean and unit variance (even approximately)

under the null hypothesis. It is merely a computationally efficient surrogate for the score test statistic; its null distribution must be investigated by other means. Asymptotics of $\mathcal{T}\Delta S(\mathbf{x}, z)$ under a large-domain limit [69] could be studied, but limit results are unlikely to hold uniformly over r. In this paper we evaluate null distributions using Monte Carlo methods.

The pseudo-sum (28) can be regarded as a functional summary statistic for the data in its own right. Its definition depends only on the choice of the statistic S, and it may have a meaningful interpretation as a nonparametric estimator of a property of the point process. The pseudo-compensator (29) might also be regarded as a functional summary statistic, but its definition involves the null model. If the null model is true, we may expect the pseudo-residual to be approximately zero. Sections 9–11 and Appendix A study particular instances of pseudo-residual diagnostics based on (28)–(30).

In the conditional case, the Papangelou conditional intensity $\lambda_{\hat{\theta}}(u, \mathbf{x})$ must be replaced by $\lambda_{\hat{\theta}}(u, \mathbf{x}^\circ | \mathbf{x}^+)$ given in (19) or (24). The integral in the definition of the pseudo-compensator (29) must be restricted to the domain W°, and the summation over data points in (28) must be restricted to points $x_i \in W^\circ$, that is, to summation over points of $\mathbf{x}^\circ$.

6.3 Residuals

A simpler surrogate for the score test is available when the canonical sufficient statistic S of the perturbing model is naturally expressible as a sum of local contributions

$$(33) \qquad S(\mathbf{x}, z) = \sum_i s(x_i, \mathbf{x}_{-i}, z).$$

Note that any statistic can be decomposed in this way unless some restriction is imposed on s; such a decomposition is not necessarily unique. We call the decomposition "natural" if $s(u, \mathbf{x}, z)$ only depends on points of $\mathbf{x}$ that are close to u, as demonstrated in the examples in Sections 9, 10 and 11 and in Appendix A.

Consider a null model with Papangelou conditional intensity $\lambda_\theta(u, \mathbf{x})$. Following [7], define the *(s-weighted) innovation* by

$$(34) \quad \mathcal{I}S(\mathbf{x}, r) = S(\mathbf{x}, z) - \int_W s(u, \mathbf{x}, z)\lambda_\theta(u, \mathbf{x})\, du,$$

which by the GNZ formula (17) has mean zero under the null model. In practice, we replace θ by an estimate $\hat{\theta}$ (e.g., the MPLE) and consider the *(s-weighted) residual*

$$(35) \quad \mathcal{R}S(\mathbf{x}, z) = S(\mathbf{x}, z) - \int_W s(u, \mathbf{x}, z)\lambda_{\hat{\theta}}(u, \mathbf{x})\, du.$$

The residual shares many properties of the score function and can serve as a computationally efficient surrogate for the score. The data-dependent integral

$$(36) \qquad \mathcal{C}S(\mathbf{x}, z) = \int_W s(u, \mathbf{x}, z)\lambda_{\hat{\theta}}(u, \mathbf{x})\, du$$

is the *(estimated) Papangelou compensator* of S. The variance of $\mathcal{R}S(\mathbf{x}, z)$ can be approximated by the innovation variance, given by the general variance formula (65) of Appendix B. The first term in (65) is the *Poincaré variance*

$$(37) \qquad \mathcal{C}^2 S(\mathbf{x}, z) = \int_W s(u, \mathbf{x}, z)^2 \lambda_{\hat{\theta}}(u, \mathbf{x})\, du.$$

Rough calculations reported in Sections 9.4 and 10.3 suggest that the Poincaré variance is likely to be the largest term in the variance for sufficiently large r. By analogy with (31) we propose to use the Poincaré variance as a surrogate for the variance of $\mathcal{R}S(\mathbf{x}, z)$, and thereby obtain a "standardized" residual

$$(38) \qquad \mathcal{T}S(\mathbf{x}, z) = \mathcal{R}S(\mathbf{x}, z)/\sqrt{\mathcal{C}^2 S(\mathbf{x}, z)}.$$

Once again $\mathcal{T}S(\mathbf{x}, z)$ is not exactly standardized, because $\mathcal{C}^2 S(\mathbf{x}, z)$ is an approximation to $\mathbb{V}\text{ar}[\mathcal{R}S(\mathbf{x}, z)]$ and because the numerator and denominator of (38) are dependent. The null distribution of $\mathcal{T}S(\mathbf{x}, z)$ must be investigated by other means.

In the conditional case, the integral in the definition of the compensator (36) must be restricted to the domain W°, and the summation over data points in (33) must be restricted to points $x_i \in W^\circ$, that is, to summation over points of $\mathbf{x}^\circ$.

7. DIAGNOSTICS FOR FIRST ORDER TREND

Consider any null model with density $f_\theta(\mathbf{x})$ and Papangelou conditional intensity $\lambda_\theta(u, \mathbf{x})$. By analogy with Section 4 we consider alternatives of the form (25) where

$$S(\mathbf{x}, z) = \sum_i s(x_i, z)$$

for some function s. The perturbing model (27) is a Poisson process with intensity $\exp(\phi s(\cdot, z))$, where z is a nuisance parameter. The score test is a test for the presence of an (extra) first order trend. The pseudo-score and residual diagnostics are both equal to

$$(39) \quad \begin{aligned} \mathcal{R}S(\mathbf{x}, z) &= \sum_i s(x_i, z) \\ &\quad - \int_W s(u, z)\lambda_{\hat{\theta}}(u, \mathbf{x})\, du. \end{aligned}$$

This is the *s*-weighted residual described in [7]. The variance of (39) can be estimated by simulation, or approximated by the Poincaré variance (37).

If Z is a real-valued covariate function on W, then we may take $s(u, z) = \mathbb{I}\{Z(u) \leq z\}$ for $z \in \mathbb{R}$, corresponding to a threshold effect (cf. Section 4.2). A plot of (39) against z was called a *lurking variable plot* in [7].

If $s(u, z) = k(u - z)$ for $z \in \mathbb{R}^2$, where k is a density function on $\mathbb{R}^2$, then

$$\mathcal{R}S(\mathbf{x}, z) = \sum_i k(x_i - z) - \int_W k(u - z)\lambda_{\hat{\theta}}(u, \mathbf{x})\, du,$$

which was dubbed the *smoothed residual field* in [7]. Examples of application of these techniques have been discussed extensively in [7].

8. INTERPOINT INTERACTION

In the remainder of the paper we concentrate on diagnostics for interpoint interaction.

8.1 Classical Summary Statistics

Following Ripley's influential paper [64], it is standard practice, when investigating association or dependence between points in a spatial point pattern, to evaluate functional summary statistics such as the K-function, and to compare graphically the empirical summaries and theoretical predicted values under a suitable model, often a stationary Poisson process ("Complete Spatial Randomness," CSR) [23, 29, 64].

The three most popular functional summary statistics for spatial point processes are Ripley's K-function, the nearest neighbor distance distribution function G and the empty space function (spherical contact distance distribution function) F. Definitions of K, G and F and their estimators can be seen in [9, 23, 29, 50]. Simple empirical estimators of these functions are of the form

$$\begin{aligned}(40)\quad \hat{K}(r) &= \hat{K}_{\mathbf{x}}(r) \\ &= \frac{1}{\hat{\rho}^2(\mathbf{x})|W|} \sum_{i \neq j} e_K(x_i, x_j)\mathbb{I}\{\|x_i - x_j\| \leq r\},\end{aligned}$$

$$\begin{aligned}(41)\quad \hat{G}(r) &= \hat{G}_{\mathbf{x}}(r) \\ &= \frac{1}{n(\mathbf{x})} \sum_i e_G(x_i, \mathbf{x}_{-i}, r)\mathbb{I}\{d(x_i, \mathbf{x}_{-i}) \leq r\},\end{aligned}$$

$$\begin{aligned}(42)\quad \hat{F}(r) &= \hat{F}_{\mathbf{x}}(r) \\ &= \frac{1}{|W|} \int_W e_F(u, r)\mathbb{I}\{d(u, \mathbf{x}) \leq r\}\, du,\end{aligned}$$

where $e_K(u, v)$, $e_G(u, \mathbf{x}, r)$ and $e_F(u, r)$ are edge correction weights, and typically $\hat{\rho}^2(\mathbf{x}) = n(\mathbf{x})(n(\mathbf{x}) - 1)/|W|^2$.

8.2 Score Test Approach

The classical approach fits naturally into the scheme of Section 6. In order to test for dependence between points, we choose a perturbing model that exhibits dependence. Three interesting examples of perturbing models are the Strauss process, the Geyer saturation model with saturation threshold 1 and the area-interaction process, with interaction potentials $V_S(\mathbf{x}, r)$, $V_G(\mathbf{x}, r)$ and $V_A(\mathbf{x}, r)$ given in (14)–(16). The nuisance parameter $r \geq 0$ determines the range of interaction. It is interesting to note that, although the Strauss density is integrable only when $\phi \leq 0$, the extended model obtained by perturbing f_θ by the Strauss density may be well-defined for some $\phi > 0$. This extended model may support alternatives that are clustered relative to the null, as originally intended by Strauss [73].

The potentials of these three models are closely related to the summary statistics $\hat{K}$, $\hat{G}$ and $\hat{F}$ in (40)–(42). Ignoring the edge correction weights $e(\cdot)$, we have

$$(43)\quad \hat{K}_{\mathbf{x}}(r) \approx \frac{2|W|}{n(\mathbf{x})(n(\mathbf{x}) - 1)} V_S(\mathbf{x}, r),$$

$$(44)\quad \hat{G}_{\mathbf{x}}(r) \approx \frac{1}{n(\mathbf{x})} V_G(\mathbf{x}, r),$$

$$(45)\quad \hat{F}_{\mathbf{x}}(r) \approx -\frac{1}{|W|} V_A(\mathbf{x}, r).$$

To draw the closest possible connection with the score test, instead of choosing the Strauss, Geyer or area-interaction process as the perturbing model, we shall take the perturbing model to be defined through (27) where S is one of the statistics $\hat{K}$, $\hat{G}$ or $\hat{F}$. We call these the *(perturbing) $\hat{K}$-model, $\hat{G}$-model* and *$\hat{F}$-model*, respectively. The score test is then precisely equivalent to comparing $\hat{K}$, $\hat{G}$ or $\hat{F}$ with its predicted expectation using (6).

Essentially $\hat{K}$, $\hat{G}$, $\hat{F}$ are renormalized versions of V_S, V_G, V_A as shown in (43)–(45). In the case of $\hat{F}$ the renormalization is not data-dependent, so the $\hat{F}$-model is virtually an area-interaction model, ignoring edge correction. For $\hat{K}$, the renormalization depends only on $n(\mathbf{x})$, and so, conditionally on $n(\mathbf{x}) = n$, the $\hat{K}$-model and the Strauss process are approximately equivalent. Similarly for $\hat{G}$, the normalization also depends only on $n(\mathbf{x})$, so, conditionally on $n(\mathbf{x}) = n$, the $\hat{G}$-model and Geyer saturation process are approximately equivalent. If we follow Ripley's [64] recommendation to

condition on n when testing for interaction, this implies that the use of the K, G or F-function is approximately equivalent to the score test of CSR against a Strauss, Geyer or area-interaction alternative, respectively.

When the null hypothesis is CSR, we saw that the extended model (25) is identical to the perturbing model, up to a change in intensity, so that the use of the $\hat{K}$-function is equivalent to testing the null hypothesis of CSR against the alternative of a $\hat{K}$-model; similarly for $\hat{G}$ and $\hat{F}$. For a more general null hypothesis, the use of the $\hat{K}$-function, for example, corresponds to adopting an alternative hypothesis that is a hybrid between the fitted model and a $\hat{K}$-model.

Note that if the edge correction weight $e_K(u, v)$ is uniformly bounded, the $\hat{K}$-model is integrable for all values of ϕ, avoiding a difficulty with the Strauss process [43].

Computation of the score test statistic (26) requires estimation or approximation of the null variance of $\hat{K}(r)$, $\hat{G}(r)$ or $\hat{F}(r)$. A wide variety of approximations is available when the null hypothesis is CSR [29, 65]. For other null hypotheses, simulation estimates would typically be used. A central limit theorem is available for $\hat{K}(r)$, $\hat{G}(r)$ and $\hat{F}(r)$ in the large-domain limit, for example, [3, 34, 35, 40, 65]. However, convergence is not uniform in r, and the normal approximation will be poor for small values of r. Instead Ripley [63] developed an exact Monte Carlo test [12, 36] based on simulation envelopes of the summary statistic under the null hypothesis.

In the following sections we develop the residual and pseudo-residual diagnostics corresponding to this approach.

9. RESIDUAL DIAGNOSTICS FOR INTERACTION USING PAIRWISE DISTANCES

This section develops residual (35) and pseudo-residual (30) diagnostics derived from a summary statistic S which is a sum of contributions depending on pairwise distances.

9.1 Residual Based on Perturbing Strauss Model

9.1.1 *General derivation.* Consider any statistic of the general "pairwise interaction" form

$$(46) \qquad S(\mathbf{x}, r) = \sum_{i < j} q(\{x_i, x_j\}, r).$$

This can be decomposed in the local form (33) with

$$s(u, \mathbf{x}, r) = \frac{1}{2} \sum_i q(\{x_i, u\}, r), \quad u \notin \mathbf{x}.$$

Hence,

$$\Delta_{x_i} S(\mathbf{x}, r) = 2s(x_i, \mathbf{x}_{-i}, r) \quad \text{and}$$

$$\Delta_u S(\mathbf{x}, r) = 2s(u, \mathbf{x}, r), \quad u \notin \mathbf{x}.$$

Consequently, the pseudo-residual and the pseudo-compensator are just twice the residual and the Papangelou compensator:

$$(47) \quad \Sigma \Delta S(\mathbf{x}, r) = 2S(\mathbf{x}, r) = \sum_{i \neq j} q(\{x_i, x_j\}, r),$$

$$\mathcal{C} \Delta S(\mathbf{x}, r) = 2\mathcal{C}S(\mathbf{x}, r)$$

$$(48)$$

$$= \int_W \sum_i q(\{x_i, u\}, r)\lambda_{\hat{\theta}}(u, \mathbf{x})\, du,$$

$$\mathcal{R} \Delta S(\mathbf{x}, z) = 2\mathcal{R}S(\mathbf{x}, r)$$

$$(49)$$

$$= 2S(\mathbf{x}, r) - 2\mathcal{C}S(\mathbf{x}, r).$$

9.1.2 *Residual of Strauss potential.* The Strauss interaction potential V_S of (14) is of the general form (46) with $q(\{x_i, x_j\}, r) = \mathbb{I}\{\|x_i - x_j\| \leq r\}$. Hence, V_S can be decomposed in the form (33) with $s(u, \mathbf{x}, r) = \frac{1}{2} t(u, \mathbf{x}, r)$, where

$$t(u, \mathbf{x}, r) = \sum_i \mathbb{I}\{\|u - x_i\| \leq r\}, \quad u \notin \mathbf{x}.$$

Hence, the Papangelou compensator of V_S is

$$(50) \qquad \mathcal{C}V_S(\mathbf{x}, r) = \frac{1}{2} \int_W t(u, \mathbf{x}, r)\lambda_{\hat{\theta}}(u, \mathbf{x})\, du.$$

9.1.3 *Case of CSR.* If the null model is CSR with intensity ρ estimated by $\hat{\rho} = n(\mathbf{x})/|W|$ (the MLE, which agrees with the MPLE in this case), the Papangelou compensator (50) becomes

$$\mathcal{C}V_S(\mathbf{x}, r) = \frac{\hat{\rho}}{2} \int_W \sum_i \mathbb{I}\{\|u - x_i\| \leq r\}\, du$$

$$= \frac{\hat{\rho}}{2} \sum_i |W \cap B(x_i, r)|.$$

Ignoring edge effects, we have $|W \cap B(x_i, r)| \approx \pi r^2$ and, applying (43), the residual is approximately

$$(51) \qquad \mathcal{R}V_S(\mathbf{x}, r) \approx \frac{n(\mathbf{x})^2}{2|W|}[\hat{K}_{\mathbf{x}}(r) - \pi r^2].$$

The term in brackets is a commonly-used measure of departure from CSR, and is a sensible diagnostic because $K(r) = \pi r^2$ under CSR.

9.2 Residual Based on Perturbing $\hat{K}$-Model

Assuming $\hat{\rho}^2(\mathbf{x}) = \hat{\rho}^2(n(\mathbf{x}))$ depends only on $n(\mathbf{x})$, the empirical K-function (40) can also be expressed as a sum of local contributions $\hat{K}_{\mathbf{x}}(r) = \sum_i k(x_i, \mathbf{x}_{-i}, r)$ with

$$k(u, \mathbf{x}, r) = \frac{t^w(u, \mathbf{x}, r)}{\hat{\rho}^2(n(\mathbf{x}) + 1)|W|}, \quad u \notin \mathbf{x},$$

where

$$t^w(u, \mathbf{x}, r) = \sum_j e_K(u, x_j) \mathbb{I}\{\|u - x_j\| \le r\}$$

is a weighted count of the points of $\mathbf{x}$ that are r-close to the location u. Hence, the compensator of the $\hat{K}$-function is

(52)
$$\mathcal{C}\hat{K}_{\mathbf{x}}(r) = \frac{1}{\hat{\rho}^2(n(\mathbf{x}) + 1)|W|}$$
$$\cdot \int_W t^w(u, \mathbf{x}, r) \lambda_{\hat{\theta}}(u, \mathbf{x}) \, du.$$

Assume the edge correction weight $e_K(u, v) = e_K(v, u)$ is symmetric; for example, this is satisfied by the Ohser–Stoyan edge correction weight [57, 58] given by $e_K(u, v) = 1/|W_u \cap W_v|$ where $W_u = \{u + v : v \in W\}$, but not by Ripley's [63] isotropic correction weight. Then the increment is, for $u \notin \mathbf{x}$,

$$\Delta_u \hat{K}_{\mathbf{x}}(r) = \frac{\hat{\rho}^2(\mathbf{x}) - \hat{\rho}^2(\mathbf{x} \cup \{u\})}{\hat{\rho}^2(\mathbf{x} \cup \{u\})} \hat{K}_{\mathbf{x}}(r)$$
$$+ \frac{2t^w(u, \mathbf{x}, r)}{\hat{\rho}^2(\mathbf{x} \cup \{u\})|W|}$$

and when $x_i \in \mathbf{x}$

$$\Delta_{x_i} \hat{K}_{\mathbf{x}}(r) = \frac{\hat{\rho}^2(\mathbf{x}_{-i}) - \hat{\rho}^2(\mathbf{x})}{\hat{\rho}^2(\mathbf{x}_{-i})} \hat{K}_{\mathbf{x}}(r)$$
$$+ \frac{2t^w(x_i, \mathbf{x}_{-i}, r)}{\hat{\rho}^2(\mathbf{x}_{-i})|W|}.$$

Assuming the standard estimator $\hat{\rho}^2(\mathbf{x}) = n(n-1)/|W|^2$ with $n = n(\mathbf{x})$, the pseudo-sum is seen to be zero, so the pseudo-residual is apart from the sign equal to the pseudo-compensator, which becomes

$$\mathcal{C}\Delta\hat{K}_{\mathbf{x}}(r) = 2\mathcal{C}\hat{K}_{\mathbf{x}}(r) - \left[\frac{2}{n-2}\int_W \lambda_{\hat{\theta}}(u, \mathbf{x}) \, du\right]\hat{K}_{\mathbf{x}}(r),$$

where $\mathcal{C}\hat{K}_{\mathbf{x}}(r)$ is given by (52). So if the null model is CSR and the intensity is estimated by $n/|W|$, the pseudo-residual is approximately $2[\hat{K}_{\mathbf{x}}(r) - \mathcal{C}\hat{K}_{\mathbf{x}}(r)]$, and, hence, it is equivalent to the residual approximated by (51). This is also the conclusion in the more general

case of a null model with an activity parameter κ, that is, where the Papangelou conditional intensity factorizes as

$$\lambda_{\theta}(u, \mathbf{x}) = \kappa \xi_{\beta}(u, \mathbf{x}),$$

where $\theta = (\kappa, \beta)$ and $\xi_{\beta}(\cdot)$ is a Papangelou conditional intensity, since the pseudo-likelihood equations then imply that $n = \int_W \lambda_{\hat{\theta}}(u, \mathbf{x}) \, du$.

