Residual analysis for spatial point processes

A. Baddeley

University of Western Australia, Perth, Australia

R. Turner

University of New Brunswick, Fredericton, Canada

J. Møller

University of Aalborg, Denmark

M. Hazelton

University of Western Australia, Perth, Australia

Summary. We define residuals for point process models fitted to spatial point pattern data, and propose diagnostic plots based on them. The residuals apply to any point process model that has a conditional intensity; the model may exhibit spatial heterogeneity, interpoint interaction and dependence on spatial covariates. Some existing *ad hoc* methods for model checking (quadrat counts, scan statistic, kernel smoothed intensity, Berman's diagnostic) are recovered as special cases.

Diagnostic tools are developed systematically, using 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. This makes it possible to adapt existing knowledge about model validation for GLM's to the spatial point process context, giving recommendations for diagnostic plots. 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.

Keywords: Berman's diagnostic, Berman-Turner device, 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.

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. Parametric inference, model selection, and goodness-of-fit testing are also feasible with Markov chain Monte Carlo (MCMC) methods.

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. Indeed it is widespread practice in the statistical analysis of spatial point pattern data to focus primarily on comparing the data with a homogeneous Poisson process ("complete spatial randomness", CSR), 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 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. 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, used routinely in survival analysis. 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 (Cox and Isham, 1980, Section 6.1; Ripley, 1988, Introduction). 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 and Grabarnik (1991).

Next, diagnostic plots are developed systematically, by exploiting an analogy between point process models and generalised linear models (GLM's). The Papangelou conditional intensity λ of the spatial point process corresponds, under this analogy, to the mean response in a GLM. The spatial point process residuals introduced in this paper correspond to the usual residuals for Poisson loglinear regression. The components of a point process model (spatial trend, dependence on spatial covariates, interaction between points of the pattern) correspond to model terms in a GLM. Thus the well-understood diagnostic plots for assessing each term in a GLM can be carried across to spatial point processes.

Section 2 presents motivating examples. Section 3 offers a review and critique of current techniques. Section 4 reviews existing theory of residuals for point processes in time and space-time. Section 5 introduces spatial point process models and the essential background for our definition of residuals. Section 6 describes the diagnostic of Stoyan and Grabarnik. Our new residuals for spatial point processes are defined in Sections 7 and 8. Properties of the residuals are studied in Section 9. Sections 10–12 develop diagnostic plots for assessing each component of a spatial point process model. Sections 13–14 discuss practical implementation and scope of the techniques.

2. Motivating examples

Figure 1 depicts the Japanese black pines data of Numata (1964), analysed by Ogata and Tanemura (1981, 1986). Dots indicate the locations of 204 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, and for positive or negative 'interaction' between trees.

One of many possible approaches to Figure 1 is to fit a parametric statistical model to the pattern. The model is a spatial point process, which may be formulated to exhibit spatial heterogeneity and/or interpoint interaction. Formal testing and model selection may then be used to decide whether heterogeneity and interaction are present, and in what form.

Practical parametric modelling of spatial point pattern data was pioneered by Besag (1975, 1978), Ripley (1977), Diggle (1978), Ogata and Tanemura (1981) and others. For

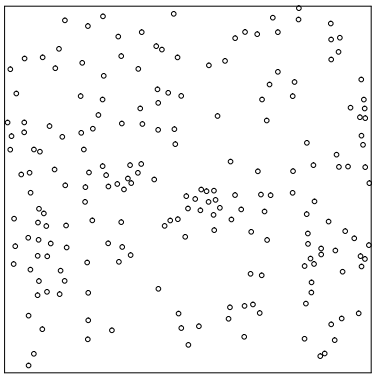

Fig. 1. Japanese black pines (seedlings and saplings) data of Numata (1964). Data kindly supplied by Professors Y. Ogata and M. Tanemura.

surveys, see Diggle (2003, Chapters 5–7) and Møller and Waagepetersen (2003a,b). Recently developed algorithms make it easy to fit a wide range of point process models to real data in applications, by pseudolikelihood methods (Baddeley and Turner, 2000, 2005a,b). Likelihood and Bayesian inference are also feasible for many models using MCMC methods (Geyer and Møller, 1994; Geyer, 1999; Møller and Waagepetersen, 2003a).

For the Japanese pines data, Ogata and Tanemura (1981, 1986) formulated several parametric models, involving heterogeneity and pairwise interaction between points. Maximum likelihood estimation was performed using a specialised numerical approximation. In their definitive analysis (Ogata and Tanemura, 1986), the Akaike information criterion favoured a 12-parameter model for the Japanese pines data, containing moderately strong heterogeneity and quite strong short-range inhibition between points.

It would be prudent to check this analysis, using a formal goodness-of-fit test and some informal validation of the fitted model. As far as we are aware, this has not been attempted. While some techniques are available for checking a point process model (see Section 3), most of them do not apply to a model involving both heterogeneity and interaction. Our goal is to provide tools for validation of a quite general point process such as this one.

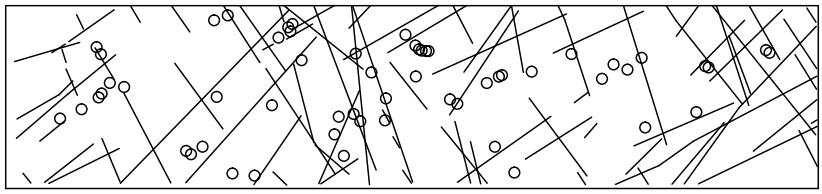

Fig. 2. Copper ore deposits (○) and lineaments (—) in an area of Queensland. Southern half of original data. North at top. Reproduced by kind permission of Dr. J. Huntington and Dr. M. Berman.

Figure 2 is a subset of data introduced and analysed by Berman (1986). It represents an intensive geological survey of a 158×35 km region in central Queensland. Dots mark the locations of 57 copper ore deposits. Line segments represent 90 geological features visible

on a 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 (no dependence on the lineaments) postulates that copper deposits are a homogeneous Poisson process. Alternative models postulate, for example, that the density of copper deposits depends on distance from the nearest lineament.

Several analyses (Berman, 1986; Berman and Turner, 1992; Foxall and Baddeley, 2002) have concluded there is no covariate effect. 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 (Foxall and Baddeley, 2002, Section 5.3).

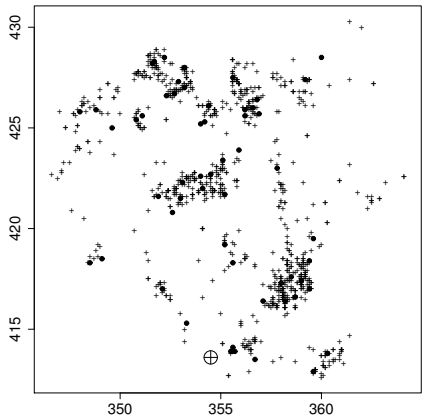

Fig. 3. Cases of cancer of the larynx (•) and lung (+) in the Chorley-South Ribble area, and the location of an industrial incinerator (⊕). Ordnance survey coordinates in km.

Figure 3 shows a spatial epidemiological dataset presented and analysed by Diggle (1990) and Diggle and Rowlingson (1994). There are two point patterns, giving the precise domicile locations of new cases of cancer of the larynx (58 cases) and of the lung (978 cases), recorded in the Chorley and South Ribble Health Authority 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 a now-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.

Diggle (1990) assumes the laryngeal cancer cases form a Poisson point process, with unknown intensity $\lambda(u)$ at spatial location u. The null model, that there is no incinerator effect, states that $\lambda(u)$ is proportional to the density of the susceptible population at u. In alternative models, $\lambda(u)$ also depends on distance from the incinerator. Diggle (1990) and Diggle and Rowlingson (1994) fit models of both types by maximum likelihood and find that the model of best fit includes an incinerator effect. Goodness-of-fit testing and informal validation of the model are carried out by transforming it to a uniform Poisson process on the real line (Diggle, 1990, Section 3.2).

The careful discussion by Diggle (1990) notes many caveats on epidemiological interpretation of the fitted model, and identifies questions for further investigation, notably the spatial clustering of disease cases. Clustered models cannot be fitted using the techniques

of Diggle (1990) and Diggle and Rowlingson (1994), which apply only to Poisson processes. While the model-checking techniques in Diggle (1990, Section 3.2) can detect the presence of clustering (as a departure from a fitted Poisson process), they cannot be used to validate a clustered point process model for the data. Thus, further analysis of the Chorley-Ribble data depends on tools for fitting and validating more general point process models.

Our goal, then, is to develop informal techniques to validate a point process model of general form that has been fitted to spatial point pattern data. The techniques should help us to recognise the presence — or the fitted model's *misspecification* — of spatial heterogeneity, interpoint interaction, and covariate effects in the data.

3. Current methods

Current techniques for checking a fitted spatial point process model are described by Van Lieshout (2000, Chapter 3), Diggle (2003, pp. 89–90, 100–103, 106, 110–111, 114, 133–143) and Møller and Waagepetersen (2003a, Chapter 4).

In his influential paper Ripley (1977) developed an exploratory analysis of interpoint interaction, *assuming the data are spatially homogeneous*. A useful summary statistic is the nonparametric estimator $\widehat{K}$ of Ripley's K function, essentially a renormalised empirical distribution of the pairwise distances between observed points. For the homogeneous Poisson process (CSR) the true value of K is known. A discrepancy between $\widehat{K}$ and the theoretical K function for CSR indicates positive or negative association between points, and suggests appropriate models. A Monte Carlo goodness-of-fit test of any fitted model can also be conducted, by comparing the values of $\widehat{K}$ for the data with those from simulations of the model (Besag and Diggle, 1977). See surveys in Cressie (1991), Diggle (2003), Møller and Waagepetersen (2003a), Ripley (1981), Ripley (1988), Stoyan et al. (1995) and Stoyan and Stoyan (1995).

Difficulties arise if we wish to validate a fitted model that also includes heterogeneity, or when we wish to detect heterogeneity in the data. The estimator $\widehat{K}$ is affected by spatial inhomogeneity as well as by spatial dependence between points. It can still be used as the basis for a Monte Carlo test of goodness-of-fit; but the interpretation of any deviations in $\widehat{K}$ is now ambiguous. Thus, in practice, the use of the K function in model *criticism* is restricted to cases where the fitted model is homogeneous and the data are still assumed to be homogeneous.

To overcome this limitation, modifications of $\widehat{K}$ (and of other statistics) have been proposed. Local Indicators of Spatial Association (LISA) (Anselin, 1995; Cressie and Collins, 2001b,a; Getis and Franklin, 1987), are localised versions of summary functions such as $\widehat{K}$. The K function has also been adapted to inhomogeneous point processes where the spatial trend is known (Baddeley et al., 2000). However, if the trend has to be estimated when estimating K, the interpretation of deviations in $\widehat{K}$ is not always clear, as shown by the conflicting examples in Diggle (2003, Sections 7.1.1 and 8.2.1).

When the fitted model is an inhomogeneous spatial Poisson process, a powerful diagnostic tool is to transform it to a Poisson process on the real line, with uniform intensity 1 on an interval (Ripley, 1982, p. 180; Brillinger and Preisler, 1986; Diggle, 1990; Schoenberg, 1999; Diggle, 2003, Section 3.2). This can be used to validate the model, as is done commonly in survival analysis (Cox and Lewis, 1966, Chapter 3; Andersen et al., 1993; Venables and Ripley, 1997, Chapter 12). Departure from unit intensity in the transformed space suggests a misspecified spatial trend, while departure from exponential distribution

of the inter-arrival times is evidence of interpoint interaction. However the *form* of departures from the model may not be easy to recognise in the transformed space. In the spatial setting, this diagnostic is restricted to Poisson models, apart from some special processes (Merzbach and Nualart, 1986; Nair, 1990, Cressie, 1991, pp. 766-770).

The advent of practical MCMC algorithms for simulating and fitting point process models (Geyer and Møller, 1994; Geyer, 1999; Møller and Waagepetersen, 2003a) has made it possible to test for the presence of spatial trend or interaction within the context of a parametric model. However, tools for model criticism are still lacking.

Finally, some writers have introduced diagnostics analogous to the residuals from a fitted GLM. The diagnostic of Stoyan and Grabarnik (1991) is described in Section 6. Lawson (1993) defined a 'deviance residual' for heterogeneous Poisson processes. For space-time point processes, Diggle et al. (1995) constructed a residual by comparing a space-time K function with the product of two K functions in time and in space, while Ogata (1988) formed residuals based on the ratio of a nonparametric intensity estimate to the model's conditional intensity. In spatial epidemiology, spatially varying relative risk may be estimated nonparametrically or modelled (semi-)parametrically; differences between these two estimates yield an estimated residual relative risk, e.g. Diggle (2003, pp. 133–143). Wartenberg (1990) canvassed exploratory methods for outliers, leverage and influence in spatial point patterns.

4. 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 (Lewis, 1972; Brillinger, 1978, 1994) and survival analysis (Andersen et al., 1993). If N_t denotes the number of arrivals (points of the process) in the time interval $[0, t]$, define the *conditional intensity*

$$\lambda(t) = \mathbb{E}[\mathrm{d}N_t \mid N_s, s < t]/\mathrm{d}t,$$

(if it exists), that is, $\lambda(t)$ is the instantaneous arrival rate of the point process given the history of the process prior to time t (Karr, 1985, p. 69 ff.). Residuals can be constructed from the fact that the *innovation* or *error process*

$$I(t) = N_t - \int_0^t \lambda(s)\, \mathrm{d}s \tag{1}$$

is a martingale with $\mathbb{E}[I(t)] = 0$ when the model is true (Karr, 1985, Theorem 2.14, p. 60). In practice, when a point process model with a parameter θ is fitted to data, the parameter estimate $\widehat{\theta}$ would be plugged in to an expression for $\lambda(t) = \lambda_\theta(t)$ to obtain a fitted conditional intensity $\widehat{\lambda}(t) = \lambda_{\widehat{\theta}}(t)$, and we compute the *raw residual process*

$$R(t) = N_t - \int_0^t \widehat{\lambda}(s)\, \mathrm{d}s.$$

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 (Lewis, 1972; Brillinger, 1978; Venables and Ripley, 1997).

