Residual analysis for spatial point processes

A. Baddeley[+], R. Turner[*], J. Møller[†] and M. Hazelton[+]

+University of Western Australia
∗ University of New Brunswick
† University of Aalborg

Abstract

We define residuals for point process models fitted to spatial point pattern data, and propose diagnostic plots based on these residuals. The techniques apply to any Gibbs point process model, which may exhibit spatial heterogeneity, interpoint interaction and dependence on spatial covariates.

Our residuals generalise the well-known residuals for point processes in time, used in signal processing and survival analysis. An important difference is that the conditional intensity or hazard rate of the temporal point process must be replaced by the Papangelou conditional intensity λ of the spatial process. Residuals are ascribed to locations in the empty background, as well as to data points of the point pattern. We obtain variance formulae, and study standardised residuals.

There is also an analogy between our spatial residuals and the usual residuals for (non-spatial) generalised linear models. The conditional intensity λ plays the role of the mean response. Elaborating this analogy gives us recommendations for diagnostic plots and model criticism using these residuals. A plot of smoothed residuals against spatial location, or against a spatial covariate, is effective in diagnosing spatial trend or covariate effects. Q–Q plots of the residuals are effective in diagnosing interpoint interaction.

Some existing ad hoc statistics of point patterns (quadrat counts, scan statistic, kernel smoothed intensity, Berman's diagnostic) are recovered as special cases.

Keywords: Berman's diagnostic, Berman-Turner device, deviance residuals, estimating equations, exponential energy marks, generalised linear models, Georgii-Nguyen-Zessin formula, K function, kernel smoothing, Ogata residual, Papangelou conditional intensity, Pearson residuals, pseudolikelihood, Q–Q plots, quadrat counts, residual plots, scan statistic, space-time point processes.
AMS Classification (1991): 62M30, 62J20.

1 Introduction

Recent work on statistical methods for spatial point pattern data has made it easy to fit a wide range of models to real data in applications [8, 7, 50]. Parametric inference, model selection, and goodness-of-fit testing are also feasible with Markov Chain Monte Carlo methods [50, 79].

However, tools for checking or criticising the fitted model are quite limited. There is currently no analogue for spatial point patterns of the comprehensive strategy for model criticism in the linear model, which uses tools such as residual plots and influence diagnostics to identify unusual or influential observations, assess model assumptions one-by-one, and recognise the form of departures from the model [4, 25, 33, 103], [76, Chap. 12]. Analogous techniques for checking a fitted spatial point process model are not available, with a few exceptions [94, Chap. 6], [29, 70, 86, 95, 96]. Indeed it is widespread practice in the statistical analysis of spatial point pattern data to focus primarily on comparing the data with the uniform Poisson process (or "complete spatial randomness", CSR [36]), which is generally the *null* model in applications, rather than the *fitted* model. The paucity of model criticism in spatial statistics is a weakness in applications, especially in areas such as spatial epidemiology [42, 43] where fitted models may invite very close scrutiny.

Accordingly, this paper sets out to develop residuals and residual plots for models fitted to spatial point patterns. Our goal is a coherent strategy for model criticism in spatial point process models, resembling the existing methods for the linear model [4]. This depends crucially on finding the right definition of residuals for a spatial point process model fitted to point pattern data. Additionally we must develop appropriate plots and transformations of the residuals for assessing each component ("assumption") of the fitted model, with a statistical rationale for each plot.

Our definition of residuals is a natural generalisation of the well-known residuals for point processes in time, developed in the 1970's for applications to signal processing [71, 19, 23, 21] and survival analysis [2, 45, 57]. It had been thought that no such generalisation exists for spatial point processes, because of the lack of a natural ordering in two-dimensional space, and that the analysis of spatial point patterns necessitated a quite different approach [26, section 6.1], [91, Introd.]. Nevertheless the generalisation from temporal to spatial point processes is straightforward after one crucial change. The key is to replace the usual conditional intensity of the process (or hazard rate of the lifetime distribution) by the Papangelou conditional intensity of the spatial process. Antecedents of this approach are to be found in the work of Stoyan & Grabarnik [96].

Our residuals may also be viewed as the counterparts of the classical residuals in the generalised linear model [25, 46, 68, 87]. This insight has its origins in a computational trick for fitting point process models in time [71, 22, 20, 21]. Dividing the time axis into small intervals, approximate the point process data by a 0–1 valued time series. The point process likelihood becomes the likelihood of a binomial GLM, which can be fitted using standard software. The conditional intensity of the point process becomes the mean response of the GLM. The residuals from the GLM fit correspond directly to residuals of the point process model. Now for spatial point processes, discretisation of the spatial domain allows us to approximate the *pseudolikelihood* [18] of the point process by the likelihood of a generalised linear model [24, 17, 8, 70]. The Papangelou conditional intensity λ of the spatial point process plays the role of the mean response in the GLM. This suggests how to define residuals, and also how to plot them. It turns out that the spatial point process residuals introduced in this paper correspond directly, under this analogy, to the usual residuals for Poisson loglinear regression. A somewhat related idea was proposed by Lawson [70].

Diagnostic plots and transformations of the residuals are also required, for assessing each component of the fitted point process model (cf. [4]). Components of a model may include spatial trend, dependence on spatial covariates, dependence between points of the pattern, and other features. Our approach is to set up a correspondence between these components of a spatial point process model and the terms in a generalised linear model [8]. Then the well-understood diagnostic plots for assessing each term in a GLM are carried across to spatial point processes.

One slightly unusual feature of our residuals is that they are ascribed both to the data points x_i (of the point pattern dataset $\mathbf{x}$) and also to locations $u \in W$ which are not points of the pattern. This is the natural generalisation of the fact that the residuals for a point process in time are a stochastic process in continuous time. However it differs from the usual situation in most other areas of applied statistics (where residuals are attributed to data points). Also in spatial statistics, Stoyan & Grabarnik [96] and Lawson [70] associate residuals or weights with the data points only.

The paper proceeds as follows. Section 2 presents some example datasets motivating the problem of model validation. Section 3 offers a review and critique of current techniques. Section 4 introduces notation and defines some point process models used in the examples. Section 5 gives some essential background for our definition of residuals, and Sections 6 and 7 defines these residuals for spatial point processes. The means, variances and interpretation of residual values are studied in Section 8. Sections 9–10 develop diagnostic plots for assessing each component of a spatial point process model: Section 9 proposes diagnostic plots for spatial trend, and Section 10 for interpoint interaction. The ability of these diagnostics to separate spatial heterogeneity and interpoint interaction is tested in Section 11. Section 12 discusses computation and implementation of the residuals. Appendix A contains the variance calculations; some other derivations are relegated to Appendices B and C.

2 Examples

Figure 1 depicts the Japanese black pines data of Numata [82], analysed by Ogata and Tanemura [84]. Dots indicate the locations of seedlings and saplings of Japanese black pine (*Pinus Thunbergii*) in a 10×10 metre square sampling region within a natural forest stand. It is of interest to assess evidence for spatial heterogeneity in the abundance of trees (which might be due to spatial variation in soil fertility and other growth conditions), and evidence for positive or negative association or 'interaction' between trees (which might be due to patterns of reproduction, competition for sunlight, etc).

The most popular approach to analysing point patterns [30, 39, 90, 91, 98] involves computing empirical summaries of the data, particularly Ripley's K function [89], essentially a renormalised empirical distribution of the distances between pairs of data points. Empirical techniques are available for measuring spatial heterogeneity, and for measuring interpoint interaction assuming the absence of spatial heterogeneity (or assuming a given form of spatial heterogeneity [5], [39, pp. 143–144]).

However, heterogeneity and clustering are often confounded or unidentifiable [13, 14, 89, 91]. Empirical techniques cannot disentangle these two features and therefore have limited scope.

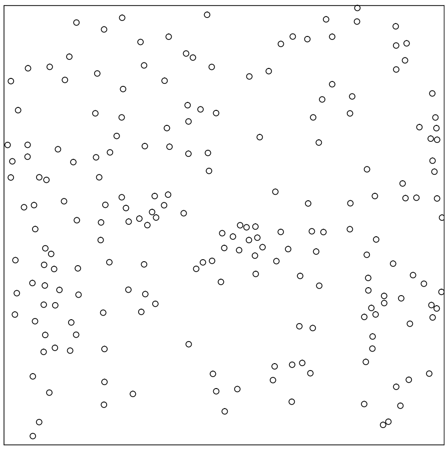

Figure 1: Japanese black pines data. 204 seedlings and saplings in a sampling rectangle 10 m $\times$ 10 m. Source: Numata [82]. Data kindly supplied by Professors Y. Ogata and M. Tanemura.

They cannot be applied easily to Figure 1.

A more powerful approach is to fit statistical models to the point pattern data. The models are point processes, which may be formulated to exhibit spatial heterogeneity and/or interpoint interaction. Formal testing and model selection may then be used to find the best model for the data, and in particular, to decide whether spatial heterogeneity and interpoint interaction are present.

Applied statistical modelling of spatial point pattern data was pioneered by Besag [18] and Ripley [89, 90]. See [39, 79] for surveys. Recently developed software will fit spatial point process models of very general form [8, 7, 6].

Ogata and Tanemura [84] developed the first practical software for fitting spatial point process models involving both spatial heterogeneity and interpoint interaction. For model selection they advocated using the Akaike Information Criterion. Applying this to the Japanese pines data (Figure 1) they found that the best model involved quite strong spatial heterogeneity and moderately strong interpoint interaction.

Textbooks on applied statistics emphasise that it is good practice to validate or check the fitted model by informal methods. Diagnostic plots help us to identify extreme or influential observations, to assess the validity of each component or 'assumption' of the model, and to detect and recognise the form of departures from the model [4, 25, 33, 103], [76, Chap. 12]. Current techniques in spatial statistics include formal tests of goodness-of-fit for spatial point process models [39, pp. 89–90, 100–103, 106, 110–111, 114, 133–143], [102, Chap. 3]. For example, the empirical K function for the point pattern data may be compared with the mean K function obtained from simulations of the fitted model using a Monte Carlo test. However, these tests are rather uninformative. There is currently no equivalent for spatial point processes of the comprehensive set of diagnostic tools available for the linear model. Our goal is to provide such tools.

Another application of point process modelling is to assess dependence on a spatial covariate. Figure 2 depicts a subset of a dataset introduced and analysed by Berman [15]. It represents an intensive geological survey of a 158 $\times$ 35 km region in central Queensland. Dots mark the locations of copper ore deposits. Line segments represent geological features visible on a

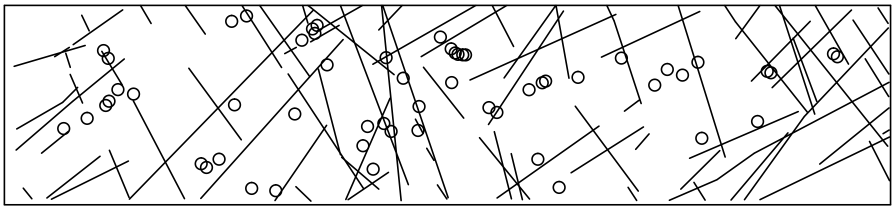

Figure 2: Copper ore deposits (○) and lineaments (—) in a region of Central Queensland. Southern half of original dataset. North at top of frame. Reproduced by kind permission of Dr J Huntington, Dr M Berman and the Royal Statistical Society.

satellite image; they are termed *lineaments* and believed to consist largely of geological faults. It would be of great interest to predict the occurrence of copper deposits from the lineament pattern.

Thus, the lineament pattern constitutes a 'spatial covariate' which might be included in the analysis. The null model (assuming no dependence on the lineaments) postulates that copper deposits are a uniform Poisson process, implying that their locations are independently and uniformly distributed in the study region. Alternative models postulate, for example, that the density of copper deposits depends on distance from the nearest lineament.

Different analyses of these data have been reported [15, 17, 47] with the general consensus that there is no covariate effect. However, what is missing is a critical assessment of the validity of the assumptions behind these analyses. The influence of different parts of the data should also be investigated, since a comparison of the analyses has fortuitously identified some influential observations [47, section 5.3].

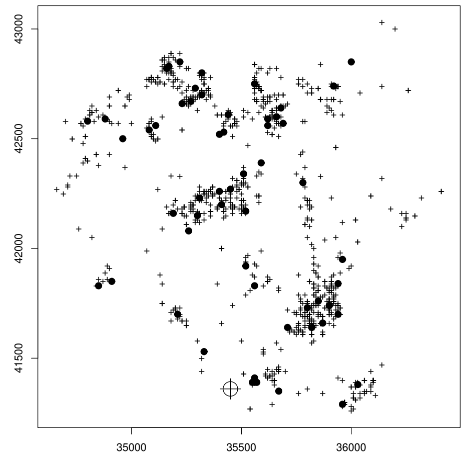

Figure 3: Cases of cancer of the larynx (•) and lung (+) in the Chorley-Ribble region, and the location of a disused industrial incinerator (⊕).

Figure 3 shows a spatial epidemiological dataset presented and analysed by Diggle [38]. There are two point patterns, giving the precise domicile locations of new cases of cancer of the larynx and of the lung, reported in the Chorley-Ribble region of Lancashire during 1974–1983. The aim is to assess evidence for an increase in the incidence of cancer of the larynx in the vicinity of the disused industrial incinerator, whose position is also indicated. The lung cancer cases serve as a surrogate for the spatially varying density of the susceptible population.

The null model (i.e. assuming no incinerator effect) is that laryngeal cancer cases are a point process with intensity proportional to the density of the susceptible population. In alternative models, the frequency of laryngeal cancer also depends on distance from the incinerator. Diggle [38] fits models of both types by maximum likelihood and finds that the model of best fit includes an incinerator effect.

The careful discussion by Diggle [38] adds many caveats on the possible epidemiological interpretations of this formal analysis, and suggests alternative models for investigation. However, further investigation would not necessarily resolve the statistical critique, because of our inability to assess individual model assumptions separately. Model criticism techniques are needed for this.

Our goal, then, is to develop informal techniques which may be used to validate a point process model that has been fitted to spatial point pattern data. The techniques should help us to recognise (the fitted model's misspecification of) spatial heterogeneity, interpoint interaction, and covariate effects in the data.

3 Current methods

The current approach to modelling spatial point pattern data was expounded in Ripley's influential paper [89] and in subsequent texts [30, 36, 39, 102, 90, 91, 98]. First, an exploratory analysis is conducted using simple summary statistics such as the K function [98, Chap. 15]. The data summaries are compared to their expected values for the homogeneous Poisson process. Since the latter is a benchmark for "complete spatial randomness" (CSR) [36], a discrepancy between the empirical and theoretical K functions (say) indicates positive or negative association between points, and suggests appropriate models. Next, models are formulated and fitted. Finally, goodness-of-fit testing or confirmation of the fitted model is performed by simulating realisations from the fitted model and comparing observed and simulated values of a summary statistic [35, 41, 51, 50, 79, 90].

Limitations of the classical approach are well-known [50, 79]. Standard estimators of the summary functions K, F, G etc assume perforce that the point process is stationary [98, Chap. 15]; however, nonstationary spatial trend is common in real datasets. It is common practice to compare real point pattern data to a completely random (Poisson) point process, whether or not this is scientifically relevant. CSR is typically the *null* model; the classical tools are not directly designed for validating a *fitted* point process model. Global goodness-of-fit tests may be performed using a summary statistic such as K, but such tests are uninformative about the type of departure from the fitted model.

The advent of practical Markov Chain Monte Carlo algorithms for simulating and fitting point process models (see [88, 50, 79]) has made it possible to test for the presence of spatial trend (say) within the context of a parametric model. However, exploratory data analysis and model criticism are still problematic.

For exploratory data analysis, many modifications of the standard summary functions have been suggested [97, 98]. Local Indicators Of Spatial Association (LISA) [3, 28, 27, 49] are essentially the individual summands in an estimator of the standard summary function. They summarise the point pattern in a neighbourhood of a given point. LISAs have been used as an exploratory technique for assessing spatial trend and for identifying outliers. The K function was recently generalised to heterogeneous point processes [5] so that the classical analysis can be adapted to situations where the null model is an heterogeneous Poisson process, for example, case-control studies in spatial epidemiology. Diggle et al [34] applied a K function to space-time point processes. Wartenberg [106] canvassed some exploratory methods for detecting outliers, leverage and influence in spatial point patterns.

For model checking, fewer techniques are available [39, pp. 89–90, 100–103, 106, 110–111, 114, 133–143], [102, Chap. 3], One is analogous to the probability integral transformation for random variables. In one dimensional time, a general point process can be transformed into a homogeneous Poisson process by a non-linear transformation of the time scale [26, p. 98], [31, Prop 13.4.III page 550]. This can be used to check a fitted point process model, by subjecting the data to the transformation dictated by the fitted model, and inspecting the transformed data for evidence of departures from the Poisson process [22] as is done commonly in survival analysis [2, 45, 57], [103, chap. 12]. For space-time point processes, transformation of the time axis serves the same purpose, although the transformed points are not uniform in space. However for spatial point processes (in two dimensions or more), analogous "Poisson-ising" transformations exist only for special models [30, pp. 766–770], [77, 29, 80, 86, 95, 105].

Some writers have proposed diagnostics which are somewhat analogous to the residuals from a fitted linear model. Stoyan & Grabarnik [96] defined a diagnostic quantity whose practical application was explored briefly by Särkkä [94, pp. 49–50]. It turns out to be very closely related to our definition of residuals. Lawson [70] defined an unrelated 'deviance residual' for heterogeneous spatial Poisson point processes. For space-time point processes, Diggle et al [34] constructed a residual by comparing their space-time K function with the product of a K function in time and a K function in space, while Ogata [83] formed residuals based on the ratio of a nonparametric estimate of intensity to a parametric estimate of conditional intensity.

In spatial epidemiology, spatially varying relative risk may be estimated nonparametrically or modelled parametrically or semiparametrically; differences between these two estimates yield an estimated residual relative risk, e.g. [39, pp. 133–143].

4 Spatial point processes

This section introduces some notation for point process models and defines the main examples of models used in our experiments.

A planar point pattern dataset

$$\mathbf{x} = \{x_1, \ldots, x_n\}$$

consists of the locations $x_1, \ldots, x_n$ of points observed in a bounded region W of the plane $\mathbb{R}^2$. Note that the number of points $n \geq 0$ is not fixed in advance, while W is fixed and known. There may also be spatial covariates. It is feasible to deal with point patterns in higher dimensions and space-time, and with points carrying additional 'marks', but we stick to two dimensions for simplicity here.

We shall fit parametric models to point pattern datasets $\mathbf{x}$. The model assumes that $\mathbf{x}$ is a realisation of a point process $\mathbf{X}$ which has a probability density $f(\mathbf{x}; \theta)$ with respect to the unit rate Poisson process on W. The model may be very general indeed; it is assumed only to be a finite Gibbs process. Assumptions are stated in Section 6.1. The method used to fit the model may also be very general.

Most models fitted in the literature are *pairwise interaction* point processes [79, 92, 97], whose density has the general form

$$f(\mathbf{x}) = a \left[\prod_{i=1}^{n} b(x_i) \right] \left[\prod_{i<j} c(x_i, x_j) \right] \tag{1}$$

where a is a normalising constant, $b(u), u \in W$ is the 'activity' and $c(u, v),\ u, v \in W$ is the 'interaction'. The activity function b can be used to model spatial variation in the abundance or intensity of points, while the interaction function c can be used to model association between points (usually negative association). Thus, one of our main goals is to be able to validate the model assumptions about the functions b and c for a pairwise interaction model fitted to data.

Three important special cases are

the homogeneous Poisson process with intensity λ points per unit area [26, 65], otherwise known as 'complete spatial randomness'. This arises when both spatial heterogeneity and interpoint interaction are absent, $b(u) \equiv \lambda$ and $c(u, v) \equiv 1$.

the heterogeneous Poisson process with intensity function $\lambda(u)$, $u \in W$. This arises when interpoint interaction is absent, $c(u, v) \equiv 1$. The intensity function $\lambda(u)$ gives the expected number of points per unit area in the vicinity of the location u, and thus models spatial heterogeneity.

the homgeneous Strauss process [62], [39, p. 75] with activity β, interaction parameter γ and interaction range r. This is obtained by setting $b(u) \equiv \beta$ and

$$c(u, v) = \begin{cases} \gamma & \text{if } ||u - v|| \le r \\ 1 & \text{otherwise} \end{cases} \tag{2}$$

This process is well-defined for $0 \le \gamma \le 1$. For $\gamma < 1$ the process exhibits negative association between points.

Figure 4 shows a realisation of the heterogeneous Poisson point process and Figure 19 a realisation of the homogeneous Strauss process.

5 Background to residuals

This section describes some essential background for defining our residuals.

5.1 Residuals in time and space-time

Residuals and diagnostics for point processes in one-dimensional time were developed in the 1970's for applications to signal processing [71, 19, 23, 21] and survival analysis [2, 45, 57].

If N_t denotes the number of arrivals (points of the process) in the time interval $[0, t]$, define the *conditional intensity*

$$\lambda_t = \mathbb{E}[\mathrm{d}N_t \mid N_s, s < t]/\mathrm{d}t, \tag{3}$$

(if it exists), that is, λ_t is the instantaneous arrival rate of the point process given the history of the process prior to time t [61, p 69 ff]. In survival analysis, λ_t is closely related to the hazard rate of the lifetime distribution.

Residuals can be constructed from the fact that the *innovation* or *error process*

$$I_t = N_t - \int_0^t \lambda_s \, \mathrm{d}s \tag{4}$$

is a martingale with $\mathbb{E}[I_t] = 0$ when the model is true [61, Thm 2.14, p 60]. In practice, when the point process model is fitted to data, the parameter estimates would be plugged in to an expression for λ_t to obtain a fitted conditional intensity $\widehat{\lambda}_t$, and we compute the *raw residual process*

$$R_t = N_t - \int_0^t \widehat{\lambda}_s \, \mathrm{d}s. \tag{5}$$

Increments of R_t are analogous to the raw residuals (observed minus fitted values) in a linear model, while increments of I_t would be analogous to the 'errors' (observed minus expected values) which are not observable. The adequacy of the fitted model can be checked by inspecting whether $R_t \approx 0$. Various plots and transformations of R_t are useful diagnostics for a fitted point process model [71, 19, 45, 103].

An alternative diagnostic technique is to transform the time axis to a new scale called operational time,

$$\tau(t) = \int_0^t \lambda_s \, \mathrm{d}s; \tag{6}$$

then the *time-transformed process*

$$J_\tau = N_t \tag{7}$$

is a homogeneous Poisson point process with intensity 1 [26, p48]. Again λ_t would be replaced by the fitted conditional intensity $\widehat{\lambda}_t$ obtained from the fitted model. The adequacy of the fitted model can be checked using diagnostics for the homogeneous Poisson process.