In conclusion, the residual diagnostics obtained from the perturbing Strauss and $\hat{K}$-models are very similar, the major difference being the data-dependent normalization of the $\hat{K}$-function; similarly for pseudo-residual diagnostics which may be effectively equivalent to the residual diagnostics. In practice, the popularity of the K-function seems to justify using the residual diagnostics based on the perturbing $\hat{K}$-model. Furthermore, due to the familiarity of the K-function, we often choose to plot the compensator(s) of the fitted model(s) in a plot with the empirical K-function rather than the residual(s) for the fitted model.

9.3 Edge Correction in Conditional Case

In the conditional case, the Papangelou conditional intensity $\lambda_{\hat{\theta}}(u, \mathbf{x})$ is known only at locations $u \in W^\circ$. The diagnostics must be modified accordingly, by restricting the domain of summation and integration to W°. Appropriate modifications are discussed in Appendices C–E.

9.4 Approximate Residual Variance Under CSR

Here we study the residual variance and the accuracy of the Poincaré variance approximation in a simple case.

We shall approximate the residual variance $\mathbb{V}\text{ar}[\mathcal{R}V_S(X, r)]$ by the innovation variance $\mathbb{V}\text{ar}[\mathcal{I}V_S(X, r)]$, that is, ignoring the effect of parameter estimation. It is likely that this approximation is conservative, because the effect of parameter estimation is typically to deflate the residual variance [7]. A more detailed investigation has been conducted in [17].

Assume the null model is CSR with intensity ρ estimated by $\hat{\rho} = n(\mathbf{x})/|W|$. The exact variance of the innovation for the Strauss canonical statistic V_S is $\mathbb{V}\text{ar}[\mathcal{I}V_S(\mathbf{X}, r)] = I_1 + I_2$ from equation (65) of Appendix B, where

$$I_1 = \frac{1}{4}\int_W \mathbb{E}[t(u, \mathbf{X}, r)^2 \lambda(u, \mathbf{X})] \, du$$
$$= \frac{\rho}{4}\int_W \mathbb{E}[t(u, \mathbf{X}, r)^2] \, du$$

and

$$I_2 = \frac{1}{4} \int_W \int_W \mathbb{E}[\mathbb{I}\{\|u - v\| \le r\} \lambda_2(u, v, \mathbf{X})] \, du \, dv$$

$$= \frac{\rho^2}{4} \int_W \int_W \mathbb{I}\{\|u - v\| \le r\} \, du \, dv$$

as $\lambda(u, \mathbf{X}) = \rho$ and $\lambda_2(u, v, \mathbf{X}) = \lambda(u, \mathbf{X})\lambda(v, \mathbf{X} \cup \{u\}) = \rho^2$. This is reminiscent of expressions for the large-domain limiting variance of $\hat{K}$ under CSR obtained using the methods of U-statistics [16, 48, 65], summarized in [29], page 51 ff. Now $Y = t(u, \mathbf{X}, r)$ is Poisson distributed with mean $\mu = \rho|B(u, r) \cap W|$ so that $\mathbb{E}(Y^2) = \mu + \mu^2$. For $u \in W_{\ominus r}$ we have $\mu = \rho \pi r^2$, so ignoring edge effects

$$I_1 \approx \frac{\rho}{4}(\nu + \nu^2)|W| \quad \text{and} \quad I_2 \approx \frac{\rho}{4}\nu|W|,$$

where $\nu = \rho \pi r^2$. Note that since ν is the expected number of points within distance r of a given point, a value of $\nu = 1$ corresponds to the scale of nearest-neighbor distances in the pattern, $r_{nn} = 1/\sqrt{\pi\rho}$. For the purposes of the K function this is a "short" distance. Hence, it is reasonable to describe I_1 as the "leading term" in the variance, since $I_1 \gg I_2$ for $\nu \gg 1$.

Meanwhile, the Poincaré variance (37) is

$$\mathcal{C}^2 V_S(\mathbf{x}, r) = \frac{n(\mathbf{x})}{4|W|} \int_W t(u, \mathbf{x}, r)^2 \, du,$$

which is an approximately unbiased estimator of I_1 by Fubini's Theorem. Hence,

$$\frac{\mathbb{E}\mathcal{C}^2 V_S(\mathbf{x}, r)}{\mathbb{V}\text{ar}[\mathcal{R}V_S(\mathbf{X}, r)]} \approx \frac{\mathbb{E}\mathcal{C}^2 V_S(\mathbf{x}, r)}{\mathbb{V}\text{ar}[\mathcal{I}V_S(\mathbf{X}, r)]}$$

$$\approx \frac{I_1}{I_1 + I_2} \approx \frac{1 + \nu}{2 + \nu}.$$

Thus, as a rule of thumb, the Poincaré variance underestimates the true variance; the ratio of means is $(1 + \nu)/(2 + \nu) \ge 1/2$. The ratio falls to 2/3 when $\nu = 1$, that is, when $r = r_{nn} = 1/\sqrt{\pi\rho}$. We can take this as a rule-of-thumb indicating the value of r below which the Poincare variance is a poor approximation to the true variance.

10. RESIDUAL DIAGNOSTICS FOR INTERACTION USING NEAREST NEIGHBOR DISTANCES

This section develops residual and pseudo-residual diagnostics derived from summary statistics based on nearest neighbor distances.

10.1 Residual Based on Perturbing Geyer Model

The Geyer interaction potential $V_G(\mathbf{x}, r)$ given by (15) is clearly a sum of local statistics (33), and its compensator is

$$\mathcal{C}V_G(\mathbf{x}, r) = \int_W \mathbb{I}\{d(u, \mathbf{x}) \le r\} \lambda_{\hat{\theta}}(u, \mathbf{x}) \, du.$$

The Poincaré variance is equal to the compensator in this case. Ignoring edge effects, $V_G(\mathbf{x}, r)$ is approximately $n(\mathbf{x})\hat{G}_{\mathbf{x}}(r)$; cf. (41).

If the null model is CSR with estimated intensity $\hat{\kappa} = n(\mathbf{x})/|W|$, then

$$\mathcal{C}V_G(\mathbf{x}, r) = \hat{\kappa}\left|W \cap \bigcup_i B(x_i, r)\right|;$$

ignoring edge effects, this is approximately $\hat{\kappa}|W|\hat{F}(r)$; cf. (42). Thus, the residual diagnostic is approximately $n(\mathbf{x})(\hat{G}(r) - \hat{F}(r))$. This is a reasonable diagnostic for departure from CSR, since $F \equiv G$ under CSR. This argument lends support to Diggle's [27], equation (5.7), proposal to judge departure from CSR using the quantity $\sup|\hat{G} - \hat{F}|$.

This example illustrates the important point that the compensator of a functional summary statistic S should not be regarded as an alternative parametric estimator of the same quantity that S is intended to estimate. In the example just given, under CSR the compensator of $\hat{G}$ is approximately $\hat{F}$, a qualitatively different and in some sense "opposite" summary of the point pattern.

We have observed that the interaction potential V_G of the Geyer saturation model is closely related to $\hat{G}$. However, the pseudo-residual associated to V_G is a more complicated statistic, since a straightforward calculation shows that the pseudo-sum is

$$\Sigma \Delta V_G(\mathbf{x}, r)$$
$$= V_G(\mathbf{x}, r) + \sum_i \sum_{j:j \ne i} \mathbb{I}\{\|x_i - x_j\| \le r \text{ and}$$
$$d(x_j, \mathbf{x}_{-i}) > r\},$$

and the pseudo-compensator is

$$\mathcal{C}\Delta V_G(\mathbf{x}, r) = \int_W \mathbb{I}\{d(u, \mathbf{x}) \le r\} \lambda_{\hat{\theta}}(u, \mathbf{x}) \, du$$
$$+ \sum_i \mathbb{I}\{d(x_i, \mathbf{x}_{-i}) > r\}$$
$$\cdot \int_W \mathbb{I}\{\|u - x_i\| \le r\} \lambda_{\hat{\theta}}(u, \mathbf{x}) \, du.$$

10.2 Residual Based on Perturbing $\hat{G}$-Model

The empirical G-function (41) can be written

$$\hat{G}_{\mathbf{x}}(r) = \sum_i g(x_i, \mathbf{x}_{-i}, r), \tag{53}$$

where

$$g(u, \mathbf{x}, r) = \frac{1}{n(\mathbf{x}) + 1} e_G(u, \mathbf{x}, r)$$

$$\cdot \mathbb{I}\{d(u, \mathbf{x}) \le r\}, \quad u \notin \mathbf{x}, \tag{54}$$

so that the Papangelou compensator of the empirical G-function is

$$\mathcal{C}\hat{G}_{\mathbf{x}}(r)$$
$$= \int_W g(u, \mathbf{x}, r) \lambda_{\hat{\theta}}(u, \mathbf{x}) \, du$$
$$= \frac{1}{n(\mathbf{x}) + 1} \int_{W \cap \bigcup_i B(x_i, r)} e_G(u, \mathbf{x}, r) \lambda_{\hat{\theta}}(u, \mathbf{x}) \, du.$$

The residual diagnostics obtained from the Geyer and $\hat{G}$-models are very similar, and we choose to use the diagnostic based on the popular $\hat{G}$-function. As with the K-function, we typically use the compensator(s) of the fitted model(s) rather than the residual(s), to visually maintain the close connection to the empirical G-function.

The expressions for the pseudo-sum and pseudo-compensator of $\hat{G}$ are not of simple form, and we refrain from explicitly writing out these expressions. For both the $\hat{G}$- and Geyer models, the pseudo-sum and pseudo-compensator are not directly related to a well-known summary statistic. We prefer to plot the pseudo-residual rather than the pseudo-sum and pseudo-compensator(s).

10.3 Residual Variance Under CSR

Again assume a Poisson process of intensity ρ as the null model. Since V_G is a sum of local statistics,

$$V_G(\mathbf{x}, r) = \sum_i \mathbb{I}\{d(x_i, \mathbf{x} \setminus x_i) \le r\},$$

we can again apply the variance formula (65) of Appendix B, which gives $\mathbb{V}\text{ar}[\mathcal{I}V_G(\mathbf{X}, r)] = L_1 + L_2$, where

$$L_1 = \rho \int_W \mathbb{P}\{d(u, \mathbf{X}) \le r\} \, du$$

and

$$L_2 = \rho^2 \int_W \int_W \mathbb{P}\{\|u - v\| \le r,$$
$$d(u, X) > r, d(v, X) > r\} \, du \, dv.$$

The Poincaré variance is equal to the compensator in this case, and is

$$\mathcal{C}^2 V_G(\mathbf{x}, r) = \int_W \mathbb{I}\{d(u, \mathbf{x}) \le r\} \lambda_{\hat{\theta}}(u, \mathbf{x}) \, du$$
$$= \frac{n(\mathbf{x})}{|W|} |W \cap U(\mathbf{x}, r)|,$$

where $U(\mathbf{x}, r) = \bigcup_i b(x_i, r)$. The Poincaré variance is an approximately unbiased estimator of the term L_1.

For $u \in W_{\ominus r}$ we have $\mathbb{P}\{d(u, \mathbf{X}) \le r\} = 1 - \exp(-\rho \pi r^2)$ so that

$$L_1 \approx \rho |W| (1 - \exp(-\rho \pi r^2)),$$

ignoring edge effects. Again, let $\nu = \rho \pi r^2$ so that $L_1 \approx \rho |W| (1 - \exp(-\nu))$. Meanwhile,

$$\mathbb{P}\{d(u, X) > r, d(v, X) > r\}$$
$$= \exp(-\rho |b(u, r) \cup b(v, r)|).$$

This probability lies between $\exp(-\nu)$ and $\exp(-2\nu)$ for all u, v. Thus (ignoring edge effects),

$$L_2 \approx \rho^2 \pi r^2 |W| \exp(-(1 + \delta)\nu)$$
$$= \rho \nu |W| \exp(-(1 + \delta)\nu),$$

where $0 \le \delta \le 1$. Hence,

$$\frac{L_2}{L_1} \le \frac{\nu e^{-\nu}}{1 - e^{-\nu}}.$$

Let $f(\nu) = \nu e^{-\nu} / (1 - e^{-\nu})$. Then $f(\nu)$ is strictly decreasing and $f(\nu) < 1$ for all $\nu > 0$ so that $L_1/(L_1 + L_2) \ge \frac{1}{2}$, that is, the variance is underestimated by at most a factor of 2. Note that $f(1.25) \approx 0.5$, so $L_1/(L_1 + L_2) \ge \frac{2}{3}$ when $r \le r_{\text{crit}}$, where $r_{\text{crit}} = \sqrt{1.25/\pi\rho}$. The conclusions and rule-of-thumb for $\mathcal{T}\hat{G}$ are similar to those obtained for $\mathcal{T}\hat{K}$ in Section 9.4.

11. DIAGNOSTICS FOR INTERACTION BASED ON EMPTY SPACE DISTANCES

11.1 Pseudo-Residual Based on Perturbing Area-Interaction Model

When the perturbing model is the area-interaction process, it is convenient to reparametrize the density, such that the canonical sufficient statistic V_A given in (16) is redefined as

$$V_A(\mathbf{x}, r) = \frac{1}{|W|} \left| W \cap \bigcup_i B(x_i, r) \right|.$$

This summary statistic is not naturally expressed as a sum of contributions from each point as in (33), so we shall only construct the pseudo-residual. Let

$$U(\mathbf{x}, r) = W \cap \bigcup_i B(x_i, r).$$

The increment

$$\begin{aligned}&\Delta_u V_A(\mathbf{x}, r) \\&= \frac{1}{|W|}\big(|U(\mathbf{x} \cup \{u\}, r)| - |U(\mathbf{x}, r)|\big), \quad u \notin \mathbf{x},\end{aligned}$$

can be thought of as "unclaimed space"—the proportion of space around the location u that is not "claimed" by the points of $\mathbf{x}$. The pseudo-sum

$$\Sigma \Delta V_A(\mathbf{x}, r) = \sum_i \Delta_{x_i} V_A(\mathbf{x}, r)$$

is the proportion of the window that has "single coverage"—the proportion of locations in W that are covered by exactly one of the balls $B(x_i, r)$. This can be used in its own right as a functional summary statistic, and it corresponds to a raw (i.e., not edge corrected) empirical estimate of a summary function $F_1(r)$ defined by

$$F_1(r) = \mathbb{P}\big(\#\{x \in \mathbf{X} | d(u, x) \le r\} = 1\big)$$

for any stationary point process $\mathbf{X}$, where $u \in \mathbb{R}^2$ is arbitrary. Under CSR with intensity ρ we have

$$F_1(r) = \rho\pi r^2 \exp(-\rho\pi r^2).$$

This summary statistic does not appear to be treated in the literature, and it may be of interest to study it separately, but we refrain from a more detailed study here.

The pseudo-compensator corresponding to this pseudo-sum is

$$\mathcal{C} \Delta V_A(\mathbf{x}, r) = \int_W \Delta_u V_A(\mathbf{x}, r) \lambda_{\hat{\theta}}(u, \mathbf{x}) \, du.$$

This integral does not have a particularly simple interpretation even when the null model is CSR.

11.2 Pseudo-Residual Based on Perturbing $\hat{F}$-Model

Alternatively, one could use a standard empirical estimator $\hat{F}$ of the empty space function F as the sum-

mary statistic in the pseudo-residual. The pseudo-sum associated with the perturbing $\hat{F}$-model is

$$\Sigma \Delta \hat{F}_\mathbf{x}(r) = n(\mathbf{x}) \hat{F}_\mathbf{x}(r) - \sum_i \hat{F}_{\mathbf{x}_{-i}}(r),$$

with pseudo-compensator

$$\mathcal{C} \Delta \hat{F}_\mathbf{x}(r) = \int_W \big(\hat{F}_{\mathbf{x} \cup \{u\}}(r) - \hat{F}_\mathbf{x}(r)\big) \lambda_{\hat{\theta}}(u, \mathbf{x}) \, du.$$

Ignoring edge correction weights, $\hat{F}_{\mathbf{x} \cup \{u\}}(r) - \hat{F}_\mathbf{x}(r)$ is approximately equal to $\Delta_u V_A(\mathbf{x}, r)$, so the pseudo-sum and pseudo-compensator associated with the perturbing $\hat{F}$-model are approximately equal to the pseudo-sum and pseudo-compensator associated with the perturbing area-interaction model. Here, we usually prefer graphics using the pseudo-compensator(s) and the pseudo-sum since this has an intuitive interpretation as explained above.

12. TEST CASE: TREND WITH INHIBITION

In Sections 12–14 we demonstrate the diagnostics on the point pattern data sets shown in Figure 1. This section concerns the synthetic point pattern in Figure 1(b).