The likelihood of the point process on the interval $[0, t]$ is (Karr, 1985, Thm 2.31, p. 71)

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

where $t_1, t_2, \ldots$ denote the successive arrival times of the point process. Hence the score $U_\theta(t)$ is closely related to the innovations:

$$U_\theta(t) = \sum_{t_i \le t} \frac{\partial}{\partial \theta} \log \lambda_\theta(t_i) - \int_0^t \frac{\partial}{\partial \theta} \lambda_\theta(s)\, \mathrm{d}s = \int_0^t \frac{\partial}{\partial \theta} \log \lambda_\theta(s)\, \mathrm{d}I_\theta(s).$$

The fact that $\mathbb{E}_\theta[U_\theta(t)] = 0$ follows from the martingale property of $I_\theta(t)$.

Analogous residuals for space-time point processes were developed for applications to earthquake modelling (Ogata, 1988; Vere-Jones, 1970).

5. Spatial point process models

5.1. Notation and assumptions

A spatial point pattern is a dataset

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

consisting of the (unordered) 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 $n = n(\mathbf{x}) \ge 0$ is not fixed in advance. Our aim is to validate a parametric spatial point process model which has been fitted to $\mathbf{x}$. The model may be very general indeed, and the method used to fit the model is arbitrary. We assume only that the model has a probability density $f_\theta(\mathbf{x})$ with respect to the unit rate Poisson process on W, and that f_θ satisfies the positivity condition

$$\text{if } f_\theta(\mathbf{x}) > 0 \text{ and } \mathbf{y} \subset \mathbf{x} \text{ then } f_\theta(\mathbf{y}) > 0 \tag{2}$$

for any finite point patterns $\mathbf{x}, \mathbf{y} \subset W$. For example, a homogeneous Poisson process with intensity β has density $f(\mathbf{x}) = \alpha\beta^{n(\mathbf{x})}$, where α represents a normalising constant throughout this paper. An inhomogeneous Poisson process with intensity function $b(u), u \in W$ has density $f(\mathbf{x}) = \alpha \prod_{i=1}^n b(x_i)$. A *pairwise interaction* point process has

$$f(\mathbf{x}) = \alpha \left[\prod_{i=1}^{n(\mathbf{x})} b(x_i)\right] \left[\prod_{i<j} c(x_i, x_j)\right] \tag{3}$$

where $b(u) \ge 0, u \in W$ is the 'activity' and $c(u, v) = c(v, u) \ge 0,\ u, v \in W$ is the 'interaction'. The activity function b can be used to model spatial variation in the abundance of points, while the interaction function c can be used to model association between points.

For *any* finite point process satisfying (2), a theorem of Ripley and Kelly (1977) states that the density can be expressed in "Gibbs" form

$$f(\mathbf{x}) = \exp\{V_0 + V_1(\mathbf{x}) + V_2(\mathbf{x}) + \ldots + V_n(\mathbf{x})\} \tag{4}$$

(where $n = n(\mathbf{x})$) for unique functions V_k called the potentials of order k. Here V_0 determines the normalising constant and V_k is of the form $V_k(\mathbf{x}) = \sum_{\mathbf{y}} v_k(\mathbf{y})$, where $v_k(\mathbf{y}) \in [-\infty, \infty)$ and the sum is over all subsets $\mathbf{y} \subseteq \mathbf{x}$ with $n(\mathbf{y}) = k$. Although the Gibbs form might not be the simplest representation of the model, we can use it to inspect the model's properties. In particular, interpoint interaction is determined by the potentials of higher order, $V_{\geq 2}(\mathbf{x}) = V_2(\mathbf{x}) + V_3(\mathbf{x}) + \ldots + V_n(\mathbf{x})$. If $V_{\geq 2}$ is identically zero, then the model reduces to a Poisson process with intensity function $b(u) = \exp(v_1(\{u\}))$.

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 conditional intensity* $\lambda(u, \mathbf{x})$ (Papangelou, 1974) which conditions on the outcome of the process at all spatial locations other than u.

Detailed theory may be consulted in Daley and Vere-Jones (1988, pp. 580–590) and Karr (1985, Section 2.6). For our purposes the following, very simplified, account is adequate. Suppose $\mathbf{X}$ is any finite point process in W with a probability density $f(\mathbf{x})$ which satisfies the analogue of the positivity condition (2). For $u \in W$ with $u \notin \mathbf{x}$, define

$$\lambda(u, \mathbf{x}) = f(\mathbf{x} \cup \{u\})/f(\mathbf{x}) \tag{5}$$

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

$$\lambda(u, \mathbf{x}) = \lambda(u, \mathbf{x} \setminus \{u\}). \tag{6}$$

Then (6) holds for all $u \in W$. Loosely speaking, $\lambda(u, \mathbf{x})\, \mathrm{d}u$ is the conditional probability that there is a point of $\mathbf{X}$ in an infinitesimal region of area $\mathrm{d}u$ containing u, given that the rest of the point process coincides with $\mathbf{x}$.

For example, the Poisson process with intensity function $b(u)$, $u \in W$ has conditional intensity $\lambda(u, \mathbf{x}) = b(u)$. The general pairwise interaction process (3) has

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

For non-Poisson processes, in general $\lambda(\cdot, \mathbf{x})$ is discontinuous at the data points x_i because of (6). For a general point process, (4) leads to a Gibbs representation

$$\log \lambda(u, \mathbf{x}) = v_1(u) + \sum_i v_2(\{u, x_i\}) + \sum_{i<j} v_3(\{u, x_i, x_j\}) + \ldots, \quad u \notin \mathbf{x}. \tag{7}$$

The Papangelou conditional intensity λ of a finite point process uniquely determines its probability density f and vice versa (because of (2)). For Markov point processes (Ripley and Kelly, 1977; Van Lieshout, 2000) 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. The normalising constant of f is eliminated in (5). Most simulation procedures are specified in terms of λ. See Møller and Waagepetersen (2003a).

It can be verified directly for finite point processes that

$$\mathbb{E}\left[\sum_{x_i\in\mathbf{X}} h(x_i, \mathbf{X}\setminus\{x_i\})\right] = \mathbb{E}\left[\int_W h(u,\mathbf{X})\lambda(u,\mathbf{X})\,\mathrm{d}u\right] \tag{8}$$

where $h(u,\mathbf{x})$ is any nonnegative function. Equation (8) and its extension to $\mathbb{R}^2$ for infinite point processes are called the *Georgii-Nguyen-Zessin (GNZ) formula* (Georgii, 1976; Nguyen and Zessin, 1979). In the present paper (8) becomes the basic identity for deriving diagnostics and residuals. We assume both sides of (8) are finite when required.

6. Stoyan–Grabarnik diagnostic

Stoyan and Grabarnik (1991) were the first to exploit the GNZ formula (8) for model checking. Assume $\lambda(\cdot,\cdot) > 0$. Take

$$h(u,\mathbf{x}) = h_B(u,\mathbf{x}) = \mathbf{1}\{u\in B\}/\lambda(u,\mathbf{x}) \tag{9}$$

where $\mathbf{1}\{\ldots\}$ is the indicator function and $B\subseteq W$ is a given subset. Then (8) becomes

$$\mathbb{E}\left[\sum_{x_i\in\mathbf{X}\cap B}\frac{1}{\lambda(x_i,\mathbf{X})}\right] = \mathbb{E}\left[\int_B 1\,\mathrm{d}u\right] = |B| \tag{10}$$

where $|B|$ denotes the area of B. This states that, if each point x_i of $\mathbf{X}$ 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 points x_i of $\mathbf{X}$ that fall in a nominated region B,

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

has expectation $\mathbb{E}M(B) = |B|$ under the model. The variance of $M(B)$ was also computed by Stoyan and Grabarnik (1991) for the case of a "stationary" pairwise interaction process (i.e. when the function b is constant, $c(u,v) = c(u-v)$ and the process is extended to $\mathbb{R}^2$). While other functions h could be substituted in (8), the judicious choice (9) made by Stoyan and Grabarnik is effectively the only one in which the integral in (8) is trivial.

Write $\lambda_\theta(s,\mathbf{x})$ for the Papangelou conditional intensity under a parametric model with density f_θ. In practice this would be replaced by a plug-in estimate $\widehat{\lambda}(u,\mathbf{x}) = \lambda_{\widehat{\theta}}(u,\mathbf{x})$. Stoyan and 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 sums $\widehat{M}(B) = \sum_i \mathbf{1}\{x_i\in B\}\widehat{m}_i$, could be used for exploratory data analysis and goodness-of-fit testing, in that (a) 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 Stoyan and Grabarnik (1991); proposal (a) was tried by Särkkä (1993, pp. 49–50) and proposal (b) by Zhuang et al. (2005).

If $\lambda(u,\mathbf{x})$ may take zero values, a few difficulties arise with the Stoyan-Grabarnik technique. For instance, the 'hard core' process obtained by setting $c(u,v) = \mathbf{1}\{\|u-v\| > \delta\}$ in (3), where $\delta > 0$, has $\lambda(u,\mathbf{x}) = b(u)$ if $\|u-\xi\| > \delta$ for all points x_i in $\mathbf{x}$, and $\lambda(u,\mathbf{x}) = 0$ otherwise. 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 (10) does not hold. We resolve this problem in Section 8.2.

7. Residuals for spatial point processes

We now start to define our spatial residuals.

7.1. Innovations

Consider a parametric model for a spatial point process $\mathbf{X}$ with density f_θ. We assume only that f_θ satisfies (2). Define the *innovation process* of the model by

$$I_\theta(B) = n(\mathbf{X} \cap B) - \int_B \lambda_\theta(u, \mathbf{X}) \, \mathrm{d}u \tag{11}$$

for any set $B \subseteq W$, where $n(\mathbf{X} \cap B)$ denotes the number of random points falling in B. This definition is closely analogous to the residuals in time and space-time, except for the use of the Papangelou conditional intensity. The innovations I_θ constitute a (random) signed measure, with a mass $+1$ at each point x_i of the spatial point process, and a negative density $-\lambda(u, \mathbf{X})$ at all other spatial locations u. They satisfy

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

by setting $h(u, \mathbf{x}) = \mathbf{1}\{u \in B\}$ in (8). Increments of the innovation process I_θ are analogous to errors in a linear model. The GNZ formula (8) can be restated as

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

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

The direct connection between the innovations and the score for point processes in time (Section 4) is lost for spatial processes, unless they are Poisson. Instead I_θ is closely related to the *pseudoscore*, the derivative of the log pseudolikelihood of the point process defined by

$$\log \mathrm{PL}(\theta, \mathbf{x}) = \sum_{x_i \in \mathbf{x}} \log \lambda_\theta(x_i, \mathbf{x}) - \int_W \lambda_\theta(u, \mathbf{x}) \, \mathrm{d}u \tag{12}$$

(Besag, 1978; Jensen and Møller, 1991) since the pseudoscore can be written

$$\frac{\partial}{\partial \theta} \log \mathrm{PL}(\theta, \mathbf{x}) = \int_W \frac{\partial}{\partial \theta} \log \lambda_\theta(u, \mathbf{x}) \, \mathrm{d}I_\theta(u). \tag{13}$$

Applying (8) to $h(u, \mathbf{x}) = \frac{\partial}{\partial \theta} \log \lambda_\theta(u, \mathbf{x})$ shows that the pseudoscore has mean zero under the model. For Poisson processes the pseudolikelihood and likelihood are equivalent, so (13) is a direct connection between the innovations and the score.

7.2. Raw residuals

Given data $\mathbf{x}$, and using a general parameter estimate $\widehat{\theta} = \widehat{\theta}(\mathbf{x})$, we define the *raw residuals*

$$R_{\widehat{\theta}}(B) = n(\mathbf{x} \cap B) - \int_B \widehat{\lambda}(u, \mathbf{x}) \, \mathrm{d}u \tag{14}$$

where $\widehat{\lambda} = \lambda_{\widehat{\theta}}$. Increments of $R_{\widehat{\theta}}$ correspond to the raw residuals in a linear model. The raw residuals $R_{\widehat{\theta}}(B)$ are a signed measure on W, with atoms of mass 1 at the data points, and a negative density $-\widehat{\lambda}(u, \mathbf{x})$ at all locations u in W. Methods of visualising these residuals are proposed in Sections 11–12.

Whereas most previous writers (Lawson, 1993; Särkkä, 1993; Stoyan and Grabarnik, 1991) 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.

8. Scaled residuals

8.1. 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 function h in the GNZ formula (8). For any nonnegative function $h(u, \mathbf{x})$, define the h-weighted *innovations*

$$I(B, h, \lambda) = \sum_{x_i \in \mathbf{X} \cap B} h(x_i, \mathbf{X} \setminus \{x_i\}) - \int_B h(u, \mathbf{x}) \lambda(u, \mathbf{X}) \, \mathrm{d}u \tag{15}$$

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 point x_i in $\mathbf{X}$, and $\mathrm{d}I(u) = -h(u, \mathbf{X})\lambda(u, \mathbf{X}) \, \mathrm{d}u$ as the innovation increment attached to a background location $u \in W$. The innovations have mean zero, from (8).