The likelihood of the point process on the interval $[0, t]$ is [61, Thm 2.31 p 71]

$$L_t = \left[\prod_{t_i \le t} \lambda_{t_i} \right] \exp\left(\int_0^t (1 - \lambda_s) \, \mathrm{d}s \right)$$

where $t_1, t_2, \ldots$ denote the successive arrival times of the point process. Hence the score function is

$$U_t^\theta = \sum_{t_i \le t} \frac{\partial}{\partial \theta} \log \lambda_{t_i} - \int_0^t \frac{\partial}{\partial \theta} \lambda_s \, \mathrm{d}s \tag{8}$$

$$= \int_0^t \frac{\partial}{\partial \theta} \log \lambda_s \, \mathrm{d}I_s, \tag{9}$$

closely related to the innovation process. The fact that $\mathbb{E}_\theta[U_t^\theta] = 0$ follows from the martingale property of I_t.

Analogous residuals for space-time point processes were developed for applications to earthquake modelling [83, 104]. Let $N_{B,t}$ denote the number of arrivals of the point process up to time $t > 0$ in spatial region $B \subset \mathbb{R}^2$. Define

$$I_{B,t} = N_{B,t} - \int_0^t \int_B \lambda_{s,u} \, \mathrm{d}s \, \mathrm{d}u \tag{10}$$

where $\lambda_{s,t}$ is the conditional intensity for the space-time point process at location s and time t; this depends on the history of the process, at all spatial locations, up to time t. Again $\mathbb{E}[I_{B,t}] = 0$ and $I_{B,t}$ is a martingale as a function of t; the score function is again closely related to $I_{s,t}$. Residual plots of various kinds have been very useful in earthquake modelling [83].

5.2 Papangelou conditional intensity

For spatial point processes, the lack of a natural ordering in two-dimensional space implies that there is no natural generalisation of the conditional intensity of a temporal or spatio-temporal process given the "past" or "history" up to time t. Instead, the appropriate counterpart for a spatial point process is the *Papangelou* [85] *conditional intensity* λ_s which conditions on the outcome of the process at all spatial locations other than s.

Detailed theory may be consulted in [31, pp. 580–590], [61, sect. 2.6] and [52, 53, 58, 59, 60, 66]. For our purposes the following, very simplified, account is adequate. Let $\mathbf{X}$ be a point process in a bounded region $W \subset \mathbb{R}^2$. Assume $\mathbf{X}$ has a probability density f with respect to the unit rate Poisson process on W, such that $f(\mathbf{x} \cup \{u\}) > 0$ implies $f(\mathbf{x}) > 0$, where u is a location in $\mathbb{R}^2$ and $\mathbf{x}$ is a point pattern (a finite subset of W). For $u \notin \mathbf{x}$ we define

$$\lambda(u, \mathbf{x}) = \frac{f(\mathbf{x} \cup \{u\})}{f(\mathbf{x})} \tag{11}$$

if $f(\mathbf{x}) > 0$, and $\lambda(u, \mathbf{x}) = 0$ otherwise. For $u \in \mathbf{x}$ we define

$$\lambda(u, \mathbf{x}) = \lambda(u, \mathbf{x} \setminus \{u\}) \tag{12}$$

so that (12) holds for all u. Then, loosely speaking, $\lambda(u, \mathbf{x}) \, \mathrm{d}u$ is the conditional probability that $\mathbf{X}$ has a point in an infinitesimal region containing u and of area $\mathrm{d}u$ given that the rest of the point process coincides with $\mathbf{x}$.

For example consider the three models mentioned in section 4. The homogeneous Poisson process of intensity $\lambda > 0$ has conditional intensity $\lambda(u, \mathbf{x}) \equiv \lambda$. The heterogeneous Poisson process with intensity function $\lambda(u)$, $u \in W$ has conditional intensity $\lambda(u, \mathbf{x}) = \lambda(u)$. The general pairwise interaction process with activity function b and pairwise interaction function c has conditional intensity

$$\lambda(u, \mathbf{x}) = b(u) \prod_{i=1}^{n} c(u, x_i), \quad u \notin \mathbf{x}. \tag{13}$$

In general $\lambda(\cdot, \mathbf{x})$ is discontinuous at the data points x_i because of (12).

The Papangelou conditional intensity λ uniquely determines the point process density f and vice versa (on finite domains only). For Markov point processes it is convenient to model

$\mathbf{X}$ by λ rather than by f, since λ plays the same role as the local characteristics do for Markov random fields when specifying local Markov properties [93]. Also most simulation procedures are specified in terms of λ. See e.g. [79].

It can be verified directly in this simple setting that

$$\mathbb{E}\left[\sum h(x_i, \mathbf{X} \setminus \{x_i\})\right] = \mathbb{E}\left[\int_W h(u, \mathbf{X})\lambda(u, \mathbf{X})\, \mathrm{d}u\right] \tag{14}$$

where $h(u, \mathbf{x})$ is any nonnegative integrable function, and the sum on the left is over all points x_i in $\mathbf{X}$. Equation (14) is called the *(Georgii-)Nguyen-Zessin formula* [48, 81].

In the more rigorous treatments (cf. [31], Definition 14.2.IV, page 580 and Proposition 14.3.III, page 590 or [61, sect. 2.6]) it is the Georgii-Nguyen-Zessin formula which defines the Papangelou conditional intensity, as the unique stochastic process λ which satisfies (14) for all integrands h which are "exvisible", a generalisation of property (12). Van der Hoeven [101] showed that this exactly parallels the theory of the conditional intensity for stochastic processes in time, where the integrand h must be "predictable" (cf. [61, sect. 2.6]).

The Georgii-Nguyen-Zessin formula (14) has often been used to derive moment equations for parameter estimation [44, 99, 100] and is intimately associated with the method of maximum pseudolikelihood [11, 40, 79]. In the present paper (14) becomes the basic identity for deriving diagnostics and residuals.

5.3 Stoyan–Grabarnik diagnostic

Stoyan & Grabarnik [96] were the first to exploit the Georgii–Nguyen–Zessin identity (14) for model checking. Assume $\lambda(u, \mathbf{x}) > 0$ for all u. Take the integrand in (14) to be

$$h(u, \mathbf{x}) = \frac{\mathbf{1}\{u \in B\}}{\lambda(u, \mathbf{x})} \tag{15}$$

where $\mathbf{1}\{\ldots\}$ is the indicator and $B \subseteq W$ is a given measurable subset. Then (14) becomes

$$\mathbb{E}\left[\sum_{x_i \in \mathbf{X} \cap B} \frac{1}{\lambda(x_i, \mathbf{X})}\right] = \mathbb{E}\left[\int_B 1\, \mathrm{d}u\right] = |B|$$

where $|B|$ denotes the area of B. This states that, if each point x_i of the dataset $\mathbf{x}$ is weighted by the reciprocal of its Papangelou conditional intensity

$$m_i = 1/\lambda(x_i, \mathbf{x})$$

(called the "exponential energy mark" by Stoyan and Grabarnik) then the total weight for all data points x_i that fall in a nominated region B,

$$M(B) = \sum_{x_i \in \mathbf{x} \cap B} m_i,$$

has expectation

$$\mathbb{E}M(B) = |B| \tag{16}$$

if the model is true. The variance of $M(B)$ was also computed in [96] for the case of a "stationary" pairwise interaction process (i.e. where the function b is constant and $c(u,v) = c(u-v)$).

While other functions h could be substituted in (14), the judicious choice (15) made by Stoyan & Grabarnik is effectively the only one in which the integral in (14) is trivial.

The conditional intensity $\lambda(u,\mathbf{x})$ depends on parameters of the model and would, in practice, be replaced by a plug-in estimate $\widehat{\lambda}(u,\mathbf{x}) = \lambda_{\widehat{\theta}}(u,\mathbf{x})$ obtained by substituting estimates of the model parameters. Stoyan & Grabarnik proposed that the fitted weights $\widehat{m}_i = 1/\widehat{\lambda}(x_i,\mathbf{x})$ associated with the data points x_i, and their sum

$$\widehat{M}(B) = \sum_{x_i \in \mathbf{x} \cap B} \widehat{m}_i,$$

could be used for exploratory data analysis and goodness-of-fit testing, in that (a) individual points or groups of points x_i with extreme values $\widehat{m}_i$ may indicate 'outliers'; (b) regions B with extreme values of $\widehat{M}(B) - |B|$ may indicate regions of irregularity; (c) the global departure $\widehat{M}(W) - |W|$ may be used to test goodness-of-fit or to test convergence of MCMC samplers. Applications were not presented in [96], and the proposals of Stoyan and Grabarnik have apparently not been pursued further, except by Särkkä [94, pp. 49–50] who experimented with proposal (a).

If $\lambda(u,\mathbf{x})$ may take zero values, a few difficulties arise with this technique. The sum $M(B)$ is still well-defined, since there is zero probability of obtaining a realisation in which $\lambda(x_i,\mathbf{X}) = 0$ for some $x_i \in \mathbf{X}$. However, equation (16) does not hold. This complicates the analysis of some processes of interest, such as hard core processes [97].

5.4 Connection to generalised linear models

Here we explain a connection between point process models and generalised linear models, which was originally developed in order to fit point process models to data, but which also provides some statistical insights.

For point processes in time, Lewis [71], Brillinger [22, 20, 21] and others implemented maximum likelihood by dividing the time axis into short intervals, approximating the point process by a 0–1 valued time series, and thereby approximating the point process likelihood by a simple binomial likelihood, which was then maximised using GLIM or similar packages. See also [72, 75, 73, 74].

For spatial point processes, Berman & Turner [17] developed a similar approach for maximising the likelihood of Poisson processes. The Berman-Turner method was extended to general Gibbs processes, for which it maximises the *pseudolikelihood*, by Clyde & Strauss [24], Lawson [69] and Baddeley & Turner [8, 9]. The likelihood or pseudolikelihood of the spatial point process model is approximated by the likelihood of a (non-spatial) loglinear Poisson regression.

Besag [18] defined the pseudolikelihood function by

$$\log \mathrm{PL}(\theta,\mathbf{x}) = \sum_{x_i \in \mathbf{x}} \log \lambda_\theta(x_i,\mathbf{x}) - \int_W \lambda_\theta(u,\mathbf{x})\, \mathrm{d}u \tag{17}$$

and proposed that θ be estimated by maximising the pseudolikelihood instead of the likelihood. For spatial Poisson point processes, the pseudolikelihood is proportional to the likelihood, while for other processes pseudolikelihood is a cheap approximation.

The Berman-Turner device relies on a special quadrature approximation of the integral in the definition of the pseudolikelihood (17). The observed point pattern $\mathbf{x}$ is first augmented by numerous "dummy" points to form a set of quadrature points u_j, $j = 1, \dots, M$. This leads to the approximation

$$\log \mathrm{PL}(\theta, \mathbf{x}) \approx \sum_j w_j \left(y_j \log \lambda_\theta(u_j, \mathbf{x}) - \lambda_\theta(u_j, \mathbf{x}) \right) \tag{18}$$

where $w_j > 0$ is a weight associated with the quadrature point u_j, and $y_j = z_j/w_j$ is the "pseudo-response" at point u_j, where z_j is the indicator equal to 1 if u_j is a data point and 0 for a dummy point. The "quadrature scheme" is simply a set of points u_j and weights w_j chosen so that (18) is a good numerical approximation to (17). For example, w_j may be the area of the Dirichlet tile associated with u_j [8].

Berman and Turner recognised that (18) is formally equivalent to the weighted loglikelihood of a Poisson loglinear regression, with responses y_j, log link function, linear predictor $\eta_j = \log \lambda_\theta(u_j, \mathbf{x})$ and weights w_j. Thus we may use standard software for generalised linear models to maximise the approximate pseudolikelihood (18). For details see [8].

Care should be taken in the statistical interpretation of this approximation, since a data-dependent quadrature scheme is used, and the formal equivalence between (18) and the likelihood of Poisson loglinear regression holds only as a function of θ for fixed $\mathbf{x}$. However, this connection is heuristically useful, is important for computation, and is equivalent to maximum likelihood if the model is an heterogeneous Poisson process.

The analogy between point processes and generalised linear models may be extended further. In the Berman-Turner device, the conditional intensity $\lambda(u; \mathbf{x})$ of the point process corresponds to the mean response of the associated Poisson GLM. The presence of interpoint interaction in the point process corresponds to dependence between the residuals in a GLM. The interpoint interaction terms in the point process density correspond roughly to the joint density of residuals in the GLM. This analogy will guide our approach to diagnostic plots for point processes.

6 Residuals for spatial point processes

This section defines residuals for spatial point processes.

6.1 Assumptions

Notation is as in section 4. We impose on $\mathbf{X}$ the minimal technical assumption that the Papangelou conditional intensity $\lambda(u, \mathbf{X})$ exists [32, 52, 53, 59, 60, 66]. Equivalently for a point process in a bounded region, the probability density f exists and $f(\mathbf{x} \cup \{u\}) > 0$ implies $f(\mathbf{x}) > 0$.

In practice there is an additional constraint, that $\lambda(u, \mathbf{x})$ be computationally tractable. Depending on computational resources, this may exclude most cluster processes and most Cox processes (mixtures of Poisson processes driven by a random intensity function). Although expressions are available for the conditional intensity in these cases [12, 78], the expressions involve expectation over the unobserved cluster centres and over the unobserved random intensity function, respectively. We do not regard the exclusion of these processes as a severe omission, and indeed it appears to be analogous to the fact that residuals are more difficult to compute in a missing data problem.

6.2 Innovation process

For a spatial point process $\mathbf{X}$ governed by a parameter θ we define the *innovation process* by

$$I_\theta(B) = N(B) - \int_B \lambda_\theta(u, \mathbf{X}) \,\mathrm{d}u \tag{19}$$

for any spatial domain B, where $N(B) = n(\mathbf{X} \cap B)$ denotes the number of points of the point process falling in B, and $\lambda_\theta(u, \mathbf{x})$ is the Papangelou conditional intensity of $\mathbf{X}$ under θ.

This definition is closely analogous to (4) and (10), (i.e. the martingale obtained by subtracting the compensator from the counting process) except for the use of the Papangelou conditional intensity in place of the conditional intensity given the past history. Whereas a point process in time has an innovation process I_t indexed by continuous time t, a spatial point process has an innovation process $I_\theta(B)$ indexed by subsets $B \subset \mathbb{R}^2$. This is a signed random measure, consisting of a mass $+1$ at each point x_i of the spatial point process, and a negative density $-\lambda(u, \mathbf{x}) \,\mathrm{d}u$ at all other spatial locations u.

If the model is correct then

$$\mathbb{E}_\theta[I_\theta(B)] = 0 \tag{20}$$

since, when the integrand in the Georgii-Nguyen-Zessin formula (14) is taken to be $h(u, \mathbf{x}) = \mathbf{1}\{u \in B\}$, we obtain

$$\mathbb{E}_\theta[N(B)] = \mathbb{E}_\theta\left[\int_B \lambda_\theta(u, \mathbf{x}) \,\mathrm{d}u\right].$$

Increments of the innovation process I_θ are analogous to errors in the linear model.

Indeed the Georgii-Nguyen-Zessin formula (14) can be restated as

$$\mathbb{E}_\theta\left[\int_B h(u, \mathbf{X} \setminus \{u\}) \,\mathrm{d}I_\theta(u)\right] = 0 \tag{21}$$

corresponding to the martingale properties of the innovations for temporal and spatiotemporal point processes.

The direct connection between the innovations and the score is lost; instead, I_θ is closely related to the *pseudoscore* (derivative of the log pseudolikelihood) of the point process. The pseudoscore satisfies

$$\begin{aligned} \frac{\partial}{\partial\theta} \log \mathrm{PL}(\theta, \mathbf{x}) &= \sum_i \frac{\partial}{\partial\theta} \log \lambda_\theta(x_i, \mathbf{x}) - \int_W \frac{\partial}{\partial\theta} \lambda_\theta(u, \mathbf{x}) \,\mathrm{d}u \\ &= \int_W \frac{\partial}{\partial\theta} \log \lambda_\theta(u, \mathbf{x}) \,\mathrm{d}I_\theta(u) \end{aligned} \tag{22}$$

assuming $\lambda_\theta(u, \mathbf{x})$ is differentiable in θ uniformly in u for each $\mathbf{x}$. The pseudoscore has mean zero, by virtue of the Georgii-Nguyen-Zessin formula (14) applied to $h(u, \mathbf{x}) = \frac{\partial}{\partial\theta} \log \lambda_\theta(u, \mathbf{x})$. This is just the statement that the maximum pseudolikelihood estimating equations are unbiased estimating equations.

6.3 Raw residuals

The conditional intensity $\lambda(u, \mathbf{x})$ depends on parameters of the model and would, in practice, be replaced by a plug-in estimate $\widehat{\lambda}(u, \mathbf{x}) = \lambda_{\widehat{\theta}}(u, \mathbf{x})$ obtained by substituting estimates of the model parameters. Then we define the *raw residual process*

$$R_{\widehat{\theta}}(B) = N(B) - \int_B \lambda_{\widehat{\theta}}(u, \mathbf{X})\, \mathrm{d}u \tag{23}$$

Increments in $R_{\widehat{\theta}}$ correspond to the raw residuals in the linear model.

For a general parameter estimator $\widehat{\theta} = \widehat{\theta}(\mathbf{X})$ the expected residual is, by (14),

$$\mathbb{E}[R_{\widehat{\theta}(\mathbf{X})}(B)] = \mathbb{E} \int_B [\lambda(u, \mathbf{X}) - \lambda_{\widehat{\theta}}(u, \mathbf{X})]\, \mathrm{d}u \tag{24}$$

where $\mathbb{E}$ is expectation with respect to the true distribution of the point process $\mathbf{X}$, and $\lambda(u, \mathbf{X})$ is its true conditional intensity.

Thus, the raw residuals have mean zero if the conditional intensity of the fitted model is an unbiased predictor of the true conditional intensity. The raw residuals are approximately zero if the conditional intensity is estimated accurately.

The innovation process I and the raw residual process $R_{\widehat{\theta}}$ are slightly unfamiliar objects. They are (random) signed measures with atoms of mass 1 at the data points, and a negative density $-\lambda(u, \mathbf{X})$ or $-\lambda_{\widehat{\theta}}(u, \mathbf{X})$ at other locations u. This is the natural generalisation of the fact that, for a point process in one dimensional time, the innovation and residual processes are random processes in continuous time, with discrete jumps of height 1 at the arrival times of the point process, and negative derivative $-\lambda_s$ at all other times.

Whereas most previous writers [70, 94, 96] have defined diagnostic values for the data points x_i only, our residuals are also ascribed to locations $u \in W$ which are not points of the pattern. This is related to an important methodological issue for point processes. In a point pattern dataset, the observed information does not consist solely of the locations of the observed points of the pattern. The *absence* of points at other locations is also informative.

Methods of visualising these residuals are proposed in Sections 9–10.

7 Scaled residuals

7.1 Scaling

In statistical modelling it is often useful to scale the raw residuals, for example to compute standardised residuals. The analogue in our context is to scale the increments of the residual measure $R_{\widehat{\theta}}$. This is done simply by making an alternative choice of the integrand h in the Georgii-Nguyen-Zessin identity (14). Let $h(u, \mathbf{x})$ be any nonnegative function satisfying

$$h(u, \mathbf{x}) = h(u, \mathbf{x} \setminus \{u\}) \tag{25}$$

for all $u \in W$ and all finite point patterns $\mathbf{x} \subset W$. Define the h-weighted innovations

$$I(B, h, \lambda) = \sum_{x_i \in B} h(x_i, \mathbf{x} \setminus \{x_i\}) - \int_B h(u, \mathbf{x})\lambda(u, \mathbf{x})\, \mathrm{d}u \tag{26}$$

for the spatial point process with Papangelou conditional intensity λ. We may interpret $\Delta I(x_i) = h(x_i, \mathbf{x} \setminus \{x_i\})$ as the innovation increment ('error') attached to a data point x_i, and $\mathrm{d}I(u) = -h(u, \mathbf{x})\lambda(u, \mathbf{x})\,\mathrm{d}u$ as the innovation increment attached to a non-data location $u \in W$. The innovations have mean zero, since we may rewrite (14) as

$$\mathbb{E}_\theta \left[\sum_{x_i \in B} h(x_i, \mathbf{X} \setminus \{x_i\}) - \int_B h(u, \mathbf{X}) \lambda_\theta(u, \mathbf{X})\,\mathrm{d}u \right] = 0 \tag{27}$$

After we fit a parametric model with parameter θ to the dataset $\mathbf{x}$, estimating the parameter by $\widehat{\theta} = \widehat{\theta}(\mathbf{x})$, we may then define the *residual* measure based on any function h by

$$\begin{aligned} R(B, h, \widehat{\theta}) &= I(B, h, \lambda_{\widehat{\theta}}) \\ &= \sum_{x_i \in B} h(x_i, \mathbf{x} \setminus \{x_i\}) - \int_B h(u, \mathbf{x}) \lambda_{\widehat{\theta}}(u, \mathbf{x})\,\mathrm{d}u. \end{aligned} \tag{28}$$

Indeed the integrand h may even depend on θ. Suppose $h_\theta(u, \mathbf{x})$ satisfies (25) for all θ. Define the *residual* measure based on h

$$\begin{aligned} R(B, \widehat{h}, \widehat{\theta}) &= I(B, \widehat{h}, \lambda_{\widehat{\theta}}) \\ &= \sum_{x_i \in B} h_{\widehat{\theta}}(x_i, \mathbf{x} \setminus \{x_i\}) - \int_B h_{\widehat{\theta}}(u, \mathbf{x}) \lambda_{\widehat{\theta}}(u, \mathbf{x})\,\mathrm{d}u \end{aligned} \tag{29}$$

where $\widehat{h}$ is an abbreviation for $h_{\widehat{\theta}}$.

Any choice of bounded integrand h in (27)–(29) yields an innovation measure with mean zero:

$$\mathbb{E}\left[I(B, h, \lambda)\right] = 0,$$

and a residual measure for which we *hope* that the mean is approximately zero when the model is true,

$$\mathbb{E}\left[R(B, \widehat{h}, \widehat{\theta})\right] \approx 0.$$

Various choices of h are discussed in sections 7.2–7.5. Statistical insight is needed to guide the choice of h and the further use of the residuals.

7.2 Raw residuals

The choice $h(u, \mathbf{x}) \equiv 1$ in (26) and (29) yields the raw innovations (19) and raw residuals (23). The raw innovations have mean zero, while the raw residuals have mean zero if the fitted conditional intensity is an unbiased estimator, as shown in equation (24).

7.3 Inverse-lambda residuals

First suppose $\lambda(u, \mathbf{x}) > 0$ always. The choice $h(u, \mathbf{x}) = 1/\lambda(u, \mathbf{x})$ in (26) yields the innovation measure

$$I(B, 1/\lambda, \lambda) = \sum_{x_i \in B} \frac{1}{\lambda(x_i, \mathbf{x})} - |B| \tag{30}$$

corresponding to the Stoyan-Grabarnik exponential energy weights, adjusted to have mean approximately zero by assigning a residual density of $-1\, \mathrm{d}u$ at the non-data locations u.

The corresponding choice $h_\theta(u, \mathbf{x}) = 1/\lambda_\theta(u, \mathbf{x})$ in (29) yields the residual measure

$$R(B, 1/\lambda_{\widehat{\theta}}, \widehat{\theta}) = \sum_{x_i \in B} \frac{1}{\lambda_{\widehat{\theta}}(x_i, \mathbf{x})} - |B|. \tag{31}$$

In order that the residuals be well defined, the estimator $\widehat{\theta}$ must have the property that $\lambda_{\widehat{\theta}}(x_i, \mathbf{x}) > 0$ for all $x_i \in \mathbf{x}$.