12.1 Data and Models

Figure 1(b) shows a simulated realization of the inhomogeneous Strauss process with first order term $\lambda(x, y) = 200 \exp(2x + 2y + 3x^2)$, interaction range $R = 0.05$, interaction parameter $\gamma = \exp(\phi) = 0.1$ and W equal to the unit square; see (13) and (14). This is an example of extremely strong inhibition (negative association) between neighboring points, combined with a spatial trend. Since it is easy to recognize spatial trend in the data (either visually or using existing tools such as kernel smoothing [28]), the main challenge here is to detect the inhibition after accounting for the trend.

We fitted four point process models to the data in Figure 1(b). They were (A) a homogeneous Poisson process (CSR); (B) an inhomogeneous Poisson process with the correct form of the first order term, that is, with intensity

$$(55) \quad \rho(x, y) = \exp(\beta_0 + \beta_1 x + \beta_2 y + \beta_3 x^2),$$

where $\beta_0, \ldots, \beta_3$ are real parameters; (C) a homogeneous Strauss process with the correct interaction range $R = 0.05$; and (D) a process of the correct form, that is, inhomogeneous Strauss with the correct interaction range $R = 0.05$ and the correct form of the first order potential (55).
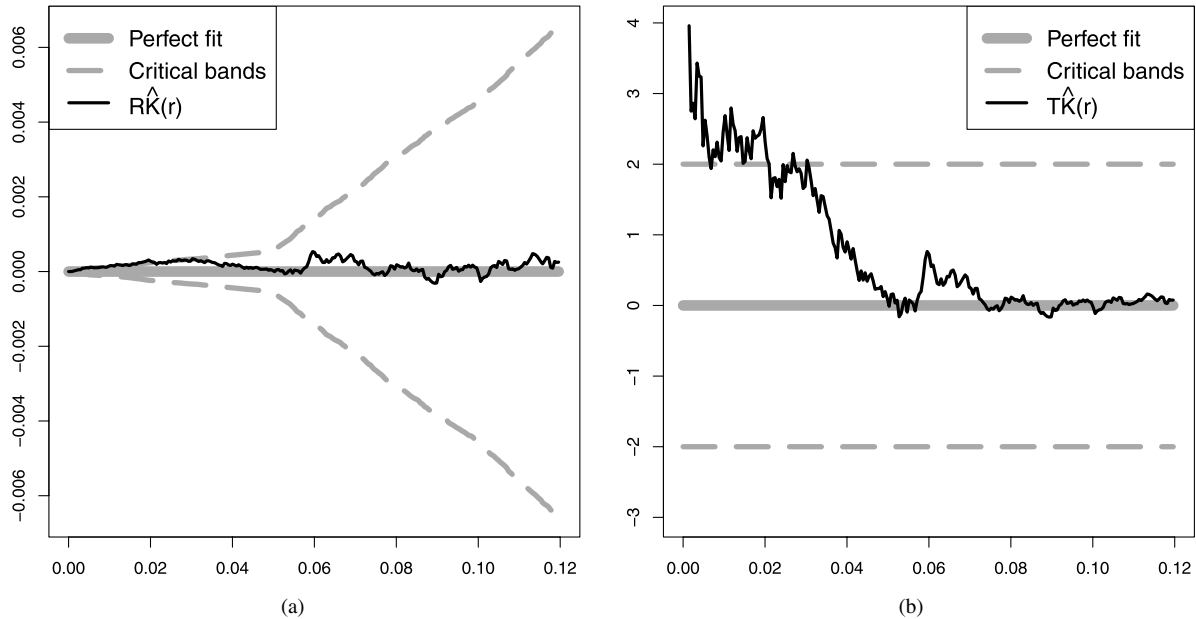
FIG. 3. *Residual diagnostics based on pairwise distances, for a model of the correct form fitted to the data in Figure 1(b). (a) Residual*
$\hat{K}$-function and two-standard-deviation limits under the fitted model of the correct form. (b) Standardized residual $\hat{K}$-function under the
fitted model of the correct form.*

12.2 Software Implementation

The diagnostics defined in Sections 9–11 were implemented in the R language, and has been publicly released in the spatstat library [6]. Unless otherwise stated, models were fitted by approximate maximum pseudo-likelihood using the algorithm of [5] with the default quadrature scheme in spatstat, having an $m \times m$ grid of dummy points where $m = \max(25, 10[1 + 2\sqrt{n(\mathbf{x})}/10])$ was equal to 40 for most of our examples. Integrals over the domain W were approximated by finite sums over the quadrature points. Some models were refitted using a finer grid of dummy points, usually 80×80. In addition to maximum pseudo-likelihood estimation, the software also supports the Huang–Ogata [37] approximate maximum likelihood.

12.3 Application of $\hat{K}$ Diagnostics

12.3.1 *Diagnostics for correct model.* First we fitted a point process model of the correct form (D). The fitted parameter values were $\hat{\gamma} = 0.217$ and $\hat{\beta} = (5.6, -0.46, 3.35, 2.05)$ using the coarse grid of dummy points, and $\hat{\gamma} = 0.170$ and $\hat{\beta} = (5.6, -0.64, 4.06, 2.44)$ using the finer grid of dummy points, as against the true values $\gamma = 0.1$ and $\beta = (5.29, 2, 2, 3)$.

Figure 2 in Section 1 shows $\hat{K}$ along with its compensator for the fitted model, together with the theoretical K-function under CSR. The empirical K-function

and its compensator coincide very closely, suggesting correctly that the model is a good fit. Figure 3(a) shows the residual $\hat{K}$-function and the two-standard-deviation limits, where the surrogate standard deviation is the square root of (37). Figure 3(b) shows the corresponding standardized residual $\hat{K}$-function obtained by dividing by the surrogate standard deviation.

Although this model is of the correct form, the standardized residual exceeds 2 for small values of r. This is consistent with the prediction in Section 9.4 that the variance approximation would be inaccurate for small r. The null model is a nonstationary Poisson process; the minimum value of the intensity is 200. Taking $\rho = 200$ and applying the rule of thumb in Section 9.4 gives

$$r_{\text{nn}} = \frac{1}{\sqrt{200\pi}} = 0.04,$$

suggesting that the Poincaré variance estimate becomes unreliable for $r \leq 0.04$ approximately.

Formal significance interpretation of the critical bands in Figure 3(b) is limited, because the null distribution of the standardized residual is not known exactly, and the values ± 2 are approximate *pointwise* critical values, that is, critical values for the score test based on fixed r. The usual problems of multiple testing arise when the test statistic is considered as a function of r; see [29], page 14. For very small r there are small-sample effects so that a normal approximation
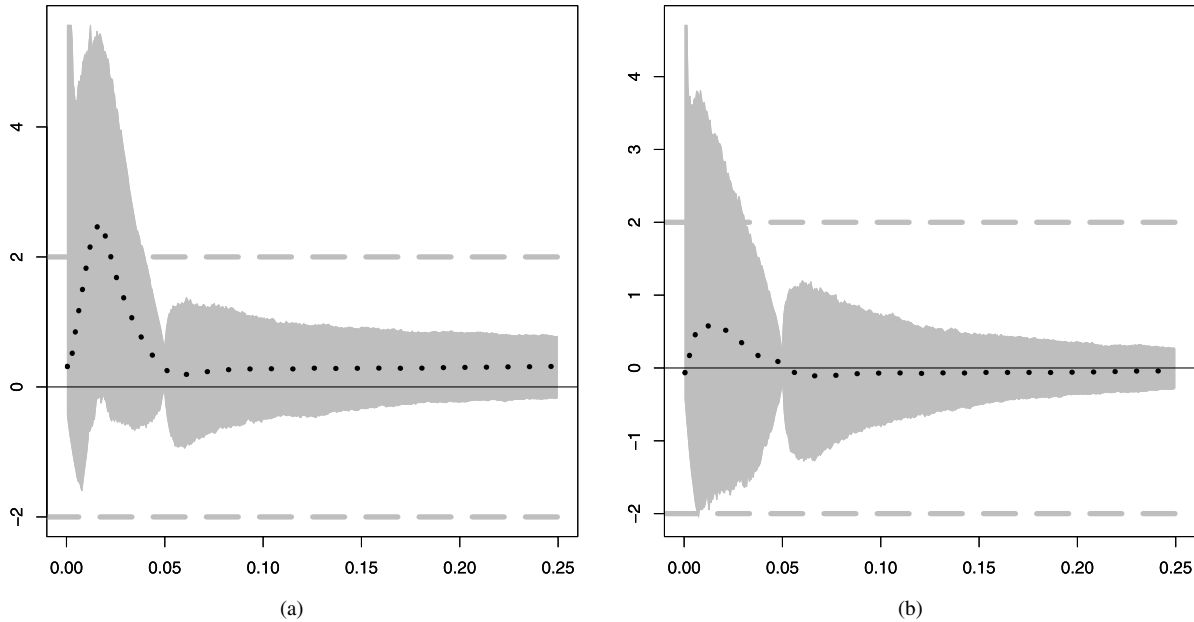
FIG. 4.   *Null distribution of standardized residual of $\hat{K}$. Pointwise* 2.5% *and* 97.5% *quantiles* (*grey shading*) *and sample mean* (*dotted lines*) *of $\mathcal{T}\hat{K}$ from* 1000 *simulated realizations of model* (D) *with estimated parameter values* (a) $\hat{\gamma} = 0.217$ *and* $\hat{\beta} = (5.6, -0.46, 3.35, 2.05)$ *using a* 40×40 *grid of dummy points*; (b) $\hat{\gamma} = 0.170$ *and* $\hat{\beta} = (5.6, -0.64, 4.06, 2.44)$ *using a* 80×80 *grid.*

to the null distribution of the standardized residual is inappropriate.

To confirm this, Figure 4 shows the pointwise 2.5% and 97.5% quantiles of the null distribution of $\mathcal{T}\hat{K}$, obtained by extensive simulation. The sample mean of the simulated $\mathcal{T}\hat{K}$ is also shown, and indicates that the expected standardized residual is nonzero for small values of r. Repeating the computation with a finer grid of quadrature points (for approximating integrals over W involved in the pseudo-likelihood and the residuals) reduces the bias, suggesting that this is a discretization artefact.

12.3.2 *Comparison of competing models*. Figure 5(a) shows the empirical K-function and its compensator for each of the models (A)–(D) in Section 12.1. Figure 5(b) shows the corresponding residual plots, and Figure 5(c) the standardized residuals. A positive or negative value of the residual suggests that the data are more clustered or more inhibited, respectively, than the model. The clear inference is that the Poisson models (A) and (B) fail to capture interpoint inhibition at range $r \approx 0.05$, while the homogeneous Strauss model (C) is less clustered than the data at very large scales, suggesting that it fails to capture spatial trend. The correct model (D) is judged to be a good fit.

The interpretation of this example requires some caution, because the residual $\hat{K}$-function of the fitted Strauss models (C) and (D) is constrained to be approx-

imately zero at $r = R = 0.05$. The maximum pseudo-likelihood fitting algorithm solves an estimating equation that is approximately equivalent to this constraint, because of (43).

It is debatable which of the presentations in Figure 5 is more effective at revealing lack of fit. A compensator plot such as Figure 5(a) seems best at capturing the main differences between competing models. It is particularly useful for recognizing a gross lack of fit. A residual plot such as Figure 5(b) seems better for making finer comparisons of model fit, for example, assessing models with slightly different ranges of interaction. A standardized residual plot such as Figure 5(c) tends to be highly irregular for small values of r, due to discretization effects in the computation and the inherent nondifferentiability of the empirical statistic. In difficult cases we may apply smoothing to the standardized residual.

12.4 Application of $\hat{G}$ Diagnostics

12.4.1 *Diagnostics for correct model*. Consider again the model of the correct form (D). The residual and compensator of the empirical nearest neighbor function $\hat{G}$ for the fitted model are shown in Figure 6. The residual plot suggests a marginal lack of fit for $r < 0.025$. This may be correct, since the fitted model parameters (Section 12.3.1) are marginally poor esti-
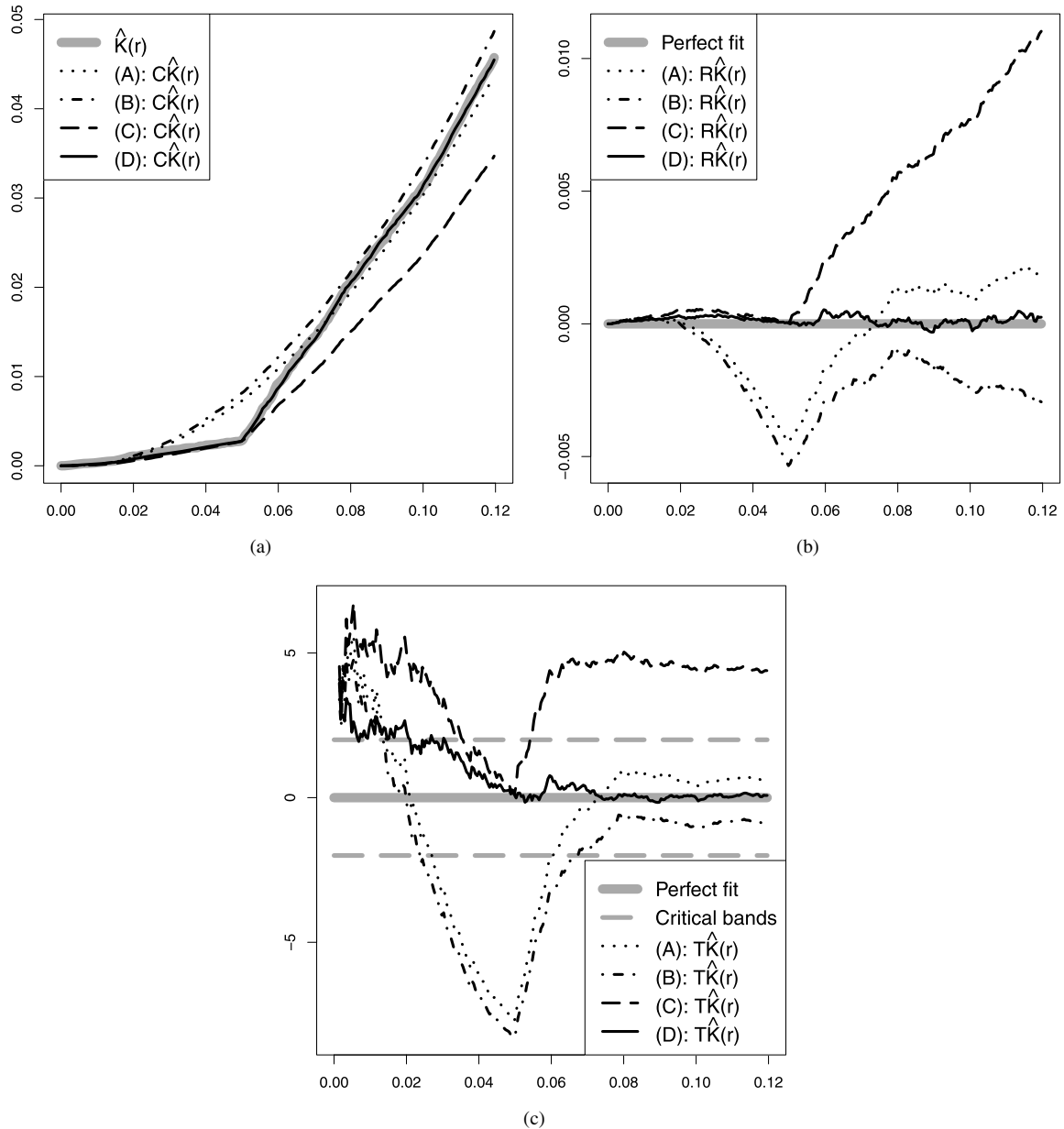
FIG. 5.   *Model diagnostics based on pairwise distances, for each of the models* (A)–(D) *fitted to the data in Figure* 1(b). (a) $\hat{K}$ *and its compensator under each model.* (b) *Residual $\hat{K}$-function (empirical minus compensator) under each model.* (c) *Standardized residual $\hat{K}$-function under each model.*

mates of the true values, in particular, of the interaction parameter. This was not reflected so strongly in the $\hat{K}$ diagnostics. This suggests that the residual of $\hat{G}$ may be particularly sensitive to lack of fit of interaction.

Applying the rule of thumb in Section 10.3, we have $r_{\text{crit}} = 0.044$, agreeing with the interpretation that the ± 2 limits are not trustworthy for $r < 0.05$ approximately.

Figure 7 shows the pointwise 2.5% and 97.5% quantiles of the null distribution of $\mathcal{T}\hat{G}$. Again, there is a

suggestion of bias for small values of r which appears to be a discretization artefact.

12.4.2 *Comparison of competing models.* For each of the four models, Figure 8(a) shows $\hat{G}$ and its Papangelou compensator. This clearly shows that the Poisson models (A) and (B) fail to capture interpoint inhibition in the data. The Strauss models (C) and (D) appear virtually equivalent in Figure 8(a).

Figure 8(b) shows the standardized residual of $\hat{G}$, and Figure 8(c) the pseudo-residual of V_G (i.e., the
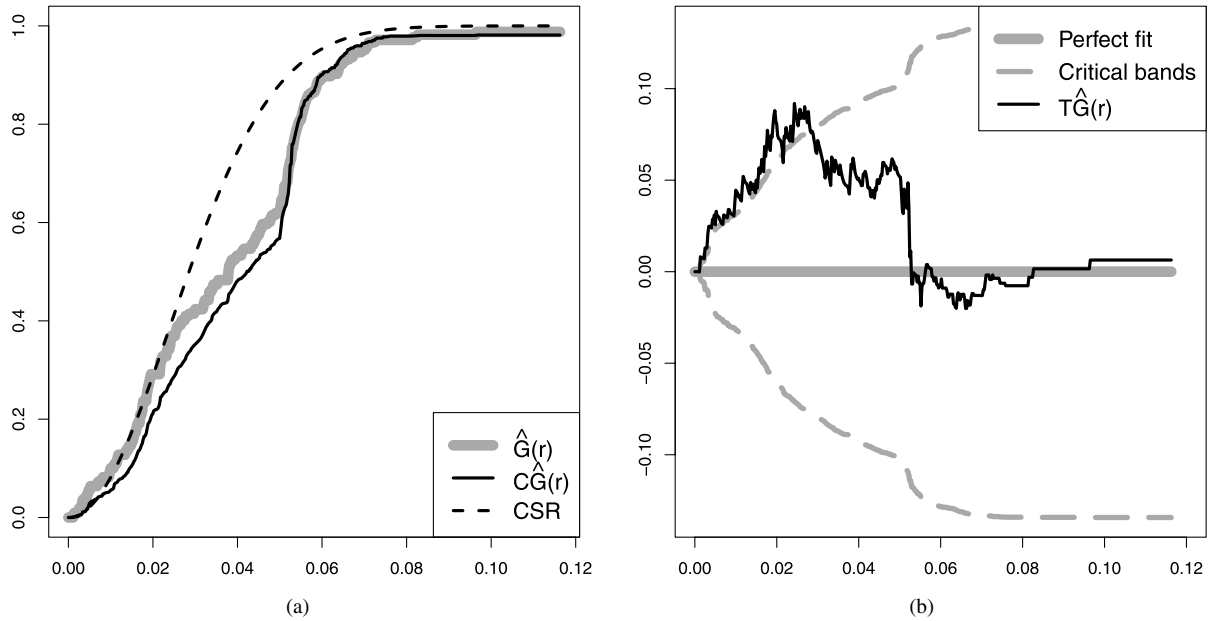
FIG. 6.   *Residual diagnostics obtained from the perturbing $\hat{G}$-model when the data pattern is a realization of an inhomogeneous Strauss process.* (a) *$\hat{G}$ and its compensator under a fitted model of the correct form, and theoretical G-function for a Poisson process.* (b) *Residual $\hat{G}$-function and two-standard-deviation limits under the fitted model of the correct form.*

pseudo-residual based on the pertubing Geyer model), with spline smoothing applied to both plots. The Strauss models (C) and (D) appear virtually equivalent in Figure 8(c). The standardized residual plot Figure 8(b) correctly suggests a slight lack of fit for

model (C) while model (D) is judged to be a reasonable fit.

