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

and M. Hazelton

University of Western Australia, Perth, Australia

[*Read before* The Royal Statistical Society *at a meeting organized by the* Research Section *on Wednesday, June 22nd, 2005*, Professor J. T. Kent *in the Chair*]

Summary. We define residuals for point process models fitted to spatial point pattern data, and we 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 and Berman's diagnostic) are recovered as special cases. Diagnostic tools are developed systematically, by using an analogy between our spatial residuals and the usual residuals for (non-spatial) generalized 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 generalized linear models 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; Generalized linear models; Georgii–Nguyen–Zessin formula; Kernel smoothing; K-function; 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 methods.

However, tools for checking or criticizing 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, to assess model assumptions one by one and to recognize the

Address for correspondence: A. Baddeley, School of Mathematics and Statistics, University of Western Australia, 35 Stirling Highway, Nedlands, WA 6009, Australia.
E-mail: adrian@maths.uwa.edu.au

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'), 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 that are 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 generalization of the well-known residuals for point processes in time, which are used routinely in survival analysis. It had been thought that no such generalization 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 quite a different approach (Cox and Isham (1980), section 6.1, and Ripley (1988), introduction). Nevertheless the generalization 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 generalized linear models (GLMs). 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 that are introduced in this paper correspond to the usual residuals for Poisson log-linear regression. The components of a point process model (spatial trend, dependence on spatial covariates and 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 (1991). 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 and 14 discuss practical implementation and scope of the techniques.

2.   Motivating examples

Fig. 1 depicts the Japanese black pines data of Numata (1964), which have been analysed by Ogata and Tanemura (1981, 1986). Dots indicate the locations of 204 seedlings and saplings of Japanese black pine (*Pinus Thunbergii*) in a sampling region $10 \times 10\,\mathrm{m}^2$ 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 Fig. 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
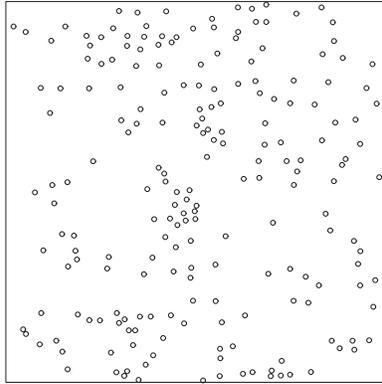
Fig. 1. Japanese black pines (seedlings and saplings) data of Numata (1964) (the data were kindly supplied by Professor Y. Ogata and Professor M. Tanemura)

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 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 by using Markov chain Monte Carlo 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 specialized 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. Although 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 quite a general point process such as this one.

Fig. 2 is a subset of data that was introduced and analysed by Berman (1986). It represents an intensive geological survey of a region in central Queensland $158 \times 35 \text{ km}^2$. Dots mark the locations of 57 copper ore deposits. Line segments represent 90 geological features which are
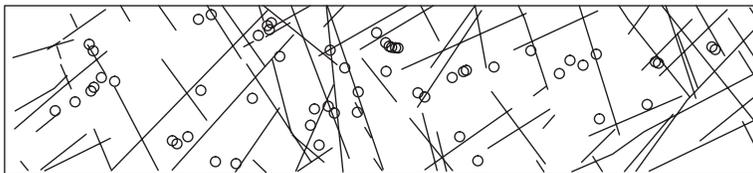
Fig. 2. Copper ore deposits ($\bigcirc$) and lineaments (———) in an area of Queensland: southern half of the original data; north at the top (the data are reproduced by kind permission of Dr J. Huntington and Dr M. Berman)
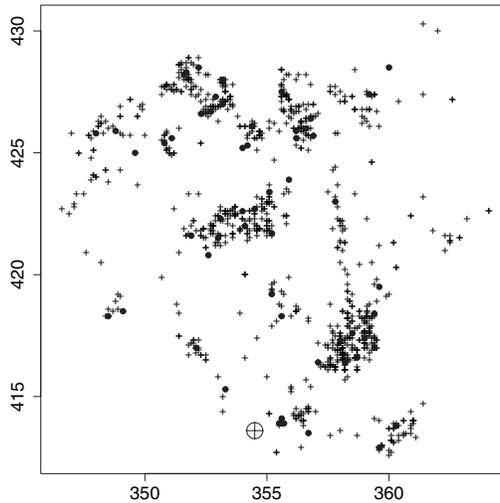
Fig. 3. Cases of cancer of the larynx (•) and lung (+) in the Chorley–south Ribble area, and the location of an industrial incinerator (⊕) (Ordnance Survey co-ordinates in kilometres)

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 the distance from the nearest lineament.

Several analyses (Berman, 1986; Berman and Turner, 1992; Foxall and Baddeley, 2002) have concluded that 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 shows a spatial epidemiological data set that was 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 near 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) assumed that 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 the distance from the incinerator. Diggle (1990) and Diggle and Rowlingson (1994) fitted models of both types by maximum likelihood and found that the model of best fit includes an incinerator effect. Goodness-of-fit testing and informal validation of the model were 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) noted many *caveats* on epidemiological interpretation of the fitted model and identified questions for further investigation, notably the spatial

clustering of cases of disease. Clustered models cannot be fitted by using the techniques of Diggle (1990) and Diggle and Rowlingson (1994), which apply only to Poisson processes. Although the model checking techniques in Diggle (1990), section 3.2, can detect 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 recognize 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), pages 89–90, 100–103, 106, 110–111, 114 and 133–143, and Møller and Waagepetersen (2003a), chapter 4.