Next suppose $\lambda(u, \mathbf{x})$ can also take zero values. The innovations (30) are well-defined, but do not have mean zero. This problem can be avoided by redefining

$$h(u, \mathbf{x}) = \begin{cases} \frac{1}{\lambda(u,\mathbf{x})} & \text{if } \lambda(u, \mathbf{x}) > 0 \\ 0 & \text{if } \lambda(u, \mathbf{x}) = 0 \end{cases}$$

which yields

$$I(B, 1/\lambda^+, \lambda) = \sum_{x_i \in B} \frac{1}{\lambda(x_i, \mathbf{x})} - \int_B \mathbf{1}\{\lambda(u, \mathbf{x}) > 0\}\, \mathrm{d}u \tag{32}$$

and the residual measure

$$R(B, 1/\widehat{\lambda}^+, \widehat{\theta}) = \sum_{x_i \in B} \frac{1}{\lambda_{\widehat{\theta}}(x_i, \mathbf{x})} - \int_B \mathbf{1}\{\lambda_{\widehat{\theta}}(u, \mathbf{x}) > 0\}\, \mathrm{d}u. \tag{33}$$

On the right hand side of (32) and (33), the area of B has been replaced by the area of the random subset of B where the conditional intensity is positive.The innovations (32) have mean zero. We shall call (33) the *inverse-λ residuals*.

7.4 Pearson residuals

By analogy with the Pearson residuals for Poisson loglinear regression, we consider the integrand

$$h(u, \mathbf{x}) = 1/\sqrt{\lambda(u, \mathbf{x})}$$

in (27), which yields the Pearson innovation measure

$$I(B, 1/\sqrt{\lambda}, \lambda) = \sum_{x_i \in \mathbf{X} \cap B} \frac{1}{\sqrt{\lambda(x_i, \mathbf{x})}} - \int_B \sqrt{\lambda(u, \mathbf{x})}\, \mathrm{d}u \tag{34}$$

and the corresponding choice $h_\theta(u, \mathbf{x}) = 1/\sqrt{\lambda_\theta(u, \mathbf{x})}$ which yields the Pearson residual measure

$$R(B, 1/\sqrt{\widehat{\lambda}}, \lambda) = \sum_{x_i \in \mathbf{X} \cap B} \frac{1}{\sqrt{\lambda_{\widehat{\theta}}(x_i, \mathbf{x})}} - \int_B \sqrt{\lambda_{\widehat{\theta}}(u, \mathbf{x})}\, \mathrm{d}u. \tag{35}$$

Data points x_i have residual mass $1/\sqrt{\widehat{\lambda}(x_i, \mathbf{x})}$, while background points u have residual density $-\sqrt{\widehat{\lambda}(u, \mathbf{x})}\, \mathrm{d}u$.

Again, the estimator $\widehat{\theta}$ must satisfy $\lambda_{\widehat{\theta}(\mathbf{x})}(x_i, \mathbf{x}) > 0$ for all $x_i \in \mathbf{x}$ in order that the Pearson residuals be well-defined.

7.5 Pseudo-score residuals

Suppose the model to be fitted has conditional intensity $\lambda_\theta(u, \mathbf{x})$ where θ is a k-dimensional vector. Take the integrand h in (27) to be

$$h(u, \mathbf{x}) = \frac{\partial}{\partial\theta} \log \lambda_\theta(u, \mathbf{x}) = \frac{\frac{\partial}{\partial\theta}\lambda_\theta(u, \mathbf{x})}{\lambda_\theta(u, \mathbf{x})}$$

assuming $\lambda_\theta(u, \mathbf{x})$ is differentiable with respect to θ uniformly in u for each fixed $\mathbf{x}$. The derivative is a k-dimensional vector. This yields

$$I(B, \frac{\partial}{\partial\theta} \log \lambda_\theta, \lambda) = \sum_{x_i \in \mathbf{X} \cap B} \frac{\partial}{\partial\theta} \log \lambda_\theta(x_i, \mathbf{x}) - \int_B \frac{\partial}{\partial\theta} \lambda_\theta(u, \mathbf{x}) \, \mathrm{d}u \tag{36}$$

$$R(B, \frac{\partial}{\partial\theta} \log \widehat{\lambda}, \widehat{\lambda}) = \sum_{x_i \in \mathbf{X} \cap B} \frac{\partial}{\partial\theta} \log \lambda_{\widehat{\theta}}(x_i, \mathbf{x}) - \int_B \frac{\partial}{\partial\theta} \lambda_{\widehat{\theta}}(u, \mathbf{x}) \, \mathrm{d}u \tag{37}$$

For any point process $\mathbf{X}$ we have

$$\mathbb{E}[I(B, \frac{\partial}{\partial\theta} \log \lambda_\theta, \lambda)] = 0$$

where λ is the true conditional intensity of $\mathbf{X}$. The residual sum $R(W, \frac{\partial}{\partial\theta} \log \widehat{\lambda}, \widehat{\lambda})$ is the pseudoscore (22) which is equated to zero when we apply the method of maximum pseudolikelihood [11, 56]. Increments of these residuals are contributions to the pseudoscore, thus corresponding to score residuals in generalised linear modelling.

In the case where the conditional intensity takes the 'loglinear' form

$$\lambda_\theta(u, \mathbf{x}) = \exp(\theta \, S(u, \mathbf{x})) \tag{38}$$

where $S(u, \mathbf{x})$ is k-dimensional vector-valued, we have

$$\begin{aligned} h(u, \mathbf{x}) &= S(u, \mathbf{x}) \\ \frac{\partial}{\partial\theta} \lambda_\theta(u, \mathbf{x}) &= S(u, \mathbf{x}) \lambda_\theta(u, \mathbf{x}) \\ R(B, \frac{\partial}{\partial\theta} \log \widehat{\lambda}, \widehat{\lambda}) &= \sum_{x_i \in \mathbf{X} \cap B} S(x_i, \mathbf{x}) - \int_B S(u, \mathbf{x}) \lambda_{\widehat{\theta}}(u, \mathbf{x}) \, \mathrm{d}u. \end{aligned} \tag{39}$$

8 Interpretation of residuals

8.1 Mean residual

Suppose a point process model with parameter θ is fitted to a realisation from a point process $\mathbf{X}$ (*not necessarily conforming to the model*) using a parameter estimator $\widehat{\theta} = \widehat{\theta}(\mathbf{x})$. Then the residuals (29) have expectation

$$\mathbb{E}\left[R(B, \widehat{h}, \widehat{\theta})\right] = \int_B \mathbb{E}\left[h_{\widehat{\theta}(\mathbf{X} \cup \{u\})}(u, \mathbf{X}) \lambda(u, \mathbf{X}) - h_{\widehat{\theta}(\mathbf{X})}(u, \mathbf{X}) \lambda_{\widehat{\theta}(\mathbf{X})}(u, \mathbf{X})\right] \, \mathrm{d}u \tag{40}$$

where $\mathbb{E}$ is the expectation for the true process $\mathbf{X}$ and $\lambda(u, \mathbf{X})$ is its true conditional intensity. This is another application of the Georgii-Nguyen-Zessin formula (14).

This yields for the raw, inverse-lambda, and Pearson residuals respectively

$$\mathbb{E}[R(B, 1, \widehat{\theta})] = \mathbb{E} \int_B [\lambda(u, \mathbf{X}) - \lambda_{\widehat{\theta}}(u, \mathbf{X})] \, \mathrm{d}u \tag{41}$$

$$\mathbb{E}\left[R(B, \widehat{\lambda}^+, \widehat{\theta})\right] = \int_B \mathbb{E}\left[\frac{\lambda(u, \mathbf{X})}{\lambda_{\widehat{\theta}(\mathbf{X} \cup \{u\})}(u, \mathbf{X})} - \mathbf{1}\{\lambda_{\widehat{\theta}(\mathbf{X})}(u, \mathbf{X}) > 0\}\right] \mathrm{d}u \tag{42}$$

$$\mathbb{E}\left[R(B, 1/\sqrt{\widehat{\lambda}}, \widehat{\theta})\right] = \int_B \mathbb{E}\left[\frac{\lambda(u, \mathbf{X})}{\sqrt{\lambda_{\widehat{\theta}(\mathbf{X} \cup \{u\})}(u, \mathbf{X})}} - \sqrt{\lambda_{\widehat{\theta}(\mathbf{X})}(u, \mathbf{X})}\right] \mathrm{d}u \tag{43}$$

where $\mathbb{E}$ is expectation with respect to the true distribution of the point process $\mathbf{X}$, and $\lambda(u, \mathbf{X})$ is its true conditional intensity. These follow from (40), provided that $\lambda_{\widehat{\theta}(\mathbf{X} \cup \{u\})}(u, \mathbf{X}) = 0$ implies $\lambda(u, \mathbf{X}) = 0$ and that we interpret $0/0 = 0$.

A diagnostic interpretation of (41) is that if we fit a point process model to a realisation from a different point process, the mean of the raw residuals is equal to the expected difference in conditional intensity between the true and fitted models. That is, a nonzero mean in the raw residuals corresponds to a bias in estimation of the (conditional) intensity. Similarly (42) shows the inverse-lambda residuals have mean zero if the fitted conditional intensity has zero relative bias as a predictor of the true conditional intensity. Deviations of the inverse-lambda residuals from zero may be interpreted as estimates of the relative bias.

8.2 Uniform Poisson case

Assume the model to be fitted is the uniform Poisson process with intensity θ, and we use the maximum likelihood estimator $\widehat{\theta} = N(W)/|W|$. Then the raw, inverse-lambda and Pearson residual measures are respectively

$$R(B) = R(B, 1, \widehat{\theta}) = N(B) - \widehat{\theta}\,|B| = N(B) - N(W)\frac{|B|}{|W|} \tag{44}$$

$$R(B, 1/\widehat{\theta}, \widehat{\theta}) = \widehat{M}(B) - |B| = |W|\frac{N(B)}{N(W)} - |B| \tag{45}$$

$$R(B, 1/\sqrt{\widehat{\theta}}, \widehat{\theta}) = N(B)\sqrt{\frac{|W|}{N(W)}} - \sqrt{N(W)|W|} \tag{46}$$

when $N(W) > 0$, and zero otherwise.

It can be verified directly that the residuals from a uniform Poisson model (44)–(46) have mean zero if the model is true.

Notice also that when $B = W$ is the entire sampling window, we get

$$R(W) = R(B, 1/\widehat{\theta}, \widehat{\theta}) = R(B, 1/\sqrt{\widehat{\theta}}, \widehat{\theta}) = 0$$

identically, *for any point pattern data*, when the fitted model is the uniform Poisson process. This is analogous to the fact that the raw residuals in simple linear regression sum to zero.

8.3 Variance of residuals

General formulae for the variances of the innovations and of the residuals, that is, for

$$\mathsf{var}\left[I(B,h,\lambda)\right] \quad \text{and} \quad \mathsf{var}\left[R(B,\widehat{h},\widehat{\theta})\right],$$

can be obtained from the results in Appendix A. The variance of $I(B,h,\lambda)$ for general h is (84). The variance of $R(B,\widehat{h},\widehat{\theta})$ in the general case can be obtained from Lemma 2.

For the variance of the innovations, these formulae are manageable. The variance of the raw innovations is given in (86) and that of the inverse-lambda innovations $I(B,1/\lambda,\theta)$ in (81). In the special case of an heterogeneous Poisson process with intensity $\lambda(u)$, these reduce to

$$\mathsf{var}\left[I(B,1,\lambda)\right] = \int_B \lambda(u)\,\mathrm{d}u \tag{47}$$

$$\mathsf{var}\left[I(B,1/\lambda,\lambda)\right] = \int_B \frac{1}{\lambda(u)}\,\mathrm{d}u \tag{48}$$

$$\mathsf{var}\left[I(B,1/\sqrt{\lambda},\theta)\right] = |B|. \tag{49}$$

The first equation is of course the variance and mean of $N(B)$. The last equation is analogous to the fact that the classical Pearson residuals are standardised, ignoring the effect of parameter estimation.

For the variance of the residuals, the formulae are cumbersome, involving characteristics of both the fitted model and the underlying point process. In the very special case where a uniform Poisson process is fitted to a realisation of a uniform Poisson process with intensity θ, the residual variances are

$$\mathsf{var}\left[R(B,1,\widehat{\theta})\right] = \theta\,|B|(1-\frac{|B|}{|W|}) \tag{50}$$

$$\mathsf{var}\left[R(B,1/\widehat{\lambda}^{+},\widehat{\theta})\right] = \frac{|B|(|W|-|B|)}{\mathbb{E}[N(W)\mathbf{1}\{N(W)>0\}]} \tag{51}$$

$$\mathsf{var}\left[R(B,1/\sqrt{\widehat{\lambda}},\widehat{\lambda})\right] = |B|(1-\frac{|B|}{|W|}). \tag{52}$$

Note that the residual variances are smaller than the corresponding innovation variances

$$\mathsf{var}\left[I(B,1,\theta)\right] = \theta\,|B| \tag{53}$$

$$\mathsf{var}\left[I(B,1/\theta,\theta)\right] = |B|/\theta \tag{54}$$

$$\mathsf{var}\left[I(B,1/\sqrt{\theta},\theta)\right] = |B|. \tag{55}$$

This is analogous to the deflation of residual variance in the linear model.

8.4 Residuals sum to zero

We saw that $R(W)=0$ when the fitted model is a uniform Poisson process. This holds in much more general contexts. Suppose we fit a point process model whose conditional intensity is of the form

$$\lambda_\theta(u,\mathbf{x}) = \exp\{\beta + \eta T(u,\mathbf{x})\}$$

where $\theta = (\beta, \eta)$, where $T(u, \mathbf{x})$ is not a function of u only. If the model is fitted to data $\mathbf{x}$ by the method of maximum pseudolikelihood, the estimating equations are

$$n(\mathbf{x}) = \int_W \lambda_{\hat{\theta}}(u, \mathbf{x}) \, \mathrm{d}u \tag{56}$$

$$\sum_{x_i \in \mathbf{x}} T(x_i, \mathbf{x}) = \int_B T(u, \mathbf{x}) \lambda_{\hat{\theta}}(u, \mathbf{x}) \, \mathrm{d}u. \tag{57}$$

The first of these equations implies that $R(W) = 0$ identically for any point pattern $\mathbf{x}$.

8.5 Uncorrelated errors

Suppose that the point process interpoint interactions have finite range r, in the sense that the conditional intensity $\lambda(u, \mathbf{x})$ depends only on points of $\mathbf{x}$ that lie within a distance r of the location u. This embraces Poisson processes, the Strauss process, and many other standard examples [102].

Let U, V be two subsets of W at least r units apart, i.e. $||u - v|| > r$ for any $u \in U$ and $v \in V$. Then $I(U)$ and $I(V)$ are uncorrelated. This is proved in Appendix C.

As a result we may expect a strong law of large numbers and a central limit theorem to hold for the asymptotic behaviour of the innovations and residuals as the sampling window W expands.

9 Diagnostic plots for spatial trend

Next we develop diagnostic plots based on the residuals. We are guided by analogy with residual plots for statistical models in general [4, 25, 33] and logistic regression models in particular [46, 68, 87].

9.1 Components of a point process model

It is of interest to check a spatial point pattern for the presence of spatial trend, dependence on spatial covariates, dependence between points of the pattern, and other features which should be incorporated into the model. Absence of *all* these features characterises the homogeneous Poisson process or "complete spatial randomness" (CSR) which is therefore the null model in most applications.

The terms 'trend' and 'interaction' can be defined rigorously for a completely general finite Gibbs point process. The Papangelou conditional intensity can be factorised as

$$\lambda(u, \mathbf{x}) = \exp\{V_1(u) + \sum_i V_2(u, x_i) + \sum_{i<j} V_3(u, x_i, x_j) + \ldots\}$$

for uniquely defined functions $V_1, V_2, \ldots$ called the potentials of order 1,2,.... The first potential $L(u) = V_1(u)$ can be interpreted as a "spatial trend" component. The higher order potential sum

$$M(u, \mathbf{x}) = \sum_i V_2(u, x_i) + \sum_{i<j} V_3(u, x_i, x_j) + \ldots$$

gives rise to dependence ("interaction") between points of the process. If M is identically zero, the model is an heterogeneous Poisson process with intensity function $\lambda(u) = \exp\{V_1(u)\}$.

In this section we develop diagnostic plots for validating the 'spatial trend' component of the model, L.

9.2 Trend example

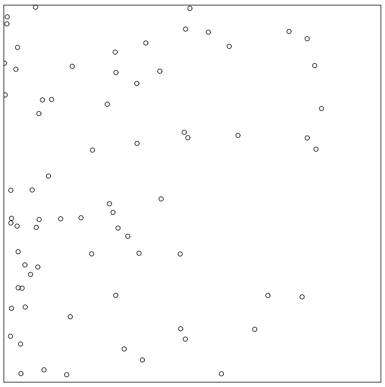

Figure 4: Example of spatial heterogeneity. Simulated realisation of the heterogeneous Poisson process with intensity $\lambda(x, y) = 300 \exp\{-3x\}$ in the unit square.

Figure 4 shows an example point pattern dataset which was simulated from the heterogeneous Poisson point process with intensity $\lambda(x, y) = 300 \exp\{-3x\}$ in the unit square. This is chosen as a fairly striking example of spatial trend which should be detected by the diagnostics.

In the following discussion we compute diagnostics for two fitted models: (a) the "correct" model, heterogeneous Poisson with intensity $\lambda(x, y) = \alpha \exp\{-\beta x\}$, where the parameters were estimated by maximum likelihood (estimates $\widehat{\alpha} = 233$, $\widehat{\beta} = 2.89$), and (b) the "null" model, the homogeneous Poisson process of intensity β (estimate $\widehat{\beta} = 73$).

Diagnostic plots will be based on both the Stoyan-Grabarnik exponential energy marks (Section 5.3) and the residuals (Sections 6–12).

9.3 Mark plot

The residual measure R has atoms at the data points $x_i \in \mathbf{x}$ and a negative density at other locations $u \in W$. A simple pictorial representation of this is shown in Figure 5. It consists of a pixel image of the density component (i.e. with grayscale proportional to the density) and a symbol plot of the atoms (i.e. a circle centred on each data point x_i with radius equal to the residual mass).

Figure 5 shows this representation for the two fitted models (a) and (b) mentioned above, using the Pearson residuals.

For the inverse-lambda residuals or the exponential energy marks, the density component is constant — at least for models with positive conditional intensity — and may be omitted, as shown in Figure 6. Interpretation is easier in this case. Note that the expansion in size of circles from left to right in the first plot is a consequence of model (a), which asserts that intensity is

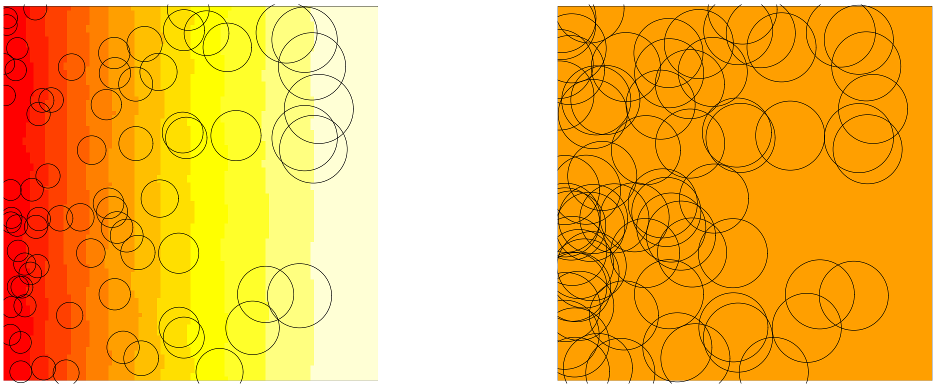

Figure 5: Mark plot based on Pearson residuals. *Left:* model of correct form, heterogeneous Poisson with intensity $\lambda(x, y) = \alpha \exp(-\beta x)$. *Right:* incorrect model, homogeneous Poisson.

decreasing from left to right. The usefulness of this plot lies mainly in the ease of seeing single points which are 'extreme' (there are none in this example, but cf. Figure 20 below).

However we can also interpret unevenness in the 'optical density' of circles as an indication that the fitted model is not correct. Thus the right panel in Figure 6 shows that the uniform Poisson model is inappropriate.

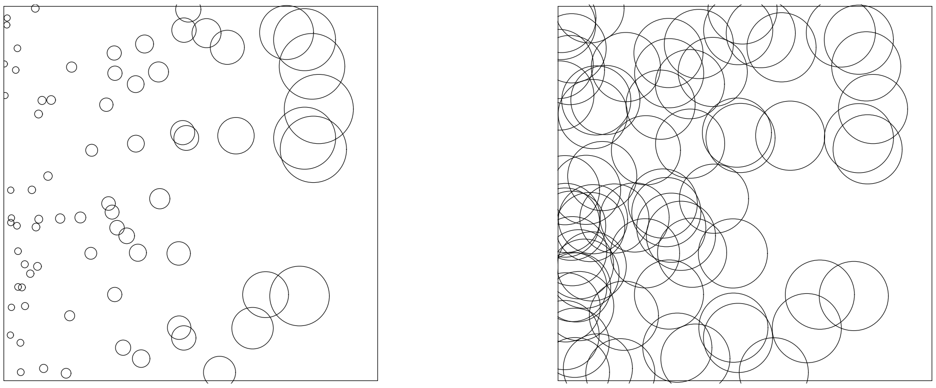

Figure 6: Mark plot for inverse-lambda residuals.

9.4 Sum of marks

The diagnostic interpretation of the residuals is based primarily on their sums over subregions B, that is, on the fact that $\mathbb{E}[R(B, h, \theta)] = 0$ for any subregion B of the window W if the model is true. One strategy is to partition W into disjoint subregions $B_1, \ldots, B_m$ (for example, dividing a rectangular window into equal squares B_k) and to evaluate $R(B_k, h, \theta)$. Deviation of these numbers from zero suggests a lack of fit.

For example, if the fitted model is the homogeneous Poisson process, the residual sum $R(B)$ for the raw, inverse-lambda, or Pearson residuals (given in equations (44), (45), (46) respectively) is simply a multiple of the number of data points falling in B. Hence this technique embraces the elementary method of quadrat counting used in spatial statistics [36, Section 2.5], [30, 97]. For other models, $R(B)$ is a weighted count with data-dependent spatially-varying weights, which have not previously been used in quadrat methods to our knowledge.