After fitting a parametric model to data $\mathbf{x}$ using a parameter estimate $\widehat{\theta} = \widehat{\theta}(\mathbf{x})$, we compute the fitted conditional intensity $\widehat{\lambda}(u, \mathbf{x}) = \lambda_{\widehat{\theta}}(u, \mathbf{x})$. The weight function h may also depend on θ, in which case we also compute $\widehat{h}(u, \mathbf{x}) = h_{\widehat{\theta}}(u, \mathbf{x})$. Then we define the h-weighted *residual* measure by

$$R(B, \widehat{h}, \widehat{\theta}) = I(B, \widehat{h}, \widehat{\lambda}) = \sum_{x_i \in \mathbf{x} \cap B} \widehat{h}(x_i, \mathbf{x} \setminus \{x_i\}) - \int_B \widehat{h}(u, \mathbf{x}) \widehat{\lambda}(u, \mathbf{x}) \, \mathrm{d}u. \tag{16}$$

The innovation measure has mean zero, $\mathbb{E}\left[I(B, h, \lambda)\right] = 0$, and we *hope* that the mean of the residual measure is approximately zero when the model is true,

$$\mathbb{E}_\theta \left[R(B, \widehat{h}, \widehat{\theta}) \right] \approx 0. \tag{17}$$

The choice $h(u, \mathbf{x}) \equiv 1$ in (15) and (16) yields the raw innovations (11) and raw residuals (14). Various other choices of h are discussed in Sections 8.2–8.4.

8.2. Inverse-lambda residuals

The choice $h(u, \mathbf{x}) = 1/\lambda(u, \mathbf{x})$ corresponds to the exponential energy weights (Section 6). Care is required if $\lambda(u, \mathbf{x})$ may take zero values. The GNZ formula (8) still holds when h

may take the value $+\infty$, provided $h(x_i, \mathbf{X} \setminus \{x_i\})$ is finite for all $x_i \in \mathbf{X}$, and we interpret $h(u, \mathbf{X})\lambda(u, \mathbf{X})$ as zero if $\lambda(u, \mathbf{X}) = 0$. Thus we obtain the innovation measure

$$I(B, 1/\lambda, \lambda) = \sum_{x_i \in \mathbf{X} \cap B} \frac{1}{\lambda(x_i, \mathbf{X})} - \int_B \mathbf{1}\{\lambda(u, \mathbf{X}) > 0\} \, \mathrm{d}u$$

which has mean zero. The corresponding choice $h_\theta(u, \mathbf{x}) = 1/\lambda_\theta(u, \mathbf{x})$ in (16) yields the residual measure

$$R(B, 1/\widehat{\lambda}, \widehat{\theta}) = \sum_{x_i \in \mathbf{x} \cap B} \frac{1}{\widehat{\lambda}(x_i, \mathbf{x})} - \int_B \mathbf{1}\{\widehat{\lambda}(u, \mathbf{x}) > 0\} \, \mathrm{d}u. \tag{18}$$

In order that the residuals be well defined, the estimator $\widehat{\theta}$ must have the property that $\lambda_{\widehat{\theta}(\mathbf{x})}(x_i, \mathbf{x}) > 0$ for all $x_i \in \mathbf{x}$ for any pattern $\mathbf{x}$. Zero values for $\lambda_{\widehat{\theta}(\mathbf{x})}(u, \mathbf{x})$ are permitted for $u \notin \mathbf{x}$. We shall call (18) the *inverse-λ residuals*. They are equivalent to the Stoyan-Grabarnik diagnostic when $\lambda_\theta(\cdot, \cdot) > 0$.

8.3. Pearson residuals

By analogy with the Pearson residuals for Poisson loglinear regression, we consider the weight function $h(u, \mathbf{x}) = 1/\sqrt{\lambda(u, \mathbf{x})}$ which yields the Pearson innovation measure

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

which has mean zero, and the corresponding Pearson residual measure

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

Again, the estimate $\widehat{\theta}$ must satisfy $\widehat{\lambda}(x_i, \mathbf{x}) > 0$ for all $x_i \in \mathbf{x}$ in order that the Pearson residuals be well-defined, but zero values for $\widehat{\lambda}(u, \mathbf{x})$ are permitted for $u \notin \mathbf{x}$.

8.4. Pseudo-score residuals

If θ is a k-dimensional vector, taking $h(u, \mathbf{x}) = \frac{\partial}{\partial \theta} \log \lambda_\theta(u, \mathbf{x})$ yields vector-valued errors

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

and vector-valued residuals

$$R(B, \frac{\partial}{\partial \theta} \log \widehat{\lambda}, \widehat{\lambda}) = \sum_{x_i \in \mathbf{x} \cap B} \left[\frac{\partial}{\partial \theta} \log \lambda_\theta(x_i, \mathbf{x}) \right]_{\theta = \widehat{\theta}} - \int_B \left[\frac{\partial}{\partial \theta} \lambda_\theta(u, \mathbf{x}) \right]_{\theta = \widehat{\theta}} \mathrm{d}u. \tag{19}$$

These residuals are increments of the pseudoscore (13), and thus correspond to the score residuals in a GLM. The residual (19) can also be interpreted as the pseudoscore in the domain B, conditional on the data outside B (Jensen and Møller, 1991).

In the case of a Strauss process model, the pseudoscore residuals are closely related to the K function. This will be explored in a subsequent paper.

9. Properties of residuals

9.1. Residuals sum to zero

The raw residuals in simple linear regression always sum to zero; a similar phenomenon occurs for our residuals. Firstly consider the homogeneous Poisson process model, fitted by maximum likelihood. The raw residual is $R_{\widehat{\theta}}(B) = n(\mathbf{x} \cap B) - n(\mathbf{x})|B|/|W|$. In particular the residual sum for the whole window W is $R_{\widehat{\theta}}(W) = 0$ for any point pattern dataset $\mathbf{x}$.

More generally, suppose we fit a point process model with no spatial trend, having conditional intensity of the common 'loglinear' form $\lambda_\theta(u, \mathbf{x}) = \exp\{\beta + \eta T(u, \mathbf{x})\}$ where $\theta = (\beta, \eta)$ and $T(u, \mathbf{x})$ is not constant. If the model is fitted by maximum pseudolikelihood, we equate (19) to zero with $B = W$, which implies $R_{\widehat{\theta}}(W) = 0$. The pseudoscore residuals (for any model with k-dimensional parameter) sum to zero over W.

9.2. Mean residual

Suppose we fit a point process model with parameter θ to a point pattern $\mathbf{x}$ using a parameter estimate $\widehat{\theta} = \widehat{\theta}(\mathbf{x})$. Assume that $\mathbf{x}$ is actually a realisation from some other point process $\mathbf{X}$, whose probability density satisfies the analogue of (2). Then the residuals (16) have true expectation

$$\mathbb{E}\left[R(B, \widehat{h}, \widehat{\theta})\right] = \int_B \mathbb{E}\left[h_{\widehat{\theta}(\mathbf{X}\cup\{u\})}(u, \mathbf{X})\lambda(u, \mathbf{X}) - h_{\widehat{\theta}(\mathbf{X})}(u, \mathbf{X})\lambda_{\widehat{\theta}(\mathbf{X})}(u, \mathbf{X})\right] \, \mathrm{d}u \tag{20}$$

by (8), where $\mathbb{E}$ is the expectation for the true process $\mathbf{X}$ and $\lambda(u, \mathbf{X})$ is its true conditional intensity. This yields for the raw, inverse-lambda, and Pearson residuals respectively

$$\mathbb{E}[R(B, 1, \widehat{\theta})] = \mathbb{E}\int_B [\lambda(u, \mathbf{X}) - \lambda_{\widehat{\theta}}(u, \mathbf{X})] \, \mathrm{d}u \tag{21}$$

$$\mathbb{E}\left[R(B, 1/\widehat{\lambda}, \widehat{\theta})\right] = \int_B \mathbb{E}\left[\frac{\lambda(u, \mathbf{X})}{\lambda_{\widehat{\theta}(\mathbf{X}\cup\{u\})}(u, \mathbf{X})} - \mathbf{1}\{\lambda_{\widehat{\theta}(\mathbf{X})}(u, \mathbf{X}) > 0\}\right] \, \mathrm{d}u \tag{22}$$

$$\mathbb{E}\left[R(B, 1/\sqrt{\widehat{\lambda}}, \widehat{\theta})\right] = \int_B \mathbb{E}\left[\frac{\lambda(u, \mathbf{X})}{\sqrt{\lambda_{\widehat{\theta}(\mathbf{X}\cup\{u\})}(u, \mathbf{X})}} - \sqrt{\lambda_{\widehat{\theta}(\mathbf{X})}(u, \mathbf{X})}\right] \, \mathrm{d}u \tag{23}$$

(provided $\lambda_{\widehat{\theta}(\mathbf{X})}(x_i, \mathbf{X}) > 0$ for all $x_i \in \mathbf{X}$). Since the true intensity of the process is $\lambda(u) = \mathbb{E}[\lambda(u, \mathbf{x})]$, a diagnostic interpretation of (21) is that the raw residuals are estimates of (negative) bias in modelling the intensity. Equation (22) has a more complex interpretation relating to relative bias in the fitted conditional intensity.

9.3. Variance of residuals

We have obtained general formulae for the variances of the innovations and residuals, that is, for $\mathsf{var}\,[I(B, h, \lambda)]$ and $\mathsf{var}\left[R(B, \widehat{h}, \widehat{\theta})\right]$, in terms of the *two-point conditional intensity*

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

See Baddeley et al. (2004, 2005). The variance of the innovations for a general weight function h is

$$\mathsf{var}\,[I(B, h, \lambda)] = \int_B \mathbb{E}\left[h(u, \mathbf{X})^2 \, \lambda(u, \mathbf{X})\right] \, \mathrm{d}u + \int_B \int_B \mathbb{E}\left[S(u, v, \mathbf{X})\right] \, \mathrm{d}u \, \mathrm{d}v \tag{24}$$

where

$$S(u,v,\mathbf{x}) = \lambda(u,\mathbf{x})\lambda(v,\mathbf{x})h(u,\mathbf{x})h(v,\mathbf{x}) + \lambda(u,v,\mathbf{x})h(v,\mathbf{x}\cup\{u\})\left[h(u,\mathbf{x}\cup\{v\}) - 2h(u,\mathbf{x})\right].$$

Substituting $h \equiv 1$, $h(u,\mathbf{x}) = 1/\lambda(u,\mathbf{x})$ or $h(u,\mathbf{x}) = 1/\sqrt{\lambda(u,\mathbf{x})}$ gives the variance of the raw, inverse-lambda or Pearson innovations, respectively. In the special case of an inhomogeneous Poisson process with intensity $\lambda(u)$, these reduce to

$$\mathsf{var}\left[I(B,1,\lambda)\right] = \int_B \lambda(u)\,\mathrm{d}u \tag{25}$$

$$\mathsf{var}\left[I(B,1/\lambda,\lambda)\right] = \int_B \frac{1}{\lambda(u)}\,\mathrm{d}u \tag{26}$$

$$\mathsf{var}\left[I(B,1/\sqrt{\lambda},\theta)\right] = |B|. \tag{27}$$

The first equation is of course the variance and mean of $n(\mathbf{X})$. The last equation is analogous to the fact that the classical Pearson residuals are standardised, ignoring the effect of parameter estimation.

It is also possible to give variance formulae under the pairwise interaction model (3). In this case

$$\lambda(u,v,\mathbf{x}) = \lambda(u,\mathbf{x}\setminus\{v\})\;\lambda(v,\mathbf{x}\setminus\{u\})\;c(u,v)$$

so the variance of the inverse-lambda innovations is

$$\mathsf{var}\left[I(B,1/\lambda,\lambda)\right] = \int_B\int_B \frac{1}{c(u,v)}\,\mathrm{d}u\,\mathrm{d}v + \int_B \mathbb{E}\left[\frac{1}{\lambda(u,\mathbf{X})}\right]\,\mathrm{d}u - |B|^2 \tag{28}$$

generalising a result of Stoyan and Grabarnik (1991).

For the variance of the residuals, the formulae are more cumbersome, involving characteristics of both the fitted model and the underlying point process (Baddeley et al., 2004, 2005). For example, suppose a Poisson process model with intensity $\lambda_\theta(u)$ is fitted to a realisation of a Poisson process with true intensity $\lambda(u)$. Then the raw residuals have variance

$$\begin{aligned}\mathsf{var}\left[R(B)\right] &= \int_B \lambda(u)\,\mathrm{d}u + \int_B\int_B \mathsf{cov}(\lambda_{\widehat{\theta}(X)}(u),\;\lambda_{\widehat{\theta}(X)}(v))\,\mathrm{d}u\,\mathrm{d}v\\ &\quad -2\int_B\int_B \mathbb{E}\left[\lambda_{\widehat{\theta}(X\cup\{u\})}(v) - \lambda_{\widehat{\theta}(X)}(v)\right]\;\lambda(u)\,\mathrm{d}v\,\mathrm{d}u.\end{aligned}$$

In the very special case where a homogeneous Poisson process is fitted to a realisation of a homogeneous Poisson process with intensity θ, the residual variances are

$$\begin{aligned}\mathsf{var}\left[R(B,1,\widehat{\theta})\right] &= \theta\,|B|(1-|B|/|W|)\\ \mathsf{var}\left[R(B,1/\widehat{\lambda},\widehat{\theta})\right] &= |B|(|W|-|B|)\mathbb{E}\left[\frac{\mathbf{1}\{n(\mathbf{X})>0\}}{n(\mathbf{X})}\right]\\ \mathsf{var}\left[R(B,1/\sqrt{\widehat{\lambda}},\widehat{\lambda})\right] &= |B|(1-|B|/|W|).\end{aligned}$$

Note that the residual variances are smaller than the corresponding innovation variances $\mathsf{var}\left[I(B,1,\theta)\right] = \theta\,|B|$, $\mathsf{var}\left[I(B,1/\theta,\theta)\right] = |B|/\theta$ and $\mathsf{var}\left[I(B,1/\sqrt{\theta},\theta)\right] = |B|$. This is analogous to the deflation of residual variance in a linear model.