In his influential paper Ripley (1977) developed an exploratory analysis of interpoint interaction, *assuming that the data are spatially homogeneous*. A useful summary statistic is the nonparametric estimator $\hat{K}$ of Ripley's K-function, essentially a renormalized empirical distribution of the pairwise distances between observed points. For the homogeneous Poisson process (complete spatial randomness) the true value of K is known. A discrepancy between $\hat{K}$ and the theoretical K-function for complete spatial randomness 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 $\hat{K}$ 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, 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 $\hat{K}$ 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 $\hat{K}$ 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 $\hat{K}$ (and of other statistics) have been proposed. Local indicators of spatial association (Anselin, 1995; Cressie and Collins, 2001a, b; Getis and Franklin, 1987) are localized versions of summary functions such as $\hat{K}$. The K-function has also been adapted to inhomogeneous point processes where the spatial trend is known (Vere-Jones and Schoenberg, 2004; Baddeley *et al.*, 2000). However, if the trend has to be estimated when estimating K, the interpretation of deviations in $\hat{K}$ 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 (1981), page 180, Brillinger and Preisler (1986), Diggle (1990), Schoenberg (1999) and 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) and Venables and Ripley (1997), chapter 12). Departure from unit intensity in the transformed space suggests a misspecified spatial trend, whereas departure from an 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 recognize 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) and Cressie (1991), pages 766–770).

The advent of practical Markov chain Monte Carlo 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 spatial trend or interaction within the context of a parametric model. However, tools for model criticism are still lacking.

Finally, some researchers have introduced diagnostics that are 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, whereas 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), pages 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 1970s 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*

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

(if it exists), i.e. $\lambda(t)$ is the instantaneous arrival rate of the point process given the history of the process prior to time t (Karr (1985), page 69). 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, page 60). In practice, when a point process model with a parameter θ is fitted to data, the parameter estimate $\hat{\theta}$ would be plugged into an expression for $\lambda(t) = \lambda_\theta(t)$ to obtain a fitted conditional intensity $\hat{\lambda}(t) = \lambda_{\hat{\theta}}(t)$, and we compute the *raw residual process*

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

Increments of $R(t)$ are analogous to the raw residuals (observed minus fitted values) in a linear model, whereas 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), theorem 2.31, page 71)

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

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:

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

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

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 modelling earthquakes (Ogata, 1988; Vere-Jones, 1970).

5.   Spatial point process models

5.1.   *Notation and assumptions*

A spatial point pattern is a data set

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

consisting of the (unordered) locations $x_1, \ldots, x_n$ of points that are observed in a bounded region W of the plane $\mathbb{R}^2$. Note that the number of points $n = n(\mathbf{x}) \geqslant 0$ 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 that is 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

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

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 normalizing 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

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

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

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

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

(where $n = n(\mathbf{x})$) for unique functions V_k called the potentials of order k. Here V_0 determines the normalizing constant and V_k is of the form $V_k(\mathbf{x}) = \Sigma_{\mathbf{y}} v_k(\mathbf{y})$, 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, $V_{\geqslant 2}(\mathbf{x}) = V_2(\mathbf{x}) + V_3(\mathbf{x}) + \ldots + V_n(\mathbf{x})$.

If $V_{\geqslant 2}$ 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 generalization of the conditional intensity of a temporal or spatiotemporal 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), pages 580–590, and Karr (1985), section 2.6. For our purposes the following, very simplified, account is adequate. Suppose that $\mathbf{X}$ 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

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

Then equation (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

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

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

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

The Papangelou conditional intensity λ of a finite point process uniquely determines its probability density f and vice versa (because of condition (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 normalizing constant of f is eliminated in definition (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

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

where $h(u, \mathbf{x})$ is any non-negative 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 equation (8) becomes the basic identity for deriving diagnostics and residuals. We assume that both sides of equation (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 that $\lambda(\cdot,\cdot) > 0$. Take

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

where $\mathbf{1}\{\dots\}$ is the indicator function and $B \subseteq W$ is a given subset. Then equation (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 $\mathbb{R}^2$). Although other functions h could be substituted in equation (8), the judicious choice (9) that was made by Stoyan and Grabarnik is effectively the only one in which the integral in equation (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 $\hat{\lambda}(u, \mathbf{x}) = \lambda_{\hat{\theta}}(u, \mathbf{x})$. Stoyan and Grabarnik (1991) proposed that the fitted weights $\hat{m}_i = 1/\hat{\lambda}(x_i, \mathbf{x})$ that are associated with the data points x_i, and their sums $\hat{M}(B) = \Sigma_i \mathbf{1}\{x_i \in B\}\hat{m}_i$, could be used for exploratory data analysis and goodness-of-fit testing, in that

(a) points x_i with extreme values $\hat{m}_i$ may indicate 'outliers';
(b) regions B with extreme values of $\hat{M}(B) - |B|$ may indicate regions of irregularity and
(c) the global departure $\hat{M}(W) - |W|$ may be used to test goodness of fit or to test convergence of Markov chain Monte Carlo samplers.

Applications were not presented in Stoyan and Grabarnik (1991); proposal (a) was tried by Särkkä (1993), pages 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 that is obtained by setting $c(u,v) = \mathbf{1}\{\|u-v\| > \delta\}$ in equation (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 realization 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 condition (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 equation (8). Increments of the innovation process I_θ are analogous to errors in a linear model. The GNZ formula (8) can be restated as

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

corresponding to the martingale properties of the innovations for temporal and spatiotemporal 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 that is 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 formula (8) to

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

shows that the pseudoscore has mean 0 under the model. For Poisson processes the pseudolikelihood and likelihood are equivalent, so equation (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 $\hat\theta = \hat\theta(\mathbf{x})$, we define the *raw residuals*

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

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

Whereas most previous researchers (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 data set, 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, e.g. to compute standardized residuals. The analogue in our context is to scale the increments of the residual measure $R_{\hat{\theta}}$. This is done simply by making an alternative choice of the function h in the GNZ formula (8). For any non-negative 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') that is 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 that is attached to a background location $u \in W$. The innovations have mean 0, from equation (8).