12.5 Application of $\hat{F}$ Diagnostics

Figure 9 shows the pseudo-residual diagnostics based on empty space distances. Both diagnostics
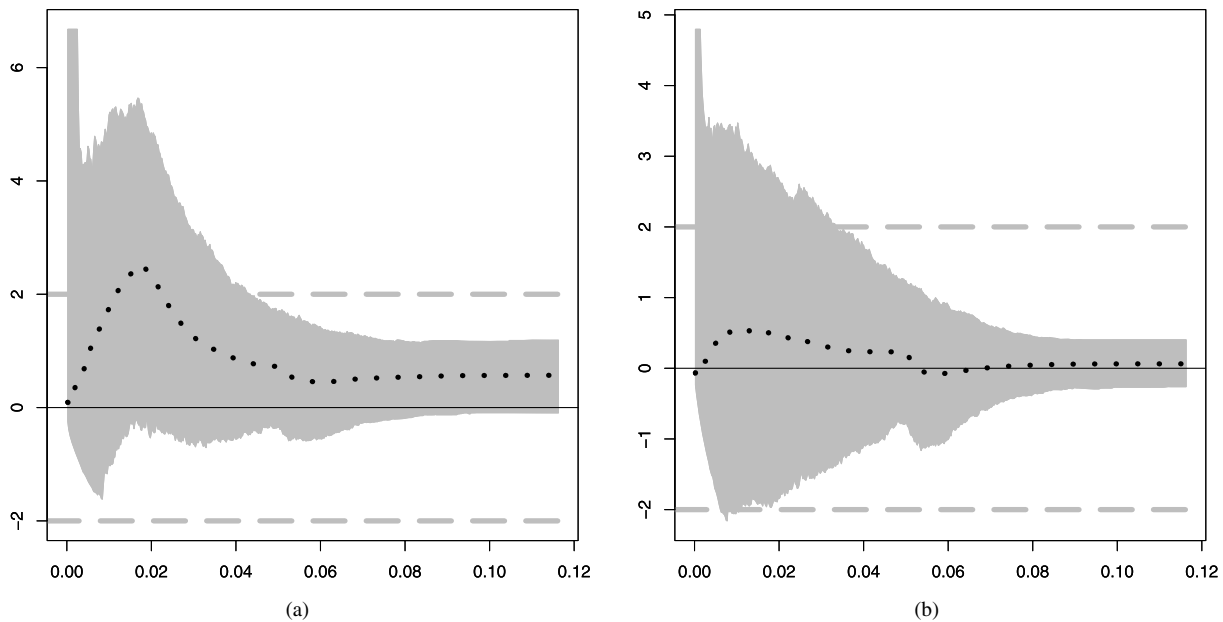
FIG. 7.   *Null distribution of standardized residual of $\hat{G}$. Pointwise 2.5% and 97.5% quantiles (grey shading) and sample mean (dotted lines) from 1000 simulated realizations of model (D) with estimated parameter values* (a) *$\hat{\gamma} = 0.217$ and $\hat{\beta} = (5.6, -0.46, 3.35, 2.05)$ using a 40×40 grid of dummy points;* (b) *$\hat{\gamma} = 0.170$ and $\hat{\beta} = (5.6, -0.64, 4.06, 2.44)$ using a 80×80 grid.*
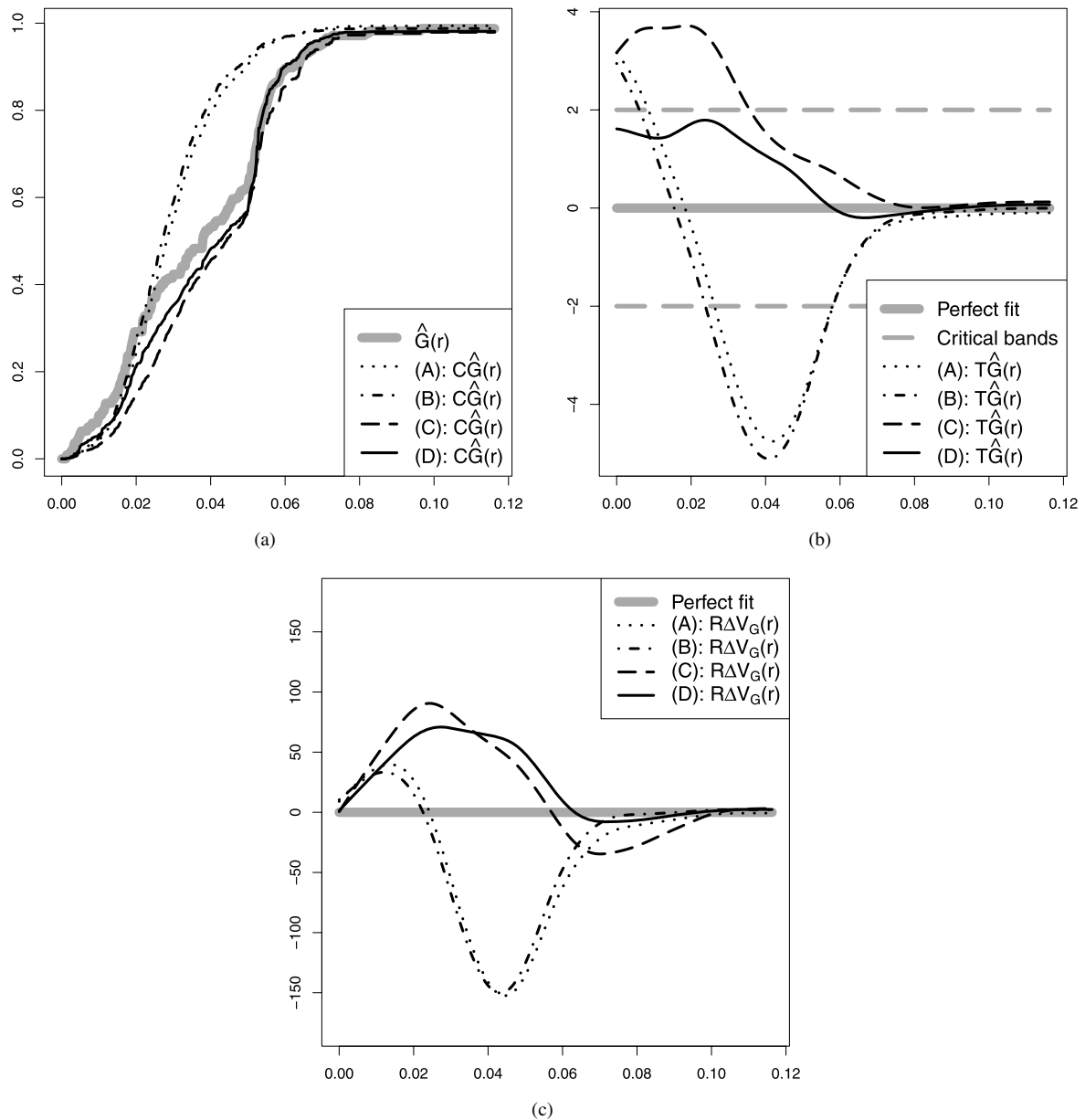
FIG. 8. *Diagnostics based on nearest neighbor distances, for the models* (A)–(D) *fitted to the data in Figure* 1(b). (a) *Compensator for* $\hat{G}$. (b) *Smoothed standardized residual of* $\hat{G}$. (c) *Smoothed pseudo-residual derived from a perturbing Geyer model.*

clearly show models (A)–(B) are poor fits to data. However, in Figure 9(a) it is hard to decide which of the models (C)–(D) provide a better fit. Despite the close connection between the area-interaction process and the $\hat{F}$-model, the diagnostic in Figure 9(b) based on the $\hat{F}$-model performs better in this particular example and correctly shows (D) is the best fit to data. In both cases it is noticed that the pseudo-sum has a much higher peak than the pseudo-compensators for the Poisson models (A)–(B), correctly suggesting that these mod-

els do not capture the strength of inhibition present in the data.

13. TEST CASE: CLUSTERING WITHOUT TREND

13.1 Data and Models

Figure 1(c) is a realization of a homogeneous Geyer saturation process [31] on the unit square, with first order term $\lambda = \exp(4)$, saturation threshold $s = 4.5$ and interaction parameters $r = 0.05$ and $\gamma = \exp(0.4) \approx$
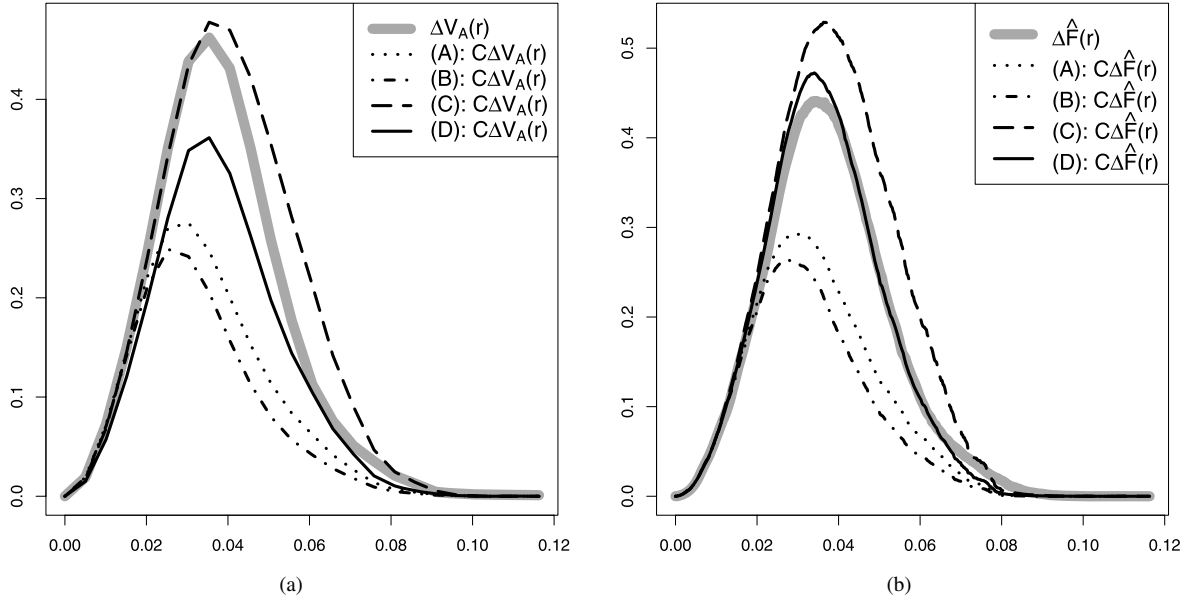
FIG. 9. *Pseudo-sum and pseudo-compensators for the models* (A)–(D) *fitted to the data in Figure* 1(b) *when the perturbing model is* (a) *the area-interaction process* (null fitted on a fine grid) *and* (b) *the $\hat{F}$-model* (null fitted on a coarse grid).

1.5, that is, the density is

$$(56) \quad f(\mathbf{x}) \propto \exp(n(\mathbf{x})\log\lambda + V_{G,s}(\mathbf{x}, r)\log\gamma),$$

where

$$V_{G,s}(\mathbf{x}, r) = \sum_i \min\left\{s, \sum_{j:j\neq i} \mathbb{I}\{\|x_i - x_j\| \leq r\}\right\}.$$

This is an example of moderately strong clustering (with interaction range $R = 2r = 0.1$) without trend. The main challenge here is to correctly identify the range and type of interaction.

We fitted three point process models to the data: (E) a homogeneous Poisson process (CSR); (F) a homogeneous area-interaction process with disc radius $r = 0.05$; (G) a homogeneous Geyer saturation process of the correct form, with interaction parameter $r = 0.05$ and saturation threshold $s = 4.5$ while the parameters λ and γ in (56) are unknown. The parameter estimates for (G) were $\log\hat{\lambda} = 4.12$ and $\log\hat{\gamma} = 0.38$.

13.2 Application of $\hat{K}$ Diagnostics

A plot (not shown) of the $\hat{K}$-function and its compensator, under each of the three models (E)–(G), demonstrates clearly that the homogeneous Poisson model (E) is a poor fit, but does not discriminate between the other models.

Figure 10 shows the residual $\hat{K}$ and the smoothed standardized residual $\hat{K}$ for the three models. These diagnostics show that the homogeneous Poisson model

(E) is a poor fit, with a positive residual suggesting correctly that the data are more clustered than the Poisson process. The plots suggests that both models (F) and (G) are considerably better fits to the data than a Poisson model. They show that (G) is a better fit than (F) over a range of r values, and suggest that (G) captures the correct form of the interaction.

13.3 Application of $\hat{G}$ Diagnostics

Figure 11 shows $\hat{G}$ and its compensator, and the corresponding residuals and standardized residuals, for each of the models (E)–(G) fitted to the clustered point pattern in Figure 1(c). The conclusions obtained from Figure 11(a) are the same as those in Section 13.2 based on $\hat{K}$ and its compensator. Figure 12 shows the smoothed pseudo-residual diagnostics based on the nearest neighbor distances. The message from these diagnostics is very similar to that from Figure 11.

Models (F) and (G) have the same range of interaction $R = 0.1$. Comparing Figures 10 and 11, we might conclude that the $\hat{G}$-compensator appears less sensitive to the *form* of interaction than the $\hat{K}$-compensator. Other experiments suggest that $\hat{G}$ is more sensitive than $\hat{K}$ to discrepancies in the *range* of interaction.

13.4 Application of $\hat{F}$ Diagnostics

Figure 13 shows the pseudo-residual diagnostics based on the empty space distances, for the three models fitted to the clustered point pattern in Figure 1(c).
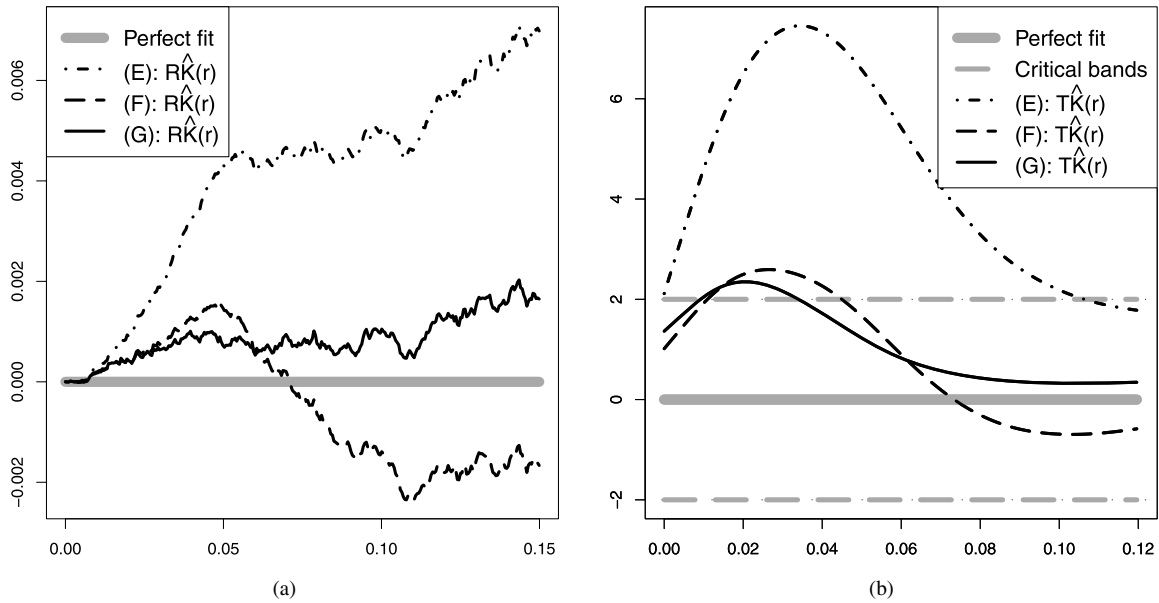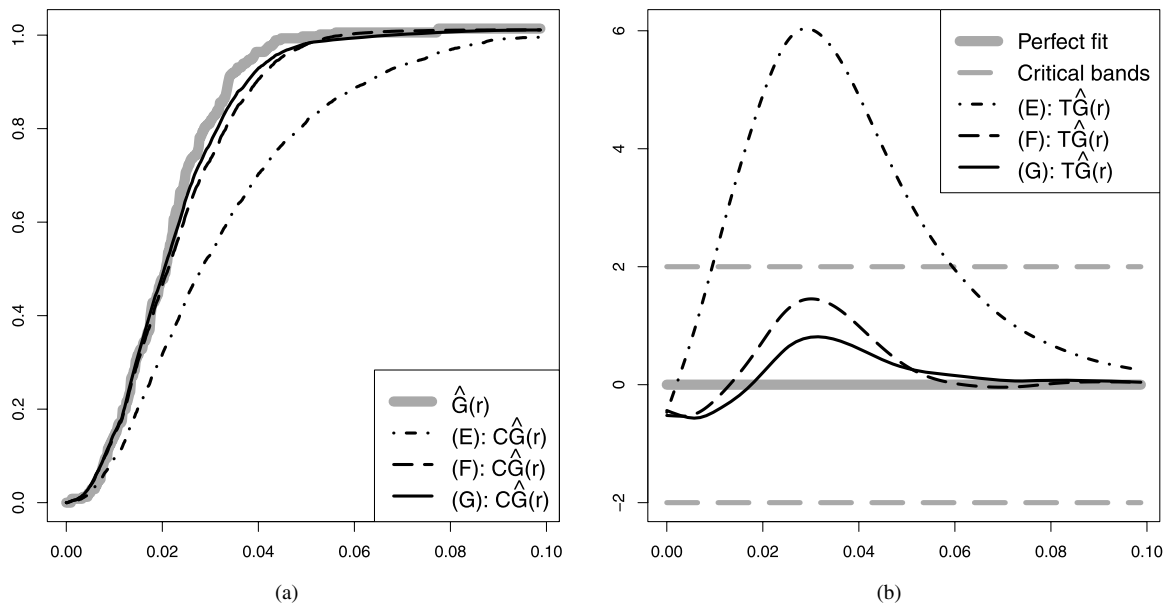
FIG. 10.   *Model diagnostics based on pairwise distances for each of the models* (E)–(G) *fitted to the data in Figure* 1(c). (a) *Residual* $\hat{K}$;
(b) *smoothed standardized residual* $\hat{K}$.