9.4. Uncorrelated errors

Residuals are easiest to interpret and use when they are independent and identically distributed. Our spatial residuals do not have independent increments. However, for the large class of Markov point processes (Van Lieshout, 2000), the residuals have conditional independence properties. Suppose that the 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. 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 it can be shown that the raw innovations $I(U) = I(U, 1, \lambda)$ and $I(V) = I(V, 1, \lambda)$ are conditionally independent given $\mathbf{X} \cap (U \cup V)^c$, and in particular $I(U)$ and $I(V)$ are uncorrelated. See Baddeley et al. (2004, 2005). We conjecture that the innovations and residuals satisfy a strong law of large numbers and a central limit theorem as the sampling window W expands.

10. Approach to diagnostic plots

10.1. Objectives

In Sections 11 and 12 we develop diagnostic plots based on the residuals. We are guided by analogy with residual plots for other statistical models (Atkinson, 1985; Collett, 1991; Davison and Snell, 1991) especially logistic regression (Fowlkes, 1987; Landwehr et al., 1984; Pregibon, 1981). A specific plot is designed for checking each component ('assumption') of the fitted model: spatial trend, dependence on spatial covariates, interaction between points of the pattern, and other effects. In particular these plots can check for the *presence* of such features when the fitted model does not include them. In general, the plots should detect *mis-specification* by the model of the true spatial trend, covariate effects and interpoint interaction in the data.

10.2. Test examples

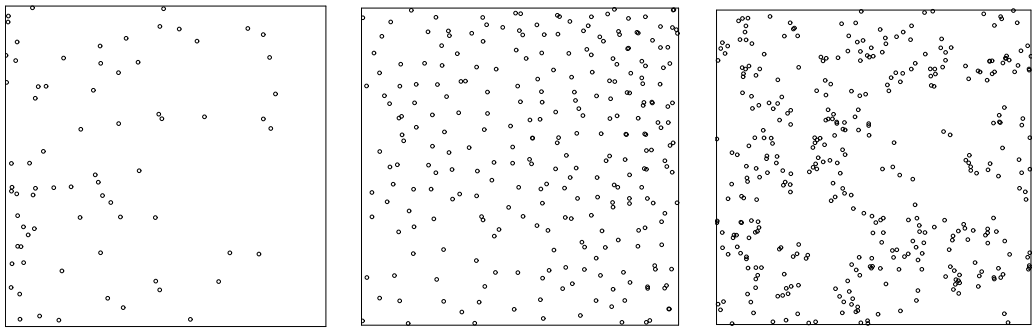

Fig. 4. Simulated patterns: (a) inhomogeneous Poisson process; (b) inhomogeneous inhibited process (Strauss process); (c) homogeneous clustered process (Geyer's saturation model).

Figure 4 shows three simulated examples that we use to test the diagnostics. The patterns contain 71, 271 and 376 points respectively in the unit square. Figure 4(a) is an example of 'trend without interaction': the Poisson process with intensity function $\lambda(x, y) =$

$300\exp\{-3x\}$. Figure 4(b) is an example of 'trend with (inhibitive) interaction': a pairwise interaction process (3) with log-quadratic activity function

$$b(x,y) = 200\exp(2x + 2y + 3x^2) \tag{29}$$

and the 'Strauss' interpoint interaction

$$c(u,v) = \begin{cases} \gamma & \text{if } ||u-v|| \le r \\ 1 & \text{otherwise} \end{cases} \tag{30}$$

with interaction range $r = 0.05$ and interaction strength $\gamma = 0.1$, corresponding to a strong negative association between points. The realisation in Figure 4(b) was generated by a Metropolis-Hastings birth-death-shift algorithm (Geyer and Møller, 1994) in a square of side 1.2 with periodic boundary conditions, then clipped to the unit square.

Figure 4(c) is an example of '(clustered) interaction without trend': a realisation of the saturation process of Geyer (1999, Section 3.9.2) which has interpoint interactions of infinite order. We used the same parameters as in Figure 3.1 of Geyer (1999), namely interaction range $r = 0.05$, saturation level $c = 4.5$, activity $\beta = \exp(4.0)$, and interaction $\gamma = \exp(0.4) \approx 1.5$. Since $\gamma > 1$ this is a clustered point process. The simulation procedure was similar to that for panel (b).

10.3. Analogy with generalised linear models

Here we explain a connection between point process models and GLM's, which provides statistical insight. For point process models in time, Lewis (1972) recognised that the discretised likelihood is formally equivalent to the likelihood of a binomial regression model, which can be maximised using standard software (Brillinger, 1988, 1994; Lindsey, 1992, 1995). For spatial Poisson point processes, Berman and Turner (1992) developed a similar approach, which was extended to non-Poisson processes by Clyde and Strauss (1991), Lawson (1992) and Baddeley and Turner (2000). In the general case, a discretised version of the log pseudolikelihood (12) is formally equivalent to the loglikelihood of a Poisson loglinear regression. The conditional intensity $\lambda(u;\mathbf{x})$ of the point process corresponds to the mean response of the loglinear regression. In the Gibbs representation (7) of the conditional intensity, the first order term v_1 corresponds to the linear predictor of a GLM, while the higher order (interaction) terms v_k are roughly analogous to the distribution of the errors in a GLM.

11. Diagnostic plots for spatial trend and covariate effects

This section proposes diagnostics for spatial trend and covariate effects. In the GLM context, useful diagnostics for covariate effects are plots of the residuals against (i) index, (ii) each explanatory variable included in the model, and (iii) explanatory variables that were not included in the model, including surrogates for a lurking variable (Atkinson, 1985, pp. 3, 34, 62ff.). Here we explore analogues of these plots.

11.1. Spatial display of residuals

To start, consider two models fitted to Figure 4(a): the "correct" model, inhomogeneous Poisson with intensity $\lambda(x,y) = \beta\exp\{-\gamma x\}$, with maximum likelihood estimates $\widehat{\beta} = 233$, $\widehat{\gamma} = 2.89$; and the "null" model, homogeneous Poisson with intensity β (MLE $\widehat{\beta} = 73$).

The residual measure R has atoms at the points $x_i \in \mathbf{x}$ and a negative density at other locations $u \in W$. A simple pictorial representation of this is the *'mark plot'* shown in Figure 5. It consists of a pixel image of the density component (i.e. with greyscale proportional to the density $\widehat{h}(u, \mathbf{x})\widehat{\lambda}(u, \mathbf{x})$) and a symbol plot of the atoms (i.e. a circle centred at each point x_i of $\mathbf{x}$ with radius equal to the residual mass $\widehat{h}(x_i, \mathbf{x} \setminus \{x_i\})$). Figure 5 shows this representation for the two fitted models using the Pearson residuals. The expansion in size of circles from left to right in the first plot is a consequence of the model.

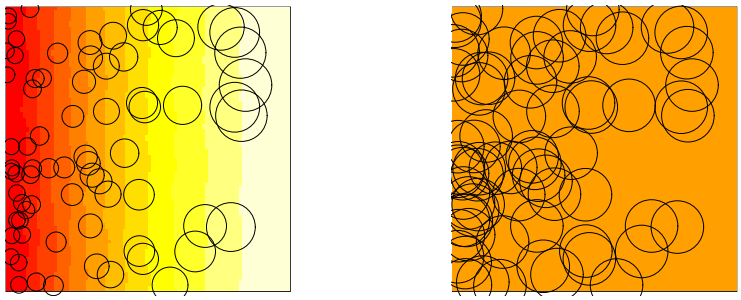

Fig. 5. Mark plot based on Pearson residuals for models fitted to Figure 4(a). *Left:* inhomogeneous Poisson model of correct form. *Right:* incorrect model, homogeneous Poisson.

The mark plot may sometimes identify 'extreme' data points (cf. Figure 14). However the diagnostic interpretation of the residuals is based primarily on their 'sums' over subregions B using (17).

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 $R(B_k, h, \theta)$. Nonzero residuals suggest a lack of fit. For example, if the fitted model is the homogeneous Poisson process, the raw residual sum $R(B) = n(\mathbf{x} \cap B) - |B|n(\mathbf{x})/|W|$ is the usual residual for the number $n(\mathbf{x} \cap B)$ of data points falling in B. Hence this technique embraces the method of quadrat counting used in spatial statistics (Diggle, 2003, Section 2.5; Cressie, 1991; Stoyan et al., 1995). For other models, $R(B)$ is a weighted count with data-dependent, spatially-varying weights. Such weights have not previously been used in quadrat methods to our knowledge.

A better approach is to smooth the residual measure. Taking a smoothing kernel $k(\cdot)$ (a probability density on $\mathbb{R}^2$), the *smoothed residual field* at location u is

$$\begin{aligned} s(u) &= e(u) \int_W k(u-v)\, \mathrm{d}R(v, \widehat{h}, \widehat{\theta}) \\ &= e(u) \left[\sum_{x_i \in \mathbf{x}} k(u-x_i)\widehat{h}(x_i, \mathbf{x} \setminus \{x_i\}) - \int_W k(u-v)\widehat{h}(v, \mathbf{x})\widehat{\lambda}(v, \mathbf{x})\, \mathrm{d}v \right] \end{aligned} \tag{31}$$

where $e(u)$ is a correction for edge effects in the window W given by $e(u)^{-1} = \int_W k(u-v)\, \mathrm{d}v$. The smoothed residual field s may be presented as a contour plot and greyscale image as shown in Figure 6. Bandwidth selection is discussed in Section 13.

The analogous quantity for the innovations has mean zero,

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

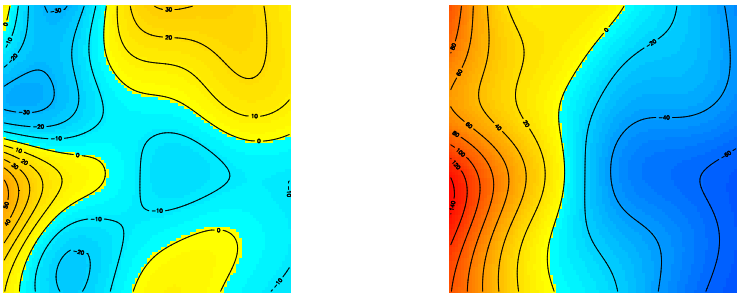

Fig. 6. Contour plots of kernel-smoothed raw residual field for two models fitted to Figure 4(a). *Left:* heterogeneous Poisson model of correct form. Range of smoothed field $[-34.3, 54.2]$ *Right:* incorrect model, homogeneous Poisson. Range of smoothed field $[-67.0, 145.6]$. Smoothing kernel: isotropic Gaussian with standard deviation 0.14. Same greyscale map in each plot.

and we hope $s(u) \approx 0$ if the fitted model is correct. For example, for the raw residuals,

$$s(u) = e(u) \left[\sum_{x_i \in \mathbf{x}} k(u - x_i) - \int_W k(u - v) \widehat{\lambda}(v, \mathbf{x}) \, \mathrm{d}v \right] = \lambda^*(u) - \lambda^\dagger(u),$$

the difference between $\lambda^*(u)$, a nonparametric, unbiased, kernel-smoothing estimator of the point process intensity function, and $\lambda^\dagger(u)$, a kernel-smoothed version of the parametric estimator of the conditional intensity. These two estimates of intensity should be approximately equal if the fitted model is correct. Positive values of $s(u)$ suggest the model underestimates the intensity.

Fowlkes (1987) proposed a smoothed residual plot for binary logistic regression in order to avoid artefacts of the binary nature of the responses. However this involved smoothing the *responses* before computing pseudo-residuals. Here it seems more appropriate to smooth the *residuals* after fitting, so that the smoothed residuals still have approximately zero mean under the model.

Kernel-smoothed estimates of the point process intensity (analogous to $\lambda^*(u)$) have been used as exploratory tools in spatial statistics (Diggle, 2003, 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.

Ogata (1988) 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.

11.2. 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 (Atkinson, 1985, p. 3, 34, 62). 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 residuals against a spatial covariate, or against one of the Cartesian coordinates (or some other coordinate), 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

$$W(z) = \{u \in W : \ Z(u) \le z\} \tag{32}$$

yielding a 'cumulative residual' function

$$A(z) = R(W(z), \widehat{h}, \widehat{\theta}). \tag{33}$$

This should be approximately zero if the fitted model is correct. For example, for the raw residuals,

$$A(z) = n(\mathbf{x} \cap W(z)) - \int_{W(z)} \widehat{\lambda}(u, \mathbf{x}) \, \mathrm{d}u. \tag{34}$$

Notice that $\widehat{H}(z) = n(\mathbf{x} \cap W(z))/n(\mathbf{x})$ is the empirical cumulative distribution function of the values of the covariate observed at the data points x_i. The function $A(z)$ in (34) is an adjustment of this empirical c.d.f. to have approximately zero mean under the model.

In Figure 4(a), the lurking variable is the x coordinate. Figure 7 shows plots of $A(z)$ against z based on the covariate $Z(x, y) = x$. Thus $A(z)$ is the "sum" of residuals in the region $W(z)$ to the left of the line $x = z$.