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

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

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

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

The choice $h(u, \mathbf{x}) \equiv 1$ in equations (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 λ 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 ∞, provided that $h(x_i, \mathbf{X} \setminus \{x_i\})$ is finite for all $x_i \in \mathbf{X}$, and we interpret $h(u, \mathbf{X}) \, \lambda(u, \mathbf{X})$ as 0 if $\lambda(u, \mathbf{X}) = 0$. Thus we obtain the innovation measure

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

which has mean 0. The corresponding choice $h_{\theta}(u, \mathbf{x}) = 1/\lambda_{\theta}(u, \mathbf{x})$ in equation (16) yields the residual measure

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

In order that the residuals be well defined, the estimator $\hat{\theta}$ must have the property that $\lambda_{\hat{\theta}(\mathbf{x})}(x_i, \mathbf{x}) > 0$ for all $x_i \in \mathbf{x}$ for any pattern $\mathbf{x}$. Zero values for $\lambda_{\hat{\theta}(\mathbf{x})}(u, \mathbf{x})$ are permitted for $u \notin \mathbf{x}$. We shall call measure (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 log-linear regression, we consider the weight function $h(u, \mathbf{x}) = 1/\sqrt{\lambda}(u, \mathbf{x})$ which yields the Pearson innovation measure

$$I\left(B, \frac{1}{\sqrt{\lambda}}, \lambda\right) = \sum_{x_i \in \mathbf{X} \cap B} \frac{1}{\sqrt{\lambda}(x_i, \mathbf{X})} - \int_B \sqrt{\lambda}(u, \mathbf{x})\, du,$$

which has mean 0, and the corresponding Pearson residual measure

$$R\left(B, \frac{1}{\sqrt{\hat{\lambda}}}, \hat{\theta}\right) = \sum_{x_i \in \mathbf{x} \cap B} \frac{1}{\sqrt{\hat{\lambda}}(x_i, \mathbf{x})} - \int_B \sqrt{\hat{\lambda}}(u, \mathbf{x})\, du.$$

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

8.4. Pseudoscore 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

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

and vector-valued residuals

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

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 Baddeley *et al.* (2006).

9. Properties of residuals

9.1. Residuals sum to 0

The raw residuals in simple linear regression always sum to 0; a similar phenomenon occurs for our residuals. Firstly consider the homogeneous Poisson process model, fitted by maximum likelihood. The raw residual is

$$R_{\hat{\theta}}(B) = n(\mathbf{x} \cap B) - n(\mathbf{x})|B|/|W|.$$

In particular the residual sum for the whole window W is $R_{\hat{\theta}}(W) = 0$ for any point pattern data set $\mathbf{x}$.

More generally, suppose that we fit a point process model with no spatial trend, having conditional intensity of the common 'log-linear' 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 measure (19) to 0 with $B = W$, which implies that $R_{\hat{\theta}}(W) = 0$. The pseudoscore residuals (for any model with k-dimensional parameter) sum to 0 over W.

9.2. Mean residual

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

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

by formula (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 and Pearson residuals respectively

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

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

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

(provided that $\lambda_{\hat{\theta}(\mathbf{X})}(x_i, \mathbf{X}) > 0$ 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 equation (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, i.e., for $\mathrm{var}\{I(B, h, \lambda)\}$ and $\mathrm{var}\{R(B, \hat{h}, \hat{\theta})\}$, 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

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

where

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

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 λ or Pearson innovations respectively. In the special case of an inhomogeneous Poisson process with intensity $\lambda(u)$, these reduce to

$$\mathrm{var}\{I(B, 1, \lambda)\} = \int_B \lambda(u)\, \mathrm{d}u, \tag{25}$$

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

$$\mathrm{var}\left\{I\left(B, \frac{1}{\sqrt{\lambda}}, \theta\right)\right\} = |B|. \tag{27}$$

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 standardized, 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 λ innovations is

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

generalizing 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 that a Poisson process model with intensity $\lambda_\theta(u)$ is fitted to a realization of a Poisson process with true intensity $\lambda(u)$. Then the raw residuals have variance

$$\mathrm{var}\{R(B)\} = \int_B \lambda(u)\mathrm{d}u + \int_B\int_B \mathrm{cov}\{\lambda_{\hat\theta(X)}(u), \lambda_{\hat\theta(X)}(v)\}\,\mathrm{d}u\,\mathrm{d}v$$
$$-\,2\int_B\int_B \mathbb{E}[\lambda_{\hat\theta(X\cup\{u\})}(v) - \lambda_{\hat\theta(X)}(v)]\,\lambda(u)\,\mathrm{d}v\,\mathrm{d}u.$$

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

$$\mathrm{var}\{R(B, 1, \hat\theta)\} = \theta|B|(1 - |B|/|W|),$$

$$\mathrm{var}\left\{R\left(B, \frac{1}{\hat\lambda}, \hat\theta\right)\right\} = |B|(|W| - |B|)\,\mathbb{E}\left[\frac{\mathbf{1}\{n(\mathbf{X}) > 0\}}{n(\mathbf{X})}\right],$$

$$\mathrm{var}\{R(B, 1/\sqrt{\hat\lambda}, \hat\lambda)\} = |B|(1 - |B|/|W|).$$

Note that the residual variances are smaller than the corresponding innovation variances $\mathrm{var}\{I(B, 1, \theta)\} = \theta|B|$, $\mathrm{var}\{I(B, 1/\theta, \theta) = |B|/\theta$ and $\mathrm{var}\{I(B, 1/\sqrt{\theta}, \theta)\} = |B|$. 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 and V be two subsets of W that are 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 *misspecification* by the model of the true spatial trend, covariate effects and interpoint interaction in the data.