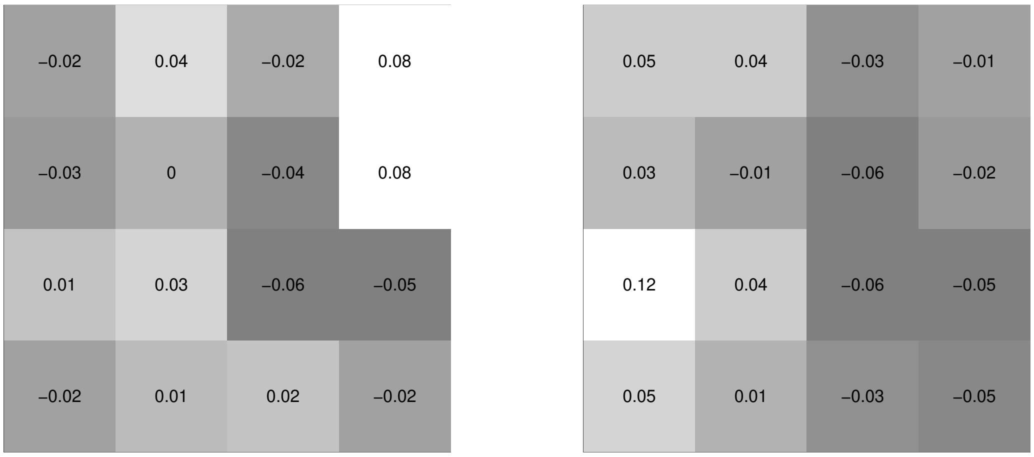

Figure 7: Residual sums for rectangular quadrats, based on Pearson residuals. *Left:* model of correct form, heterogeneous Poisson with intensity $\lambda(x, y) = \alpha \exp(-\beta x)$. *Right:* incorrect model, homogeneous Poisson.

Figure 7 depicts the values obtained by dividing W into equal quadrats (rectangles) and evaluating the Pearson residual for each quadrat. The display for the incorrect (homogeneous Poisson) model shows clear evidence of a trend from left to right in the values of these sums, correctly suggesting that the fitted model is inappropriate.

9.5 Kernel smoothed plot

A better approach is to apply kernel smoothing to the residual measure. Intuitively, the interpretation of the right-hand plot in Figure 6 is that the uneven 'density of ink' reflects a lack of fit due to spatial heterogeneity. This visual effect would be improved by spatially smoothing the 'ink'. In Figure 5 the pixel image must be interpreted as 'negative ink' to be smoothed and combined with the 'positive ink' to yield a smoothed residual density.

If $k(\cdot)$ is a smoothing kernel, assumed to be a probability density on $\mathbb{R}^2$, define the *smoothed residual field*

$$\begin{aligned} s(u) &= e(u) \int_W k(u - v) \, \mathrm{d}R(v, \widehat{h}, \widehat{\theta}) && (58) \\ &= e(u) \sum_i k(u - x_i) h(x_i, \mathbf{x} \setminus \{x_i\}) - e(u) \int_W k(u - v) h(v, \mathbf{x}) \lambda_{\widehat{\theta}}(v, \mathbf{x}) \, \mathrm{d}v. && (59) \end{aligned}$$

where $e(u)$ is a correction for edge effects in the window W given by

$$e(u) = \left[\int_W k(u - v) \, \mathrm{d}v \right]^{-1}. \qquad (60)$$

The smoothed residual field s may be presented as an image, contour map or perspective plot in the window W, as shown in Figures 8 and 9.

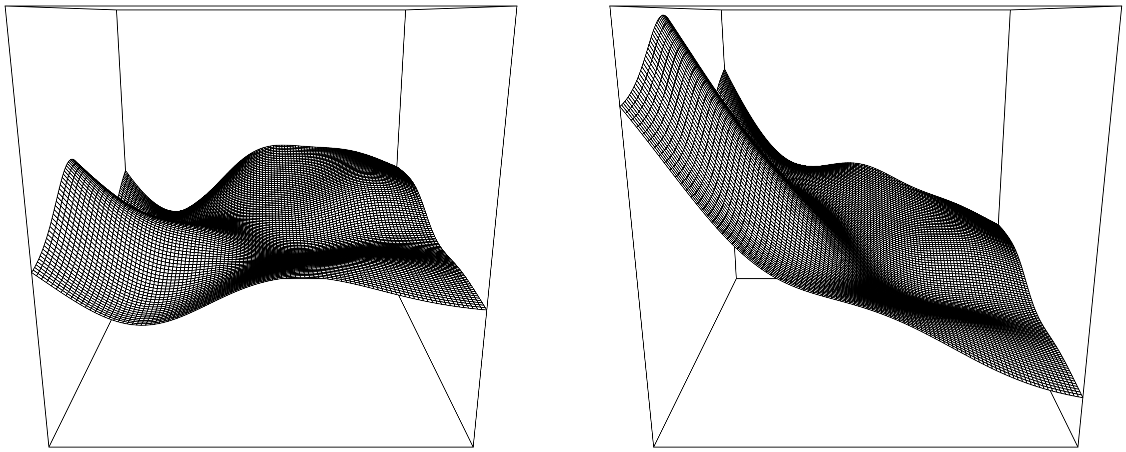

Figure 8: Perspective plots of kernel-smoothed residual field for two models fitted to Figure 4. Plotted on the same scale. *Left:* model of correct form, heterogeneous Poisson with intensity $\lambda(x, y) = \alpha \exp(-\beta x)$. Range of smoothed field $[0.2, 2.1]$. *Right:* incorrect model, homogeneous Poisson. Range of smoothed field $[0.1, 3.0]$. Smoothing kernel: isotropic Gaussian with standard deviation 0.14.

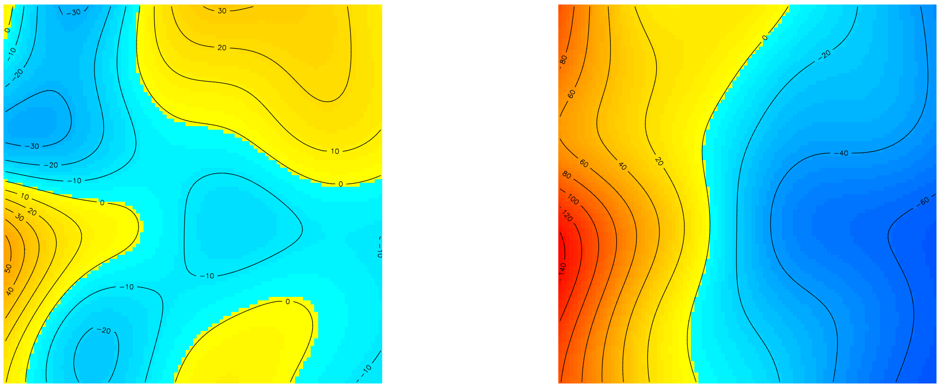

Figure 9: Contour/image plots of surfaces in Figure 8 drawn using the same greyscale map in each plot.

From previous results it is clear that the analogous quantity for the innovations has mean zero,

$$\mathbb{E}\left[\int_W k(u-v)\,\mathrm{d}I(v,h,\lambda)\right] = 0$$

and we expect $s(u) \approx 0$ if the fitted model is correct.

For the raw residuals, the smoothed residual field is

$$\begin{aligned} s(u) &= e(u)\left[\sum_i k(u-x_i) - \int_W k(u-v)\lambda_{\widehat{\theta}}(v,\mathbf{x})\,\mathrm{d}v\right] \\ &= \lambda^*(u) - \lambda^\dagger(u) \end{aligned} \tag{61}$$

the difference between $\lambda^*(u)$, a nonparametric, kernel-smoothing estimate of the point process intensity, and $\lambda^\dagger(u)$, a kernel-smoothed version of the parametric estimate of the conditional intensity, which is an unbiased estimate of the intensity. Thus the smoothed raw residual field compares nonparametric and parametric estimates of intensity, which should be approximately equal if the fitted model is correct. Additionally an observed deviation of $s(u)$ from zero is an estimate of the (negative) bias in estimation of the point process intensity.

This might be compared to the work of Fowlkes [46], who proposed a smoothed residual plot for binary logistic regression in order to avoid features that are solely the result of the binary nature of the responses. However Fowlkes' plots were obtained by smoothing the *binary responses* before computing pseudo-residuals. Here it is more appropriate to smooth the *residuals* after fitting, since this smoothing can simply be interpreted as a smoothing of the residual measure.

Kernel-smoothed estimates of the point process intensity (analogous to $\lambda^*(u)$) have been used as exploratory tools in spatial statistics [39, section 8.2]. The technique described here introduces model-dependent, data-dependent and spatially-varying weights on the data points and centering of the smoothed estimate.

As usual, it is not obvious how to choose the bandwidth for the kernel k. Given that the smoothed residual field for the raw residuals has an interpretation closely related to that of a kernel smoothed estimate of the point process intensity, it would be reasonable to follow established practice for that technique [16, 37], [98, pp. 237–238] and minimise an estimate of the mean square error. Alternatively there are rules of thumb for bandwidth selection for the pair correlation function [98, p. 285]. However the bandwidths used in this paper were instead chosen by cross-validation of kernel smoothing of the discretised residual measure, as described in Section 12.

Ogata [83] proposed a residual for space-time point processes, defined as the ratio of a kernel-smoothed estimate of the local space-time intensity to a parametric estimate of the spatial intensity. This might be regarded as analogous to the smoothed residual field of the inverse-lambda residuals.

9.6 Lurking variable plots

In linear modelling, if we suspect the data may depend on a covariate that was not included in the model, the usual diagnostic is a plot of the residuals against the covariate [4, p. 3]. Any systematic pattern in this plot indicates a departure from the model, and suggests the appropriate modification of the linear predictor.

For point process models, by analogy, we may plot the exponential energy weights against a spatial covariate, or against one of the Cartesian coordinates (or some other given directions), to investigate the presence of spatial trend (or to assess whether the true spatial trend differs from that specified by the fitted model).

For a spatial covariate $Z(u)$ defined at each location $u \in W$, we may evaluate the residual measure on each sub-level set

$$W(z) = \{u \in W : \ Z(u) \leq z\}$$

yielding a 'cumulative residual' function

$$C(z) = R(W(z), \widehat{h}, \widehat{\theta}). \tag{62}$$

This should be approximately zero if the fitted model is correct.

For example, for the raw residuals,

$$C(z) = N(W(z)) - \int_{W(z)} \lambda_{\widehat{\theta}}(u, \mathbf{x}) \, \mathrm{d}u. \tag{63}$$

Notice that $\widehat{F}(z) = N(W(z))/N(W)$ is the empirical cumulative distribution function of the values of the covariate observed at the data points x_i. The function $C(z)$ is an adjustment of this empirical d.f. to have approximately zero mean when the model is correct.

In the example dataset we have been using (Figure 4), the lurking variable is the x coordinate. Figure 10 shows plots of $C(z)$ against z based on the covariate $Z(x, y) = x$. Thus $C(z)$ is the sum of residuals in the region $W(z)$ to the left of the line $x = z$.

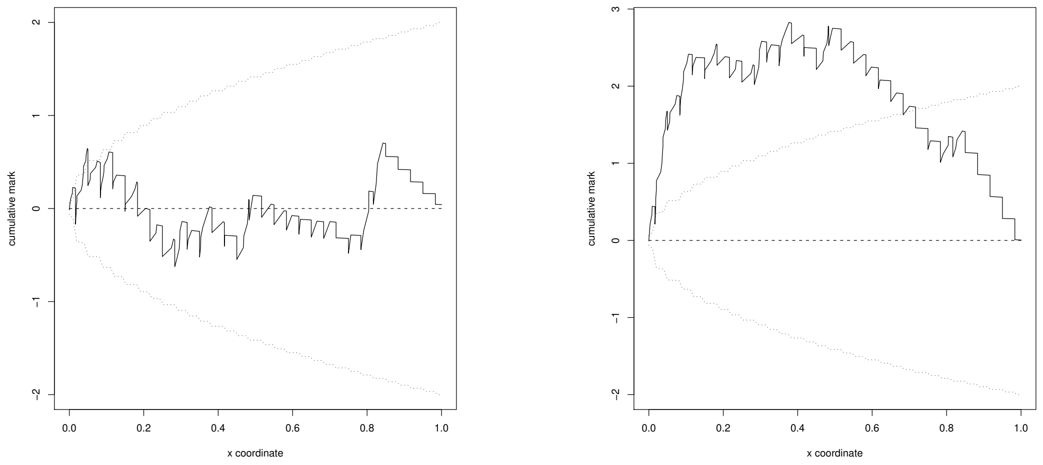

Figure 10: Lurking variable plots for the x coordinate, for two models fitted to the data in Figure 4. Cumulative style, Pearson residuals. *Left:* model of correct form, heterogeneous Poisson with intensity loglinear in x, fitted to data. The plot correctly suggests a good fit. *Right:* incorrect model, homogeneous Poisson, fitted to data. The plot correctly suggests a poor fit and strong effect of x covariate.

The dotted envelopes in Figure 10 are 2σ limits based on variance calculations for the heterogeneous Poisson process. That is, the variance of $C(z)$ is estimated by evaluating (49) for $B = W(z)$, where $W(z)$ is as just described, substituting the fitted Poisson intensity $\widehat{\lambda}(u)$ for $\lambda(u)$. This is a conservative (i.e. positively biased) estimate of the residual variance, because of the phenomenon of variance deflation.

The 2σ limits have the usual significance interpretation assuming a central limit theorem applies. The glaring violation of these bounds by the graph in the right panel of Figure 10 is ample evidence that the model is inappropriate. In the left panel there is also a slight excursion beyond the limits for small x, but this should not be invested with formal significance since it relates to a small sample.

This diagnostic plot is very effective in the example shown in Figure 10. For the incorrect fitted model, a homogeneous Poisson process, the function $C(z)$ deviates markedly from zero, and the form of the deviation is suggestive of an exponential function of x. For a fitted model of the correct form (heterogeneous Poisson with $\lambda(x, y) = \alpha \exp(-\beta x)$ with α, β estimated) the function $C(z)$ is approximately zero.

Alternatively we may plot approximate derivatives of $C(z)$. Let

$$c(z) = \int_W k_1(Z(u) - z) \, \mathrm{d}R(u, \widehat{h}, \widehat{\theta}) \tag{64}$$

$$= \sum_i k_1(Z(x_i) - z) h(x_i, \mathbf{x} \setminus \{x_i\}) - \int_W k_1(Z(u) - z) h(u, \mathbf{x}) \lambda_{\widehat{\theta}}(u, \mathbf{x}) \, \mathrm{d}u \tag{65}$$

where k_1 is a one-dimensional smoothing kernel (a probability density on $\mathbb{R}$). Thus $c(z)$ is the differenceen between two unnormalised kernel-smoothed estimates of the empirical distribution of covariate values, one based on the data points $x_i \in \mathbf{x}$ and the other on the background points $u \in W$. If the fitted model is correct then $c(z) \approx 0$.

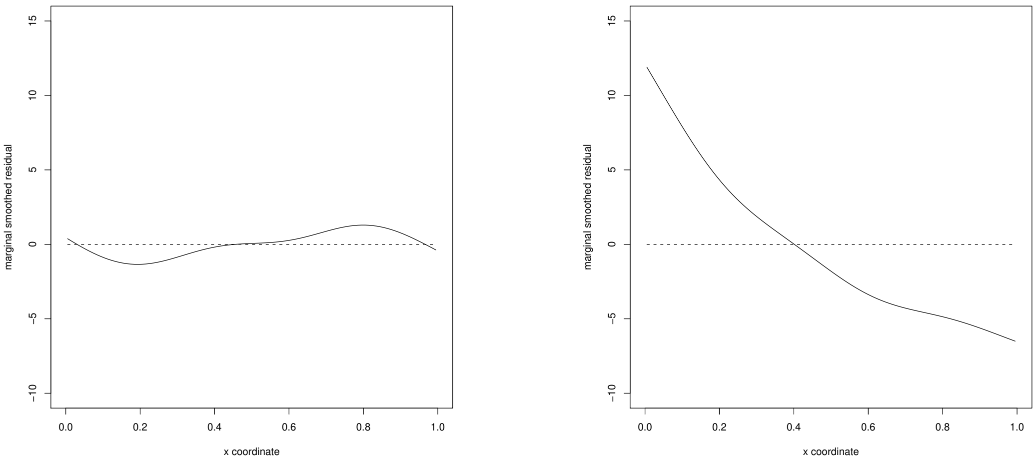

Figure 11: "Density style" lurking variable plots for the x coordinate, for two models fitted to the data in Figure 4. *Left:* model of correct form, heterogeneous Poisson with intensity loglinear in x, fitted to data. The plot correctly suggests a good fit. *Right:* incorrect model, homogeneous Poisson, fitted to data. The plot correctly suggests a poor fit and strong effect of x covariate.

If the covariate Z is one of the Cartesian coordinates, the function c can be computed rapidly from the smoothed residual field s. Consider the case of the x coordinate. If the one-dimensional kernel k_1 is the marginal of the two-dimensional smoothing kernel k,

$$k_1(x) = \int k(x, y) \, \mathrm{d}y,$$

then

$$c(x) = \int_W s(x, y) \, \mathrm{d}y$$

is simply the corresponding marginal of the two-dimensional kernel smoothed residual field s. It may be computed to a good approximation by summing values of s computed on a rectangular grid.

Plots of $c(z)$ for our example are shown in Figure 11. The results are even more striking.

9.7 Four-panel plot

Figure 12 shows our standard presentation of the diagnostic plots for spatial trend. The top left panel is the mark plot. The bottom right panel is a contour and image plot of the smoothed residual field, using a colour mapping which always represents the value $s(u) = 0$ by the same colour, for easy interpretation.

The bottom left panel is a lurking variable plot for the x coordinate, its x axis aligned correctly with the x coordinates in the mark plot. The top right panel is a lurking variable plot for the y coordinate, rotated 90 degrees anticlockwise, the y axis aligned correctly with the y coordinates in the mark plot. We have found that the combination of these two lurking variable plots is often sufficient to draw attention to a spatial trend when it is present.

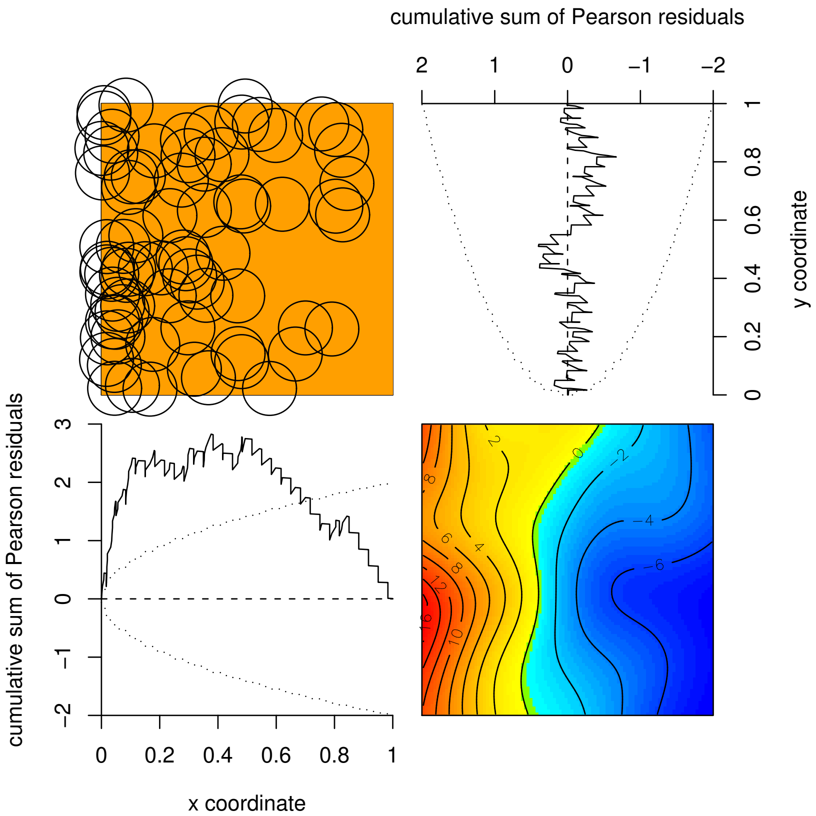

Figure 12: Standard presentation of the diagnostic plots. Top left: mark plot. Top right: lurking variable plot for y coordinate. Bottom left: lurking variable plot for x coordinate. Bottom right: contour plot of smoothed mark field. Data from Figure 4; fitted model is uniform Poisson; Pearson residuals.

9.8 Spatial covariate example

Figure 13 shows a synthetic example of a point pattern dataset which depends on a spatial covariate. It is a realisation of an heterogeneous Poisson process with intensity 100 inside the region R delineated by the dotted lines, and intensity 10 outside this region. The spatial covariate Z is the indicator function of this region.

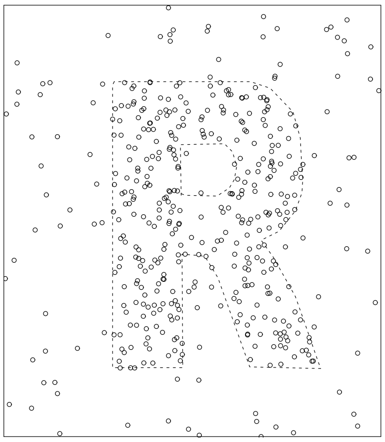

Figure 13: Synthetic data depending on a spatial covariate. Poisson point process with different intensity inside and outside the region shown by the dotted lines.

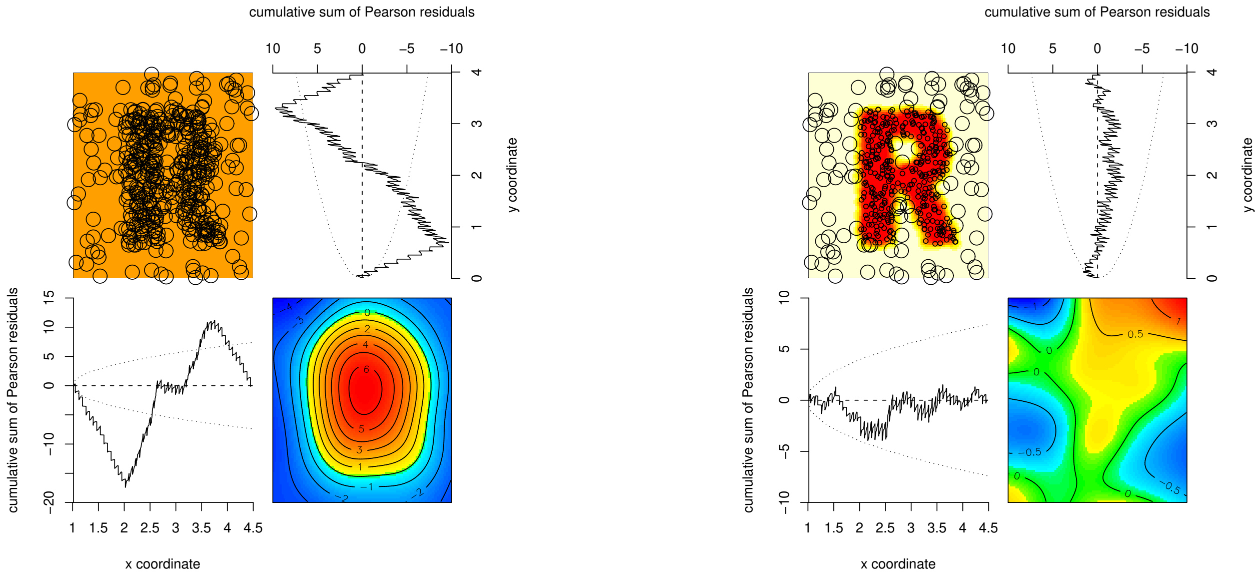

Figure 14: Pearson residual diagnostic plots for models fitted to the point pattern data of Figure 13. *Left:* uniform Poisson process. *Right:* heterogeneous Poisson process with intensity dependent on covariate Z.