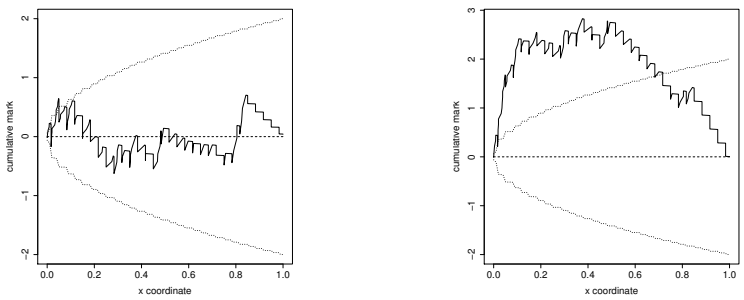

Fig. 7. Lurking variable plots for the x coordinate, for two models fitted to the data in Figure 4(a). Cumulative style, Pearson residuals. *Left:* model of correct form, inhomogeneous Poisson process with intensity loglinear in x, fitted to data. *Right:* incorrect model, homogeneous Poisson process, fitted to data. Solid line: empirical curve $A(x)$. Dotted lines: pointwise 2σ limits based on (27).

The dotted envelopes in Figure 7 are 2σ limits based on the variance of the innovations under an inhomogeneous Poisson process. This is an overestimate of the residual variance, because of variance deflation. We use $\mathsf{var}\,[A(z)] \approx \mathsf{var}\,[I(W(z))]$ where $W(z)$ is given in (32). This variance can be estimated using (25)–(27), substituting the fitted Poisson intensity $\widehat{\lambda}(u)$ for $\lambda(u)$. In this case we used the Pearson residuals, which are standardised so that $\mathsf{var}\,[I(B)] = |B|$ regardless of λ. Thus the dotted limits in the left and right panels of Figure 7 are identical.

The 2σ limits have the usual significance interpretation (pointwise) assuming a central limit theorem applies. The glaring violation of these bounds by the graph in the right panel of Figure 7 is ample evidence that a homogeneous trend 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 the normal approximation may be inaccurate for

small x (since it relates to a small subset of the data). The lurking variable plot is very effective in this trivial example.

Alternatively we may plot an approximate derivative of $A(z)$, such as

$$\begin{aligned} a(z) &= \int_W k_1(Z(u) - z) \, \mathrm{d}R(u, \widehat{h}, \widehat{\theta}) \\ &= \sum_{x_i \in \mathbf{x}} k_1(Z(x_i) - z) \widehat{h}(x_i, \mathbf{x} \setminus \{x_i\}) - \int_W k_1(Z(u) - z) \widehat{h}(u, \mathbf{x}) \widehat{\lambda}(u, \mathbf{x}) \, \mathrm{d}u \end{aligned} \quad (35)$$

where k_1 is a one-dimensional smoothing kernel (a probability density on $\mathbb{R}$). If the fitted model is correct we expect $a(z) \approx 0$.

11.3. Four-panel plot

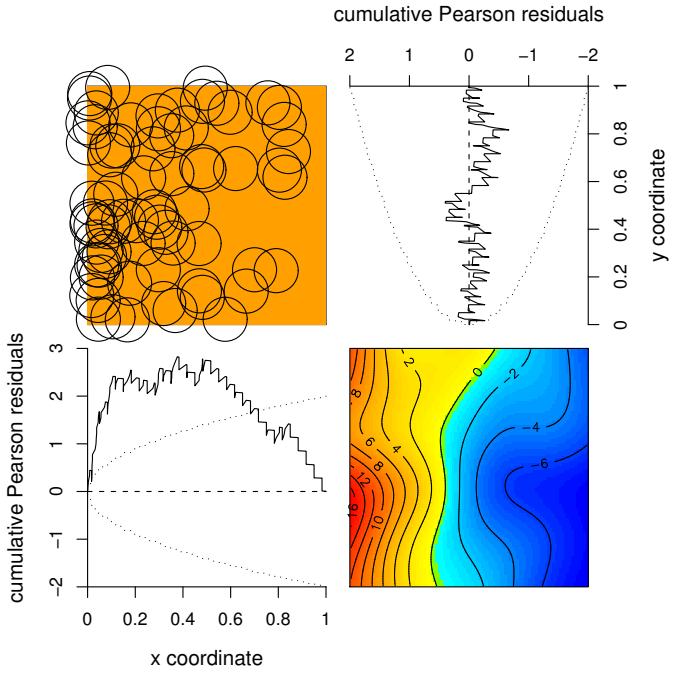

Fig. 8. Standard presentation of the diagnostic plots. Top left: mark plot. Top right: lurking variable plot for y coordinate. Bottom left: lurking variable plot for x coordinate. Bottom right: contour plot of smoothed mark field. Data from Figure 4(a); fitted model is homogeneous Poisson; Pearson residuals.

Figure 8 shows our standard presentation of the diagnostic plots for spatial trend. At the top left is the mark plot. The bottom right panel is a contour and image plot of the smoothed residual field, rendered so that the value $s(u) = 0$ is always represented by the same greyscale, for easy interpretation.

At the bottom left is a lurking variable plot for the x coordinate, its x axis aligned with the x coordinates in the mark plot. At top right is a lurking variable plot for the y

coordinate, rotated 90 degrees anticlockwise, the y axis aligned 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.

11.4. Trend in the presence of interaction

We now turn to the more challenging example in Figure 4(b), a simulation of an inhomogeneous Strauss process with log-quadratic activity function (29) and pair interaction (30). Figure 9 shows the four-panel diagnostic plots for two models fitted to these data. Both models have the correct interpoint interaction (30) with fixed range $r = 0.05$ but with γ estimated. On the left is a fitted model of the correct form, with activity

$$b(x, y) = \beta_0 \exp(\beta_1 x + \beta_2 y + \beta_3 x^2) \tag{36}$$

where β_i are estimated by maximum pseudolikelihood (MPL) yielding $(\widehat{\beta}_0, \widehat{\beta}_1, \widehat{\beta}_2, \widehat{\beta}_3) = (180, 1.53, 1.90, 1.35)$ and $\widehat{\gamma} = 0.24$; on the right is an incorrect model with homogeneous trend $b(x, y) \equiv \beta$ where β is estimated with MPLE $\widehat{\beta} = 1099$, $\widehat{\gamma} = 0.34$.

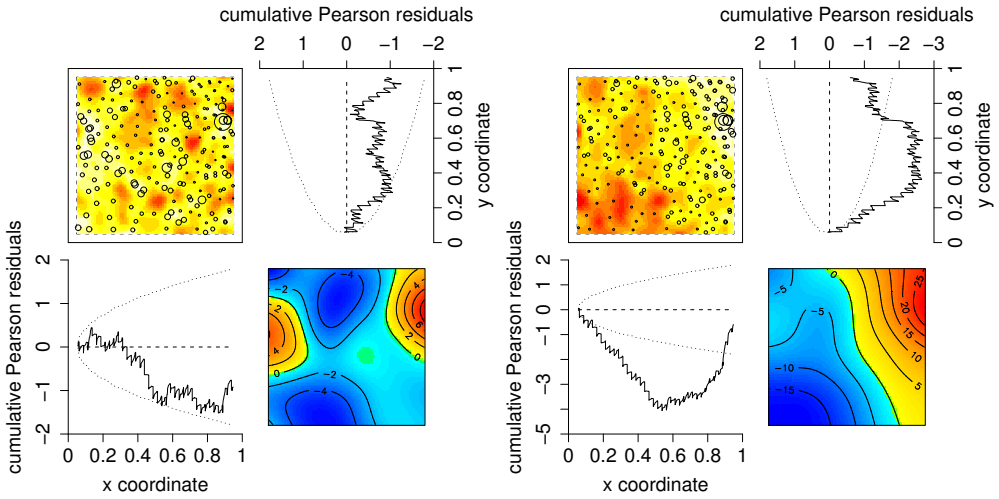

Fig. 9. Four-panel diagnostic plots for (*Left*) a model of the correct form (inhomogeneous Strauss process with log-quadratic activity), and (*Right*) model with incorrect trend, homogeneous Strauss process, fitted to the data in Figure 4(b). Dotted lines are 2σ limits for a Poisson model.

In computing residuals from a Strauss model, we encounter an 'edge effect' problem. Suppose the data are only a partially observed realisation of a point process $\mathbf{Y}$ in a larger bounded region S containing W, and let $\lambda(u, \mathbf{x})$ denote the Papangelou conditional intensity for $\mathbf{Y}$. Then

$$\lambda(u, \mathbf{x}) = b(u)\gamma^{t(u,\mathbf{x})} \tag{37}$$

depends on $t(u, \mathbf{x})$, the number of points of $\mathbf{x}$ within a distance r of the location u. When u lies close to the boundary of W, this number is not observable. To avoid this, we compute and plot 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 boundary of W. When W is the unit square, $W_{\ominus r} = [r, 1-r]^2$. We caution that substantial bias may occur if edge effects are ignored.

The dotted lines in Figure 9 are the 2σ limits for the innovations under the *Poisson* model and are shown for indicative purposes only. These underestimate the innovation variance for a Strauss model. We are still developing algorithms for computing residual variance in non-Poisson models.

The left panel in Figure 9 indicates that the correct model is a tolerably good fit, although it (correctly) suggests the trend is underestimated. In the right panel, the lurking variable plot for the x coordinate (bottom left) shows a distinctive and persistent dip, which strongly indicates that the homogeneous model is inappropriate. The mark plots show a few data points with large residual mass. Before interpreting these as 'outliers' one should remember the exponential form of the conditional intensity (37) and the small value of the interaction parameter $\widehat{\gamma} \sim 0.25$. 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 $\widehat{\gamma}^{-1/2} \sim 2$. This phenomenon is even more exaggerated for the inverse-lambda residuals, where each extra neighbour increases the residual mass by a factor of $\widehat{\gamma}^{-1} \sim 4$. This sensitivity is analogous to the high variance of the Horvitz-Thompson estimator (Horvitz and Thompson, 1952) when some of the sampling units have small probabilities of being selected. 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.

11.5. Applications

11.5.1. Japanese pines data

Ogata and Tanemura (1986) fitted an inhomogeneous pairwise interaction model to Figure 1. The activity function b was log-cubic (i.e. $\log b(u)$ was a cubic polynomial in the Cartesian coordinates of u) and c was the 'soft-core' interaction

$$c(u, v) = \exp\left\{ -\left(\frac{\sigma}{||u - v||} \right)^{2/\kappa} \right\}, \quad u \neq v \tag{38}$$

where $0 \leq \sigma < \infty$ and $0 < \kappa < 1$ are parameters with $\kappa = 0.5$ fixed. See also Baddeley and Turner (2000).

The left side of Figure 10 shows the diagnostic plots for the Ogata-Tanemura model. It suggests good agreement between the trend in the fitted model and in the data. The right side shows the diagnostics for the *homogeneous* soft-core model, i.e. with no trend term but with soft-core interaction as in the Ogata-Tanemura model. This shows clearly that the homogeneous model misspecifies the trend. A similar diagnostic plot for the soft-core model with *log-linear* trend conveys the same message.

11.5.2. Queensland copper data

A common thread in the analyses of Figure 2 (Berman, 1986; Berman and Turner, 1992; Foxall and Baddeley, 2002) is to assess the dependence of the point pattern of copper deposits on proximity to the lineaments. Let the covariate $Z(u)$ be the distance from the location u to the nearest lineament. This can easily be computed analytically for all locations in a fine pixel grid. See Figure 11.

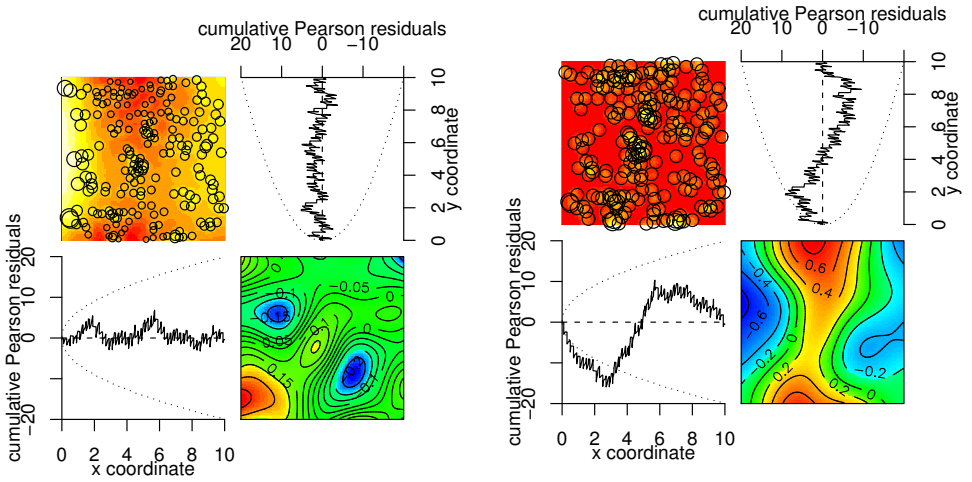

Fig. 10. Diagnostics for two models fitted to the Japanese pines data of Figure 1. *Left:* soft core with log-cubic trend. *Right:* homogeneous soft-core.

Fig. 11. Spatial covariate for the copper data of Figure 2. Thick lines: observed lineaments. Thin lines: contours of distance to the nearest lineament.