10.2.   Test examples

Figs 4(a), 4(b) and 4(c) show three simulated examples that we use to test the diagnostics. The patterns contain 71, 271 and 376 points respectively in the unit square. Fig. 4(a) is an example of 'trend without interaction': the Poisson process with intensity function $\lambda(x, y) = 300 \exp(-3x)$. Fig. 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

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

with interaction range $r = 0.05$ and interaction strength $\gamma = 0.1$, corresponding to a strong negative association between points. The realization in Fig. 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.

Fig. 4(c) is an example of '(clustered) interaction without trend': a realization 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 Fig. 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 Fig. 4(b).
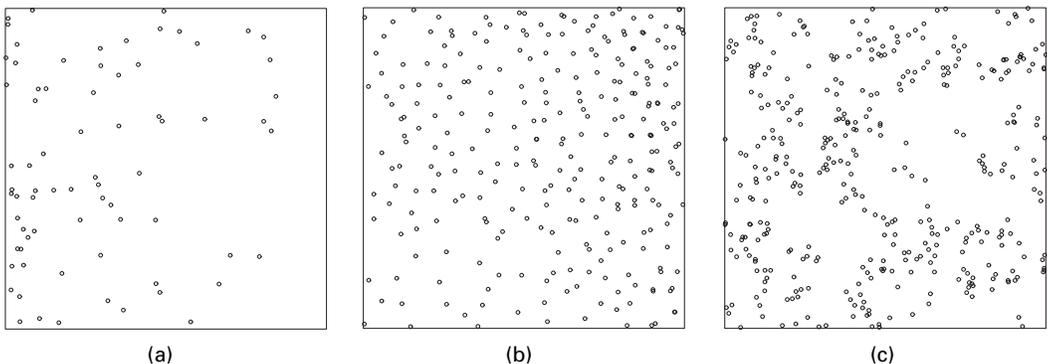
Fig. 4.   Simulated patterns: (a) inhomogeneous Poisson process; (b) inhomogeneous inhibited process (Strauss process); (c) homogeneous clustered process (Geyer's saturation model)

10.3.   Analogy with generalized linear models

Here we explain a connection between point process models and GLMs, which provides statistical insight. For point process models in time, Lewis (1972) recognized that the discretized likelihood is formally equivalent to the likelihood of a binomial regression model, which can be maximized by 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 discretized version of the log-pseudolikelihood (12) is formally equivalent to the log-likelihood of a Poisson log-linear regression. The conditional intensity $\lambda(u; \mathbf{x})$ of the point process corresponds to the mean response of the log-linear regression. In the Gibbs representation (7) of the conditional intensity, the first-order term v_1 corresponds to the linear predictor of a GLM, whereas 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

  (a) index,
  (b) each explanatory variable included in the model and
  (c) explanatory variables that were not included in the model, including surrogates for a lurking variable (Atkinson (1985), pages 3, 34, 62 and following feature). Here we explore analogues of these plots.

11.1.   Spatial display of residuals

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

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* that is shown in Fig. 5.
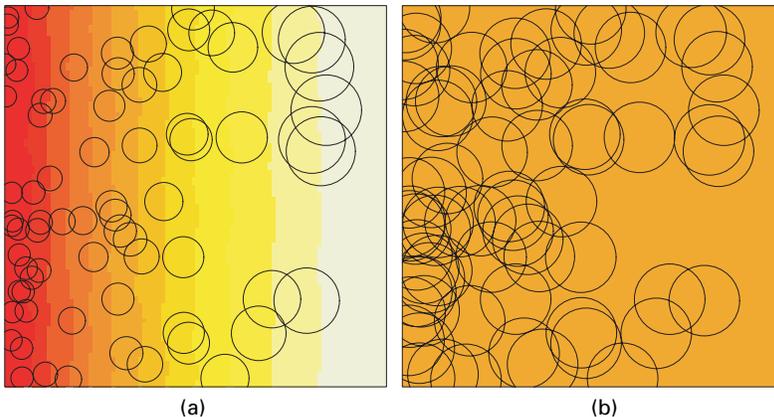
Fig. 5. Mark plot based on Pearson residuals for models fitted to Fig. 4(a): (a) inhomogeneous Poisson model of the correct form; (b) incorrect model, homogeneous Poisson

It consists of a pixel image of the density component (i.e. with grey scale proportional to the density $\hat{h}(u, \mathbf{x})\,\hat{\lambda}(u, \mathbf{x})$) 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 $\hat{h}(x_i, \mathbf{x}\backslash\{x_i\})$). Fig. 5 shows this representation for the two fitted models by using the Pearson residuals. The expansion in size of circles from left to right in Fig. 5(a) is a consequence of the model.

The mark plot may sometimes identify 'extreme' data points (see Fig. 14 in Section 11.5.3). However, the diagnostic interpretation of the residuals is based primarily on their 'sums' over subregions B by using equation (17).