In this case diagnostics based on the area-interaction process and the $\hat{F}$-model are very similar, as expected due to the close connection between the two diagnostics. Here it is very noticeable that the pseudo-compensator for the Poisson model has a higher peak than the pseudo-sum, which correctly indicates that the data is more clustered than a Poisson process.

14. TEST CASE: JAPANESE PINES

14.1 Data and Models

Figure 1(a) shows the locations of seedlings and saplings of Japanese black pine, studied by Numata [53, 54] and analyzed extensively by Ogata and Tanemura [55, 56]. In their definitive analysis [56] the fit-

FIG. 11.   *Model diagnostics based on nearest neighbor distances for each of the models* (E)–(G) *fitted to the data in Figure* 1(c). (a) $\hat{G}$ *and its compensator under each model*; (b) *smoothed standardized residual* $\hat{G}$.
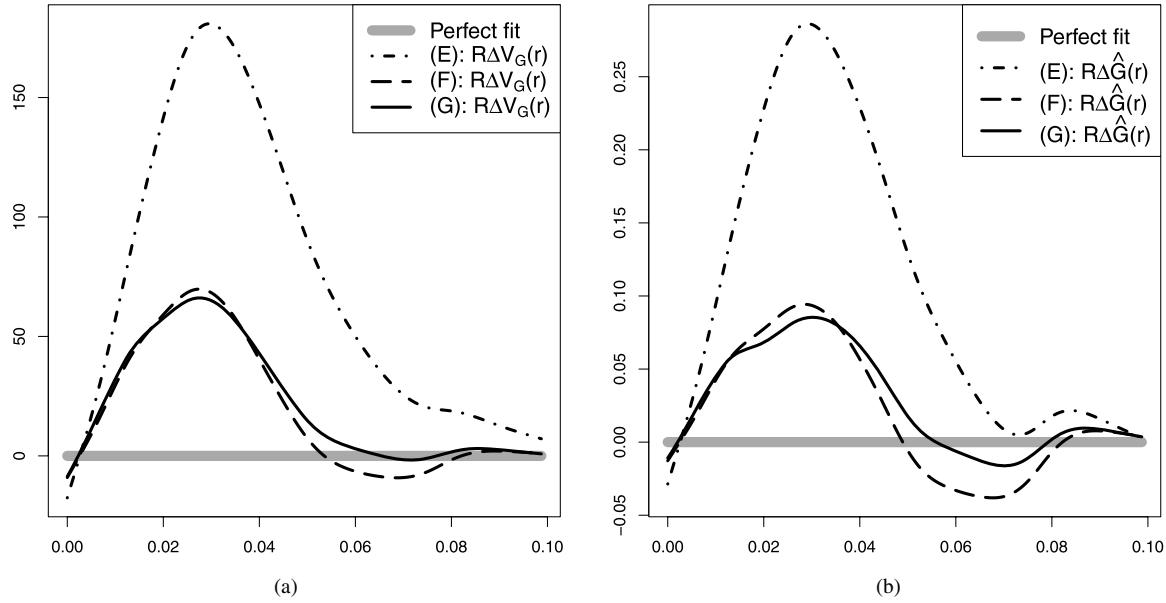
FIG. 12. *Smoothed pseudo-residuals for each of the models* (E)–(G) *fitted to the clustered point pattern in Figure* 1(c) *when the perturbing model is* (a) *the Geyer saturation model with saturation* 1 *and* (b) *the* $\hat{G}$*-model.*

ted model was an inhomogeneous "soft core" pair-wise interaction process with log-cubic first order term $\lambda_\beta(x, y) = \exp(P_\beta(x, y))$, where P_β is a cubic polynomial in x and y with coefficient vector β, and density

$$f_{(\beta,\sigma^2)}(\mathbf{x}) = c_{(\beta,\sigma^2)} \exp\Big(\sum_i P_\beta(x_i) $$
(57)
$$ - \sum_{i<j} (\sigma^4/\|x_i - x_j\|^4) \Big),$$

where σ^2 is a positive parameter.

Here we evaluate three models: (H) an inhomogeneous Poisson process with log-cubic intensity; (I) a homogeneous soft core pairwise interaction process, that is, when $P_\beta(x, y)$ in (57) is replaced by a real parameter; (J) the Ogata–Tanemura model (57). For more detail on the data set and the fitted inhomogeneous soft core model, see [7, 56].
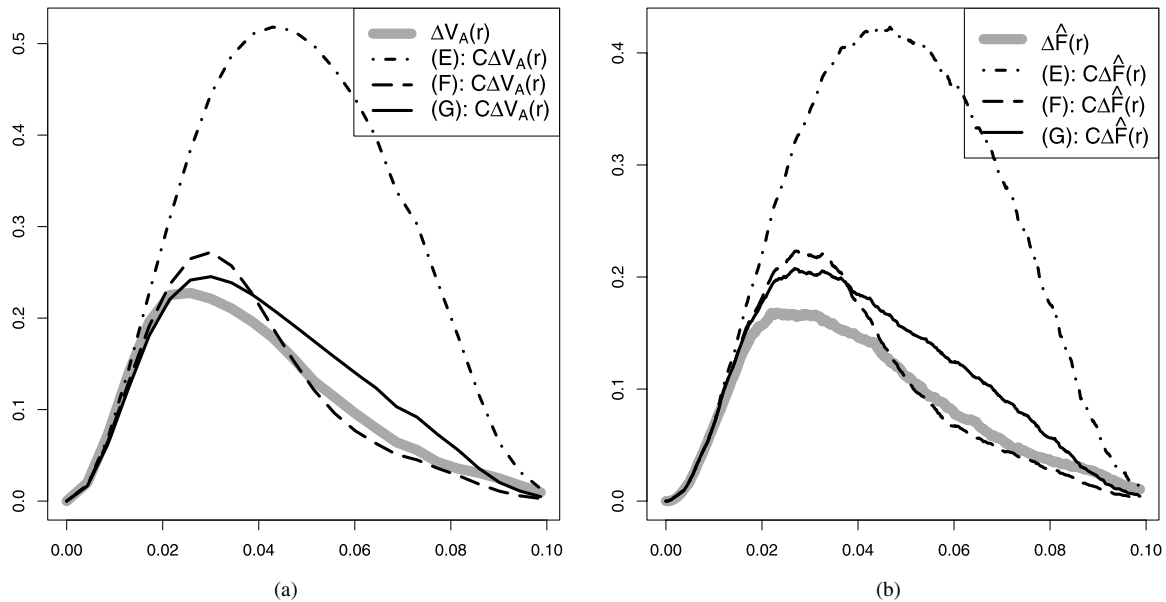

FIG. 13. *Pseudo-sum and pseudo-compensators for the models* (E)–(G) *fitted to the clustered point pattern in Figure* 1(c) *when the perturbing model is* (a) *the area-interaction process and* (b) *the* $\hat{F}$*-model.*
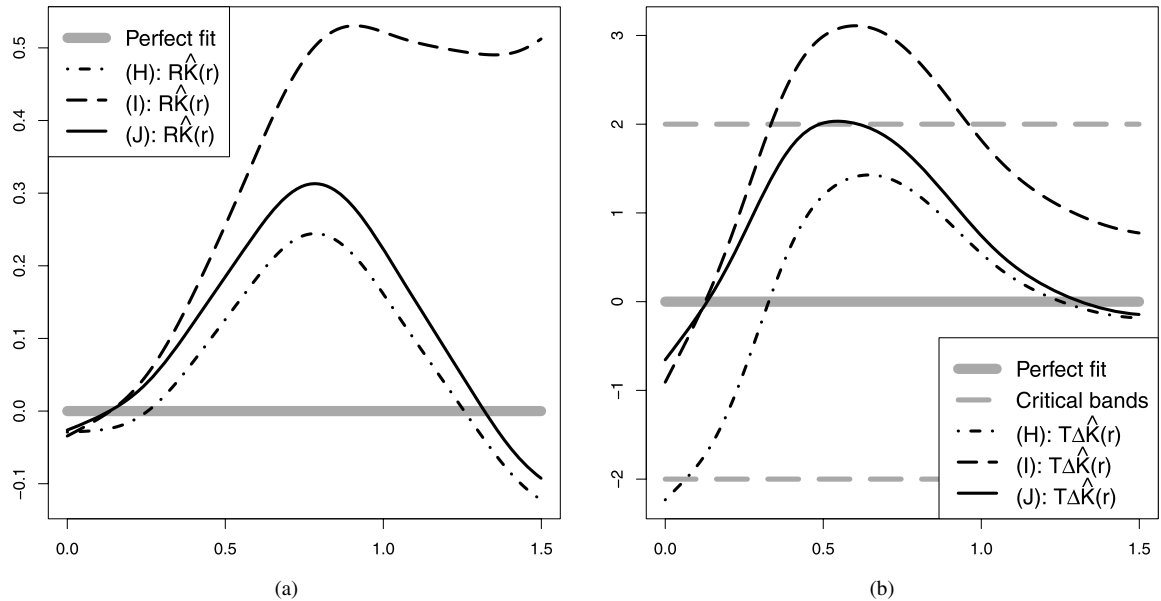
FIG. 14. *Model diagnostics based on pairwise distances for each of the models* (H)–(J) *fitted to the Japanese pines data in Figure* 1(a). (a) *Smoothed residual* $\hat{K}$; (b) *smoothed standardized residual* $\hat{K}$.

A complication in this case is that the soft core process (57) is not Markov, since the pair potential $c(u, v) = \exp(-\sigma^4/\|u - v\|^4)$ is always positive. Nevertheless, since this function decays rapidly, it seems reasonable to apply the residual and pseudo-residual diagnostics, using a cutoff distance R such that $|\log c(u, v)| \leq \varepsilon$ when $\|u - v\| \leq R$, for a specified tolerance ε. The cutoff depends on the fitted parameter value σ^2. We chose $\varepsilon = 0.0002$, yielding $R = 1$. Estimated interaction parameters were $\hat{\sigma}^2 = 0.11$ for model (I) and $\hat{\sigma}^2 = 0.12$ for model (J).

14.2 Application of $\hat{K}$ Diagnostics

A plot (not shown) of $\hat{K}$ and its compensator for each of the models (H)–(J) suggests that the homogeneous soft core model (I) is inadequate, while the inhomogeneous models (H) and (J) are reasonably good fits to the data. However, it does not discriminate between the models (H) and (J).

Figure 14 shows smoothed versions of the residual and standardized residual of $\hat{K}$ for each model. The Ogata–Tanemura model (J) is judged to be the best fit.

14.3 Application of $\hat{G}$ diagnostics

Finally, for each of the models (H)–(J) fitted to the Japanese pines data in Figure 1(a), Figure 15(a) shows $\hat{G}$ and its compensator. The conclusions are the same as those based on $\hat{K}$ shown in Figure 14. Figure 16 shows the pseudo-residuals when using either a per-

turbing Geyer model [Figure 16(a)] or a perturbing $\hat{G}$-model [Figure 16(b)]. Figures 16(a)–(b) tell almost the same story: the inhomogeneous Poisson model (H) provides the worst fit, while it is difficult to discriminate between the fit for the soft core models (I) and (J). In conclusion, considering Figures 14, 15 and 16, the Ogata–Tanemura model (J) provides the best fit.

14.4 Application of $\hat{F}$ diagnostics

Finally, the empty space pseudo-residual diagnostics are shown in Figure 17 for the Japanese Pines data in Figure 1(a). This gives a clear indication that the Ogata–Tanemura model (J) is the best fit to the data, and the data pattern appears to be too regular compared to the Poisson model (H) and not regular enough for the homogeneous softcore model (I).

15. SUMMARY OF TEST CASES

In this section we discuss which of the diagnostics we prefer to use based on their behavior for the three test cases in Sections 12–14.

Typically, the various diagnostics supplement each other well, and it is recommended to use more than one diagnostic when validating a model. It is well known that $\hat{K}$ is sensitive to features at a larger scale than $\hat{G}$ and $\hat{F}$. Compensator and pseudo-compensator plots are informative for gaining an overall picture of model validity, and tend to make it easy to recognize a poor fit
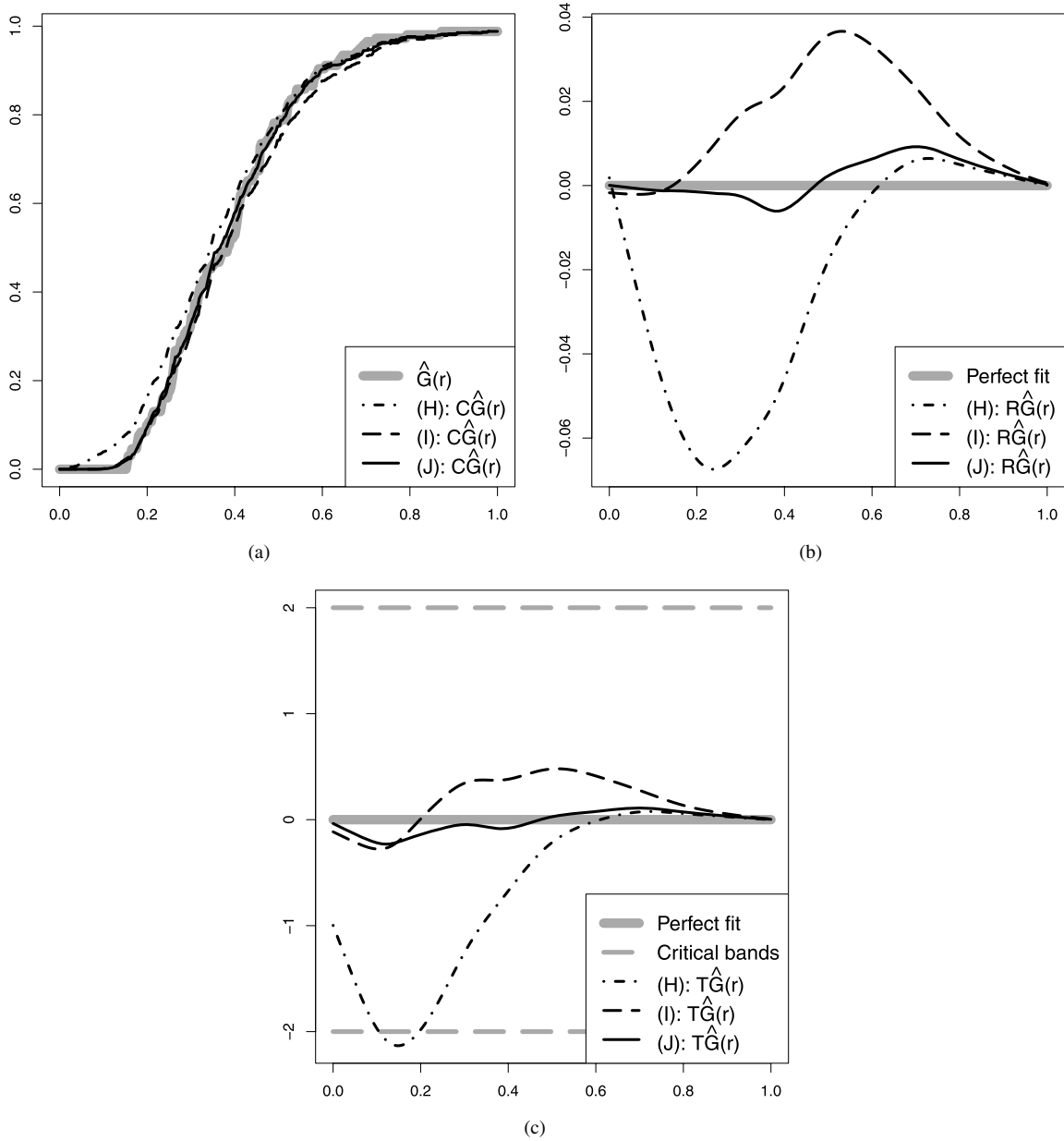
FIG. 15. *Model diagnostics based on nearest neighbour distances for each of the models* (H)–(J) *fitted to the Japanese pines data in Figure* 1(a). (a) $\hat{G}$ *and its compensator*; (b) *smoothed residual* $\hat{G}$; (c) *smoothed standardised residual* $\hat{G}$.

when comparing competing models. To compare models which fit closely, it may be more informative to use (standardized) residuals or pseudo-residuals. We prefer to use the standardized residuals, but it is important not to over-interpret the significance of departure from zero.