Because of the fine spatial structure of this covariate, the four-panel plot is not useful: instead a lurking variable plot for the covariate is appropriate. First we fit the null model of a homogeneous Poisson process. For this model, and for the raw residuals, (34) becomes

$$A(z) = n(\mathbf{x} \cap W(z)) - n(\mathbf{x})|W(z)|/|W| = n(\mathbf{x}) \left[\widehat{H}(z) - H_0(z)\right]$$

where $\widehat{H}(z) = n(\mathbf{x} \cap W(z))/n(\mathbf{x})$ is the empirical c.d.f. of the values of the covariate at the data points, and $H_0(z) = |W(z)|/|W|$ is the empirical c.d.f. of the covariate at all locations in W. Berman (1986) proposed comparing the two c.d.f.'s $\widehat{H}$ and H_0 by plotting them against z, as shown in the left panel of Figure 12, and comparing various moments. This is equivalent to a lurking variable plot for the exponential energy marks (for the homogeneous Poisson model). The right panel shows a lurking variable plot of the Pearson residuals for the homogeneous Poisson model. Both plots suggest the model is adequate. Our technique has the advantage that any fitted model may be treated in the same way.

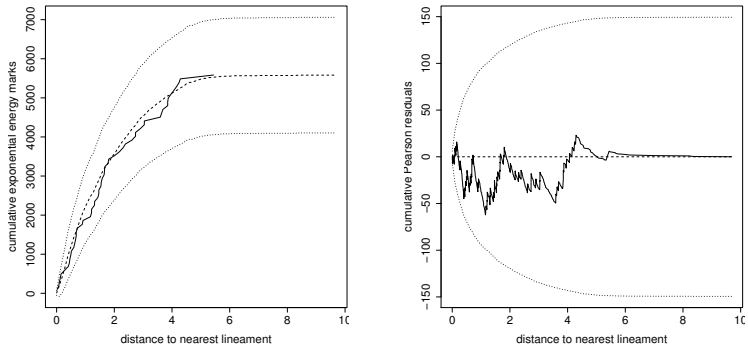

Fig. 12. Lurking variable plots for the homogeneous Poisson point process fitted to the Queensland copper data. *Left:* Berman's plot: unnormalised empirical c.d.f. of distance to nearest lineament at the observed points of the process. *Right:* cumulative Pearson residuals against distance to nearest lineament. Pointwise 2σ limits based on equations (25) and (27) respectively.

11.5.3. Chorley-Ribble data

The null model considered by Diggle (1990) states that the point pattern of cases of cancer of the larynx in Figure 3 is an inhomogeneous Poisson process with intensity $\lambda(u) = \beta\,\rho(u)$ where $\beta > 0$ is a parameter and $\rho(u)$ 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, using an isotropic Gaussian kernel with standard deviation $\sigma = 0.15$ km as chosen by Diggle (1990).

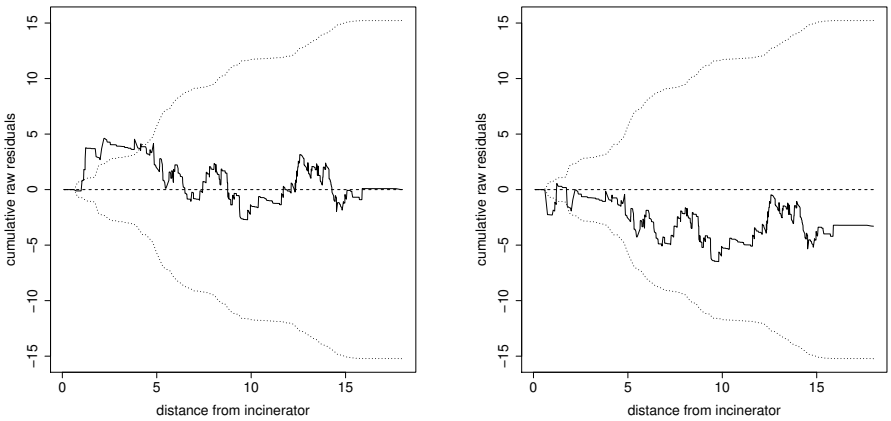

Fig. 13. Lurking variable plots of the raw residuals against distance from the incinerator, for the Chorley-Ribble data. *Left:* null model. *Right:* model fitted by Diggle and Rowlingson (1994).

Figure 13 shows lurking variable plots of the raw residuals against distance from the incinerator, for two models fitted to the Chorley-Ribble data. The left panel is for the null model of no incinerator effect. It suggests strongly that the null model is not correct, and that there is increased intensity near the incinerator. Possible explanations discussed by Diggle (1990) include clustering of disease cases (e.g. due to correlation within families) as

well as a carcinogenic effect from the incinerator.

The right panel in Figure 13 is for the model of Diggle (1990) that includes an incinerator effect. It is an inhomogeneous Poisson process with intensity $\lambda(u) = \beta\rho(u)\delta(u,\theta)$, where $\delta(u,\theta)$ is a parametric function of the distance from u to the incinerator (Diggle, 1990, equation (6)). We used the parameter values θ obtained by Diggle and Rowlingson (1994). The right panel gives a slight suggestion that the model fitted by Diggle and Rowlingson (1994) overestimates the intensity of cases, at distances close to zero and beyond 5 km. The graph strays outside the 2σ limits close to the origin (distance $= 0$) but, again, these limits probably have less than 95% coverage near the origin because the expected number of cases is small. A more conclusive assessment of significance could be obtained by simulation. The left panel of Figure 13 can also be used to suggest the functional form of the incinerator effect term $\delta(u,\theta)$ as alternatives to equation (6) of Diggle (1990).

The lurking variable plot, for the raw residuals and for a Poisson fitted model, is a plot of $n(\mathbf{x})\widehat{F}(d) - \widehat{T}(d)$ against d, where $\widehat{F}(d)$ is the empirical c.d.f. of the distances d_i, and $\widehat{T}(d) = \int_{W(d)} \widehat{\lambda}(u)\, \mathrm{d}u$. This is closely related to the model-checking technique used in Berman (1986) and Diggle (1990, Section 3.2). The lurking variable plots for the inverse-lambda and Pearson residuals also have direct interpretations.

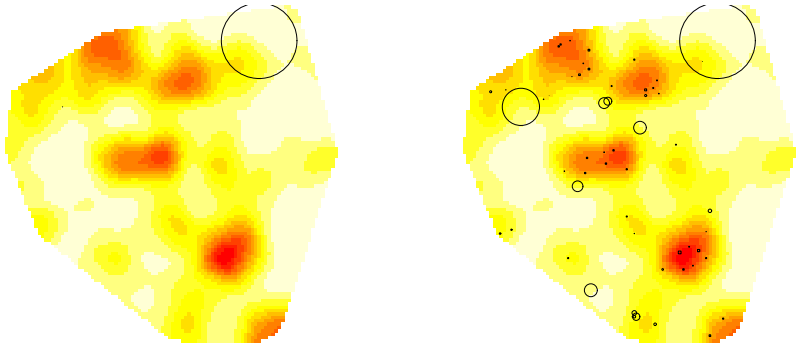

Fig. 14. *Left:* mark plot of Pearson residuals for the null model fitted to the Chorley-Ribble data, showing a huge residual at the location $(360, 428)$. *Right:* radii proportional to log residuals.

Diagnostic plots for the inverse-lambda and Pearson residuals reveal some other inconsistencies between the data and the *fitted* null model. Figure 14 shows an apparent 'outlier'. This is a case of laryngeal cancer where the kernel estimate $\widehat{\rho}(u)$ is very low. This raises questions about the appropriateness of the estimator $\widehat{\rho}$, rather than necessarily indicating an anomalous observation.

The greatest advantage of the lurking variable plot is that it is equally applicable to non-Poisson point process models. Diggle (1990) mentioned several competing explanations for the observed elevated incidence of laryngeal cancer near the incinerator, including outliers and clustering. It is feasible to assess alternative models by computing the analogue of Figure 13. For example, to assess whether the raised incidence of laryngeal cancer near the incinerator could be attributable to clustering of disease cases, we fitted a heterogeneous version of the Geyer saturation process model (Geyer, 1999) to the Chorley-Ribble data, with first order term proportional to $\rho(u)$. This model allows for either positive or negative association between points. The fitted model was, however, very close to a Poisson process, and the lurking variable plots were indistinguishable from the left panel of Figure 13. This

suggests that clustering (as fitted in this model) does not explain the observations. Further investigation will be reported elsewhere.

12. Diagnostic plots for interpoint interaction

12.1. Q–Q plots

Next we develop residual plots to validate the interpoint interaction component of a model. Under the analogy between point processes and GLM's, interpoint interaction in a point process is analogous to the distribution of residuals in a GLM. The most appropriate tool for assessing the distributional assumptions in a GLM 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 residual 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, and sorting them to obtain the order statistics. This is done for the data and for 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.

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 $s_{[1]} \le s_{[2]} \le \ldots \le s_{[J]}$. 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 $s_j^{(n)} = s(u_j, \mathbf{x}^{(n)}, \widehat{\theta}^{(n)})$, and obtain order statistics $s_{[1]}^{(n)} \le \ldots \le s_{[J]}^{(n)}$. 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. A rationale for using expected quantiles is offered in Gnanadesikan and Wilk (1970). The Q–Q plot is a scatterplot of the data quantiles $s_{[j]}$ against the mean quantiles e_j. To gauge the significance of any deviations we may add critical intervals for $s_{[j]}$, of pointwise significance level α, obtained as the sample quantiles, of probability $\alpha/2$ and $1-\alpha/2$, of $\{s_{[j]}^{(1)}, \ldots, s_{[j]}^{(N)}\}$.

We caution that substantial bias and other artefacts in the Q–Q plot may occur if edge effects are ignored. The residuals should only be evaluated in the eroded window $W_{\ominus r}$ where r is the range of interpoint interaction.

12.2. Test example

Figure 4(c) is a homogeneous clustered pattern generated by Geyer's saturation process. Figure 15 shows Q–Q plots, based on the Pearson residuals, for models fitted to this pattern. The left panel of Figure 15 is for a model of the correct form, the homogeneous saturation process, with the irregular parameters r, c fixed at their correct values. This suggests good agreement between the fitted model and the data. The right panel is for the (incorrect) homogeneous Poisson model. It shows clear disagreement between the model and the data. The smoothed residual field for the data has a heavier left tail, and higher variability, than the smoothed residual field for simulations from the fitted Poisson model. This is consistent

with a clustered point pattern. Note the different scales of the left and right panels, and the wider prediction interval in the left panel.

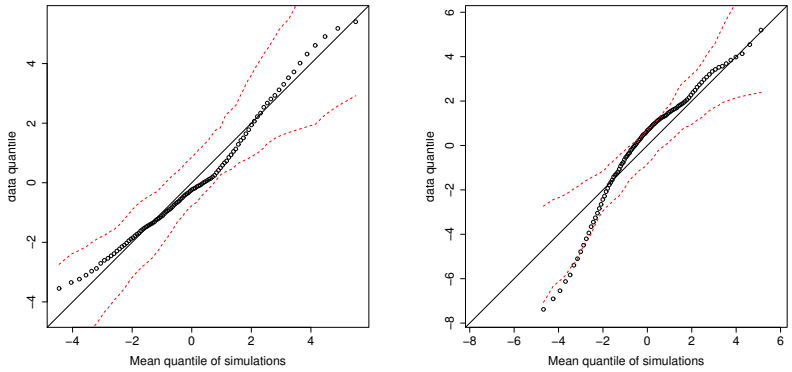

Fig. 15. Q–Q plots for spatial point process models fitted to Figure 4(c). *Left*: model of the correct form, homogeneous saturation process. *Right:* homogeneous Poisson process. Dotted lines show pointwise 95% critical intervals obtained by simulation.

For each Q–Q plot in Figure 15, the smoothed residual field was evaluated at a 40×40 grid of pixels, and percentiles of these values obtained. The expected percentiles were estimated (using the 5% trimmed sample mean) from 100 simulated realisations of the fitted model. For the saturation model (left panel), each realisation was obtained from the same simulation algorithm as the original data. Total computation time for Figure 15 was approximately 70 minutes on a 2.0 GHz Linux desktop. For investigative work, adopting a 25×25 pixel grid, using only 40 simulated realisations and 10^4 Metropolis-Hastings iterations per realisation, produces a practically useful Q–Q plot in one minute.

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 in general; it is done for the homogeneous Poisson process in Baddeley et al. (2005). The null distribution of values $s(u)$ is heavily skewed, justifying the use of the trimmed mean in our Monte Carlo calculations.

12.3. Spatial rationale for interpreting Q–Q plots

Qualitative interpretation of the Q–Q plots requires us to understand the information conveyed in the smoothed residual field $s(u)$ in (31). For heuristic purposes, suppose the fitted model is CSR, and take the smoothing kernel k to be the uniform density on the disc of radius r centred at the origin, $k(u) = \mathbf{1}\{||u|| \le r\}/(\pi r^2)$. Ignore edge effects by restricting attention to locations u in $W_{\ominus r}$. Then the raw residual field is

$$s(u) = \frac{t(u, \mathbf{x})}{\lambda \pi r^2} - 1$$

where again $t(u, \mathbf{x})$ is the number of points of $\mathbf{x}$ within a distance r of the location u. This is known to contain information about interpoint interaction. The maximum value of $t(u, \mathbf{x})$ over all locations $u \in W_{\ominus r}$ is the *scan statistic*, a well-known summary statistic used for detecting clustering (Kulldorff, 1999). The sum of $t(x_i, \mathbf{x})$ over all data points $x_i \in \mathbf{x}$ is

proportional to an estimate of the K function. The zero fraction of $t(u, \mathbf{x})$ is

$$|\{u \in W_{\ominus r} : \ t(u, \mathbf{x}) = 0\}|/|W_{\ominus r}| = 1 - \widehat{F}(r)$$

where $\widehat{F}$ is the empirical empty space function of $\mathbf{x}$ (using the border method edge correction). The latter is a popular summary statistic for detecting interpoint interaction (Ripley, 1977; Diggle, 2003; Møller and Waagepetersen, 2003a; Stoyan et al., 1995).