One strategy is to partition W into disjoint subregions $B_1, \ldots, B_m$ (e.g. dividing a rectangular window into equal squares B_k) and to evaluate $R(B_k, h, \theta)$. Non-zero 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 that is used in spatial statistics (Diggle (2003), section 2.5, Cressie (1991) and 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

$$s(u) = e(u) \int_W k(u - v)\,\mathrm{d}R(v, \hat{h}, \hat{\theta})$$
$$= e(u)\left[\sum_{x_i \in \mathbf{x}} k(u - x_i)\,\hat{h}(x_i, \mathbf{x}\backslash\{x_i\}) - \int_W k(u - v)\,\hat{h}(v, \mathbf{x})\,\hat{\lambda}(v, \mathbf{x})\,\mathrm{d}v \right] \qquad (31)$$

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 grey scale image as shown in Fig. 6. Bandwidth selection is discussed in Section 13.

The analogous quantity for the innovations has mean 0,

$$\mathbb{E}\left[\int_W k(u - v)\,\mathrm{d}I(v, h, \lambda) \right] = 0,$$
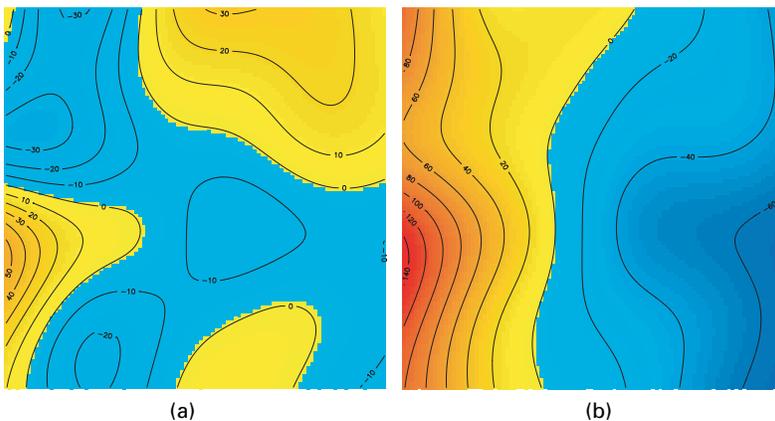
Fig. 6. Contour plots of the kernel-smoothed raw residual field for two models fitted to Fig. 4(a): (a) heterogeneous Poisson model of the correct form (range of smoothed field, [−34.3, 54.2]); (b) incorrect model, homogeneous Poisson (range of smoothed field, [−67.0, 145.6]); smoothing kernel, isotropic Gaussian with standard deviation 0.14; same grey scale map in each plot

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

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

the difference between $\lambda^*(u)$, a nonparametric 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 that the model underestimates the intensity.

Fowlkes (1987) proposed a smoothed residual plot for binary logistic regression to avoid artefacts of the binary nature of the responses. However, this involved smoothing the *responses* before computing pseudoresiduals. 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 that is described here introduces model-dependent, data-dependent and spatially varying weights on the data points and centring 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 λ residuals.

11.2. Lurking variable plots

In linear modelling, if we suspect that 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), pages 3, 34 and 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 co-ordinates (or some other co-ordinate), 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 sublevel set

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

yielding a 'cumulative residual' function

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

This should be approximately 0 if the fitted model is correct. For example, for the raw residuals,

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

Note that $\hat{H}(z) = n\{\mathbf{x} \cap W(z)\}/n(\mathbf{x})$ is the empirical cumulative distribution function (CDF) of the values of the covariate that is observed at the data points x_i. The function $A(z)$ in equation (34) is an adjustment of this empirical CDF to have approximately zero mean under the model.
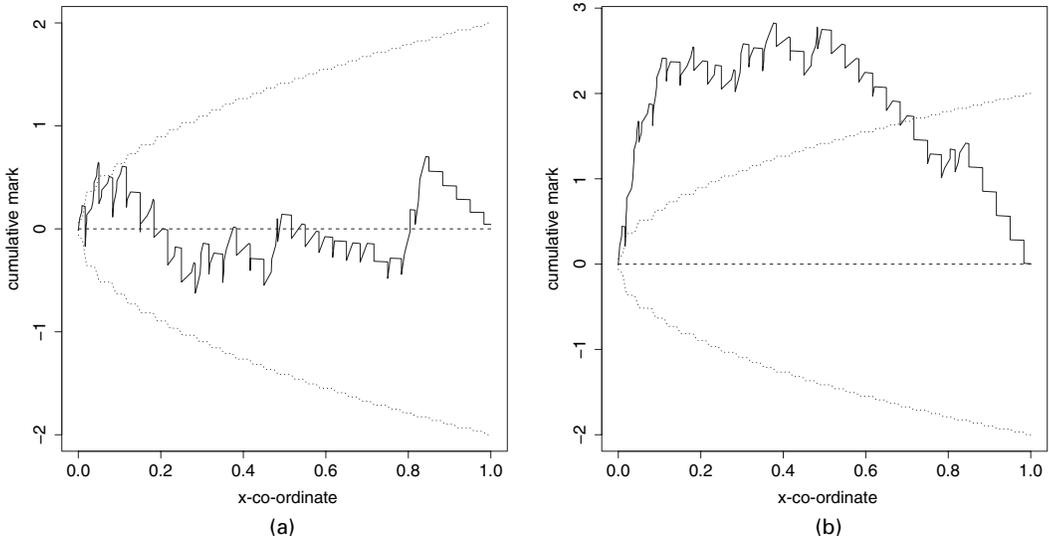
Fig. 7. Lurking variable plots for the x-co-ordinate, for two models fitted to the data in Fig. 4(a) (cumulative style, Pearson residuals; ——, empirical curve $A(x)$; ·······, pointwise 2σ-limits based on equation (27)): (a) model of the correct form, an inhomogeneous Poisson process with intensity log-linear in x, fitted to data; (b) incorrect model, a homogeneous Poisson process, fitted to data

In Fig. 4(a), the lurking variable is the x-co-ordinate. Fig. 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 $x = z$.