Based on the test cases here, it is not clear whether diagnostics based on pairwise distances, nearest neighbor distances, or empty space distances are preferable. However, for each of these we prefer to work with compensators and residuals rather than pseudo-

compensators and pseudo-residuals when possible (i.e., it is only necessary to use pseudo-versions for diagnostics based on empty space distances). For instance, for the first test case (Section 12) the best compensator plot is that in Figure 5(a) based on pairwise distances ($\hat{K}$ and $\mathcal{C}\hat{K}$) which makes it easy to identify the correct model. On the other hand, in this test case the best residual type plot is that in Figure 8(b) based on nearest neighbor distances ($\mathcal{T}\hat{G}$) where the correct model is the only one within the critical bands. However, in the third test case (Section 14) the best
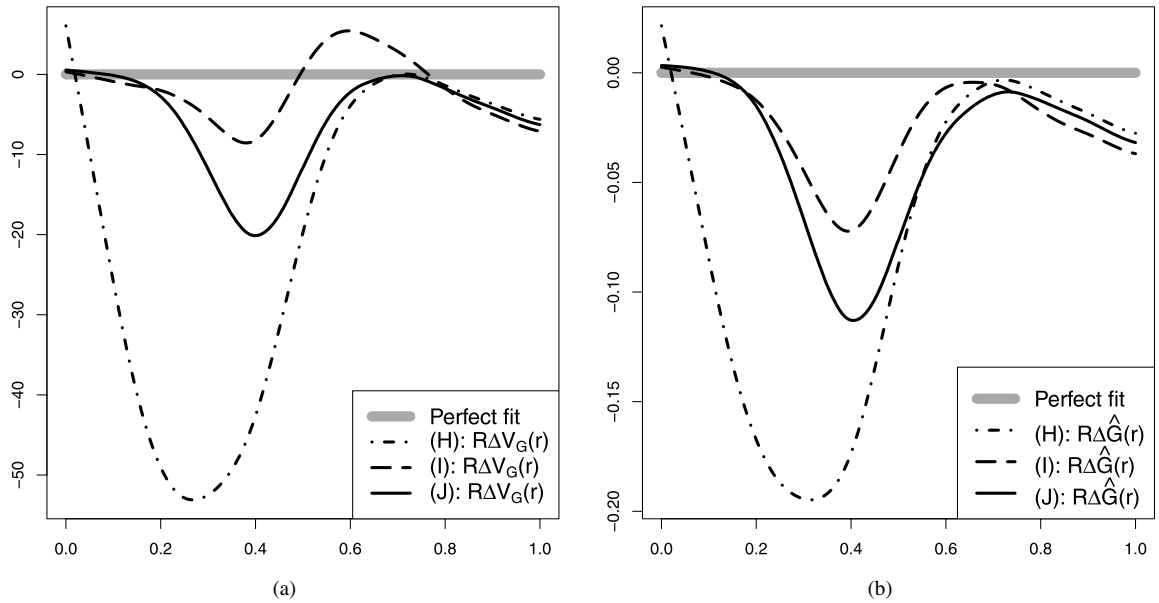
FIG. 16.   *Smoothed pseudo-residuals for each of the models* (H)–(J) *fitted to the Japanese pines data in Figure* 1(a) *when the perturbing model is* (a) *the Geyer saturation model with saturation* 1 (*null fitted on a fine grid*) *and* (b) *the* $\hat{G}$-*model.*

compensator plot is one of the plots in Figure 17 with pseudo-compensators based on empty space distances ($\Sigma \Delta V_A$ and $\mathcal{C} \Delta V_A$ or $\Sigma \Delta \hat{F}$ and $\mathcal{C} \Delta \hat{F}$, respectively) which clearly indicates which model is correct.

In the first and third test cases (Sections 12 and 14), which both involve inhomogeneous models, it is clear

that $\hat{K}$ and its compensator are more sensitive to lack of fit in the first order term than $\hat{G}$ and its compensator [compare, e.g., the results for the homogeneous model (C) in Figures 5(a) and 8(b)]. It is our general experience that diagnostics based on $\hat{K}$ are particularly well suited to assess the presence of interaction and to identify the general form of interaction. Diagnostics based
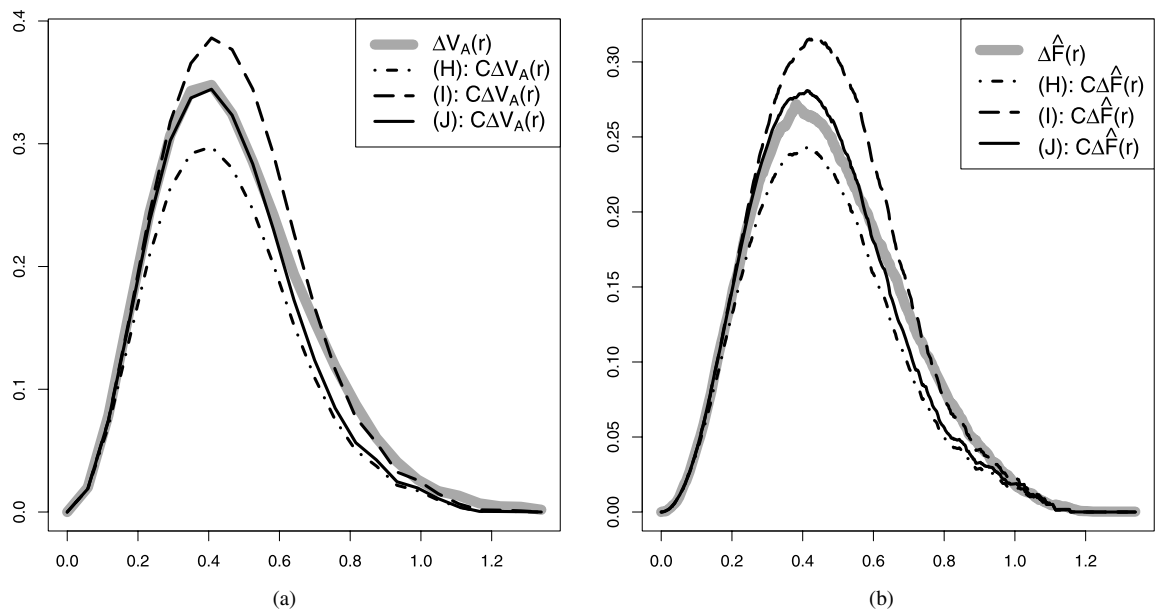

FIG. 17.   *Pseudo-sum and pseudo-compensators for the models* (H)–(J) *fitted to the real data pattern in Figure* 1(a) *when the perturbing model is* (a) *the area-interaction process and* (b) *the* $\hat{F}$-*model.*

on $\hat{K}$ and, in particular, on $\hat{G}$ seem to be good for assessing the range of interaction.

Finally, it is worth mentioning the computational difference between the various diagnostics (timed on a 2.5 GHz laptop). The calculations for $\hat{K}$ and $\mathcal{C}\hat{K}$ used in Figure 2 are carried out in approximately five seconds, whereas the corresponding calculations for $\hat{G}$ and $\mathcal{C}\hat{G}$ only take a fraction of a second. For $\Sigma\Delta\hat{F}$ and $\mathcal{C}\Delta\hat{F}$, for example, the calculations take about 45 seconds.

16. POSSIBLE EXTENSIONS

The definition of residuals and pseudo-residuals should extend immediately to marked point processes. For space–time point processes, residual diagnostics can be defined using the spatiotemporal conditional intensity (i.e., given the past history). Pseudo-residuals are unnecessary because the likelihood of a general space–time point process is a product integral (Mazziotto–Szpirglas identity). In the space–time case there is a martingale structure in time, which gives more hope of rigorous asymptotic results in the temporal (long-run) limit regime.

Residuals can be derived from many other summary statistics. Examples include third-order and higher-order moments (Appendix A.1), tessellation statistics (Appendix A.2), and various combinations of F, G and K.

In the definition of the extended model (25) the canonical statistic S could have been allowed to depend on the nuisance parameter θ, but this would have complicated our notation and some analysis.

APPENDIX A: FURTHER DIAGNOSTICS

In this appendix we present other diagnostics which we have not implemented in software, and which therefore are not accompanied by experimental results.

A.1 Third and Higher Order Functional Summary Statistics

While the intensity and K-function are frequently-used summaries for the first and second order moment properties of a spatial point process, third and higher order summaries have been less used [49, 67, 70, 72].

Statistic of order k. For a functional summary statistic of kth order, say,

$$(58) \quad S(\mathbf{x}, r) = \sum_{\{x_{i_1}, \ldots, x_{i_k}\} \subseteq \mathbf{x}} q(\{x_{i_1}, \ldots, x_{i_k}\}, r),$$

we obtain

$$
\begin{aligned}
(59) \quad &\Sigma\Delta S(\mathbf{x}, r) \\
&= k! S(\mathbf{x}, r) \\
&= k! \sum_{\{x_{i_1}, \ldots, x_{i_k}\} \subseteq \mathbf{x}} q(\{x_{i_1}, \ldots, x_{i_k}\}, r),
\end{aligned}
$$

$$
\begin{aligned}
&\mathcal{C}\Delta S(\mathbf{x}, r) \\
&= k! \mathcal{C} S(\mathbf{x}, r) \\
(60) \quad &= (k - 1)! \\
&\quad \cdot \int_W \sum_{\{x_{i_1}, \ldots, x_{i_{k-1}}\} \subseteq \mathbf{x}} q(\{x_{i_1}, \ldots, x_{i_{k-1}}, u\}, r) \\
&\qquad \cdot \lambda_{\hat{\theta}}(u, \mathbf{x}) \, du,
\end{aligned}
$$

$$
\begin{aligned}
(61) \quad &\mathsf{PU}(\hat{\theta}, r) \\
&= k! \mathcal{R} S(\mathbf{x}, r) = k! S(\mathbf{x}, r) - k! \mathcal{C} S(\mathbf{x}, r),
\end{aligned}
$$

where $i_1, i_2, \ldots$ are pairwise distinct in the sums in (59)–(60). In this case again, pseudo-residual diagnostics are equivalent to those based on residuals.

Third order example. For a stationary and isotropic point process (i.e., when the distribution of $\mathbf{X}$ is invariant under translations and rotations), the intensity and K-function of the process completely determine its first and second order moment properties. However, even in this case, the simplest description of third order moments depends on a three-dimensional vector specified from triplets (x_i, x_j, x_k) of points from $\mathbf{X}$ such as the lengths and angle between the vectors $x_i - x_j$ and $x_j - x_k$. This is often considered too complex, and instead one considers a certain one-dimensional property of the triangle $T(x_i, x_j, x_k)$ as exemplified below, where $L(x_i, x_j, x_k)$ denotes the largest side in $T(x_i, x_j, x_k)$.

Let the canonical sufficient statistic of the perturbing density (27) be

$$
\begin{aligned}
(62) \quad S(\mathbf{x}, r) &= V_T(\mathbf{x}, r) \\
&= \sum_{i < j < k} \mathbb{I}\{L(x_i, x_j, x_k) \le r\}.
\end{aligned}
$$

The perturbing model is a special case of the *triplet interaction point process* studied in [31]. It is also a special case of (58) with

$$q(\{x_i, x_j, x_k\}, r) = \mathbb{I}\{L(x_i, x_j, x_k) \le r\};$$

residual and pseudo-residual diagnostics are equivalent and given by (59)–(61).

A.2 Tessellation Functional Summary Statistics

Some authors have suggested the use of tessellation methods for characterizing spatial point processes [38]. A planar tessellation is a subdivision of planar region such as W or the entire plane $\mathbb{R}^2$. For example, consider the Dirichlet tessellation of W generated by $\mathbf{x}$, that is, the tessellation with cells

$$C(x_i|\mathbf{x}) = \{u \in W \,|\, \|u - x_i\| \leq \|u - x_j\|$$
$$\text{for all } x_j \text{ in } \mathbf{x}\},$$
$$i = 1, \ldots, n.$$

Suppose the canonical sufficient statistic of the perturbing density (27) is

$$(63) \quad S(\mathbf{x}, r) = V_O(\mathbf{x}, r) = \sum_i \mathbb{I}\{|C(x_i|\mathbf{x})| \leq r\}.$$

This is a sum of local contributions as in (33), although not of local statistics in the sense mentioned in Section 6.3, since $\mathbb{I}\{|C(x_i|\mathbf{x})| \leq r\}$ depends on those points in $\mathbf{x}_{-i}$ which are Dirichlet neighbors to x_i and such points may of course not be r-close to x_i (unless r is larger than the diameter of W). We call this perturbing model a *soft Ord process*; Ord's process as defined in [10] is the limiting case $\phi \to -\infty$ in (27), that is, when r is the lower bound on the size of cells. Since $V_O(\mathbf{x}) \leq n(\mathbf{x})$, the perturbing model is well-defined for all $\phi \in \mathbb{R}$.

Let $\sim_{\mathbf{x}}$ denote the Dirichlet neighbor relation for the points in $\mathbf{x}$, that is, $x_i \sim_{\mathbf{x}} x_j$ if $C(x_i|\mathbf{x}) \cap C(x_j|\mathbf{x}) \neq \varnothing$. Note that $x_i \sim_{\mathbf{x}} x_i$. Now,

$$\Delta_u S(\mathbf{x}, r) = \mathbb{I}\{|C(u|\mathbf{x} \cup \{u\})| \leq r\}$$
$$(64) \qquad + \sum_{v \neq u : v \sim_{\mathbf{x} \cup \{u\}} u} [\mathbb{I}\{|C(v|\mathbf{x} \cup \{u\})| \leq r\}$$
$$- \mathbb{I}\{|C(v|\mathbf{x} \setminus \{u\})| \leq r\}]$$

depends not only on the points in $\mathbf{x}$ which are Dirichlet neighbors to u (with respect to $\sim_{\mathbf{x} \cup \{u\}}$) but also on the Dirichlet neighbors to those points (with respect to $\sim_{\mathbf{x} \cup \{u\}}$ or with respect to $\sim_{\mathbf{x} \setminus \{u\}}$). In other words, if we define the iterated Dirichlet neighbor relation by that $x_i \sim_{\mathbf{x}}^2 x_j$ if there exists some x_k such that $x_i \sim_{\mathbf{x}} x_k$ and $x_j \sim_{\mathbf{x}} x_k$, then $t(u, \mathbf{x})$ depends on those points in $\mathbf{x}$ which are iterated Dirichlet neighbors to u with respect to $\sim_{\mathbf{x} \cup \{u\}}$ or with respect to $\sim_{\mathbf{x} \setminus \{u\}}$. The pseudo-sum associated to the soft Ord process is

$$\Sigma \Delta V_O(\mathbf{x}, r) = V_O(\mathbf{x}, r)$$
$$+ \sum_i \sum_{j \neq i : x_j \sim_{\mathbf{x}} x_i} [\mathbb{I}\{|C(x_j|\mathbf{x})| \leq r\}$$
$$- \mathbb{I}\{|C(x_j|\mathbf{x}_{-i})| \leq r\}]$$

and from (29) and (64) we obtain the pseudo-compensator. From (36) and (63), we obtain the Papangelou compensator

$$\mathcal{C}V_O(\mathbf{x}, r) = \int_W \mathbb{I}\{|C(u|\mathbf{x} \cup \{u\})| \leq r\} \lambda_{\hat{\theta}}(u, \mathbf{x}) \, du.$$

Many other examples of tessellation characteristics may be of interest. For example, often the Delaunay tessellation is used instead of the Dirichlet tessellation. This is the dual tessellation to the Dirichlet tessellation, where the Delaunay cells generated by $\mathbf{x}$ are given by those triangles $T(x_i, x_j, x_k)$ such that the disc containing x_i, x_j, x_k in its boundary does not contain any further points from $\mathbf{x}$ (strictly speaking we need to assume a regularity condition, namely, that $\mathbf{x}$ has to be in general quadratic position; for such details, see [10]). For instance, the summary statistic $t(\mathbf{x}, r)$ given by the number of Delaunay cells $T(x_i, x_j, x_k)$ with $L(x_i, x_j, x_k) \leq r$, where $L(u, v, w)$ is the length of the triangle with vertices u, v, w, is a kind of third order statisticis related to (62) but concerns only the maximal cliques of Dirichlet neighbors (assuming again the general quadratic position condition). The corresponding perturbing model has not been studied in the literature, to the best of our knowledge.

APPENDIX B: VARIANCE FORMULAE

This appendix concerns the variance of diagnostic quantities of the form

$$I = \sum_i h(x_i, \mathbf{X}_{-i}) - \int_W h(u, \mathbf{x}) \lambda_{\theta}(u, \mathbf{X}) \, du,$$

$$R = \sum_i h(x_i, \mathbf{X}_{-i}) - \int_W h(u, \mathbf{x}) \lambda_{\hat{\theta}}(u, \mathbf{X}) \, du,$$

where $h(\cdot)$ is a functional for which these quantities are almost surely finite, $\mathbf{X}$ is a point process on W with Papangelou conditional intensity $\lambda_{\theta}(u, \mathbf{X})$ and $\hat{\theta}$ is an estimate of θ (e.g., the MPLE).

B.1 General Identity

Exact formulae for the variance of the innovation I and residual R are given in [4]. Expressions for $\mathbb{V}\text{ar}R$ are unwieldy [4], Section 6, but to a first approximation we may ignore the effect of estimating θ and consider the variance of I. Suppressing the dependence on θ, this is ([4], Proposition 4),

$$\mathbb{V}\text{ar}I = \int_W \mathbb{E}[h(u, \mathbf{X})^2 \lambda(u, \mathbf{X})] \, du$$
$$(65) \qquad + \int_{W^2} \mathbb{E}[A(u, v, \mathbf{X}) + B(u, v, \mathbf{X})] \, du \, dv,$$

where

$$A(u, v, \mathbf{X}) = \Delta_u h(v, \mathbf{X}) \Delta_v h(u, \mathbf{X}) \lambda_2(u, v, \mathbf{X}),$$

$$B(u, v, \mathbf{X}) = h(u, \mathbf{X}) h(v, \mathbf{X})$$
$$\cdot \{\lambda(u, \mathbf{X}) \lambda(v, \mathbf{X}) - \lambda_2(u, v, \mathbf{X})\},$$

where $\lambda_2(u, v, \mathbf{x}) = \lambda(u, \mathbf{x}) \lambda(v, \mathbf{x} \cup \{u\})$ is the second order Papangelou conditional intensity. Note that for a Poisson process $B(u, v, \mathbf{X})$ is identically zero.

B.2 Pseudo-Score

Let $S(\mathbf{x}, z)$ be a functional summary statistic with function argument z. Take $h(u, \mathbf{X}) = \Delta_u S(\mathbf{x}, z)$. Then the innovation I is the pseudo-score (23), and the variance formula (65) becomes

$$\mathbb{V}\text{ar}[\text{PU}(\theta)]$$

$$= \int_W \mathbb{E}[(\Delta_u S(\mathbf{X}, z))^2 \lambda(u, \mathbf{X})] \, du$$

$$(66) \quad + \int_{W^2} \mathbb{E}[(\Delta_u \Delta_v S(\mathbf{X}, z))^2 \lambda_2(u, v, \mathbf{X})] \, du \, dv$$

$$+ \int_{W^2} \mathbb{E}[\Delta_u S(\mathbf{x}, z) \Delta_v S(\mathbf{x}, z)$$
$$\cdot \{\lambda(u, \mathbf{X}) \lambda(v, \mathbf{X}) - \lambda_2(u, v, \mathbf{X})\}] \, du \, dv,$$

where for $u \neq v$ and $\{u, v\} \cap \mathbf{x} = \varnothing$,

$$\Delta_u \Delta_v S(\mathbf{x}, z) = S(\mathbf{x} \cup \{u, v\}, z) - S(\mathbf{x} \cup \{u\}, z)$$
$$- S(\mathbf{x} \cup \{v\}, z) + S(\mathbf{x}, z)$$

satisfies $\Delta_u \Delta_v S(\mathbf{x}, z) = \Delta_v \Delta_u S(\mathbf{x}, z)$.