Rules for qualitative interpretation of the Q–Q plots are therefore very similar to the established rules for interpreting $\widehat{F}$. If the data pattern is more clustered than the model, the empirical distribution of $s(u)$ should have heavier tails than the reference distribution, especially in the left tail. Figure 15 is an example. If the pattern is more inhibited (less clustered) than the model, the empirical distribution of $s(u)$ should have lighter tails than the reference distribution, especially in the right tail. See Figure 17. Since it is the tails of $s(u)$ which are of primary interest, Q–Q plots are the appropriate tool.

12.4. Range of interpoint interaction

Here we test the ability of the Q–Q plots to detect an *incorrectly specified* interpoint interaction. Recall that the data in Figure 4(c) were generated by a saturation process with interaction range $r = 0.05$. Figure 16 shows Q–Q plots for fitted models in which r has been underestimated or overestimated with an error of 0.01. They clearly show that these models have incorrectly estimated the interaction. Similar results were obtained for under- and over-estimation of the interaction range r in the Strauss process.

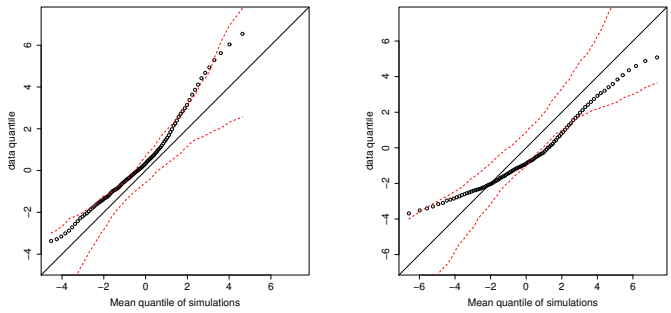

Fig. 16. Q–Q plots for fitted models in which the range of interaction has been underestimated (Left: $r = 0.04$) or overestimated (Right, $r = 0.06$) in the Geyer saturation model with $r = 0.05$. Fitted to data in Figure 4(c). Pearson residuals.

12.5. Interaction in presence of trend

The synthetic pattern in Figure 4(b) has log-quadratic spatial trend and a Strauss interpoint interaction. Figure 17 shows the diagnostic plots for a model with the correct form of spatial trend, but no interpoint interaction, that is, an inhomogeneous Poisson process with log-quadratic intensity function (36). The trend plots show no evidence of departure, while the Q–Q plot demonstrates that this model is quite inappropriate.

Figure 18 shows residual Q–Q plots for two models fitted to Figure 4(b) with the correct form of the interaction. The left panel is for a model with a trend of the correct form, and

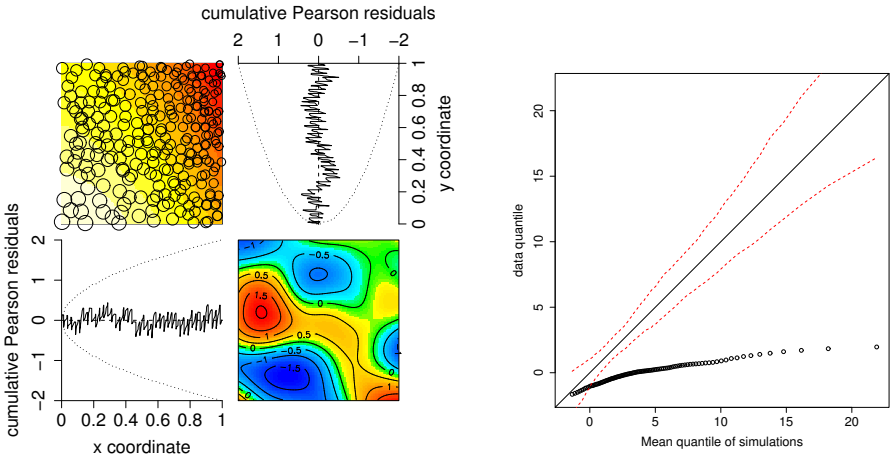

Fig. 17. Diagnostics for a Poisson process with log-quadratic intensity fitted to the data in Figure 4(b).

suggests that the interaction is modelled correctly. The right panel is for a homogeneous process, i.e. $b(u)$ is constant. It suggests the observed residuals have heavier left and right tails than the reference distribution. However this is an artefact of the large spatial variation in intensity, and shows that gross misspecification of the spatial trend can affect the Q–Q plots.

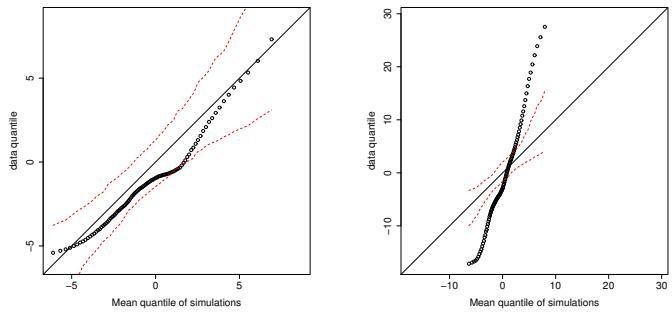

Fig. 18. Q–Q plots for models with the correct interaction (Strauss) fitted fitted to the data in Figure 4(b). *Left:* trend of the correct form (log-quadratic). *Right:* trend of the incorrect form (homogeneous).

12.6. Applications

12.6.1. Japanese pines data

The left panel of Figure 19 shows a Q–Q plot for the Ogata-Tanemura model, based on the Pearson residuals. It suggests the soft core interaction is a good fit. The right panel is for a homogeneous soft core model. Although the trend is misspecified, the Q–Q plot still suggests a reasonable fit to the interaction. The conclusion from our analysis here and in Section 11.5.1 is that the Ogata-Tanemura model is a good fit to the Japanese pines data.

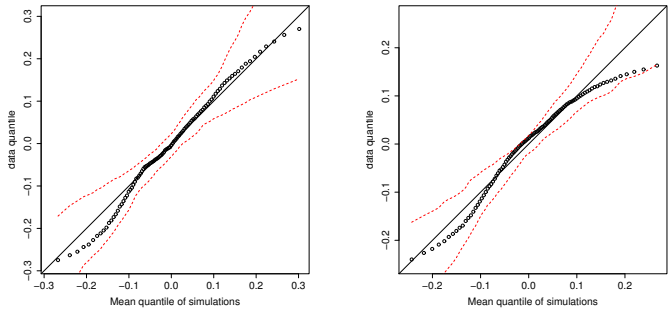

Fig. 19. Q–Q plots for soft-core models fitted to the Japanese pines data of Figure 1. *Left:* Ogata-Tanemura model, log cubic trend. *Right:* homogeneous trend.

12.6.2. Queensland copper data

Although the lurking variable plots in Figure 12 suggest that a homogeneous Poisson model for the copper deposits is adequate, these plots are only designed to assess spatial trend. The copper deposits are clustered, as can be shown using standard techniques such as the K function or G function, or using our Q–Q plot for a homogeneous Poisson model.

We fitted a homogeneous saturation model to the copper data by maximum profile pseudolikelihood (Baddeley and Turner, 2000) to obtain $\widehat{r} = 1.19$ km, $\widehat{c} = 2$, $\widehat{\gamma} = 2.86$, reflecting quite strong clustering. Diagnostics for this model are shown in Figure 20. The Q–Q plot suggests that the data may be even more clustered than the fitted model, assuming the trend is correctly specified; alternatively a covariate effect may be present. Models involving both covariate effects and strong clustering are investigated by Baddeley and Turner (2005a).

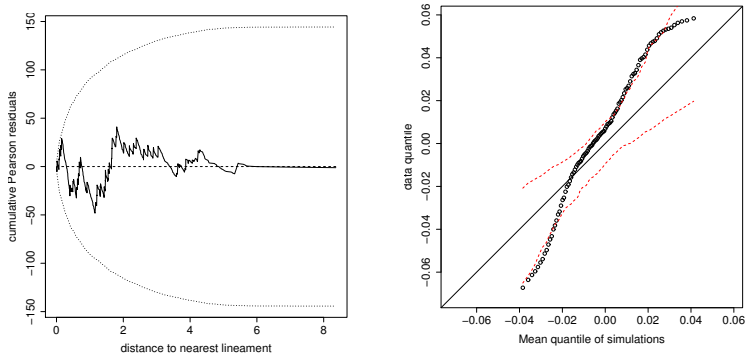

Fig. 20. Diagnostics for a homogeneous saturation model fitted to the Queensland copper data. *Left:* lurking variable plot against distance to nearest lineament. *Right:* Q–Q plot. Pearson residuals.

12.6.3. Cells data

An extreme example of interaction is the 'biological cells' dataset (Ripley, 1977) shown in the left panel of Figure 21. It is often modelled by a hard core process (Ripley, 1981; Diggle,

2003) with hard core radius 0.04 (the window is the unit square). This corresponds to a homogeneous Strauss process with interaction range $r = 0.08$ and interaction parameter $\gamma = 0$. A hard core model was fitted by maximum pseudolikelihood with $r = 0.08$ fixed, yielding $\widehat{\beta} = 201, 142$ with and without edge correction. Four-panel plots of the two models (not shown) indicate clearly that $\widehat{\beta} = 142$ is an underestimate, while $\widehat{\beta} = 201$ appears reasonable.

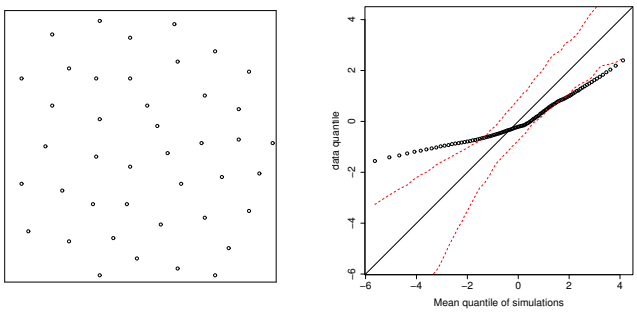

Fig. 21. *Left:* Cells data. *Right:* Q–Q plot for fitted Strauss process model. Pearson residuals.

The Q–Q plot in the right panel of Figure 21 is for the model with $\widehat{\beta} = 201$. The empirical quantiles have a lighter left tail, suggesting that the data are more tightly packed than the fitted model. However, in this extreme case, both the fitting algorithm (maximum pseudolikelihood) and simulation algorithm (Metropolis-Hastings) are known to have poor performance, and alternatives should be explored (cf. Mase et al., 2001).

In practice, a hard core model is often fitted using algorithms for fitting the Strauss process, since the hard core model is obtained by setting $\gamma = 0$ in the Strauss model. This introduces a further difficulty with the inverse-lambda residuals. Since the residual density $\widehat{h}(u, \mathbf{x})\widehat{\lambda}(u, \mathbf{x}) = \mathbf{1}\{\widehat{\lambda}(u, \mathbf{x}) > 0\}$ is a discontinuous function of $\widehat{\lambda}(u, \mathbf{x})$, the residuals are unstable to numerical error when $\widehat{\lambda}(u, \mathbf{x}) \approx 0$, that is, when $\widehat{\gamma} \approx 0$. The remedy is to constrain γ to be exactly zero.

13. Computation

This section explains our strategy for computing the spatial residuals. The software implementation has been incorporated in our library `spatstat` (Baddeley and Turner, 2005a,b) in the R package (R Development Core Team, 2004).

13.1. Discretised residual measure

Some discretisation of the residual measure is required for computation. For this paper we fitted models by maximum pseudolikelihood using the device of Berman and Turner (1992), although this is not essential to the technique. It is therefore convenient to use the same discretisation device to compute the residual measure.

In the Berman-Turner device, the observed point pattern $\mathbf{x}$ is first augmented by numerous "dummy" points to form a set of quadrature points u_j, $j = 1, \ldots, M$. Weights w_j are then associated with the quadrature points u_j so that integrals of the form $\int_B g(u)\, \mathrm{d}u$ are

well approximated by finite sums $\sum_j \mathbf{1}\{u_j \in B\} w_j g(u_j)$. The residual measure $R(B, \widehat{h}, \widehat{\theta})$ will be approximated by

$$R^*(B) = \sum_{u_j \in B} (y_j - \widehat{\lambda}_j) w_j \widehat{h}_j \tag{39}$$

where $\widehat{h}_j = h_{\widehat{\theta}}(u_j, \mathbf{x} \setminus \{u_j\})$, $\widehat{\lambda}_j = \lambda_{\widehat{\theta}}(u_j, \mathbf{x})$ and $y_j = z_j/w_j$, where z_j is the indicator equal to 1 if u_j is a data point and 0 for a dummy point. The y_j correspond to the responses in the associated Poisson loglinear regression. An individual summand in (39),

$$r_j^h = (y_j - \widehat{\lambda}_j) w_j \widehat{h}_j = (z_j - \widehat{\lambda}_j w_j) \widehat{h}_j, \tag{40}$$

may be regarded as a residual for the region surrounding the quadrature point u_j.

Note that the residuals attached to the data points x_i have different meanings in the continuous theory (Section 7) and in the discrete approximation (40). For example, the raw residuals defined in Section 7.2 attach unit mass to each data point x_i, while the discretised raw residual $r_j^1 = z_j - \widehat{\lambda}_j w_j$ is not equal to 1 for a data point; rather it approximates the residual sum over a region associated with u_j.

In other statistical modelling contexts where a weighted likelihood is used, the Pearson residuals would usually be defined as $r_j^* = \sqrt{w_j}\left(y_j/\widehat{\lambda}_j^{1/2} - \widehat{\lambda}_j^{1/2}\right)$ so that the sum of squared Pearson residuals equals the Pearson X^2 statistic. However, this does not hold here.