The dotted envelopes in Fig. 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 $\mathrm{var}\{A(z)\} \approx \mathrm{var}[I\{W(z)\}]$ where $W(z)$ is given in equation (32). This variance can be estimated by using equations (25)–(27), substituting the fitted Poisson intensity $\hat{\lambda}(u)$ for $\lambda(u)$. In this case we used the Pearson residuals, which are standardized so that $\mathrm{var}\{I(B)\} = |B|$ regardless of λ. Thus the dotted limits in Figs 7(a) and 7(b) are identical.

The 2σ-limits have the usual interpretation of significance (pointwise), assuming that a central limit theorem applies. The glaring violation of these bounds by Fig. 7(b) is ample evidence that a homogeneous trend is inappropriate. In Fig. 7(a) 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.
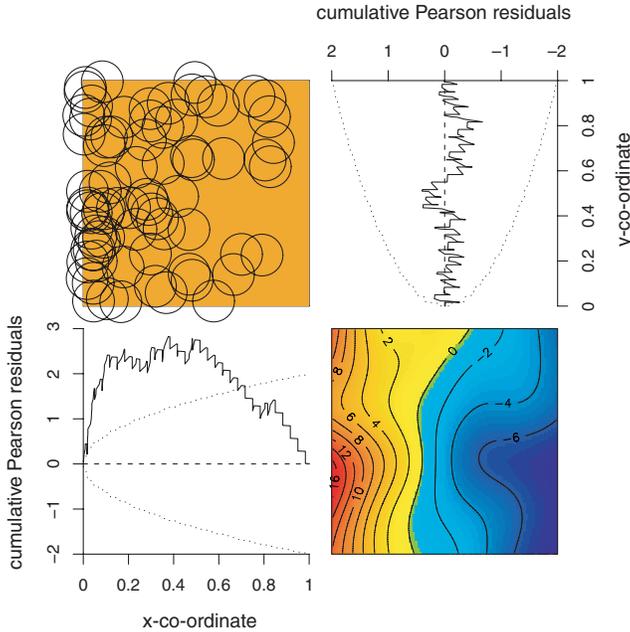
Alternatively we may plot an approximate derivative of $A(z)$, such as

$$a(z) = \int_W k_1\{Z(u) - z\}\, \mathrm{d}R(u, \hat{h}, \hat{\theta})$$

$$= \sum_{x_i \in \mathbf{x}} k_1\{Z(x_i) - z\}\, \hat{h}(x_i, \mathbf{x} \setminus \{x_i\}) - \int_W k_1\{Z(u) - z\}\, \hat{h}(u, \mathbf{x})\, \hat{\lambda}(u, \mathbf{x})\, \mathrm{d}u \qquad (35)$$

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 shows our standard presentation of the diagnostic plots for spatial trend. At the top left

Fig. 8. Standard presentation of the diagnostic plots: top left, mark plot; top right, lurking variable plot for the *y*-co-ordinate; bottom left, lurking variable plot for the *x*-co-ordinate; bottom right, contour plot of the smoothed mark field; the data are from Fig. 4(a); the fitted model is homogeneous Poisson; Pearson residuals

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 grey scale, for easy interpretation.

At the bottom left is a lurking variable plot for the *x*-co-ordinate, its *x*-axis aligned with the *x*-co-ordinates in the mark plot. At the top right is a lurking variable plot for the *y*-co-ordinate, rotated 90° anticlockwise, the *y*-axis aligned with the *y*-co-ordinates 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 Fig. 4(b), a simulation of an inhomogeneous Strauss process with log-quadratic activity function (29) and pair interaction (30). Fig. 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. In Fig. 9(a) is a fitted model of the correct form, with activity

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

where β_i are estimated by maximum pseudolikelihood yielding $(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3) = (180, 1.53, 1.90, 1.35)$ and $\hat{\gamma} = 0.24$; in Fig. 9(b) is an incorrect model with homogeneous trend $b(x, y) \equiv \beta$ where β and γ are estimated with maximum pseudolikelihood estimate $\hat{\beta} = 1099$ and $\hat{\gamma} = 0.34$.