APPENDIX C: MODIFIED EDGE CORRECTIONS

Appendices C–E describe modifications to the standard edge corrected estimators of $K(r)$ and $G(r)$ required in the conditional case (Section 2.3) because the Papangelou conditional intensity $\lambda(u, \mathbf{x})$ can or should only be evaluated at locations $u \in W^\circ$ where $W^\circ \subset W$. Corresponding compensators are also given.

Assume the point process is Markov and we are in the conditional case as described in Section 5.4. Consider an empirical functional statistic of the form

$$(67) \qquad S_W(\mathbf{x}, r) = \sum_{x_i \in \mathbf{x}} s_W(x_i, \mathbf{x} \setminus \{x_i\}, r)$$

with compensator (in the unconditional case)

$$CS_W(\mathbf{x}, r) = \int_W s_W(u, \mathbf{x}, r) \lambda_{\hat{\theta}}(u, \mathbf{x}) \, du.$$

We explore two different strategies for modifying the edge correction.

In the *restriction approach*, we replace W by W° and $\mathbf{x}$ by $\mathbf{x}^\circ = \mathbf{x} \cap W^\circ$, yielding

$$S_{W^\circ}(\mathbf{x}, r) = \sum_{x_i \in \mathbf{x}^\circ} s_{W^\circ}(x_i, \mathbf{x}^\circ \setminus \{x_i\}, r),$$

$$(68)$$
$$CS_{W^\circ}(\mathbf{x}, r) = \int_{W^\circ} s_{W^\circ}(u, \mathbf{x}^\circ, r) \lambda_{\hat{\theta}}(u, \mathbf{x}^\circ | \mathbf{x}^+) \, du.$$

Data points in the boundary region W^+ are ignored in the calculation of the empirical statistic S_{W°. The boundary configuration $\mathbf{x}^+ = \mathbf{x} \cap W^+$ contributes only to the estimation of θ and the calculation of the Papangelou conditional intensity $\lambda_{\hat{\theta}}(\cdot, \cdot | \mathbf{x}^+)$. This has the advantage that the modified empirical statistic (68) is identical to the standard statistic S computed on the subdomain W°; it can be computed using existing software, and requires no new theoretical justification. The disadvantage is that we lose information by discarding some of the data.

In the *reweighting approach* we retain the boundary points and compute

$$S_{W^\circ, W}(\mathbf{x}, r) = \sum_{x_i \in \mathbf{x}^\circ} s_{W^\circ, W}(x_i, \mathbf{x} \setminus \{x_i\}, r),$$

$$CS_{W^\circ, W}(\mathbf{x}, r) = \int_{W^\circ} s_{W^\circ, W}(u, \mathbf{x}, r) \lambda_{\hat{\theta}}(u, \mathbf{x}^\circ | \mathbf{x}^+) \, du,$$

where $s_{W^\circ, W}(\cdot)$ is a modified version of $s_W(\cdot)$. Boundary points contribute to the computation of the modified summary statistic $S_{W^\circ, W}$ and its compensator. The modification is designed so that $S_{W^\circ, W}$ has properties analogous to S_W.

The K-function and G-function of a point process $\mathbf{Y}$ in $\mathbb{R}^2$ are defined [63, 64] under the assumption that $\mathbf{Y}$ is second order stationary and strictly stationary, respectively. The standard estimators $\hat{K}_W(r)$ and $\hat{G}_{\mathbf{x}}(r)$ of the K-function and G-function, respectively, are designed to be approximately pointwise unbiased estimators when applied to $\mathbf{X} = \mathbf{Y} \cap W$.

We do not necessarily assume stationarity, but when constructing modified summary statistics $\hat{K}_{W^\circ, W}(r)$ and $\hat{G}_{W^\circ, W}(r)$, we shall require that they are also approximately pointwise unbiased estimators of $K(r)$ and $G(r)$, respectively, when $\mathbf{Y}$ is stationary. This greatly simplifies the interpretation of plots of $\hat{K}_{W^\circ, W}(r)$ and $\hat{G}_{W^\circ, W}(r)$ and their compensators.

APPENDIX D: MODIFIED EDGE CORRECTIONS FOR THE K-FUNCTION

D.1 Horvitz–Thompson Estimators

The most common nonparametric estimators of the K-function [9, 57, 63] are continuous Horvitz–

Thompson type estimators [8, 20] of the form

$$\hat{K}(r) = \hat{K}_W(r)$$

(69)

$$= \frac{1}{\hat{\rho}^2(\mathbf{x})|W|} \sum_{i \neq j} e_W(x_i, x_j) \mathbb{I}\{\|x_i - x_j\| \leq r\}.$$

Here $\hat{\rho}^2 = \hat{\rho}^2(\mathbf{x})$ should be an approximately unbiased estimator of the squared intensity ρ^2 under stationarity. Usually $\hat{\rho}^2(\mathbf{x}) = n(n-1)/|W|^2$ where $n = n(\mathbf{x})$.

The term $e_W(u, v)$ is an edge correction weight, depending on the geometry of W, designed so that the double sum in (69), say, $\hat{Y}(r) = \hat{\rho}^2(\mathbf{x})|W|\hat{K}(r)$, is an unbiased estimator of $Y(r) = \rho^2|W|K(r)$. Popular examples are the Ohser–Stoyan translation edge correction with

$$e_W(u, v) = e_W^{\text{trans}}(u, v)$$

(70)

$$= \frac{|W|}{|W \cap (W + (u - v))|}$$

and Ripley's isotropic correction with

$$e_W(u, v) = e_W^{\text{iso}}(u, v)$$

(71)

$$= \frac{2\pi \|u - v\|}{\text{length}(\partial B(u, \|u - v\|) \cap W)}.$$

Estimators of the form (69) satisfy the local decomposition (67) where

$$s_W(u, \mathbf{x}, r) = \frac{1}{\hat{\rho}^2(\mathbf{x} \cup \{u\})|W|}$$

$$\cdot \sum_j e_W(u, x_j) \mathbb{I}\{\|u - x_j\| \leq r\}, \quad u \notin \mathbf{x}.$$

Now we wish to modify (69) so that the outer summation is restricted to data points x_i in $W^\circ \subset W$, while retaining the property of unbiasedness for stationary and isotropic point processes. The *restriction estimator* is

$$\hat{K}_{W^\circ}(r)$$

(72)

$$= \frac{1}{\hat{\rho}^2(\mathbf{x}^\circ)|W^\circ|}$$

$$\cdot \sum_{x_i \in \mathbf{x}^\circ} \sum_{x_j \in \mathbf{x}^\circ_{-i}} e_{W^\circ}(x_i, x_j) \mathbb{I}\{\|x_i - x_j\| \leq r\},$$

where the edge correction weight is given by (70) or (71) with W replaced by W°. A more efficient alternative is to replace (69) by the *reweighting estimator*

$$\hat{K}_{W^\circ, W}(r)$$

(73)

$$= \frac{1}{\hat{\rho}^2(\mathbf{x})|W^\circ|}$$

$$\cdot \sum_{x_i \in \mathbf{x}^\circ} \sum_{x_j \in \mathbf{x}_{-i}} e_{W^\circ, W}(x_i, x_j) \mathbb{I}\{\|x_i - x_j\| \leq r\},$$

where $e_{W^\circ, W}(u, v)$ is a modified version of $e_W(\cdot)$ constructed so that the double sum in (73) is unbiased for $Y(r)$. Compared to the restriction estimator (72), the reweighting estimator (73) contains additional contributions from point pairs (x_i, x_j) where $x_i \in \mathbf{x}^\circ$ and $x_j \in \mathbf{x}^+$.

The modified edge correction factor $e_{W^\circ, W}(\cdot)$ for (73) is the Horvitz–Thompson weight [9] in an appropriate sampling context. Ripley's [63, 64] isotropic correction (71) is derived assuming isotropy, by Palm conditioning on the location of the first point x_i, and determining the probability that x_j would be observed inside W after a random rotation about x_i. Since the constraint on x_j is unchanged, no modification of the edge correction weight is required, and we take $e_{W^\circ, W}(\cdot) = e_W(\cdot)$ as in (71). Note, however, that the denominator in (73) is changed from $|W|$ to $|W^\circ|$.

The Ohser–Stoyan [58] translation correction (70) is derived by considering two-point sets (x_i, x_j) sampled under the constraint that both x_i and x_j are inside W. Under the modified constraint that $x_i \in W^\circ$ and $x_j \in W$, the appropriate edge correction weight is

$$e_{W^\circ, W}(u, v) = e_{W^\circ, W}(u - v)$$

$$= \frac{|W \cap (W^\circ + (u - v))|}{|W^\circ|}$$

so that $1/e_{W^\circ, W}(z)$ is the fraction of locations u in W° such that $u + z \in W$.

D.2 Border Correction

A slightly different creature is the border corrected estimator [using usual intensity estimator $\hat{\rho} = n(\mathbf{x})/|W|$]

$$\hat{K}_W(r) = \frac{|W|}{n(\mathbf{x})n(\mathbf{x} \cap W_{\ominus r})}$$

$$\cdot \sum_{x_i \in \mathbf{x}} \sum_{x_j \in \mathbf{x}_{-i}} \mathbb{I}\{x_i \in W_{\ominus r}\} \mathbb{I}\{\|x_i - x_j\| \leq r\}$$

with compensator (in the unconditional case)

$$\mathcal{C}\hat{K}_W(r) = \int_{W_{\ominus r}} \frac{|W| \sum_{x_j \in \mathbf{x}} \mathbb{I}\{\|u - x_j\| \leq r\}}{(n(\mathbf{x}) + 1)(n(\mathbf{x} \cap W_{\ominus r}) + 1)}$$

$$\cdot \lambda_{\hat{\theta}}(u, \mathbf{x}^\circ | \mathbf{x}^+) \, du.$$

The *restriction estimator* is

$$\hat{K}_{W^\circ}(r) = \frac{|W^\circ|}{n(\mathbf{x}^\circ)n(\mathbf{x} \cap W^\circ_{\ominus r})}$$

$$\cdot \sum_{x_i \in \mathbf{x}^\circ} \sum_{x_j \in \mathbf{x}^\circ_{-i}} \mathbb{I}\{x_i \in W^\circ_{\ominus r}\} \mathbb{I}\{\|x_i - x_j\| \leq r\}$$

and the compensator is

$$C\hat{K}_{W^\circ}(r) = \int_{W^\circ_{\ominus r}} \frac{|W^\circ| \sum_{x_j \in \mathbf{x}^\circ} \mathbb{I}\{\|u - x_j\| \le r\}}{(n(\mathbf{x}^\circ) + 1)(n(\mathbf{x} \cap W^\circ_{\ominus r}) + 1)}$$
$$\cdot \lambda_{\hat\theta}(u, \mathbf{x}^\circ | \mathbf{x}^+) \, du.$$

Typically, $W^\circ = W_{\ominus R}$, so $W^\circ_{\ominus r}$ is equal to $W_{\ominus(R+r)}$. The *reweighting estimator* is

$$\hat{K}_{W^\circ, W}(r) = \frac{|W|}{n(\mathbf{x}) n(\mathbf{x}^\circ \cap W_{\ominus r})}$$
$$\cdot \sum_{x_i \in \mathbf{x}^\circ} \sum_{x_j \in \mathbf{x}_{-i}} \mathbb{I}\{x_i \in W_{\ominus r}\} \mathbb{I}\{\|x_i - x_j\| \le r\}$$

and the compensator is

$$C\hat{K}_{W^\circ, W}(r) = \int_{W^\circ \cap W_{\ominus r}} \frac{|W| \sum_{x_j \in \mathbf{x}} \mathbb{I}\{\|u - x_j\| \le r\}}{(n(\mathbf{x}) + 1)(n(\mathbf{x}^\circ \cap W_{\ominus r}) + 1)}$$
$$\cdot \lambda_{\hat\theta}(u, \mathbf{x}^\circ | \mathbf{x}^+) \, du.$$

Usually, $W^\circ = W_{\ominus R}$, so $W^\circ \cap W_{\ominus r}$ is equal to $W_{\ominus \max(R,r)}$. From this we conclude that when using border correction we should always use the reweighting estimator since the restriction estimator discards additional information and neither the implementation nor the interpretation is easier.

APPENDIX E: MODIFIED EDGE CORRECTIONS FOR NEAREST NEIGHBOR FUNCTION G

E.1 Hanisch Estimators

Hanisch [32] considered estimators for $G(r)$ of the form $\hat{G}_W(r) = \hat{D}_\mathbf{x}(r)/\hat\rho$, where $\hat\rho$ is some estimator of the intensity ρ, and

$$(74) \qquad \hat{D}_\mathbf{x}(r) = \sum_{x_i \in \mathbf{x}} \frac{\mathbb{I}\{x_i \in W_{\ominus d_i}\} \mathbb{I}\{d_i \le r\}}{|W_{\ominus d_i}|},$$

where $d_i = d(x_i, \mathbf{x} \setminus \{x_i\})$ is the nearest neighbor distance for x_i. If $\hat\rho$ were replaced by ρ, then $\hat{G}_W(r)$ would be an unbiased, Horvitz–Thompson estimator of $G(r)$. See [71], pages 128–129, [9]. Hanisch's recommended estimator D_4 is the one in which $\hat\rho$ is taken to be

$$\hat{D}_\mathbf{x}(\infty) = \sum_{x_i \in \mathbf{x}} \frac{\mathbb{I}\{x_i \in W_{\ominus d_i}\}}{|W_{\ominus d_i}|}.$$

This is sensible because $\hat{D}_\mathbf{x}(\infty)$ is an unbiased estimator of ρ and is positively correlated with $\hat{D}_\mathbf{x}(r)$. The resulting estimator $\hat{G}_W(r)$ can be decomposed in the form (67) where

$$s_W(u, \mathbf{x}, r) = \frac{\mathbb{I}\{u \in W_{\ominus d(u,\mathbf{x})}\} \mathbb{I}\{d(u, \mathbf{x}) \le r\}}{\hat{D}_{\mathbf{x} \cup \{u\}}(\infty) |W_{\ominus d(u,\mathbf{x})}|}$$

for $u \notin \mathbf{x}$, where $d(u, \mathbf{x})$ is the ("empty space") distance from location u to the nearest point of $\mathbf{x}$. Hence, the corresponding compensator is

$$C\hat{G}_W(r) = \int_W \frac{\mathbb{I}\{u \in W_{\ominus d(u,\mathbf{x})}\} \mathbb{I}\{d(u, \mathbf{x}) \le r\}}{\hat{D}_{\mathbf{x} \cup \{u\}}(\infty) |W_{\ominus d(u,\mathbf{x})}|}$$
$$\cdot \lambda_{\hat\theta}(u, \mathbf{x}) \, du.$$

This is difficult to evaluate, since the denominator of the integrand involves a summation over all data points: $D_{\mathbf{x} \cup \{u\}}(\infty)$ is not related in a simple way to $D_\mathbf{x}(\infty)$. Instead, we choose $\hat\rho$ to be the conventional estimator $\hat\rho = n(\mathbf{x})/|W|$. Then

$$\hat{G}_W(r) = \frac{|W|}{n(\mathbf{x})} \hat{D}_\mathbf{x}(r),$$

which can be decomposed in the form (67) with

$$s_W(u, \mathbf{x}, r) = \frac{|W|}{n(\mathbf{x}) + 1} \frac{\mathbb{I}\{u \in W_{\ominus d(u,\mathbf{x})}\} \mathbb{I}\{d(u, \mathbf{x}) \le r\}}{|W_{\ominus d(u,\mathbf{x})}|}$$

for $u \notin \mathbf{x}$, so that the compensator is

$$C\hat{G}_W(r) = \frac{|W|}{n(\mathbf{x}) + 1} \int_W \frac{\mathbb{I}\{u \in W_{\ominus d(u,\mathbf{x})}\} \mathbb{I}\{d(u, \mathbf{x}) \le r\}}{|W_{\ominus d(u,\mathbf{x})}|}$$
$$(75) \qquad \cdot \lambda_{\hat\theta}(u, \mathbf{x}) \, du.$$

In the *restriction estimator* we exclude the boundary points and take $d_i^\circ = d(x_i, \mathbf{x}^\circ_{-i})$, effectively replacing the data set $\mathbf{x}$ by its restriction $\mathbf{x}^\circ = \mathbf{x} \cap W^\circ$:

$$\hat{G}_{W^\circ}(r) = \frac{|W^\circ|}{n(\mathbf{x}^\circ)} \sum_{x_i \in \mathbf{x}^\circ} \frac{\mathbb{I}\{x_i \in W^\circ_{\ominus d_i^\circ}\} \mathbb{I}\{d_i^\circ \le r\}}{|W^\circ_{\ominus d_i^\circ}|}.$$

The compensator is (75) but computed for the point pattern $\mathbf{x}^\circ$ in the window W°:

$$C\hat{G}_{W^\circ}(r)$$
$$= \frac{|W^\circ|}{n(\mathbf{x}^\circ) + 1} \int_{W^\circ} \frac{\mathbb{I}\{u \in W^\circ_{\ominus d(u,\mathbf{x}^\circ)}\} \mathbb{I}\{d(u, \mathbf{x}^\circ) \le r\}}{|W^\circ_{\ominus d(u,\mathbf{x}^\circ)}|}$$
$$\cdot \lambda_{\hat\theta}(u, \mathbf{x}^\circ | \mathbf{x}^+) \, du.$$

In the usual case $W^\circ = W_{\ominus R}$, we have $W^\circ_{\ominus d} = W_{\ominus(R+d)}$.

In the *reweighting estimator* we take $d_i = d(x_i, \mathbf{x} \setminus \{x_i\})$. To retain the Horvitz–Thompson property, we must replace the weights $1/|W_{\ominus d_i}|$ in (74) by $1/|W^\circ \cap W_{\ominus d_i}|$. Thus, the modified statistics are

$$(76) \quad \hat{G}_{W^\circ, W}(r) = \frac{|W|}{n(\mathbf{x})} \sum_{x_i \in \mathbf{x}^\circ} \frac{\mathbb{I}\{x_i \in W_{\ominus d_i}\} \mathbb{I}\{d_i \le r\}}{|W^\circ \cap W_{\ominus d_i}|}$$

and

$$\mathcal{C}\hat{G}_{W^\circ, W}(r)$$

$$(77) \quad = \frac{|W|}{n(\mathbf{x}) + 1} \int_{W^\circ} \frac{\mathbb{I}\{u \in W_{\ominus d(u, \mathbf{x})}\}\mathbb{I}\{d(u, \mathbf{x}) \le r\}}{|W^\circ \cap W_{\ominus d(u, \mathbf{x})}|}$$
$$\cdot \lambda_{\hat{\theta}}(u, \mathbf{x}^\circ | \mathbf{x}^+) \, du.$$

In the usual case where $W^\circ = W_{\ominus R}$, we have $W^\circ \cap W_{\ominus d_i} = W_{\ominus \max(R, d_i)}$.