The left panel of Figure 14 shows diagnostic plots based on Pearson residuals for the uniform Poisson process fitted to the point pattern in Figure 13. It indicates clearly that the model is inappropriate. The mark plot suggests the form of the spatial heterogeneity.

The right panel of Figure 14 shows the corresponding plot for the heterogeneous Poisson process with intensity

$$\lambda(u) = \beta_1 + \beta_2 Z(u)$$

fitted to Figure 13. It suggests a good fit.

9.9 Applications

9.9.1 Queensland copper data

The Queensland copper data are shown in Figure 2. A common thread in the analyses of this dataset [15, 17, 47] is to assess the dependence of the point pattern of copper deposits on the distance to the nearest lineament.

Let the covariate $Z(u)$ be the minimum distance from the location u to the nearest lineament. This can easily be computed analytically for all locations in a fine pixel grid. The covariate is depicted in Figure 15.

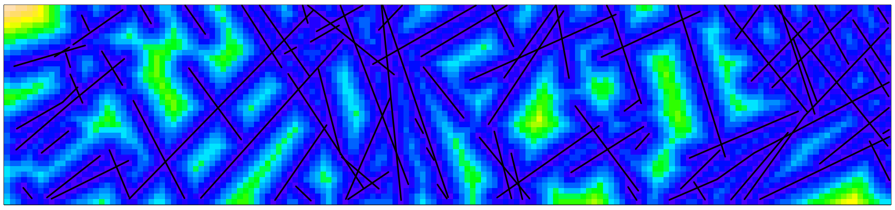

Figure 15: Image plot of the spatial covariate: the distance from each location in the Queensland study region to the nearest lineament in Figure 2.

First we fit the null model of a uniform Poisson process. For this model, and for the raw residuals, the lurking variable plot (63) becomes

$$\begin{aligned} C(z) &= N(W(z)) - \frac{N(W)}{|W|}|W(z)| \\ &= N(W)\,[F(z) - G(z)] \end{aligned}$$

where $F(z) = N(W(z))/N(W)$ is the empirical c.d.f. of the values of the covariate at the data points, and $G(z) = |W(z)|/|W|$ is the empirical c.d.f. of the covariate at all locations in W. Similarly for the other types of residuals.

Berman [15] proposed comparing the two c.d.f.'s F and G by plotting them against z, as shown in the left panel of Figure 16, and comparing various moments. Our lurking variable plot is essentially the same diagnostic plot, at least for the raw residuals and the uniform Poisson model. The right panel shows a lurking variable plot of the Pearson residuals for the uniform Poisson model. Our technique has the advantage that any fitted model may be treated in the same way.

9.9.2 Chorley-Ribble data

The Chorley-Ribble data are shown in Figure 3. The null model considered by Diggle [38] states that the point pattern of cases of cancer of the larynx is a heterogeneous Poisson process with intensity

$$\lambda(u) = \eta\rho(u) \tag{66}$$

where $\rho(u)$ is the spatially-varying density of the susceptible population. Cases of lung cancer served as a surrogate for the susceptible population, and ρ was estimated by kernel smoothing the point pattern of lung cancer cases.

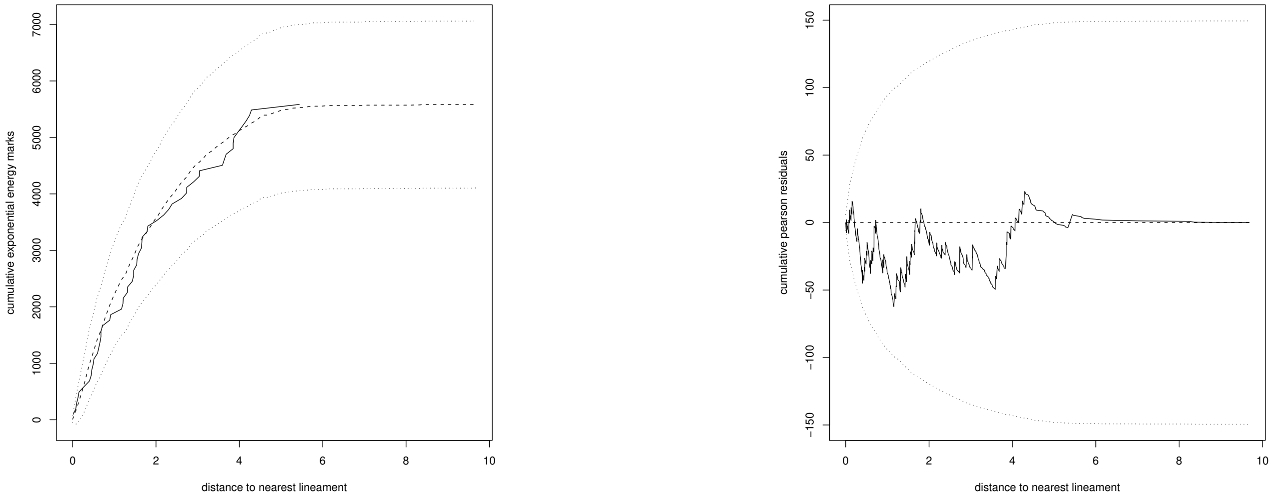

Figure 16: Lurking variable plots for the homogeneous Poisson point process fitted to the Queensland copper data. *Left:* cumulative exponential energy weights against distance to nearest lineament. Equivalent to empirical c.d.f. of distance to nearest lineament at the observed points of the process. *Right:* cumulative Pearson residuals against distance to nearest lineament. Solid lines: empirical curve. Dashed lines: expectation for fitted homogeneous Poisson point process model. Dotted lines: pointwise 2σ limits based on equations (48) and (49) respectively.

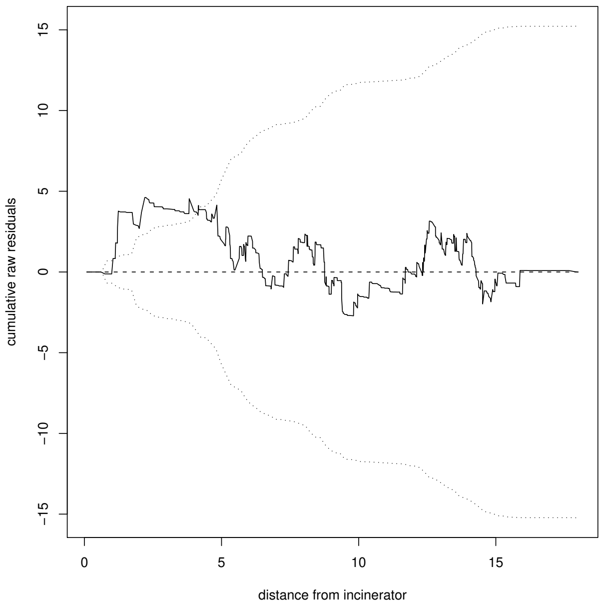

Figure 17: Lurking variable plot of the raw residuals from the null model against distance from the incinerator, for the Chorley-Ribble data.

We fitted the null model (66) using the kernel-smoothed estimate of ρ using a Gaussian kernel with $\sigma = 0.15$ km, identical to that chosen by Diggle [38]. Figure 17 is a lurking variable plot of the raw residuals from this model against distance from the incinerator. This plot indicates a deviation from the fitted model and suggests there is increased intensity near the incinerator.

It should be emphasised that no formal conclusions ought to be drawn from this plot, which is intended primarily to draw attention to features of the data that require explanation. Possible explanations discussed by Diggle [38] include clustering of disease cases (e.g. due to correlation within families) as well as a carcinogenic effect from the incinerator. However, it is feasible to assess each such explanation by the same technique, by fitting the appropriate model and computing the analogue of Figure 17.

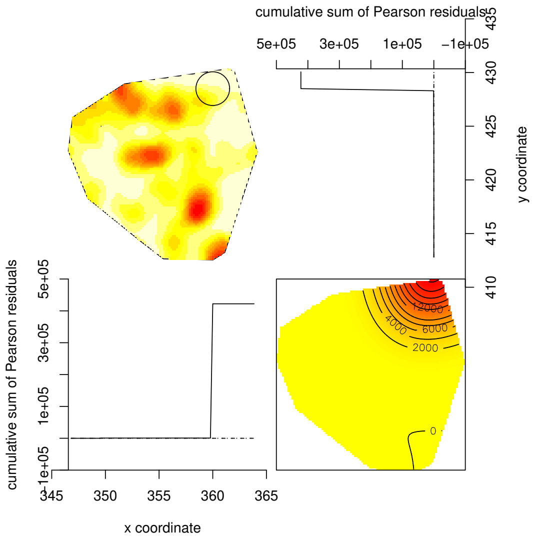

Figure 18: Trend diagnostic plots using the Pearson residuals for the null model, showing a huge residual at the location $(360, 428)$.

Diagnostic plots for the inverse-lambda and Pearson residuals reveal some other inconsistencies between the data and the fitted model. Figure 18 shows an apparent 'outlier'. This is a case of cancer of the larynx occurring in a region where the kernel-smoothed estimate $\widehat{\rho}(u)$ of $\rho(u)$ is very low. This raises questions about the appropriateness of the estimator $\widehat{\rho}$, rather than necessarily indicating an anomalous observation.

10 Diagnostic plots for interpoint interaction

Next we develop residual plots to validate the interpoint interaction component of a model.

10.1 Strauss example

Figure 19 shows a realisation of the Strauss process [62], [39, p. 75], with interaction parameter $\gamma = 0.1$, interaction radius $r = 0.1$ and activity $\beta = 100$ in the unit square. It was obtained using a perfect sampler based on spatial birth-death processes as in [63, 64] using `C` code kindly supplied by Kasper K. Berthelsen. In order to obtain a pattern exhibiting approximate stationarity, the Strauss realisation was generated in the window $[-1, 2]^2$ and subsequently clipped to the unit square $W = [0, 1]^2$.

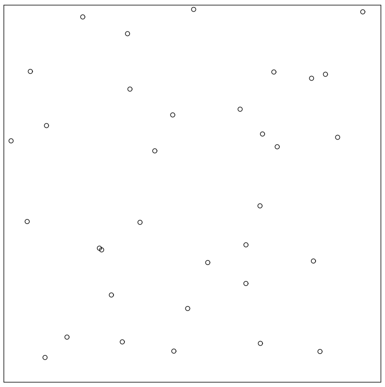

Figure 19: Example of interpoint interaction. Simulated realisation of the Strauss process with activity $\beta = 100$, interaction parameter $\gamma = 0.1$ and interaction range $r = 0.1$ in the uint square.

This process has been chosen as an example of strong interpoint interaction which should be picked up by the diagnostics. The pattern contains 34 points, which is typical of classical datasets considered in [36, 90].

We shall compare two models for this dataset:

(a) fitted model of correct form, the stationary Strauss process with known interaction range $r = 0.1$ and other parameters fitted by maximum pseudolikelihood using the Berman-Turner device [8] (parameter estimates $\widehat{\beta} = 101$, $\widehat{\gamma} = 0.24$);

(b) incorrect fitted model, the homogeneous Poisson process with intensity λ (estimate $\widehat{\lambda} = 34$).

Since we are seeking to validate the interpoint interaction part of the fitted model, the main objective is to develop a plot which shows in this example that (b) is not an appropriate model while (a) is appropriate.

Edge effects arise when fitting the Strauss process, because its conditional intensity

$$\lambda(x_i, \mathbf{X}) = \beta \gamma^{t(x_i, \mathbf{X})} \tag{67}$$

depends on $t(x_i, \mathbf{X})$, the number of points of $\mathbf{X} \setminus \{x_i\}$ within a distance r of x_i. When x_i lies close to the boundary of W, this number is not observable. Strategies for dealing with edge effects are surveyed in [10, 91, 97]. The only relevant strategy here seems to be the 'border method' in which we compute residuals only for those locations u lying inside the clipped window

$$W_{\ominus r} = \{u \in W : \ d(u, W^c) \geq r\}$$

where $d(u, W^c)$ is the distance from u to the complement of W. In our example, W is the unit square and $W_{\ominus r} = [r, 1-r]^2$.

Figure 20 shows the standard diagnostic plots for spatial trend from Section 9.7, applied to the fitted Strauss process model (a). The plots suggest that this is an adequate model, although they call attention to a region of large residuals near the location $(0.6, 0.6)$.

The mark plot (top left panel) shows that one data point has a relatively large residual mass. Before interpreting this as an 'outlier' (which it cannot be in this synthetic example) one should

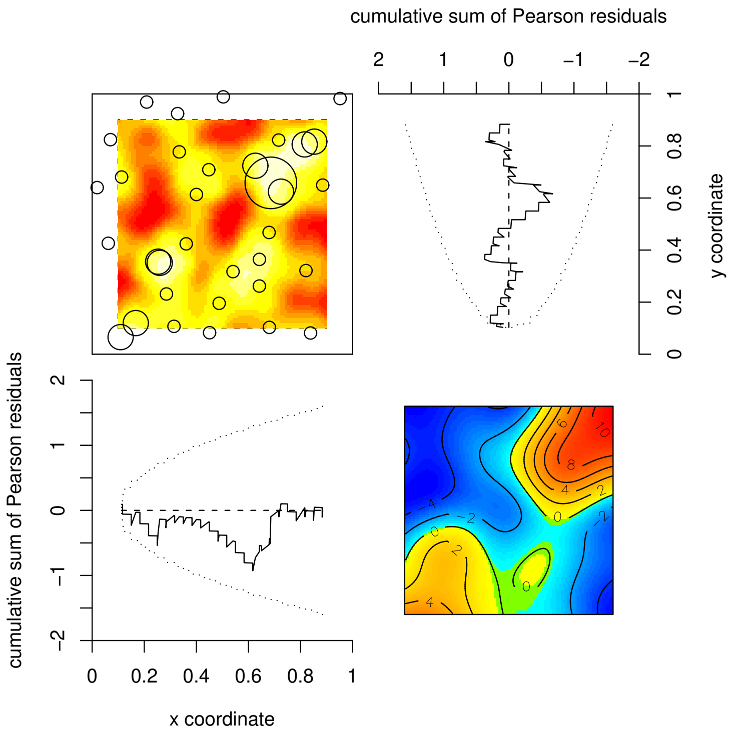

Figure 20: Diagnostic plots for the Strauss point process model (with r known), fitted to the data in Figure 19.

remember the form of the conditional intensity (67) and the very small value of the interaction parameter γ, estimated at $\widehat{\gamma} = 0.24$. An increase of 1 in the number $t(x_i, \mathbf{x})$ of nearby data points causes the Pearson residual mass to increase by a factor of $0.24^{-1/2} = 2.0$. This phenomenon is even more exaggerated for the inverse-lambda residuals or the exponential energy marks, where each extra neighbour increases the residual mass by a factor of $0.24^{-1} = 4.2$. This sensitivity is analogous to the high variance of the Horvitz-Thompson [55] estimator when some events have very low probability.

The raw residuals have no such sensitivity. This suggests that the raw residuals may be the best tool for investigating outliers in the context of strong interpoint inhibition.

10.2 Inappropriateness of trend plots for assessing interaction

The diagnostic plots proposed in Section 9 for detecting spatial trend are not suitable for detecting interpoint interaction.

Figure 21 shows the corresponding plots when the fitted model is the homogeneous Poisson process. They suggest erroneously that the homogeneous Poisson is a good fit to these data.

Equations (44)–(46) gave the form of the residuals for a fitted uniform Poisson model. It is clear that the raw residuals (44) will have mean zero under any stationary process, i.e. whenever $\mathbb{E}[N(B)] = \theta\,|B|$. Thus the trend plots for a fitted uniform Poisson model may be expected to be flat whenever the data come from a stationary point process.

Indeed, Figure 21 even suggests that the Poisson model is a better fit than the Strauss model. This phenomenon is easy to explain. For the raw residuals, for example, the smoothed

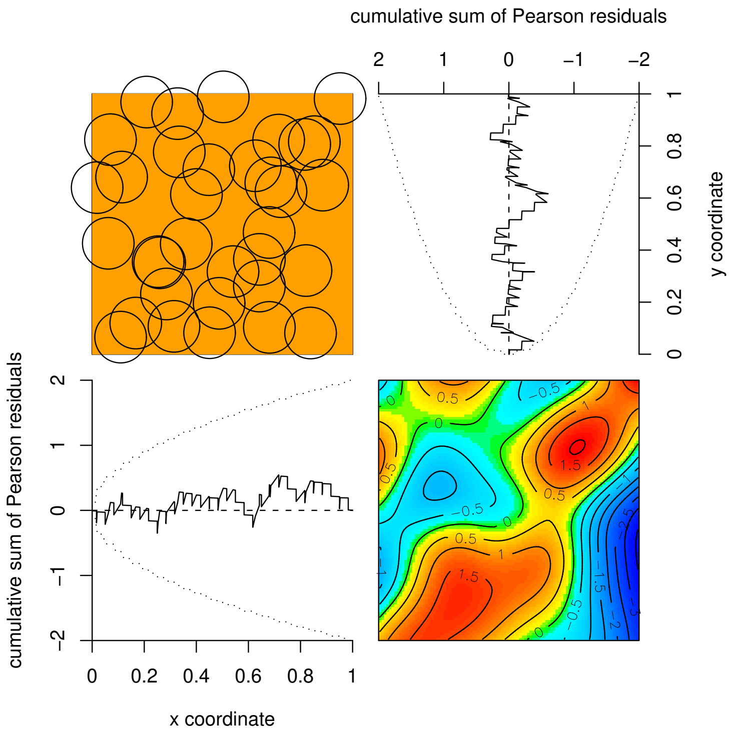

Figure 21: Diagnostic plots for an incorrect model, the uniform Poisson process, fitted to the data in Figure 19. Pearson residuals.

residual field $s(u)$ is expressed in equation (61) as the difference between $\lambda^*(u)$, a nonparametric, kernel-smoothing estimate of the point process intensity, and $\lambda^\dagger(u)$, a kernel-smoothed version of the parametric predictor of the conditional intensity. For the Strauss data in Figure 19 the estimate $\lambda^*(u)$ is roughly constant, because the Strauss process is stationary. The Poisson process has constant intensity so the parametric estimate $\lambda^\dagger(u)$ for a Poisson model is constant. Thus $s(u)$ is close to zero in Figure 21.

Indeed, the effect of interpoint inhibition in the Strauss data (Figure 19) is to make $\lambda^*(u)$ a flatter surface than it would be for a typical realisation of the uniform Poisson process. Thus $s(u)$ is closer to zero in Figure 21 than it is in Figure 20.

10.3 Information in the residuals

The smoothed residual field $s(u)$ does contain useful information about the presence of interpoint interaction in the data.

For example, assume the fitted model is homogeneous Poisson, and take the smoothing kernel k to be the uniform density in the disc of radius r centred at the origin:

$$k(u) = \frac{1}{\pi r^2}\mathbf{1}\{||u|| \leq r\}.$$

We will ignore edge effects by restricting attention to locations u in $W_{\ominus r}$. Let $B(u,r)$ denote the ball with centre u and radius r. Then

$$s(u) = \frac{N(\mathbf{x} \cap B(u,r))}{\lambda \pi r^2}$$

is the (rescaled) number of points of $\mathbf{x}$ within a distance r of the location u.

We have the following connections. First the maximum value of $s(u)$ over all locations $u \in W_{\ominus r}$ is the *scan statistic* [67, 1], a well-known summary statistic used for detecting clustering.

Second, the zero fraction of s,

$$Z = \frac{|\{u \in W_{\ominus r} : \ s(u) = 0\}|}{|W|}$$

is equal to

$$|\{u \in W_{\ominus r} : \ \mathbf{x} \cap B(u, r) = \emptyset\}|/|W_{\ominus r}| = 1 - \hat{F}(r)$$

where $\hat{F}$ is the empirical empty space function of $\mathbf{x}$ (using the border method edge correction). This is a popular spatial summary statistic, often used for detecting interpoint inhibition, see e.g. [39, 97].

The distribution of $s(u)$ under a homogeneous Poisson process is that of $S = N/(\lambda \pi r^2)$ where $N \sim \mathsf{Poisson}(\lambda \pi r^2)$, ignoring the effect of estimating λ. The distribution of S for a point process with interpoint interaction is usually over- or under-dispersed with respect to the Poisson.

10.4 Q–Q plots

It could have been expected that plots of the residuals against a covariate would fail to detect interpoint interaction. In Section 5.4 we drew an analogy between point processes and generalised linear models, in which the intensity of a point process is analogous to the mean response of the GLM, and interpoint interaction in a point process is analogous to the joint distribution of residuals in a GLM. The plots developed in Section 9 are analogous to plots of residuals against a covariate in a GLM, which are of use in detecting systematic departures from the assumed form of the linear predictor. Such plots are not very useful for assessing the distributional assumptions, except with regard to outliers.

In the GLM context the most appropriate tool for assessing the distributional assumptions is a summary of the empirical distribution of the residuals, such as a Q–Q plot. We therefore propose a Q–Q plot comparing empirical quantiles of the smoothed mark field $s(u)$ to the corresponding *expected* empirical quantiles for $s(u)$ under the fitted model (estimated by Monte Carlo).

In practice, this would be achieved by computing the values $s(u_j)$ at a fine grid of locations u_j in W, sorting them to obtain the order statistics, generating a large number of simulated realisations from the fitted model. To each simulated dataset we fit the same model, performing similar calculations, and taking the sample mean of the order statistics in the simulated arrays. We supply such a plot with pointwise prediction intervals.

Details are as follows. Denote by $s(u; \mathbf{x}; \widehat{\theta})$ the value of the smoothed residual field at the location $u \in W$ computed for the model with fitted parameter $\widehat{\theta}$ on the dataset $\mathbf{x}$. Let $u_j, \ j = 1, \ldots, J$ be fixed locations in W. After fitting the model to the original dataset $\mathbf{x}$ we compute the values of the smoothed residual field s at these locations, $s_j = s(u_j, \mathbf{x}, \widehat{\theta})$, and sort them to obtain their order statistics $s_{[1]} \leq s_{[2]} \leq \ldots \leq s_{[J]}$. We then generate N independent simulated realisations of the fitted model $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(N)}$. For each $n = 1, \ldots, N$ we fit the model to $\mathbf{x}^{(n)}$ with parameter estimate $\widehat{\theta}^{(n)}$, compute the smoothed residual field values

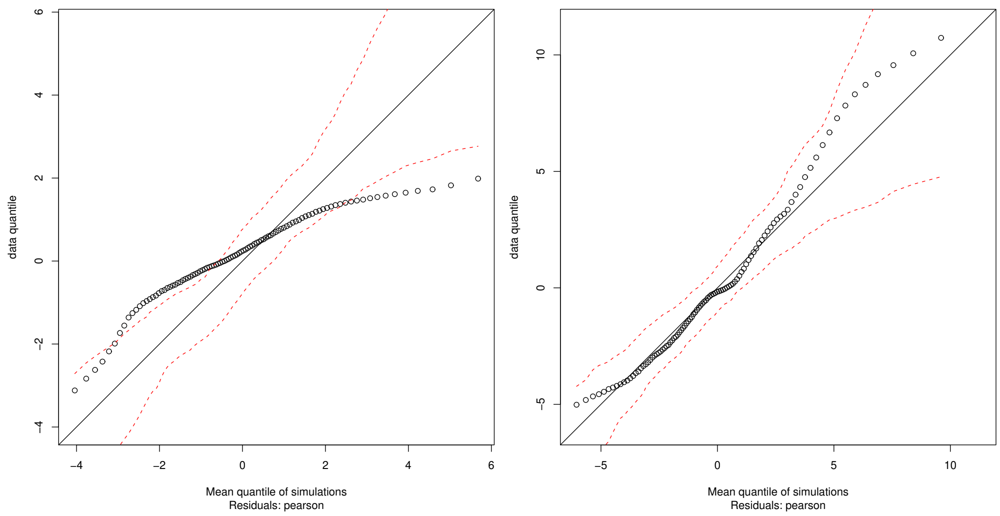

Figure 22: Q–Q plots for fitted spatial point processes: quantiles of the smoothed Pearson residual field for the data, plotted against corresponding expected quantiles under the fitted model (estimated by simulation). Dotted lines show pointwise 95% prediction intervals. Both plots are for models fitted to the Strauss example data of Figure 19. *Left:* incorrect model, homogeneous Poisson. Plot indicates fitted model is inappropriate. *Right:* fitted model of correct form, stationary Strauss ($r = 0.1$). Plot indicates agreement of fitted model with data.