In computing residuals from a Strauss model, we encounter an 'edge effect' problem. Suppose that the data are only a partially observed realization of a point process **Y** in a larger bounded
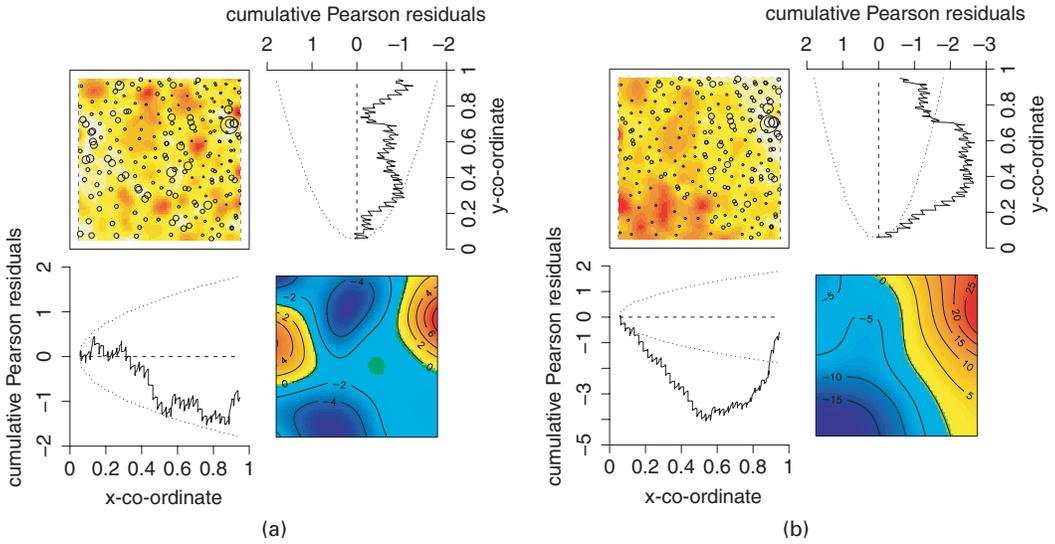
Fig. 9.   Four-panel diagnostic plots for (a) a model of the correct form (an inhomogeneous Strauss process with log-quadratic activity) and (b) a model with incorrect trend (a homogeneous Strauss process), fitted to the data in Fig. 4(b): $\cdots\cdots$, 2σ-limits for a Poisson model

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

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

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 Fig. 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.

Fig. 9(a) indicates that the correct model is a tolerably good fit, although it (correctly) suggests that the trend is underestimated. In Fig. 9(b) the lurking variable plot for the *x*-co-ordinate (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 $\hat{\gamma} \sim 0.25$. 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 $\hat{\gamma}^{-1/2} \sim 2$. This phenomenon is even more exaggerated for the inverse λ residuals, where each extra neighbour increases the residual mass by a factor of $\hat{\gamma}^{-1} \sim 4$. 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 Fig. 1. The activity function b was log-cubic (i.e. $\log\{b(u)\}$ was a cubic polynomial in the Cartesian co-ordinates of u) and c was the 'soft core' interaction

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

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

Fig. 10(a) shows the diagnostic plots for the Ogata–Tanemura model. It suggests good agreement between the trend in the fitted model and in the data. Fig. 10(b) 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 Fig. 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: Fig. 11.

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, equation (34) becomes

$$A(z) = n\{\mathbf{x} \cap W(z)\} - n(\mathbf{x})|W(z)|/|W| = n(\mathbf{x})\{\hat{H}(z) - H_0(z)\}$$
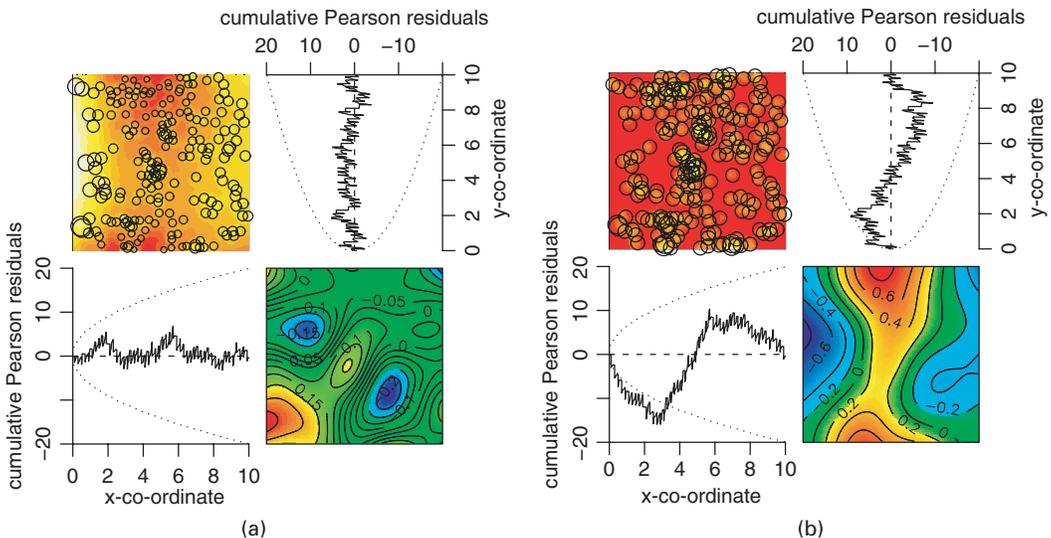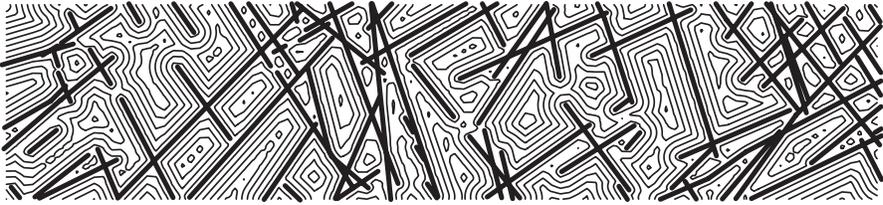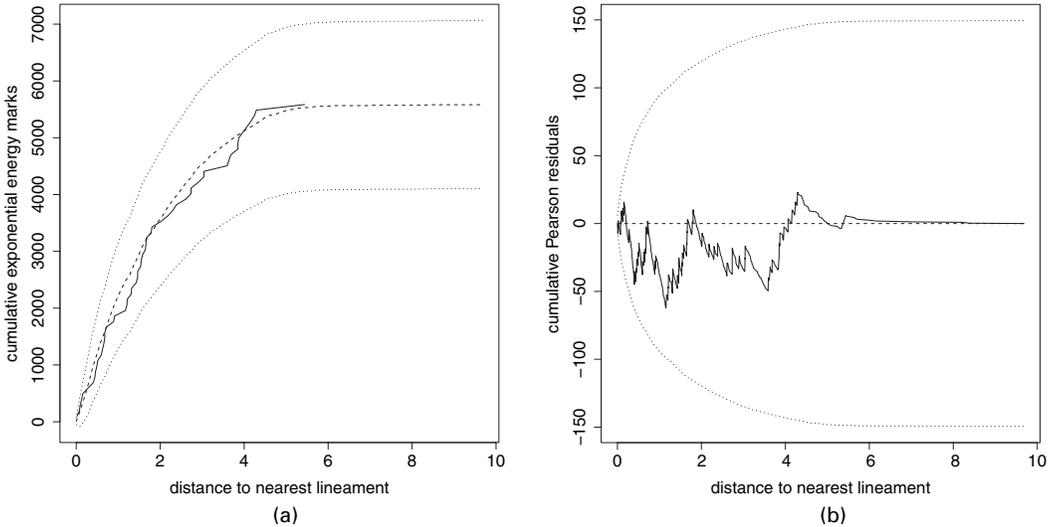