Optionally, we may also replace $|W|/n(\mathbf{x})$ in (76) by $|W^\circ|/n(\mathbf{x} \cap W^\circ)$, and, correspondingly, replace $\frac{|W|}{n(\mathbf{x})+1}$ in (77) by $|W^\circ|/(n(\mathbf{x} \cap W^\circ) + 1)$.

E.2 Border Correction Estimator

The classical border correction estimate of G is

$$(78) \quad \begin{aligned} \hat{G}_W(r) &= \frac{1}{n(\mathbf{x} \cap W_{\ominus r})} \\ &\quad \cdot \sum_{x_i \in \mathbf{x}} \mathbb{I}\{x_i \in W_{\ominus r}\}\mathbb{I}\{d(x_i, \mathbf{x}_{-i}) \le r\} \end{aligned}$$

with compensator (in the unconditional case)

$$(79) \quad \begin{aligned} \mathcal{C}\hat{G}_W(r) &= \frac{1}{1 + n(\mathbf{x} \cap W_{\ominus r})} \\ &\quad \cdot \int_{W_{\ominus r}} \mathbb{I}\{d(u, \mathbf{x}) \le r\}\lambda_{\hat{\theta}}(u, \mathbf{x}) \, du. \end{aligned}$$

In the conditional case, the Papangelou conditional intensity $\lambda_{\hat{\theta}}(u, \mathbf{x})$ must be replaced by $\lambda_{\hat{\theta}}(u, \mathbf{x}^\circ | \mathbf{x}^+)$ given in (24). The *restriction estimator* is obtained by replacing W by W° and $\mathbf{x}$ by $\mathbf{x}^\circ$ in (78)–(79), yielding

$$\begin{aligned} \hat{G}_{W^\circ}(r) &= \frac{1}{n(\mathbf{x} \cap W_{\ominus r}^\circ)} \\ &\quad \cdot \sum_{x_i \in \mathbf{x}^\circ} \mathbb{I}\{x_i \in W_{\ominus r}^\circ\}\mathbb{I}\{d(x_i, \mathbf{x}_{-i}^\circ) \le r\}, \end{aligned}$$

$$\begin{aligned} \mathcal{C}\hat{G}_{W^\circ}(r) &= \frac{1}{1 + n(\mathbf{x} \cap W_{\ominus r}^\circ)} \\ &\quad \cdot \int_{W_{\ominus r}^\circ} \mathbb{I}\{d(u, \mathbf{x}^\circ) \le r\}\lambda_{\hat{\theta}}(u, \mathbf{x}^\circ | \mathbf{x}^+) \, du. \end{aligned}$$

Typically, $W^\circ = W_{\ominus R}$ so that $W_{\ominus r}^\circ = W_{\ominus(R+r)}$. The *reweighting estimator* is obtained by restricting x_i and u in (78)–(79) to lie in W°, yielding

$$\begin{aligned} \hat{G}_{W^\circ, W}(r) &= \frac{1}{n(\mathbf{x}^\circ \cap W_{\ominus r})} \\ &\quad \cdot \sum_{x_i \in \mathbf{x}^\circ} \mathbb{I}\{x_i \in W_{\ominus r}\}\mathbb{I}\{d(x_i, \mathbf{x}_{-i}) \le r\}, \end{aligned}$$

$$\begin{aligned} \mathcal{C}\hat{G}_{W^\circ, W}(r) &= \frac{1}{1 + n(\mathbf{x}^\circ \cap W_{\ominus r})} \\ &\quad \cdot \int_{W^\circ \cap W_{\ominus r}} \mathbb{I}\{d(u, \mathbf{x}) \le r\} \\ &\quad \cdot \lambda_{\hat{\theta}}(u, \mathbf{x}^\circ | \mathbf{x}^+) \, du. \end{aligned}$$

In the usual case where $W^\circ = W_{\ominus R}$, we have $W^\circ \cap W_{\ominus r} = W_{\ominus \max(R, r)}$. Again, the reweighting approach is preferable to the restriction approach.

The border corrected estimator $\hat{G}(r)$ has relatively poor performance and sample properties [38], page 209. Its main advantage is its computational efficiency in large data sets. Similar considerations should apply to its compensator.

ACKNOWLEDGMENTS

REFERENCES

[1] ALM, S. E. (1998). Approximation and simulation of the distributions of scan statistics for Poisson processes in higher dimensions. *Extremes* **1** 111–126. MR1652932

[2] ATKINSON, A. C. (1982). Regression diagnostics, transformations and constructed variables (with discussion). *J. Roy. Statist. Soc. Ser. B* **44** 1–36. MR0655369

[3] BADDELEY, A. (1980). A limit theorem for statistics of spatial data. *Adv. in Appl. Probab.* **12** 447–461. MR0569436

[4] BADDELEY, A., MØLLER, J. and PAKES, A. G. (2008). Properties of residuals for spatial point processes. *Ann. Inst. Statist. Math.* **60** 627–649. MR2434415

[5] BADDELEY, A. and TURNER, R. (2000). Practical maximum pseudolikelihood for spatial point patterns (with discussion). *Aust. N. Z. J. Stat.* **42** 283–322. MR1794056

[6] BADDELEY, A. and TURNER, R. (2005). Spatstat: An R package for analyzing spatial point patterns. *J. Statist. Software* **12** 1–42.

[7] BADDELEY, A., TURNER, R., MØLLER, J. and HAZELTON, M. (2005). Residual analysis for spatial point processes (with discussion). *J. R. Stat. Soc. Ser. B Stat. Methodol.* **67** 617–666. MR2210685

[8] BADDELEY, A. J. (1993). Stereology and survey sampling theory. *Bull. Int. Statist. Inst.* **50** 435–449.

[9] BADDELEY, A. J. (1999). Spatial sampling and censoring. In *Stochastic Geometry* (*Toulouse*, 1996). *Monogr. Statist. Appl. Probab.* **80** 37–78. Chapman & Hall/CRC, Boca Raton, FL. MR1673114

[10] BADDELEY, A. J. and MØLLER, J. (1989). Nearest-neighbour Markov point processes and random sets. *Int. Stat. Rev.* **57** 89–121.

[11] BADDELEY, A. J. and VAN LIESHOUT, M. N. M. (1995). Area-interaction point processes. *Ann. Inst. Statist. Math.* **47** 601–619. MR1370279

[12] BARNARD, G. (1963). Discussion of "The spectral analysis of point processes" by M. S. Bartlett. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **25** 294.

[13] BERMAN, M. (1986). Testing for spatial association between a point process and another stochastic process. *J. Roy. Statist. Soc. Ser. C* **35** 54–62.

[14] BESAG, J. (1978). Some methods of statistical analysis for spatial data. *Bull. Int. Statist. Inst.* **44** 77–92.

[15] CHEN, C. (1983). Score tests for regression models. *J. Amer. Statist. Assoc.* **78** 158–161.

[16] CHETWYND, A. G. and DIGGLE, P. J. (1998). On estimating the reduced second moment measure of a stationary spatial point process. *Aust. N. Z. J. Stat.* **40** 11–15. MR1628212

[17] COEURJOLLY, J. F. and LAVANCIER, F. (2010). Residuals for stationary marked Gibbs point processes. Preprint. Available at http://arxiv.org/abs/1002.0857.

[18] CONNIFFE, D. (2001). Score tests when a nuisance parameter is unidentified under the null hypothesis. *J. Statist. Plann. Inference* **97** 67–83. MR1851375

[19] COOK, R. D. and WEISBERG, S. (1983). Diagnostics for heteroscedasticity in regression. *Biometrika* **70** 1–10. MR0742970

[20] CORDY, C. B. (1993). An extension of the Horvitz–Thompson theorem to point sampling from a continuous universe. *Statist. Probab. Lett.* **18** 353–362. MR1247446

[21] COX, D. R. (1972). The statistical analysis of dependencies in point processes. In *Stochastic Point Processes*: *Statistical Analysis*, *Theory, and Applications* (*Conf.*, *IBM Res. Center*, *Yorktown Heights*, *N.Y.*, 1971) 55–66. Wiley-Interscience, New York. MR0375705

[22] COX, D. R. and HINKLEY, D. V. (1974). *Theoretical Statistics*. Chapman & Hall, London. MR0370837

[23] CRESSIE, N. A. C. (1991). *Statistics for Spatial Data*. Wiley, New York. MR1127423

[24] DALEY, D. J. and VERE-JONES, D. (2003). *An Introduction to the Theory of Point Processes*. *Volume I*: *Elementary Theory and Methods*, 2nd ed. Springer, New York. MR1950431

[25] DAVIES, R. B. (1977). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* **64** 247–254. MR0501523

[26] DAVIES, R. B. (1987). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* **74** 33–43. MR0885917

[27] DIGGLE, P. J. (1979). On parameter estimation and goodness-of-fit testing for spatial point patterns. *Biometrika* **35** 87–101.

[28] DIGGLE, P. J. (1985). A kernel method for smoothing point process data. *J. Roy. Statist. Soc. Ser. C* **34** 138–147.

[29] DIGGLE, P. J. (2003). *Statistical Analysis of Spatial Point Patterns*, 2nd ed. Hodder Arnold, London.

[30] GEORGII, H.-O. (1976). Canonical and grand canonical Gibbs states for continuum systems. *Comm. Math. Phys.* **48** 31–51. MR0411497

[31] GEYER, C. (1999). Likelihood inference for spatial processes. In *Stochastic Geometry* (*Toulouse*, 1996). *Monogr. Statist. Appl. Probab.* **80** 79–140. Chapman & Hall/CRC, Boca Raton, FL. MR1673118

[32] HANISCH, K. H. (1984). Some remarks on estimators of the distribution function of nearest neighbour distance in stationary spatial point processes. *Math. Operationsforsch. Statist. Ser. Statist.* **15** 409–412. MR0756346

[33] HANSEN, B. E. (1996). Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica* **64** 413–430. MR1375740

[34] HEINRICH, L. (1988). Asymptotic behaviour of an empirical nearest-neighbour distance function for stationary Poisson cluster processes. *Math. Nachr.* **136** 131–148. MR0952468

[35] HEINRICH, L. (1988). Asymptotic Gaussianity of some estimators for reduced factorial moment measures and product densities of stationary Poisson cluster processes. *Statistics* **19** 87–106. MR0921628

[36] HOPE, A. C. A. (1968). A simplified Monte Carlo significance test procedure. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **30** 582–598.

[37] HUANG, F. and OGATA, Y. (1999). Improvements of the maximum pseudo-likelihood estimators in various spatial statistical models. *J. Comput. Graph. Statist.* **8** 510–530.

[38] ILLIAN, J., PENTTINEN, A., STOYAN, H. and STOYAN, D. (2008). *Statistical Analysis and Modelling of Spatial Point Patterns*. Wiley, Chichester. MR2384630

[39] JENSEN, J. L. and MØLLER, J. (1991). Pseudolikelihood for exponential family models of spatial point processes. *Ann. Appl. Probab.* **1** 445–461. MR1111528

[40] JOLIVET, E. (1981). Central limit theorem and convergence of empirical processes for stationary point processes. In *Point Processes and Queuing Problems* (*Colloq.*, *Keszthely*, 1978). *Colloq. Math. Soc. János Bolyai* **24** 117–161. North-Holland, Amsterdam. MR0617406

[41] KALLENBERG, O. (1978). On conditional intensities of point processes. *Z. Wahrsch. Verw. Gebiete* **41** 205–220. MR0461654

[42] KALLENBERG, O. (1984). An informal guide to the theory of conditioning in point processes. *Internat. Statist. Rev.* **52** 151–164. MR0967208

[43] KELLY, F. P. and RIPLEY, B. D. (1976). A note on Strauss's model for clustering. *Biometrika* **63** 357–360. MR0431375

[44] KULLDORFF, M. (1999). Spatial scan statistics: Models, calculations, and applications. In *Scan Statistics and Applications* 303–322. Birkhäuser, Boston, MA. MR1697758

[45] KUTOYANTS, Y. A. (1998). *Statistical Inference for Spatial Poisson Processes*. *Lecture Notes in Statist.* **134**. Springer, New York. MR1644620

[46] LAST, G. and PENROSE, M. (2011). Poisson process Fock space representation, chaos expansion and covariance inequalities. *Probab. Theory Related Fields* **150** 663–690.

[47] LAWSON, A. B. (1993). On the analysis of mortality events around a prespecified fixed point. *J. Roy. Statist. Soc. Ser. A* **156** 363–377.

[48] LOTWICK, H. W. and SILVERMAN, B. W. (1982). Methods for analysing spatial processes of several types of points. *J. Roy. Statist. Soc. Ser. B* **44** 406–413. MR0693241

[49] MØLLER, J., SYVERSVEEN, A. R. and WAAGEPETER-SEN, R. P. (1998). Log Gaussian Cox processes. *Scand. J. Statist.* **25** 451–482. MR1650019

[50] MØLLER, J. and WAAGEPETERSEN, R. P. (2004). *Statistical Inference and Simulation for Spatial Point Processes. Monogr. Statist. Appl. Probab.* **100**. Chapman & Hall/CRC, Boca Raton, FL. MR2004226

[51] MØLLER, J. and WAAGEPETERSEN, R. P. (2007). Modern statistics for spatial point processes. *Scand. J. Statist.* **34** 643–684. MR2392447

[52] NGUYEN, X.-X. and ZESSIN, H. (1979). Integral and differential characterizations of the Gibbs process. *Math. Nachr.* **88** 105–115. MR0543396

[53] NUMATA, M. (1961). Forest vegetation in the vicinity of Choshi—Coastal flora and vegetation at Choshi, Chiba prefecture, IV (in Japanese). *Bull. Choshi Mar. Lab.* **3** 28–48.

[54] NUMATA, M. (1964). Forest vegetation, particularly pine stands in the vicinity of Choshi—Flora and vegetation in Choshi, Chiba prefecture, VI (in Japanese). *Bull. Choshi Mar. Lab.* **6** 27–37.

[55] OGATA, Y. and TANEMURA, M. (1981). Estimation of interaction potentials of spatial point patterns through the maximum likelihood procedure. *Ann. Inst. Statist. Math.* **33** 315–338.

[56] OGATA, Y. and TANEMURA, M. (1986). Likelihood estimation of interaction potentials and external fields of inhomogeneous spatial point patterns. In *Pacific Statistical Congress* (I. S. Francis, B. J. F. Manly and F. C. Lam, eds.) 150–154. Elsevier, Amsterdam.

[57] OHSER, J. (1983). On estimators for the reduced second moment measure of point processes. *Math. Operationsforsch. Statist. Ser. Statist.* **14** 63–71. MR0697340

[58] OHSER, J. and STOYAN, D. (1981). On the second-order and orientation analysis of planar stationary point processes. *Biometrical J.* **23** 523–533. MR0635658

[59] PAPANGELOU, F. (1974). The conditional intensity of general point processes and an application to line processes. *Z. Wahrsch. Verw. Gebiete* **28** 207–226. MR0373000

[60] PREGIBON, D. (1982). Score tests in GLIM with applications. In *GLIM* 82: *Proceedings of the International Conference on Generalized Linear Models. Lecture Notes in Statist.* **14**. Springer, New York.

[61] RADHAKRISHNA RAO, C. (1948). Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation. *Proc. Cambridge Philos. Soc.* **44** 50–57. MR0024111

[62] RATHBUN, S. L. and CRESSIE, N. (1994). Asymptotic properties of estimators for the parameters of spatial inhomogeneous Poisson point processes. *Adv. in Appl. Probab.* **26** 122–154. MR1260307

[63] RIPLEY, B. D. (1976). The second-order analysis of stationary point processes. *J. Appl. Probab.* **13** 255–266. MR0402918

[64] RIPLEY, B. D. (1977). Modelling spatial patterns (with discussion). *J. Roy. Statist. Soc. Ser. B* **39** 172–212. MR0488279

[65] RIPLEY, B. D. (1988). *Statistical Inference for Spatial Processes*. Cambridge Univ. Press, Cambridge. MR0971986

[66] RIPLEY, B. D. and KELLY, F. P. (1977). Markov point processes. *J. Lond. Math. Soc.* (2) **15** 188–192. MR0436387

[67] SCHLADITZ, K. and BADDELEY, A. J. (2000). A third order point process characteristic. *Scand. J. Statist.* **27** 657–671. MR1804169

[68] SILVAPULLE, M. J. (1996). A test in the presence of nuisance parameters. *J. Amer. Statist. Assoc.* **91** 1690–1693. MR1439111

[69] STEIN, M. L. (1995). An approach to asymptotic inference for spatial point processes. *Statist. Sinica* **5** 221–234. MR1329294

[70] STILLINGER, D. K., STILLINGER, F. H., TORQUATO, S., TRUSKETT, T. M. and DEBENEDETTI, P. G. (2000). Triangle distribution and equation of state for classical rigid disks. *J. Statist. Phys.* **100** 49–72. MR1791553

[71] STOYAN, D., KENDALL, W. S. and MECKE, J. (1987). *Stochastic Geometry and Its Applications*. Wiley, Chichester. MR0895588

[72] STOYAN, D. and STOYAN, H. (1995). *Fractals, Random Shapes and Point Fields*. Wiley, Chichester.

[73] STRAUSS, D. J. (1975). A model for clustering. *Biometrika* **62** 467–475. MR0383493

[74] VAN LIESHOUT, M. N. M. (2000). *Markov Point Processes and Their Applications*. Imperial College Press, London. MR1789230

[75] WALD, A. (1941). Some examples of asymptotically most powerful tests. *Ann. Math. Statist.* **12** 396–408. MR0006683

[76] WALLER, L., TURNBULL, B., CLARK, L. C. and NASCA, P. (1992). Chronic Disease Surveillance and testing of clustering of disease and exposure: Application to leukaemia incidence and TCE-contaminated dumpsites in upstate New York. *Environmetrics* **3** 281–300.

[77] WANG, P. C. (1985). Adding a variable in generalized linear models. *Technometrics* **27** 273–276. MR0797565

[78] WIDOM, B. and ROWLINSON, J. S. (1970). New model for the study of liquid–vapor phase transitions. *J. Chem. Phys.* **52** 1670–1684.

[79] WU, L. (2000). A new modified logarithmic Sobolev inequality for Poisson point processes and several applications. *Probab. Theory Related Fields* **118** 427–438. MR1800540