$s_j^{(n)} = s(u_j, \mathbf{x}^{(n)}, \widehat{\theta}^{(n)})$, and obtain order statistics $s_{[1]}^{(n)} \leq \ldots \leq s_{[J]}^{(n)}$. The *sample mean* of the jth order statistic

$$e_j = \frac{1}{N} \sum_{n=1}^{N} s_{[j]}^{(n)}$$

is computed for each j. Thus, we are estimating the *expected* jth quantile of s under the model fitted to the original data. These values necessarily satisfy $e_1 \leq \ldots \leq e_J$. A rationale for using expected quantiles is offered in [54]. The Q–Q plot is a scatterplot of the data quantiles $s_{[j]}$ against the mean quantiles e_j.

For inference, we may add $100(1-\alpha)\%$ prediction intervals (a_j, b_j) obtained by

$$\begin{aligned} a_j &= q(\alpha/2, \{s_{[j]}^{(1)}, \ldots, s_{[j]}^{(N)}\}) \\ b_j &= q(1-\alpha/2, \{s_{[j]}^{(1)}, \ldots, s_{[j]}^{(N)}\}) \end{aligned}$$

where $q(\beta, S)$ denotes the probability β quantile of a set of values S.

Figure 22 shows Q–Q plots for the Strauss pattern in Figure 19. The left panel is for the (incorrect) homogeneous Poisson model. It shows clear disagreement between the model and the data. The smoothed mark field for the data has a shorter right tail, and a lower variability, than the smoothed mark field for simulations from the fitted model. This is consistent with an inhibitive point pattern.

The right panel of Figure 22 is for a Strauss point process model of the correct form (with $r = 0.1$ known) fitted to the data by maximum pseudolikelihood. This suggests good agreement between the fitted model and the data. Note the different scales of the left and right panels, and the wider prediction interval in the right panel.

Following are some notes about computation of the Q–Q plots. For each point pattern, the smoothed mark field was evaluated at a 40×40 grid of pixels, and the percentiles of these values obtained. The expected percentiles were estimated (using the 5% trimmed sample mean) from 1000 simulated realisations of the fitted Strauss process. Each Strauss realisation was obtained by running a simple birth-death-shift Metropolis-Hastings algorithm [79] for 100,000 steps from a random initial state. Total computation time for Figure 22 was approximately 10 minutes on a 2.0 GHz Linux desktop. For investigative work, this can be accelerated by reducing the simulation parameters. Adopting a 25×25 pixel grid and using only 100 simulated realisations of the fitted model produces a practically useful Q–Q plot in less than 30 seconds.

A possible alternative to Monte Carlo simulation would be analytic evaluation of the distribution of $s(u)$ for a given model. This seems to be difficult; it is done for the uniform Poisson model in Appendix B. This shows that the null distribution of values $s(u)$ is heavily skewed, justifying the use of the trimmed mean in our Monte Carlo calculations.

10.5 Range of interpoint interaction

Next we test the ability of the Q–Q plots to detect an *incorrectly specified* interpoint interaction.

The Strauss example data in Figure 19 were generated by a Strauss process with interaction range $r = 0.1$. To this dataset we fitted Strauss processes with r incorrectly set at 0.05 and at 0.15. The Q–Q plots for these incorrect models are shown in Figure 23. They suggest moderately strong disagreement between the data and the fitted model with regard to interpoint interaction.

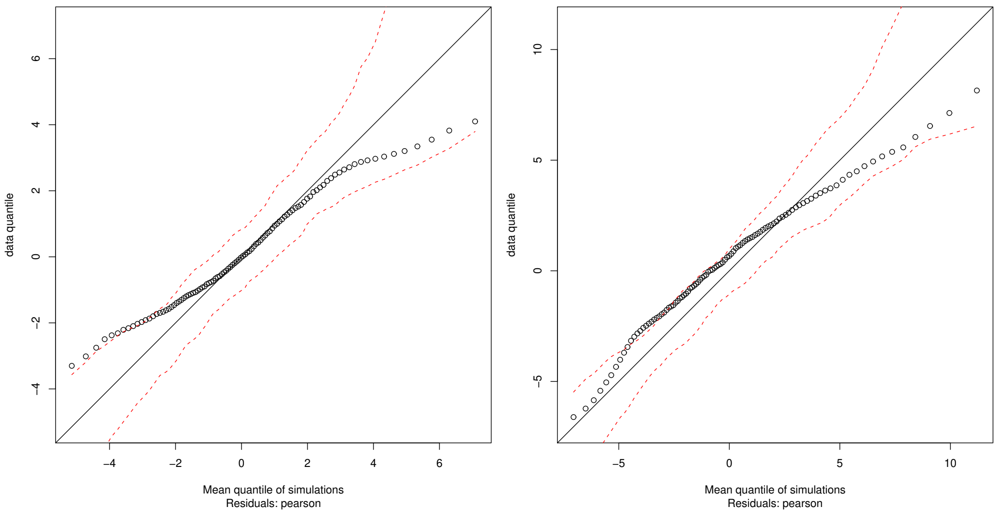

Figure 23: Q–Q plots for Strauss process models with $r = 0.05$ (Left) and $r = 0.15$ (Right) fitted to the Strauss data in Figure 19 which were generated with $r = 0.1$. The plots correctly suggest that the models do not adequately fit the interpoint interaction in the data.

10.6 Application examples

The left panel of Figure 24 shows the 'Swedish pines' data of Strand and Ripley [90] which is frequently modelled as a realisation of a homogeneous Strauss process [39, 90, 91]. An

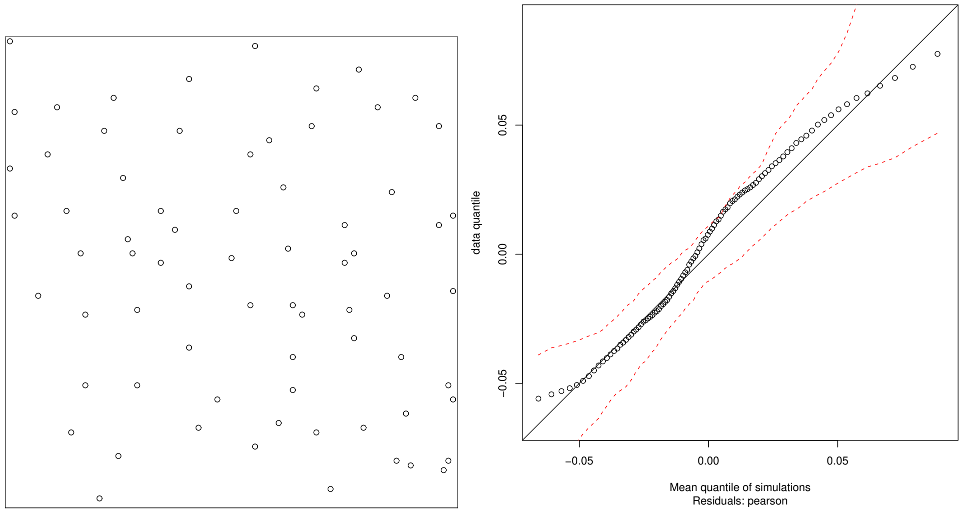

Figure 24: *Left:* Swedish pines data. *Right:* Q–Q plot for fitted Strauss process model.

interaction range of $r = 7$ (the window dimensions are 96×100) is frequently suggested.

We fitted the homogeneous Strauss model with $r = 7$ to these data, obtaining $\widehat{\gamma} = 0.2472$. The Q–Q plot is shown in the right panel of Figure 24. It suggests an adequate, but possibly marginal, fit.

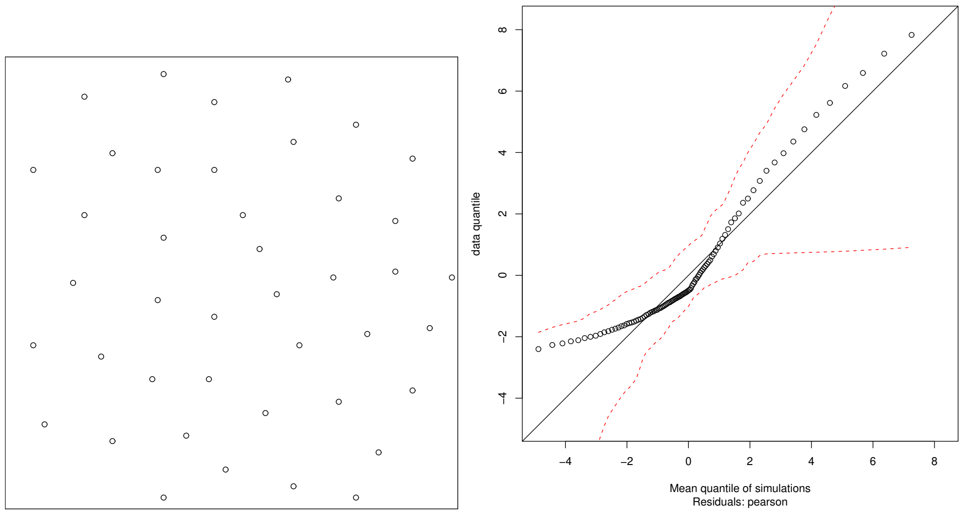

Figure 25: *Left:* Cells data. *Right:* Q–Q plot for fitted Strauss process model.

The left panel of Figure 25 shows Ripley's 'biological cells' dataset [89], which is frequently modelled as a realisation of a hard core process [39, 89, 90, 91] with hard core radius 0.05 (the window is the unit square). This corresponds to a Strauss process with interaction range $r = 0.1$ and interaction parameter $\gamma = 0$. The right panel of the Figure shows the Q-Q plot for a fitted Strauss process model with $r = 0.1$ fixed and $\widehat{\gamma} = 0.007$, very close to a hard-core process. The Q–Q plot suggests good agreement between data and model.

11 Interaction and trend

A more stringent trial for our diagnostic plots is to analyse data containing both spatial trend and interpoint interaction, and to fit models in which either component is misspecified. It is well known that spatial trend and clustering are often confounded, cf. [13, 14, 5].

11.1 Synthetic data

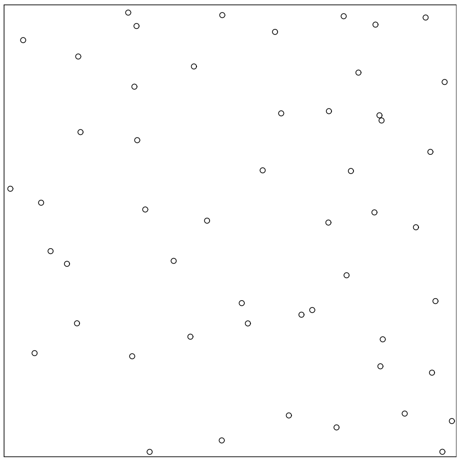

Figure 26: Simulated data from an heterogeneous Strauss process with log-quadratic activity function.

Figure 26 shows a synthetic dataset obtained from an heterogeneous Strauss process. The activity function was log-quadratic,

$$b(x, y) = 100 \exp(x + y + x^2) \tag{68}$$

and the interpoint interaction had range $r = 0.1$ and strength $\gamma = 0.1$. This pattern consists of 52 points in the unit square.

Figure 27 shows the diagnostic plots for a model of the correct form fitted to the data in Figure 26. The model has activity function

$$b(x, y) = \exp(\alpha_0 + \alpha_1 x + \alpha_2 y + \alpha_3 x^2) \tag{69}$$

and Strauss interpoint interaction with fixed range $r = 0.1$ and parameters β, γ. The plots suggest a good fit.

To assess the spatial trend, we first fit a model with the same interpoint interaction but with no spatial trend. Figure 28 shows the diagnostic plots for a stationary Strauss process with fixed interaction range $r = 0.1$. The trend plots show clearly that this model is inappropriate.

Next we fit an heterogeneous Strauss model with misspecified trend. The trend is loglinear,

$$b(x, y) = \exp(\alpha_0 + \alpha_1 x + \alpha_2 y)$$

and again the interpoint interaction is Strauss with $r = 0.1$. Figure 29 shows the diagnostic plots which suggest that this model has marginally incorrect spatial trend.

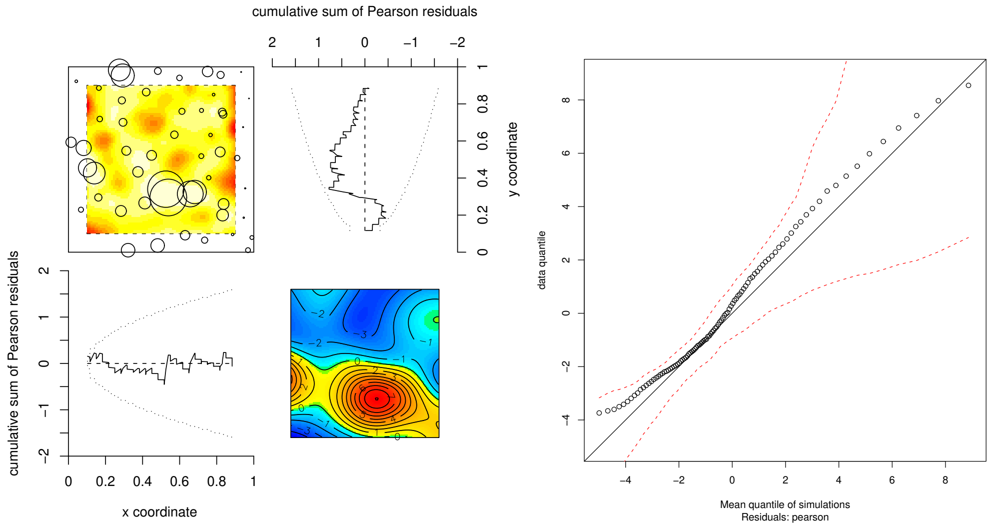

Figure 27: Diagnostics for a model of the correct form (heterogeneous Strauss process with log-quadratic activity) fitted to the data in Figure 26.

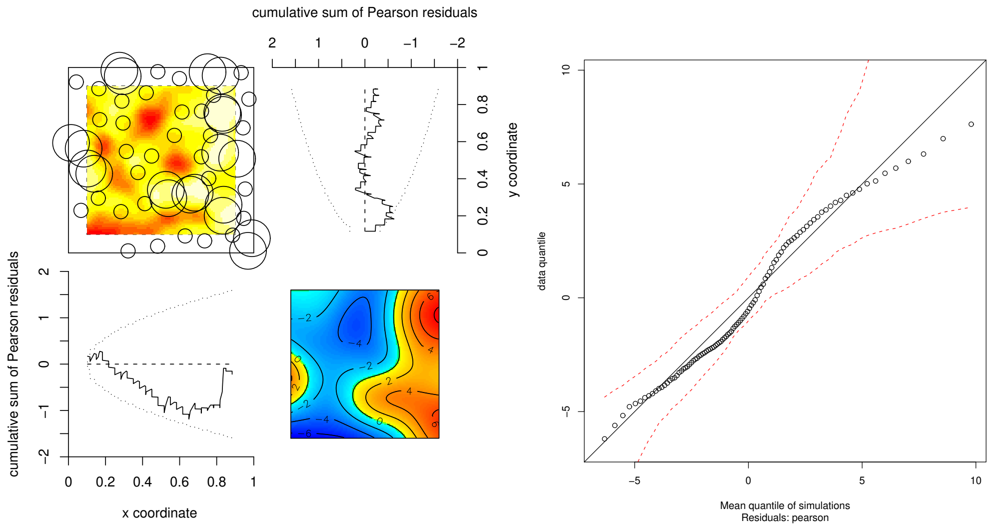

Figure 28: Diagnostics for a homogeneous Strauss process fitted to the data in Figure 26.

Turning to the interpoint interaction, we first fit a model with the correct form of spatial trend, but no interpoint interaction. An heterogeneous Poisson process with log-quadratic intensity function (69) was fitted to the data in Figure 26. The diagnostic plots are shown in Figure 30. The Q–Q plot demonstrates that this model is quite inappropriate.

Finally we try a model with an incorrectly specified interpoint interaction and the correct spatial trend. We fit an heterogeneous Strauss process with interaction range $r = 0.05$ and log-quadratic activity function (69) to the data in Figure 26. The diagnostic plots are shown in Figure 31. The Q–Q plot indicates that the interpoint interaction is misspecified.

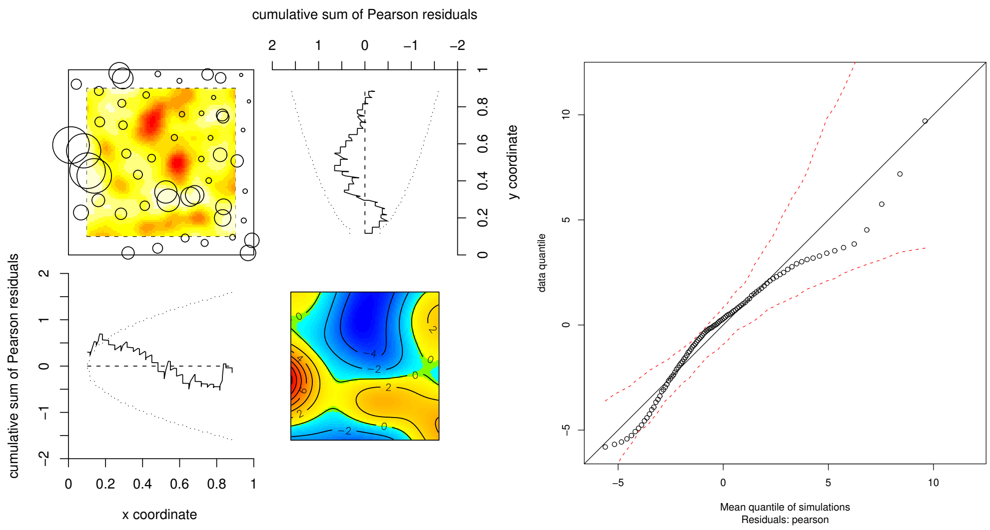

Figure 29: Diagnostics for an heterogeneous Poisson process with loglinear intensity fitted to the data in Figure 26.

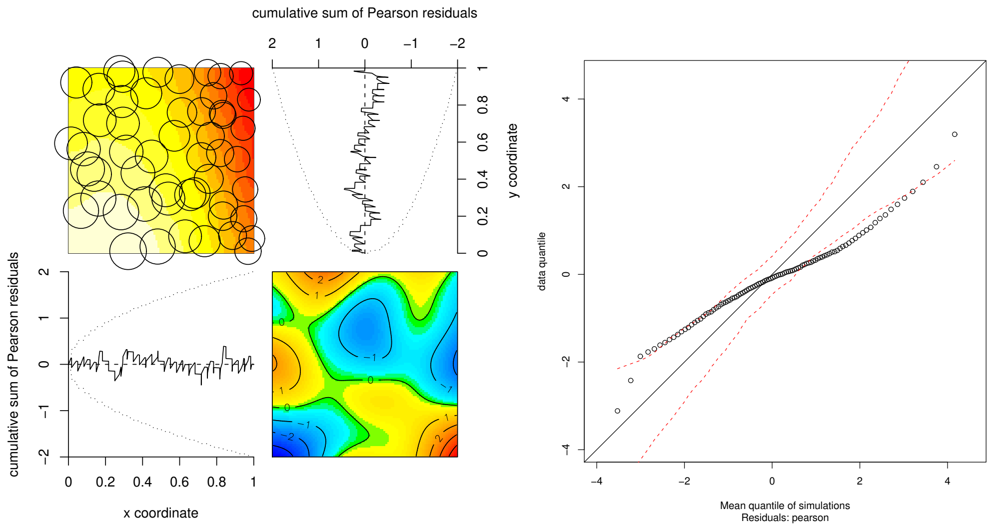

Figure 30: Diagnostics for an heterogeneous Poisson process with log-quadratic intensity fitted to the data in Figure 26.

11.2 Application

The Japanese pines data are shown in Figure 1. Ogata and Tanemura [84] fitted a pairwise interaction model involving both a spatially varying activity function b and a nontrivial interaction function c. The activity was log-polynomial (i.e. $\log b(u)$ was a polynomial in the Cartesian coordinates of u) and c was the 'soft-core' interaction

$$c(u,v) = \exp\left\{ - \left(\frac{\sigma}{||u-v||} \right)^{2/\kappa} \right\} \quad u \neq v \tag{70}$$

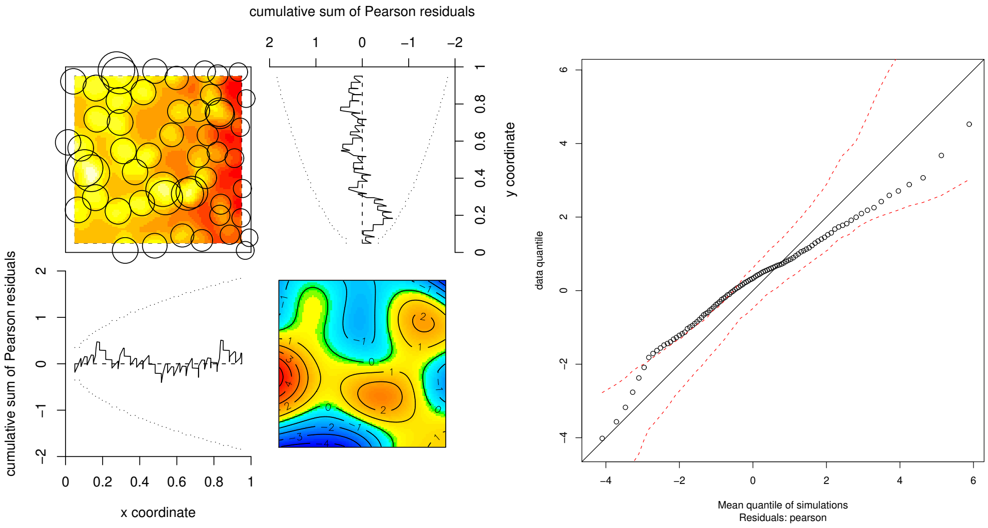

Figure 31: Diagnostics for a model of the correct form, but incorrect interaction range r, fitted to the data in Figure 26.

where $0 \leq \sigma < \infty$ and $0 < \kappa < 1$ are parameters. They chose a cubic polynomial as giving the optimal fit to these data.