13.2. Bandwidth selection

The bandwidth for the kernel k in (31) could be chosen in several ways. Since the smoothed raw residual field is closely related to a kernel-smoothed estimate of the point process intensity, we could minimise mean square error (Diggle, 1985; Berman and Diggle, 1989; Stoyan and Stoyan, 1995, pp. 237–238). Alternatively there are rules of thumb for bandwidth selection for the pair correlation function (Stoyan and Stoyan, 1995, p. 285).

The bandwidths in this paper were chosen by cross-validation (Wand and Jones, 1995) of the discretised residuals. Let $r_j = r_j^h$ denote one of the discretised residuals defined above. Recall that r_j is a quadrature approximation to the residual integral in a region surrounding u_j. By the usual rationale for nonparametric regression, we should apply cross-validation to the values $s_j = r_j/w_j$. Let $k_b(u) = b^2 k_1(u/b)$ be the kernel of bandwidth $b > 0$, where k_1 is a fixed probability density on $\mathbb{R}^2$. Then we choose b to minimise $C(b) = \sum_i (s_i - \widehat{s}_i)^2$ where $\widehat{s}_i = \sum_{j \neq i} a_{ij} s_j / \sum_{j \neq i} a_{ij}$ with $a_{ij} = k_b(u_j - u_i)$.

14. Scope of application

The residuals are defined for *any* point process model with a density satisfying the positivity condition (2). A practical constraint is that the conditional intensity must be computable. Our current software does not handle two important families of models for clustered patterns, the independent cluster processes and Cox processes. However, the absence of examples of these models here does not reflect any fundamental limitation of the method.

For a Cox process $\mathbf{X}$ driven by a random intensity function $\Lambda(u)$, $u \in W$ (that is, $\mathbf{X}$ is conditionally a Poisson process with intensity function Λ) the conditional intensity is

$$\lambda(u, \mathbf{x}) = \mathbb{E}[\Lambda(u) | \mathbf{X} = \mathbf{x}], \quad u \notin \mathbf{x}. \tag{41}$$

The right hand side can be estimated by MCMC methods (Møller et al., 1998; Møller, 2003; Møller and Waagepetersen, 2003a) so that the residuals can then be computed. The same calculation is required to implement many MCMC techniques for this model.

The trend plots (Section 11) will have the greatest benefit for models where it is easy to compute the conditional intensity. This includes (inhomogeneous) pairwise interaction processes like the Strauss process, and some infinite-order interactions like Geyer's saturation model. For such models the trend plots do not require any simulation.

The Q–Q plots depend on extensive simulation, and therefore compete on a more even footing with *ad hoc* methods in spatial statistics; except that the Q–Q plot is already a familiar tool, and that the residuals are constructed so that they reflect intrinsically any deviation from the specific model in question.

Our technique requires that the fitted model be a fully-specified point process. It does not immediately apply to partial likelihood techniques, like that of Diggle and Rowlingson (1994), which only estimate some of the model's parameters. However, the missing parameters can usually be filled in by standard estimators, and our technique then applies.

Acknowledgements

We thank David Brillinger, Peter Green, Peter Guttorp, Tony Pakes, Dietrich Stoyan, Paul Switzer, Berwin Turlach, David Vere-Jones, Rick Vitale, Joe Whittaker and the referees for illuminating comments. This research was supported by the Australian Research Council (Large Grant A69941083), the Danish Natural Science Research Council, and NSERC.

References

Andersen, P., Ø. Borgan, R. Gill, and N. Keiding (1993). *Statistical Models Based on Counting Processes*. New York: Springer-Verlag.

Anselin, L. (1995). Local indicators of spatial association – LISA. *Geographical Analysis 27*, 93–115.

Atkinson, A. (1985). *Plots, Transformations and Regression*. Number 1 in Oxford Statistical Science Series. Oxford University Press/ Clarendon.

Baddeley, A., J. Møller, M. Hazelton, and R. Turner (2005). Residuals for spatial point processes using the Papangelou conditional intensity. In preparation.

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

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

Baddeley, A. and R. Turner (2005a). Modelling spatial point patterns in R. In A. Baddeley, J. Mateu, and D. Stoyan (Eds.), *Case Studies in Spatial Point Pattern Modelling*. Springer-Verlag. To appear.

Baddeley, A. and R. Turner (2005b). Spatstat: an R package for analyzing spatial point patterns. *Journal of Statistical Software 12*(6), 1–42. URL: `www.jstatsoft.org`, ISSN: 1548-7660.

Baddeley, A., R. Turner, J. Møller, and M. Hazelton (2004). Residual analysis for spatial point processes. Research Report 2004/08, School of Mathematics and Statistics, University of Western Australia.

Berman, M. (1986). Testing for spatial association between a point process and another stochastic process. *Applied Statistics 35*, 54–62.

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

Berman, M. and T. Turner (1992). Approximating point process likelihoods with GLIM. *Applied Statistics 41*, 31–38.

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

Besag, J. (1978). Some methods of statistical analysis for spatial data. *Bulletin of the International Statistical Institute 44*, 77–92.

Besag, J. and P. Diggle (1977). Simple Monte Carlo tests for spatial pattern. *Applied Statistics 26*, 327–333.

Brillinger, D. (1978). Comparative aspects of the study of ordinary time series and of point processes. In *Developments in Statistics*, pp. 33–133. Academic Press.

Brillinger, D. (1988). Maximum likelihood analysis of spike trains of interacting nerve cells. *Biological Cybernetics 59*, 189–200.

Brillinger, D. (1994). Time series, point processes, and hybrids. *Canadian Journal of Statistics 22*, 177–206.

Brillinger, D. and H. Preisler (1986). 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*.

Clyde, M. and D. Strauss (1991). Logistic regression for spatial pair-potential models. In A. Possolo (Ed.), *Spatial Statistics and Imaging*, Volume 20 of *Lecture Notes - Monograph series*, Chapter II, pp. 14–30. Institute of Mathematical Statistics. ISBN 0-940600-27-7.

Collett, D. (1991). *Modelling Binary Data*. London: Chapman and Hall.

Cox, D. and P. Lewis (1966). *The Statistical Analysis of Series of Events*. London: Methuen.

Cox, D. R. and V. Isham (1980). *Point Processes*. London: Chapman and Hall.

Cressie, N. (1991). *Statistics for Spatial Data*. New York: John Wiley and Sons.

Cressie, N. and L. Collins (2001a). Analysis of spatial point patterns using bundles of product density LISA functions. *Journal of Agricultural, Biological and Environmental Statistics 6*, 118–135.

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

Daley, D. and D. Vere-Jones (1988). *An Introduction to the Theory of Point Processes*. New York: Springer Verlag.

Davison, A. and E. Snell (1991). Residuals and diagnostics. In D. Hinkley, N. Reid, and E. Snell (Eds.), *Statistical Theory and Modelling (in honour of Sir David Cox FRS)*, Chapter 4, pp. 83–106. London: Chapman and Hall.

Diggle, P. (1978). On parameter estimation for spatial point processes. *Journal of the Royal Statistical Society, series B 40*, 178–181.

Diggle, P. (1990). 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.

Diggle, P. (2003). *Statistical Analysis of Spatial Point Patterns* (Second ed.). London: Arnold.

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

Diggle, P. and B. Rowlingson (1994). A conditional approach to point process modelling of elevated risk. *Journal of the Royal Statistical Society, series A 157*, 433–440.

Diggle, P. J. (1985). A kernel method for smoothing point process data. *Applied Statistics 34*, 138–147.

Fowlkes, E. (1987). Some diagnostics for binary logistic regression via smoothing. *Biometrika 74*, 503–515.

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

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

Getis, A. and J. Franklin (1987). Second-order neighbourhood analysis of mapped point patterns. *Ecology 68*, 473–477.

Geyer, C. (1999). Likelihood inference for spatial point processes. In O. Barndorff-Nielsen, W. Kendall, and M. van Lieshout (Eds.), *Stochastic Geometry: Likelihood and Computation*, Number 80 in Monographs on Statistics and Applied Probability, Chapter 3, pp. 79–140. Boca Raton, Florida: Chapman and Hall / CRC.

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

Gnanadesikan, R. and M. Wilk (1970). A probability plotting procedure for general analysis of variance. *Journal of the Royal Statistical Society, series B 32*, 88–101.

Horvitz, D. and D. Thompson (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association 47*, 663–685.

Jensen, J. and J. Møller (1991). Pseudolikelihood for exponential family models of spatial point processes. *Annals of Applied Probability 1*, 445–461.

Karr, A. (1985). *Point Processes and their Statistical Inference*. New York: Dekker.

Kulldorff, M. (1999). Spatial scan statistics: models, calculations, and applications. In J. Glaz and N. Balakrishnan (Eds.), *Recent Advances on Scan Statistics*, pp. 303–322. Boston: Birkhauser.

Landwehr, J., D. Pregibon, and A. Shoemaker (1984). Graphical methods for assessing logistic regression models. *Journal of the American Statistical Association 79*, 61–83.

Lawson, A. (1992). On fitting non-stationary Markov point process models on GLIM. In Y. Dodge and J. Whittaker (Eds.), *Computational statistics*, Volume 1, Heidelberg-New York, pp. 35–40. Physica/Springer.

Lawson, A. (1993). A deviance residual for heterogeneous spatial Poisson processes. *Biometrics 49*, 889–897.

Lewis, P. (1972). Recent results in the statistical analysis of univariate point processes. In P. Lewis (Ed.), *Stochastic point processes*, pp. 1–54. New York: Wiley.

Lindsey, J. (1992). *The Analysis of Stochastic Processes using GLIM*. Berlin: Springer.

Lindsey, J. (1995). Fitting parametric counting processes by using linear models. *Applied Statistics 44*, 201–212.

Mase, S., J. Møller, D. Stoyan, R. Waagepetersen, and G. Döge (2001). Packing densities and simulated tempering for hard core Gibbs point processes. *Annals of the Institute of Statistical Mathematics 53*, 661–680.

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

Møller, J. (2003). Shot noise Cox processes. *Advances in Applied Probability 35*, 614–640.

Møller, J., A.-R. Syversveen, and R. Waagepetersen (1998). Log Gaussian Cox processes. *Scandinavian Journal of Statistics 25*, 451–482.

Møller, J. and R. Waagepetersen (2003a). *Statistical Inference and Simulation for Spatial Point Processes*. Boca Raton: Chapman and Hall/CRC.

Møller, J. and R. P. Waagepetersen (2003b). An introduction to simulation-based inference for spatial point processes. In J. Møller (Ed.), *Spatial Statistics and Computational Methods*, Lecture Notes in Statistics 173, pp. 143–198. Springer-Verlag, New York.

Nair, M. (1990). Random space change for multiparameter point processes. *Annals of Probability 18*, 1222–1231.

Nguyen, X. and H. Zessin (1979). Integral and differential characterizations of Gibbs processes. *Mathematische Nachrichten 88*, 105–115.

Numata, M. (1964). Forest vegetation, particularly pine stands in the vicinity of Choshi — flora and vegetation in Choshi, Chiba prefecture, VI (in Japanese). *Bulletin of the Choshi Marine Laboratory* (6), 27–37. Chiba University.

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

Ogata, Y. and M. Tanemura (1981). Estimation of interaction potentials of spatial point patterns through the maximum likelihood procedure. *Annals of the Institute of Statistical Mathematics B 33*, 315–338.

Ogata, Y. and M. Tanemura (1986). Likelihood estimation of interaction potentials and external fields of inhomogeneous spatial point patterns. In I. Francis, B. Manly, and F. Lam (Eds.), *Pacific Statistical Congress*, pp. 150–154. Elsevier.

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

Pregibon, D. (1981). Logistic regression diagnostics. *Annals of Statistics 9*, 705–724.

R Development Core Team (2004). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-00-3.

Ripley, B. (1977). Modelling spatial patterns (with discussion). *Journal of the Royal Statistical Society, series B 39*, 172–212.

Ripley, B. (1981). *Spatial Statistics*. New York: John Wiley and Sons.

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

Ripley, B. and F. Kelly (1977). Markov point processes. *Journal of the London Mathematical Society 15*, 188–192.

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

Schoenberg, F. (1999). Transforming spatial point processes into Poisson processes. *Stochastic Processes and their Applications 81*, 155–164.

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

Stoyan, D., W. Kendall, and J. Mecke (1995). *Stochastic Geometry and its Applications* (Second ed.). Chichester: John Wiley and Sons.

Stoyan, D. and H. Stoyan (1995). *Fractals, Random Shapes and Point Fields*. Chichester: John Wiley and Sons.

van Lieshout, M. (2000). *Markov Point Processes and their Applications*. Imperial College Press.

Venables, W. and B. Ripley (1997). *Modern Applied Statistics with S-Plus* (second ed.). New York: Springer.

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

Vere-Jones, D. and F. P. Schoenberg (2004). Rescaling marked point processes. *Australian and New Zealand Journal of Statistics 46*, 133–143.

Wand, M. and M. Jones (1995). *Kernel Smoothing*. London: Chapman and Hall.

Wartenberg, D. (1990). Exploratory spatial analyses: outliers, leverage points, and influence functions. In D. Griffith (Ed.), *Spatial Statistics: Past, Present and Future*, pp. 133–162. Ann Arbor, Michigan, USA: Institute of Mathematical Geography.

Zhuang, J., P. Chang, Y. Ogata, and Y. Chen (2005). A study on the background and clustering seismicity in the Taiwan region based on a point process model. *Journal of Geophysical Research*, to appear.