Figure 32 shows the diagnostic plots for the same model, using Pearson residuals. They suggest good agreement between the fitted model and the data. However, the same can be said for Figure 33 which shows the diagnostic plots for a simpler model, in which $\log b$ is a quadratic rather than a cubic function of the Cartesian coordinates. Other plots (not shown here) suggest that all sub-models are inappropriate.

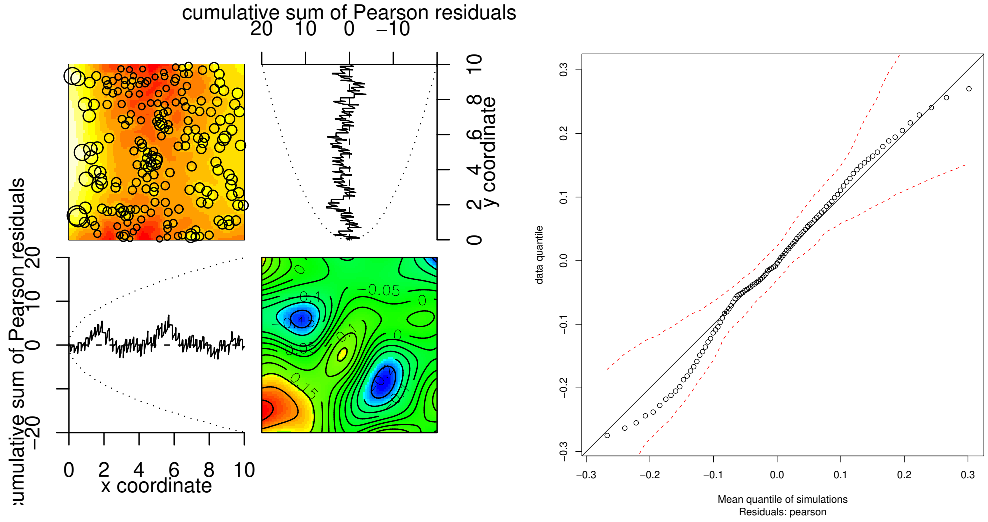

Figure 32: Diagnostics for the Ogata-Tanemura model (soft core with log-cubic trend) fitted to the Japanese pines data of Figure 1.

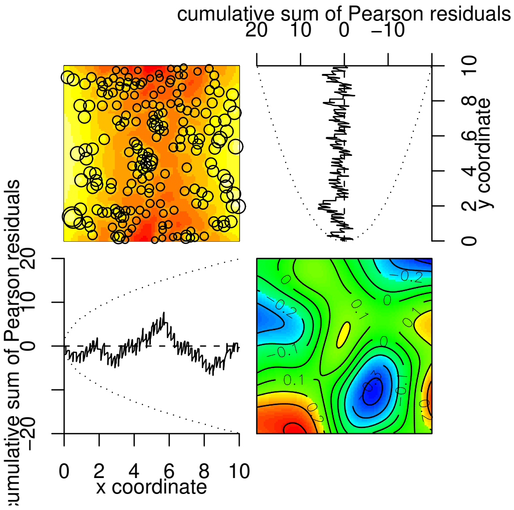

Figure 33: Diagnostics for the sub-model with log-quadratic trend.

12 Computation

This section explains the strategy we employed for computing the spatial residuals defined in section 6.

Some discretisation of the residual measure is required in order to compute it. For the present paper we fitted models by maximum pseudolikelihood using the Berman-Turner device, although this is not essential to the technique. It is therefore convenient to use the discretisation inherent in the Berman-Turner device to compute the residual measure.

Assume that we have decided on a quadrature scheme consisting of quadrature points u_j (comprising both data and dummy points) and quadrature weights w_j. No assumptions are made about this quadrature scheme except that we implicitly assume that integrals of the form $\int_B g(u)\, du$ are well approximated by their quadrature approximation $\sum_j \mathbf{1}\{u_j \in B\} g(u_j)$.

The residual measure $R(B, \widehat{h}, \widehat{\theta})$ will be approximated by

$$R^*(B) = \sum_j (y_j - \widehat{\lambda}_j) w_j \widehat{h}_j \tag{71}$$

where the sum is over all j such that $u_j \in B$, and we write $\widehat{h}_j = h_{\widehat{\theta}}(u_j, \mathbf{x} \setminus \{u_j\})$ and $\widehat{\lambda}_j = \lambda_{\widehat{\theta}}(u_j, \mathbf{x})$. As in the previous section, $y_j = z_j / w_j$, where z_j is the indicator equal to 1 if u_j is a data point and 0 for a dummy point. The values y_j correspond to the responses in the associated Poisson loglinear regression. An individual summand in this approximation,

$$r_j^h = (y_j - \lambda_j) w_j h_j = (z_j - \lambda_j w_j) h_j, \tag{72}$$

may be regarded as a residual attached to the quadrature point u_j (or as an approximation to the residual measure for the Dirichlet tile associated with u_j). The residuals r_j are summed without weights to obtain the approximate residual measure R^*.

For the raw, inverse-lambda and Pearson residuals, the discrete approximation (71) is

$$r_j^{(1)} \quad = \quad (z_j - \widehat{\lambda}_j w_j) \tag{73}$$

$$r_j^{(1/\lambda)} = \frac{z_j}{\sqrt{\widehat{\lambda}_j}} - w_j \tag{74}$$

$$r_j^{(1/\sqrt{\lambda})} = \frac{z_j - \widehat{\lambda}_j w_j}{\sqrt{\widehat{\lambda}_j}}. \tag{75}$$

Note that the residuals attached to the data points x_i are different in the continuous theory (Section 6) and in the discrete approximation. For example, the raw residuals defined in Section 6.3 attach a unit mass to each data point x_i, while the discretised raw residuals r_j^1 defined in (73) above are equal to $1 - w_j$ for a data point. This discrepancy occurs because in the discrete approximation, the residual attached to each quadrature point is an approximation to the residual measure (sum of all residuals) in the tile associated with u_j.

In other statistical modelling contexts where a weighted likelihood is used, the Pearson residuals would typically be defined as

$$r_j^* = \sqrt{w_j}\left(\frac{y_j}{\sqrt{\widehat{\lambda}_j}} - \sqrt{\widehat{\lambda}_j}\right)$$

so that the sum of the squared Pearson residuals equals the Pearson X^2 statistic. However, this does not apply in our case.

The pseudoscore residuals (Section 7.5) have discretised counterpart

$$s_j z_j - s_j \lambda_j w_j = s_j r_j^{(1)}$$

where $s_j = (\mathrm{d}/\mathrm{d}\theta) \log \lambda_\theta(u_j, \mathbf{x})$ and $r_j^{(1)}$ are again the raw residuals. In the common case where the conditional intensity is loglinear, equation (38), we have $s_j = S(u_j, \mathbf{x})$ so that the pseudoscore residuals are the raw residuals weighted by the "sufficient statistic" S.

Software for computing and plotting the residuals from a Gibbs point process model has been incorporated in our library `spatstat` [7] in the statistical package R.

Acknowledgements

We are most grateful to Kasper K. Berthelsen for providing code for perfect simulation of the Strauss process. We thank David Brillinger, Peter Green, Peter Guttorp, Tony Pakes, Dietrich Stoyan, Paul Switzer, Berwin Turlach, David Vere-Jones, Rick Vitale and Joe Whittaker for illuminating comments. This research was supported by the Australian Research Council (Large Grant A69941083 *Extrapolating and interpolating spatial patterns*) and by NSERC and MaPhySto.

References

[1] S.E. Alm. Approximation and simulation of the distributions of scan statistics for Poisson processes in higher dimensions. *Extremes*, 1(1):111–126, 1988.

[2] P.K. Andersen, Ø. Borgan, R.D. Gill, and N. Keiding. *Statistical Models Based on Counting Processes.* Springer-Verlag, New York, 1993.

[3] L. Anselin. Local indicators of spatial association – LISA. *Geographical analysis*, 27:93–115, 1995.

[4] A.C. Atkinson. *Plots, Transformations and Regression.* Number 1 in Oxford Statistical Science Series. Oxford University Press/ Clarendon, 1985.

[5] A. Baddeley, J. Møller, and R. Waagepetersen. Non- and semiparametric estimation of interaction in inhomogeneous point patterns. *Statistica Neerlandica*, 54(3):329–350, November 2000.

[6] A. Baddeley and R. Turner. Fitting spatial point process models in `spatstat`. In preparation.

[7] A. Baddeley and R. Turner. `Spatstat:` spatial point pattern analysis in R. In preparation.

[8] A. Baddeley and R. Turner. Practical maximum pseudolikelihood for spatial point patterns (with discussion). *Australian and New Zealand Journal of Statistics*, 42(3):283–322, 2000.

[9] A. Baddeley and R. Turner. Addendum to: "Practical maximum pseudolikelihood for spatial point patterns (with discussion)". *Australian and New Zealand Journal of Statistics*, 44:503, 2002.

[10] A.J. Baddeley. Spatial sampling and censoring. In O.E. Barndorff-Nielsen, W.S. Kendall, and M.N.M. van Lieshout, editors, *Stochastic Geometry: Likelihood and Computation*, chapter 2, pages 37–78. Chapman and Hall, London, 1998.

[11] A.J. Baddeley. Time-invariance estimating equations. *Bernoulli*, 6(5):783–808, 2000.

[12] A.J. Baddeley, M.N.M. van Lieshout, and J. Møller. Markov properties of cluster processes. *Advances in Applied Probability*, 28:346–355, 1996.

[13] M.S. Bartlett. A note on spatial pattern. *Biometrics*, 20:891–892, 1964.

[14] M.S. Bartlett. The spectral analysis of two-dimensional point processes. *Biometrika*, 51:299–311, 1964.

[15] M. Berman. Testing for spatial association between a point process and another stochastic process. *Applied Statistics*, 35:54–62, 1986.

[16] M. Berman and P.J. Diggle. Estimating weighted integrals of the second-order intensity of a spatial point process. *Journal of the Royal Statistical Society, series B*, 51:81–92, 1989.

[17] M. Berman and T.R. Turner. Approximating point process likelihoods with GLIM. *Applied Statistics*, 41:31–38, 1992.

[18] J. Besag. Statistical analysis of non-lattice data. *The Statistician*, 24:179–195, 1975.

[19] D.R. Brillinger. Comparative aspects of the study of ordinary time series and of point processes. In *Developments in Statistics*, pages 33–133. Academic Press, 1978.

[20] D.R. Brillinger. Maximum likelihood analysis of spike trains of interacting nerve cells. *Biological Cybernetics*, 59:189–200, 1988.

[21] D.R. Brillinger. Time series, point processes, and hybrids. *Canadian Journal of Statistics*, 22:177–206, 1994.

[22] D.R. Brillinger and H.K. Preisler. Two examples of quantal data analysis: a) multivariate point process, b) pure death process in an experimental design. In *Proceedings, XIII International Biometric Conference, Seattle*, 1986.

[23] D.R. Brillinger and J.P. Segundo. Empirical examination of the threshold model of neuron firing. *Biological Cybernetics*, 35:213–220, 1979.

[24] M. Clyde and D. Strauss. Logistic regression for spatial pair-potential models. In A. Possolo, editor, *Spatial Statistics and Imaging*, volume 20 of *Lecture Notes - Monograph series*, chapter II, pages 14–30. Institute of Mathematical Statistics, 1991. ISBN 0-940600-27-7.

[25] D. Collett. *Modelling Binary Data*. Chapman and Hall, London, 1991.

[26] D. R. Cox and V. Isham. *Point Processes*. Chapman and Hall, London, 1980.

[27] N. Cressie and L.B. Collins. Analysis of spatial point patterns using bundles of product density lisa functions. *J Agric Biol Environ Stat*, 6:118–135, 2001.

[28] N. Cressie and L.B. Collins. Patterns in spatial point locations: local indicators of spatial association in a minefield with clutter. *Naval Research Logistics*, 48:333–347, 2001.

[29] N. Cressie and G.M. Laslett. Random set theory and problems of modeling. *SIAM Review*, 28:557–574, 1987.

[30] N.A.C. Cressie. *Statistics for Spatial Data*. John Wiley and Sons, New York, 1991.

[31] D.J. Daley and D. Vere-Jones. *An introduction to the theory of point processes*. Springer Verlag, New York, 1988.

[32] D.J. Daley and D. Vere-Jones. *An introduction to the theory of point processes*. Springer, 2nd ed. edition, 2003.

[33] A.C. Davison and E.J. Snell. Residuals and diagnostics. In D.V. Hinkley, N. Reid, and E.J. Snell, editors, *Statistical theory and modelling (in honour of Sir David Cox FRS)*, chapter 4, pages 83–106. Chapman and Hall, London, 1991.

[34] P. Diggle, A. Chetwynd, R. Haggkvist, and S. Morris. Second order analysis of space-time clustering. *Statistical Methods in Medical Research*, 4:124–136, 1995.

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

[36] P.J. Diggle. *Statistical Analysis of Spatial Point Patterns.* Academic Press, London, 1983.

[37] P.J. Diggle. A kernel method for smoothing point process data. *Applied Statistics*, 34:138–147, 1985.

[38] P.J. Diggle. A point process modelling approach to raised incidence of a rare phenomenon in the vicinity of a prespecified point. *Journal of the Royal Statistical Society, series A*, 153:349–362, 1990.

[39] P.J. Diggle. *Statistical Analysis of Spatial Point Patterns.* Arnold, second edition, 2003.

[40] P.J. Diggle, T. Fiksel, P. Grabarnik, Y. Ogata, D. Stoyan, and M. Tanemura. On parameter estimation for pairwise interaction processes. *International Statistical Review*, 62:99–117, 1994.

[41] P.J. Diggle and R.J. Gratton. Monte Carlo methods of inference for implicit statistical models. *Journal of the Royal Statististical Society, Series B*, 46:193–227, 1984. (with discussion).

[42] P. Elliott and J. Wakefield. Disease clusters: should they be investigated, and if so, when and how? *Journal of the Royal Statistical Society, Series A*, 164:3–12, 2001.

[43] P. Elliott, J. Wakefield, N. Best, and D. Briggs, editors. *Spatial Epidemiology: Methods and Applications.* Oxford University Press, Oxford, 2000.

[44] T. Fiksel. Estimation of parameterized pair potentials of marked and non-marked Gibbsian point processes. *Elektronische Informationsverarbeitung u. Kybernetika*, 20:270–278, 1984.

[45] T.R. Fleming and D.P. Harrington. *Counting processes and survival analysis.* Wiley, New York, 1991.

[46] E.B. Fowlkes. Some diagnostics for binary logistic regression via smoothing. *Biometrika*, 74:503–515, 1987.

[47] R. Foxall and A. Baddeley. Nonparametric measures of association between a spatial point process and a random set, with geological applications. *Applied Statistics*, 51(2):165–182, 2002.

[48] H.-O. Georgii. Canonical and grand canonical Gibbs states for continuum systems. *Communications of Mathematical Physics*, 48:31–51, 1976.

[49] A. Getis and J. Franklin. Second-order neighbourhood analysis of mapped point patterns. *Ecology*, 68:473–477, 1987.

[50] C.J. Geyer. Likelihood inference for spatial point processes. In O.E. Barndorff-Nielsen, W.S. Kendall, and M.N.M. van Lieshout, editors, *Stochastic Geometry: Likelihood and Computation*, number 80 in Monographs on Statistics and Applied Probability, chapter 3, pages 79–140. Chapman and Hall / CRC, Boca Raton, Florida, 1999.

[51] C.J. Geyer and J. Møller. Simulation procedures and likelihood inference for spatial point processes. *Scandinavian Journal of Statistics*, 21(4):359–373, 1994.

[52] E. Glötzl. Bemerkungen zu einer Arbeit von O.K. Kozlov. *Mathematische Nachrichten*, 94:277–289, 1980.

[53] E. Glötzl. Lokale Energien und Potentiale für Punktprozessen. *Mathematische Nachrichten*, 96:195–206, 1980.

[54] R. Gnanadesikan and M.B. Wilk. A probability pllotting procedure for general analysis of variance. *Journal of the Royal Statistical Society, series B*, 32:88–101, 1970.

[55] D.G. Horvitz and D.J. Thompson. A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47:663–685, 1952.

[56] J.L. Jensen and J. Møller. Pseudolikelihood for exponential family models of spatial point processes. *Annals of Applied Probability*, 1:445–461, 1991.

[57] J.D. Kalbfleisch and R.L. Prentice. *The Statistical Analysis of Failure Time Data*. Wiley, 1980.

[58] O. Kallenberg. On conditional intensities of point processes. *Zeitschrift fuer Wahrscheinlichkeitstheorie und verwandte Gebiete*, 41:205–220, 1978.

[59] O. Kallenberg. *Random Measures*. Akademie Verlag/Academic Press, Berlin/New York, third edition, 1983.

[60] O. Kallenberg. An informal guide to the theory of conditioning in point processes. *International Statistical Review*, 52:151–164, 1984.

[61] A.F. Karr. *Point processes and their statistical inference*. Dekker, New York, 1985.

[62] F.P. Kelly and B.D. Ripley. On Strauss's model for clustering. *Biometrika*, 63:357–360, 1976.

[63] W. S. Kendall. Perfect simulation for the area-interaction point process. In L. Accardi and C.C. Heyde, editors, *Probability Towards 2000*, pages 218–234. Springer Lecture Notes in Statistics 128, Springer Verlag, New York, 1998.

[64] W. S. Kendall and J. Møller. Perfect simulation using dominating processes on ordered spaces, with application to locally stable point processes. *Advances in Applied Probability*, 32:844–865, 2000.

[65] J.F.C. Kingman. *Poisson Processes*. Oxford University Press, 1993.

[66] O.K. Kozlov. Gibbsian description of point random fields. *Theory of Probability and its Applications*, 21:339–355, 1976.

[67] M. Kulldorff. Spatial scan statistics: models, calculations, and applications. In J. Glaz and N. Balakrishnan, editors, *Recent Advances on Scan Statistics*, pages 303–322. Birkhauser, Boston, 1999.

[68] J.M. Landwehr, D. Pregibon, and A.C. Shoemaker. Graphical methods for assessing logistic regression models. *Journal of the American Statistical Association*, 79:61–83, 1984.

[69] A. Lawson. On fitting non-stationary Markov point process models on GLIM. In Y. Dodge and J. Whittaker, editors, *COMPSTAT: Proceedings 10th Symposium on Computational Statistics*, volume 1, pages 35–40. Physica Verlag, 1992.

[70] A.B. Lawson. A deviance residual for heterogeneous spatial Poisson processes. *Biometrics*, 49:889–897, 1993.

[71] P.A.W. Lewis. Recent results in the statistical analysis of univariate point processes. In P.A.W. Lewis, editor, *Stochastic point processes*, pages 1–54. Wiley, New York, 1972.

[72] J.K. Lindsey. *The Analysis of Stochastic Processes using GLIM.* Springer, Berlin, 1992.

[73] J.K. Lindsey. Fitting parametric counting processes by using linear models. *Applied Statistics*, 44:201–212, 1995.

[74] J.K. Lindsey. Fitting bivariate intensity functions, with an application to modelling delays in reporting acquired immune deficiency syndrome. *Journal of the Royal Statistical Society, Series A*, 159:125–131, 1996.

[75] J.K. Lindsey and G. Mersch. Fitting and comparing probability distributions with log linear models. *Computational Statistics and Data Analysis*, 13:373–384, 1992.

[76] P. McCullagh and J.A. Nelder. *Generalized linear models.* Chapman and Hall, London, second edition, 1989.

[77] E. Merzbach and D. Nualart. A characterization of the spatial Poisson process and changing time. *Annals of Probability*, 14:1380–1390, 1986.

[78] J. Møller. Shot noise Cox processes. *Advances in Applied Probability*, 35, 2003. To appear.

[79] J. Møller and R.P. Waagepetersen. *Statistical Inference and Simulation for Spatial Point Processes.* Chapman and Hall/CRC, Boca Raton, 2003.

[80] M.G. Nair. Random space change for multiparameter point processes. *Annals of Probability*, 18(3):1222–1231, 1990.

[81] X. X. Nguyen and H. Zessin. Integral and differential characterizations of Gibbs processes. *Mathematische Nachrichten*, 88:105–115, 1979.

[82] M. Numata. Forest vegetation, particularly pine stems in the vicinity of Choshi — coastal flora and vegetation in Choshi, Chiba prefecture, IV (in Japanese). *Bulletin of the Choshi Marine Laboratory*, (6):27–37, 1964. Chiba University.

[83] Y. Ogata. Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical Association*, 83:9–27, 1988.

[84] Y. Ogata and M. Tanemura. Likelihood estimation of interaction potentials and external fields of inhomogeneous spatial point patterns. In I.S. Francis, B.J.F. Manly, and F.C. Lam, editors, *Pacific Statistical Congress*, pages 150–154. Elsevier, 1986.

[85] F. Papangelou. The conditional intensity of general point processes and an application to line processes. *Zeitschrift fuer Wahscheinlichkeitstheorie und verwandte Gebiete*, 28:207–226, 1974.

[86] R. D. Peng, F. P. Schoenberg, and J. Woods. Multi-dimensional point process models for evaluating a wildfire hazard index. Submitted, 2003.

[87] D. Pregibon. Logistic regression diagnostics. *Annals of Statistics*, 9:705–724, 1981.

[88] B.D. Ripley. The second-order analysis of stationary point processes. *jap*, 13:255–266, 1976.

[89] B.D. Ripley. Modelling spatial patterns (with discussion). *Journal of the Royal Statistical Society, Series B*, 39:172 – 212, 1977.

[90] B.D. Ripley. *Spatial Statistics.* John Wiley and Sons, New York, 1981.

[91] B.D. Ripley. *Statistical Inference for Spatial Processes.* Cambridge University Press, 1988.

[92] B.D. Ripley. Gibbsian interaction models. In D.A. Griffiths, editor, *Spatial Statistics: Past, Present and Future*, pages 1–19. Image, New York, 1989.

[93] B.D. Ripley and F.P. Kelly. Markov point processes. *Journal of the London Mathematical Society*, 15:188–192, 1977.

[94] A. Särkkä. *Pseudo-likelihood approach for pair potential estimation of Gibbs processes.* Number 22 in Jyväskylä Studies in Computer Science, Economics and Statistics. University of Jyväskylä, 1993.

[95] F. Schoenberg. Transforming spatial point processes into Poisson processes. *Stochastic Processes and their Applications*, 81:155–164, 1999.

[96] D. Stoyan and P. Grabarnik. Second-order characteristics for stochastic structures connected with Gibbs point processes. *Mathematische Nachrichten*, 151:95–100, 1991.

[97] D. Stoyan, W.S. Kendall, and J. Mecke. *Stochastic Geometry and its Applications.* John Wiley and Sons, Chichester, second edition, 1995.

[98] D. Stoyan and H. Stoyan. *Fractals, Random Shapes and Point Fields.* Wiley, 1995.

[99] R. Takacs. Estimator for the pair potential of a Gibbsian point process. *Statistics*, 17:429–433, 1986.

[100] R. Takacs and T. Fiksel. Interaction pair-potentials for a system of ants' nests. *Biometrical Journal*, 28:1007–1013, 1986.

[101] P.C.T. van der Hoeven. Une projection de processus ponctuels. *Zeitschrift fuer Wahscheinlichkeitstheorie und verwandte Gebiete*, 61:483–499, 1982.

[102] M.N.M. van Lieshout. *Markov point processes and their applications.* Imperial College Press, 2000.

[103] W.N. Venables and B.D. Ripley. *Modern Applied Statistics with S-Plus.* Springer, second edition, 1997.

[104] D. Vere-Jones. Stochastic models for earthquake occurrence (with discussion). *Journal of the Royal Statistical Society, series B*, 32:1–62, 1970.

[105] D. Vere-Jones and F. Paik Schoenberg. Rescaling marked point processes. *Australian and New Zealand Journal of Statistics*, 46(1):133–143, 2004.

[106] D. Wartenberg. Exploratory spatial analyses: outliers, leverage points, and influence functions. In D.A. Griffith, editor, *Spatial statistics: past, present and future*, pages 133–162. Institute of Mathematical Geography, Ann Arbor, Michigan, USA, 1990.

APPENDICES

A Variance calculations

In this Appendix we derive general formulae for the variance of the innovations and residuals. We assume $W \subset \mathbb{R}^d$ is a compact set with nonempty interior, and $\mathbf{X}$ is a Gibbs point process in W with probability density f with respect to the unit rate Poisson process.

A.1 Two-point conditional intensity

Let $\mathbf{X} \otimes \mathbf{X}$ be the point process on $W \times W$ consisting of all pairs (x_i, x_j) of *distinct* points of $\mathbf{X}$. Assume $\mathbf{X} \otimes \mathbf{X}$ is also a Gibbs process. Then $\mathbf{X} \otimes \mathbf{X}$ has a Papangelou conditional intensity $\lambda((u,v), \mathbf{X} \otimes \mathbf{X})$ which we call the *two-point conditional intensity* of $\mathbf{X}$, and denote by $\lambda(u, v; \mathbf{X})$. It is defined to satisfy the Georgii-Nguyen-Zessin formula in the form

$$\mathbb{E}\left[\sum_{x_i \neq x_j} h(x_i, x_j, \mathbf{X} \setminus \{x_i, x_j\})\right] = \mathbb{E}\left[\int\int h(u, v, \mathbf{X}) \lambda(u, v, \mathbf{X}) \, \mathrm{d}u \, \mathrm{d}v\right] \tag{76}$$

for any nonnegative measurable function $h(u, v, \mathbf{x})$, where the sum is over all ordered pairs of distinct points x_i, x_j of the process $\mathbf{X}$.

If $\mathbf{X}$ has probability density f then the two-point conditional intensity can be written

$$\lambda(u, v, \mathbf{x}) = \lambda(u, \mathbf{x} \setminus \{v\}) \lambda(v, \mathbf{x} \cup \{u\}) = f(\mathbf{x} \cup \{u, v\}) / f(\mathbf{x} \setminus \{u, v\}).$$

Note that $\lambda(u, v, \mathbf{x})$ is symmetric in u and v. In the statistical physics interpretation, $\log \lambda(u, v, \mathbf{x})$ is the energy required to add the two points u and v to the configuration $\mathbf{x}$.

Later we use the fact that a pairwise interaction Gibbs process with one-point conditional intensity (13) has

$$\lambda(u, v, \mathbf{x}) = \lambda(u, \mathbf{x} \setminus \{v\}) \, \lambda(v, \mathbf{x} \setminus \{u\}) \, c(u, v). \tag{77}$$

A.2 Variance of a sum

From (14) and (76) we easily get the following:

Lemma 1 (Variance of a sum) *For any nonnegative measurable function h,*

$$\begin{aligned} \mathsf{var}\left[\sum h(x_i, \mathbf{X} \setminus \{x_i\})\right] &= \int\int \mathbb{E}\left[h(u, \mathbf{X} \cup \{v\}) h(v, \mathbf{X} \cup \{u\}) \lambda(u, v, \mathbf{X})\right] \, \mathrm{d}u \, \mathrm{d}v \\ &\quad + \int \mathbb{E}\left[h(u, \mathbf{X})^2 \, \lambda(u, \mathbf{X})\right] \, \mathrm{d}u \\ &\quad - \left(\int \mathbb{E}\left[h(u, \mathbf{X}) \lambda(u, \mathbf{X})\right] \, \mathrm{d}u\right)^2 \end{aligned} \tag{78}$$

Proof:

Expand

$$\begin{aligned}
\mathsf{var}\left[\sum h(x_i, \mathbf{X} \setminus \{x_i\})\right] &= \mathbb{E}\left[\sum_{x_i \neq x_j} h(x_i, \mathbf{X} \setminus \{x_i\}) h(x_j, \mathbf{X} \setminus \{x_j\})\right] \\
&\quad + \mathbb{E}\left[\sum h(x_i, \mathbf{X} \setminus \{x_i\})^2\right] \\
&\quad - \left(\mathbb{E}\left[\sum h(x_i, \mathbf{X} \setminus \{x_i\})\right]\right)^2.
\end{aligned}$$

The second and third terms on the right hand side can be evaluated directly using (14), while the first term follows from (76). □

For example, for a Poisson process with intensity function $\lambda(u)$, (78) yields

$$\begin{aligned}
\mathsf{var}\left[\sum h(x_i, \mathbf{X} \setminus \{x_i\})\right] &= \int\int \mathbb{E}\left[h(u, \mathbf{X} \cup \{v\}) h(v, \mathbf{X} \cup \{u\}) \lambda(u)\lambda(v)\right] \, \mathrm{d}u \, \mathrm{d}v \\
&\quad + \int \mathbb{E}\left[h(u, \mathbf{X})^2\right] \lambda(u) \, \mathrm{d}u \\
&\quad - \left(\int \mathbb{E}\left[h(u, \mathbf{X}) \lambda(u)\right] \, \mathrm{d}u\right)^2
\end{aligned}$$

which in the special case $h(u, \mathbf{x}) = h(u)$ reduces to

$$\mathsf{var}\left[\sum h(x_i)\right] = \int h(u)^2 \lambda(u) \, \mathrm{d}u$$

as expected by the independence properties of the Poisson process.

Lemma 1 also gives a connection between the two-point conditional intensity and the second moments of the point process $\mathbf{X}$. Taking $h(u, \mathbf{x}) = \mathbf{1}\{u \in B\}$ in (78), we obtain

$$\mathsf{var}\left[n(\mathbf{X} \cap B)\right] = \int_B \int_B \mathbb{E}\left[\lambda(u, v, \mathbf{X})\right] \, \mathrm{d}u \, \mathrm{d}v + \mu(B) - \mu(B)^2 \tag{79}$$

where

$$\mu(B) = \mathbb{E}n(\mathbf{X} \cap B) = \int_B \mathbb{E}\lambda(u, \mathbf{X}) \, \mathrm{d}u$$

is the intensity measure. In particular we get the well-known result that the pair correlation function [79], [98, p. 262 ff.] of $\mathbf{X}$ equals

$$g(u, v) = \frac{\mathbb{E}\left[\lambda(u, v, \mathbf{X})\right]}{\mathbb{E}\left[\lambda(u, \mathbf{X})\right] \mathbb{E}\left[\lambda(v, \mathbf{X})\right]}. \tag{80}$$

Applying (78) to the sum of the Stoyan-Grabarnik exponential energy marks, we find that

$$\begin{aligned}
\mathsf{var}\left[\sum_{x_i \in B} \frac{1}{\lambda(x_i, \mathbf{X})}\right] &= \int_B \int_B \mathbb{E}\left[\frac{\lambda(u, v, \mathbf{X})}{\lambda(u, \mathbf{X} \cup \{v\}) \lambda(v, \mathbf{X} \cup \{u\})}\right] \, \mathrm{d}u \, \mathrm{d}v \\
&\quad + \int_B \mathbb{E}\left[\frac{1}{\lambda(u, \mathbf{X})}\right] \, \mathrm{d}u - |B|^2
\end{aligned} \tag{81}$$

For a pairwise interaction process (77) yields

$$\mathsf{var}\left[\sum_{x_i \in B} \frac{1}{\lambda(x_i, \mathbf{X})}\right] = \int_B \int_B \frac{1}{c(u,v)} \,\mathrm{d}u\,\mathrm{d}v + \int_B \mathbb{E}\left[\frac{1}{\lambda(u, \mathbf{X})}\right] \,\mathrm{d}u - |B|^2.$$

This was derived in [96] in the "stationary" case where $b(u) \equiv \beta$, $c(u,v) = c(u-v)$. For a Poisson process with intensity function $\lambda(\cdot)$ we obtain

$$\mathsf{var}\left[\sum_{x_i \in B} \frac{1}{\lambda(x_i)}\right] = \int_B \frac{1}{\lambda(u)} \,\mathrm{d}u. \tag{82}$$

A.3 Variance of innovations

The technique used in section A.2 can also be used to compute the variance of terms like

$$\sum h(x_i, \mathbf{X} \setminus \{x_i\}) - \int_W g(u, \mathbf{X}) \,\mathrm{d}u.$$

Lemma 2 *For any nonnegative measurable functions $h(u, \mathbf{x})$ and $g(u, \mathbf{x})$,*

$$\begin{aligned}
& \mathsf{var}\left[\sum h(x_i, \mathbf{X} \setminus \{x_i\}) - \int_W g(u, \mathbf{X}) \,\mathrm{d}u\right] \\
= & \int\int \mathbb{E}\left[h(u, \mathbf{X} \cup \{v\}) h(v, \mathbf{X} \cup \{u\}) \lambda(u, v, \mathbf{X})\right] \,\mathrm{d}u\,\mathrm{d}v \\
& + \int \mathbb{E}\left[h(u, \mathbf{X})^2 \,\lambda(u, \mathbf{X})\right] \,\mathrm{d}u \\
& - \left(\int \mathbb{E}\left[h(u, \mathbf{X}) \lambda(u, \mathbf{X})\right] \,\mathrm{d}u\right)^2 \\
& + \int_W \int_W \mathsf{cov}\left(g(u, \mathbf{X}), g(v, \mathbf{X})\right) \,\mathrm{d}u\,\mathrm{d}v \\
& -2 \int_W \int_W \mathbb{E}\left[h(u, \mathbf{X}) g(v, \mathbf{X} \cup \{u\}) \lambda(u, \mathbf{X})\right] \,\mathrm{d}u\,\mathrm{d}v \\
& +2 \left(\int_W \mathbb{E}\left[h(u, \mathbf{X}) \lambda(u, \mathbf{X})\right] \,\mathrm{d}u\right) \left(\int_W \mathbb{E}\, g(u, \mathbf{X}) \,\mathrm{d}u\right).
\end{aligned} \tag{83}$$

In the special case $g(u, \mathbf{x}) = \lambda(u, \mathbf{x}) h(u, \mathbf{x})$ relevant to the Georgii-Nguyen-Zessin formula (14), this becomes

$$\begin{aligned}
& \mathsf{var}\left[\sum h(x_i, \mathbf{X} \setminus \{x_i\}) - \int_W \lambda(u, \mathbf{x}) h(u, \mathbf{X}) \,\mathrm{d}u\right] \\
= & \int\int \mathbb{E}\left[h(u, \mathbf{X} \cup \{v\}) h(v, \mathbf{X} \cup \{u\}) \lambda(u, v, \mathbf{X})\right] \,\mathrm{d}u\,\mathrm{d}v \\
& + \int \mathbb{E}\left[h(u, \mathbf{X})^2 \,\lambda(u, \mathbf{X})\right] \,\mathrm{d}u
\end{aligned}$$

$$
\begin{aligned}
&+\int_W \int_W \mathbb{E}\left[h(u,\mathbf{X})h(v,\mathbf{X})\lambda(u,\mathbf{X})\lambda(v,\mathbf{X})\right] \,\mathrm{d}u\,\mathrm{d}v \\
&-2\int_W \int_W \mathbb{E}\left[h(u,\mathbf{X})h(v,\mathbf{X}\cup\{u\})\lambda(u,v,\mathbf{X})\right] \,\mathrm{d}u\,\mathrm{d}v.
\end{aligned} \tag{84}
$$

Proof:

Expanding the variance using standard properties of variances,

$$
\begin{aligned}
& \mathsf{var}\left[\sum h(x_i,\mathbf{X}\setminus\{x_i\}) - \int_W g(u,\mathbf{X})\,\mathrm{d}u\right] \\
= \; & \mathsf{var}\left[\sum h(x_i,\mathbf{X}\setminus\{x_i\})\right] \\
& + \mathsf{var}\left[\int_W g(u,\mathbf{X})\,\mathrm{d}u\right] \\
& -2\mathsf{cov}\left(\sum h(x_i,\mathbf{X}\setminus\{x_i\}), \int_W g(u,\mathbf{X})\,\mathrm{d}u\right) \\
= \; & \mathsf{var}\left[\sum h(x_i,\mathbf{X}\setminus\{x_i\})\right] \\
& + \int_W \int_W \mathsf{cov}\left(g(u,\mathbf{X}), g(v,\mathbf{X})\right)\,\mathrm{d}u\,\mathrm{d}v \\
& -2\mathbb{E}\left[\sum h(x_i,\mathbf{X}\setminus\{x_i\}) \int_W g(u,\mathbf{X})\,\mathrm{d}u\right] \\
& +2\mathbb{E}\left[\sum h(x_i,\mathbf{X}\setminus\{x_i\})\right] \mathbb{E}\left[\int_W g(u,\mathbf{X})\,\mathrm{d}u\right].
\end{aligned}
$$

The first term above is now expanded using Lemma 1. The third term is evaluated by putting

$$
k(v,\mathbf{x}) = h(v,\mathbf{x}) \int_W g(u,\mathbf{x}\cup\{v\})\,\mathrm{d}u \qquad (\text{for } v \notin \mathbf{x})
$$

so that

$$
k(x_i,\mathbf{x}\setminus\{x_i\}) = h(x_i,\mathbf{x}\setminus\{x_i\}) \int_W g(u,\mathbf{x})\,\mathrm{d}u, \qquad x_i \in \mathbf{x}
$$

and applying (14) gives

$$
\begin{aligned}
& \mathbb{E}\left[\sum h(x_i,\mathbf{X}\setminus\{x_i\}) \int_W g(u,\mathbf{X})\,\mathrm{d}u\right] \\
= \; & \mathbb{E}\left[\sum k(x_i,\mathbf{X}\setminus\{x_i\})\right] \\
= \; & \int_W \mathbb{E}\left[k(u,\mathbf{X})\lambda(u,\mathbf{X})\right]\,\mathrm{d}u \\
= \; & \int_W \mathbb{E}\left[h(u,\mathbf{X})\lambda(u,\mathbf{X}) \int_W g(v,\mathbf{X}\cup\{u\})\,\mathrm{d}v\right]\,\mathrm{d}u
\end{aligned}
$$

Thereby (83) is verified. The special case (84) follows trivially. □

For the Pearson innovations (34), setting $h(u,\mathbf{x}) = \mathbf{1}\{u \in B\}/\sqrt{\lambda(u,\mathbf{x})}$ in (84) we get

$$
\begin{aligned}
&\mathsf{var}\left[\sum_{x_i \in B} \frac{1}{\sqrt{\lambda(x_i,\mathbf{X})}} - \int_B \sqrt{\lambda(u,\mathbf{X})}\,\mathrm{d}u\right] \\
= \; & \int_B\int_B \mathbb{E}\left[\frac{\lambda(u,v,\mathbf{X})}{\sqrt{\lambda(u,\mathbf{X}\cup\{v\})\,\lambda(v,\mathbf{X}\cup\{u\})}}\right]\mathrm{d}u\,\mathrm{d}v \\
&+|B| \\
&+\int_B\int_B \mathbb{E}\left[\sqrt{\lambda(u,\mathbf{X})}\,\sqrt{\lambda(v,\mathbf{X})}\right]\,\mathrm{d}u\,\mathrm{d}v \\
&-2\int_B\int_B \mathbb{E}\left[\sqrt{\lambda(u,v,\mathbf{X})}\right]\,\mathrm{d}u\,\mathrm{d}v.
\end{aligned}
$$

For an heterogeneous Poisson process, this reduces to

$$
\mathsf{var}\left[\sum_B \frac{1}{\sqrt{\lambda(x_i)}} - \int_B \sqrt{\lambda(u)}\,\mathrm{d}u\right] = |B| \tag{85}
$$

as expected by analogy with Pearson residuals in the classical context.

For the raw innovations we have

$$
\begin{aligned}
\mathsf{var}\left[N(B) - \int_B \lambda(u,\mathbf{x})\,\mathrm{d}u\right] \;=\; & \mathbb{E}\left[\int_B\int_B \lambda(u,v,\mathbf{X})\,\mathrm{d}u\,\mathrm{d}v\right] \\
&+\mathbb{E}\int_B \lambda(u,\mathbf{X})\,\mathrm{d}u \\
&+\mathbb{E}\left[\int_B\int_B \lambda(u,\mathbf{X})\lambda(v,\mathbf{X})\,\mathrm{d}u\,\mathrm{d}v\right] \\
&-2\mathbb{E}\left[\int_B\int_B \lambda(u,\mathbf{X})\lambda(v,\mathbf{X}\cup\{u\})\,\mathrm{d}u\,\mathrm{d}v\right].
\end{aligned} \tag{86}
$$

For the heterogeneous Poisson process, this reduces to

$$
\mathsf{var}\left[N(B) - \int_B \lambda(u,\mathbf{x})\,\mathrm{d}u\right] = \int_B \lambda(u)\,\mathrm{d}u \tag{87}
$$

as expected.

B Null distribution of smoothed residual field

Here we calculate the distribution of the value of the smoothed residual field $s(u)$ at one location u for the uniform Poisson process. We ignore the effect of parameter estimation, that is, we consider the kernel-smoothed innovation measure, rather than the kernel-smoothed residual measure. Edge effects will also be ignored.

For the uniform Poisson process, the raw innovations, inverse-lambda innovations and Pearson innovations are all proportional. Consider the raw innovations. The smoothed raw innovation field is

$$s(u) = \sum_i k(u - x_i) - \lambda \tag{88}$$

where $k(\cdot)$ is the smoothing kernel and λ is the intensity of the Poisson process. Take the isotropic Gaussian kernel

$$k(u) = \frac{1}{2\pi\sigma^2} \exp\{-\frac{||u||^2}{2\sigma^2}\}$$

so that

$$s(u) = \frac{1}{2\pi\sigma^2} \sum_i \exp\{-\frac{||u - x_i||^2}{2\sigma^2}\} - \lambda.$$

For a uniform Poisson process of intensity λ the values $||u - x_i||^2$ are the event times T_i of a homogeneous Poisson process of intensity $\lambda\pi$ on $\mathbb{R}_+$:

$$S = \frac{1}{2\pi\sigma^2} \sum_i \exp\{-\frac{T_i}{2\sigma^2}\} - \lambda$$

Since the differences between successive event times T_i are exponentially distributed with rate $\lambda\pi$ we can represent this as

$$S = \frac{\lambda}{\mu} \sum_i \left(\prod_{j=1}^{i} U_j \right)^{1/\mu} - \lambda \tag{89}$$

where U_j are i.i.d. uniform $[0, 1]$ r.v.'s and $\mu = 2\lambda\pi\sigma^2$.

We are grateful to A.G. Pakes for the following solution. Let $X = (\mu(1 + S/\lambda))^\mu$ be the sum in (89). Then X satisfies the distributional equivalence

$$X \equiv U(1 + X)$$

where U is a uniform $[0, 1]$ random variable independent of X. This leads to a recursion for the c.d.f. $F(x)$ of X,

$$F(x) = x \left[C - \int_0^{x-1} F(z)/(1 + z)^2 dz \right] \tag{90}$$

where $C = \exp(-\gamma)$, where $\gamma \approx 0.5772$ is Euler's constant. In principle we can solve this iteratively for each interval $[n, n + 1]$. First for $x \in [0, 1]$ the integral is zero and

$$F(x) = Cx, \quad 0 \leq x \leq 1.$$

Next for $x \in [1, 2]$

$$F(x) = C[2x - x \log x - 1]$$

Subsequent intervals yield more intricate expressions which may nevertheless be computed numerically. The resulting probability density function is plotted in Figure 34. The 95% and 99% quantiles of X are 2.33 and 3.09 respectively. See also [1].

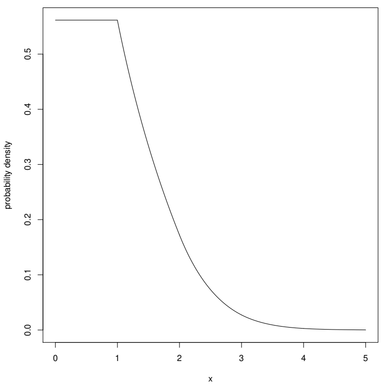

Figure 34: Probability density function of $X = (\mu(1 + S/\lambda))^{\mu}$ where $S = s(0)$ is a typical value of the kernel-smoothed raw innovation field for a uniform Poisson process of rate λ, using a Gaussian kernel with standard deviation σ, and where $\mu = 2\lambda\pi\sigma^2$.

C Proof that errors are uncorrelated

To prove the statement in Section 8.5 we use another property of the Papangelou conditional intensity $\lambda(u, \mathbf{X})$. For a region $B \subset W$, the conditional distribution of $\mathbf{X} \cap B$ given $\mathbf{X} \cap B^c$ (where c denotes complement) has effectively the same conditional intensity as the original process,

$$\lambda_{\mathbf{X}\cap B|\mathbf{X}\cap B^c}(u, \mathbf{X} \cap B) = \lambda_{\mathbf{X}}(u, (\mathbf{X} \cap B) \cup (\mathbf{X} \cap B^c)).$$

It follows that the raw innovations of $\mathbf{X}$ have *conditional* mean zero given $\mathbf{X} \cap B^c$,

$$\mathbb{E}\left[I(B) \mid \mathbf{X} \cap B^c\right] = 0.$$

Suppose that the point process interpoint interactions have finite range r, in the sense that the conditional intensity $\lambda(u, \mathbf{x})$ depends only on points of $\mathbf{x}$ that lie within a distance r of the location u. Let U, V be two subsets of W at least r units apart, i.e. $||u - v|| > r$ for any $u \in U$ and $v \in V$.

Taking $B = U$ we find that $I(V) = N(V) - \int_V \lambda(u, \mathbf{X})\, \mathrm{d}u$ depends only on $\mathbf{X} \cap B^c$ so that

$$\begin{aligned}
\mathbb{E}[I(U)I(V)] &= \mathbb{E}\left(\mathbb{E}\left[I(U)I(V) \mid \mathbf{X} \cap B^c\right]\right) \\
&= \mathbb{E}\left(I(V)\, \mathbb{E}\left[I(U) \mid \mathbf{X} \cap B^c\right]\right) \\
&= 0
\end{aligned}$$

so that $I(U)$ and $I(V)$ are uncorrelated.

`$Revision: 5.5 $  $Date: 2004/06/09 01:04:58 $`