Fig. 10.   Diagnostics for two models fitted to the Japanese pines data of Fig. 1: (a) soft core with a log-cubic trend; (b) homogeneous soft core

Fig. 11. Spatial covariate for the copper data of Fig. 2: ———, observed lineaments; ———, contours of distance to the nearest lineament


Fig. 12. Lurking variable plots for the homogeneous Poisson point process fitted to the Queensland copper data (pointwise 2σ-limits based on equations (25) and (27) respectively): (a) Berman's plot, unnormalized empirical CDF of distance to the nearest lineament at the observed points of the process; (b) cumulative Pearson residuals against the distance to the nearest lineament

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

11.5.3. Chorley–Ribble data
The null model that was considered by Diggle (1990) states that the point pattern of cases of cancer of the larynx in Fig. 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).
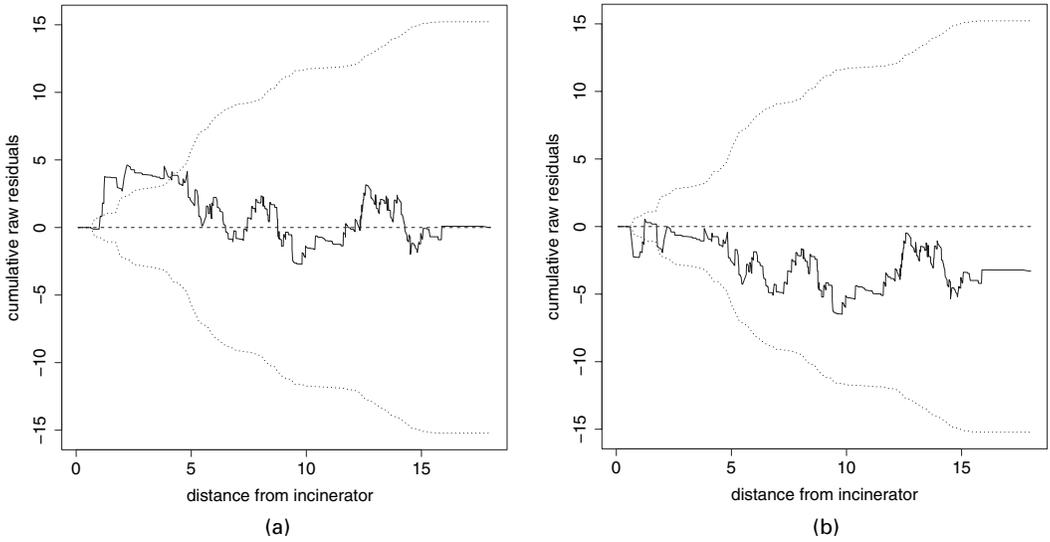
Fig. 13.   Lurking variable plots of the raw residuals against distance from the incinerator, for the Chorley–Ribble data: (a) null model; (b) model fitted by Diggle and Rowlingson (1994)

Fig. 13 shows lurking variable plots of the raw residuals against distance from the incinerator, for two models fitted to the Chorley–Ribble data. Fig. 13(a) 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 that were discussed by Diggle (1990) include clustering of cases of disease (e.g. due to correlation within families) as well as a carcinogenic effect from the incinerator.

Fig. 13(b) 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 that were obtained by Diggle and Rowlingson (1994). Fig. 13(b) gives a slight suggestion that the model that was fitted by Diggle and Rowlingson (1994) overestimates the intensity of cases, at distances close to 0 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. Fig. 13(a) 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 $n(\mathbf{x}) \hat{F}(d) - \hat{T}(d)$ against d, where $\hat{F}(d)$ is the empirical CDF of the distances d_i, and $\hat{T}(d) = \int_{W(d)} \hat{\lambda}(u) \, du$. This is closely related to the model checking technique that was used in Berman (1986) and Diggle (1990), section 3.2. The lurking variable plots for the inverse λ and Pearson residuals also have direct interpretations.

Diagnostic plots for the inverse λ and Pearson residuals reveal some other inconsistencies between the data and the *fitted* null model. Fig. 14 shows an apparent 'outlier'. This is a case of laryngeal cancer where the kernel estimate $\hat{\rho}(u)$ is very low. This raises questions about the appropriateness of the estimator $\hat{\rho}$, 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

(a)   (b)
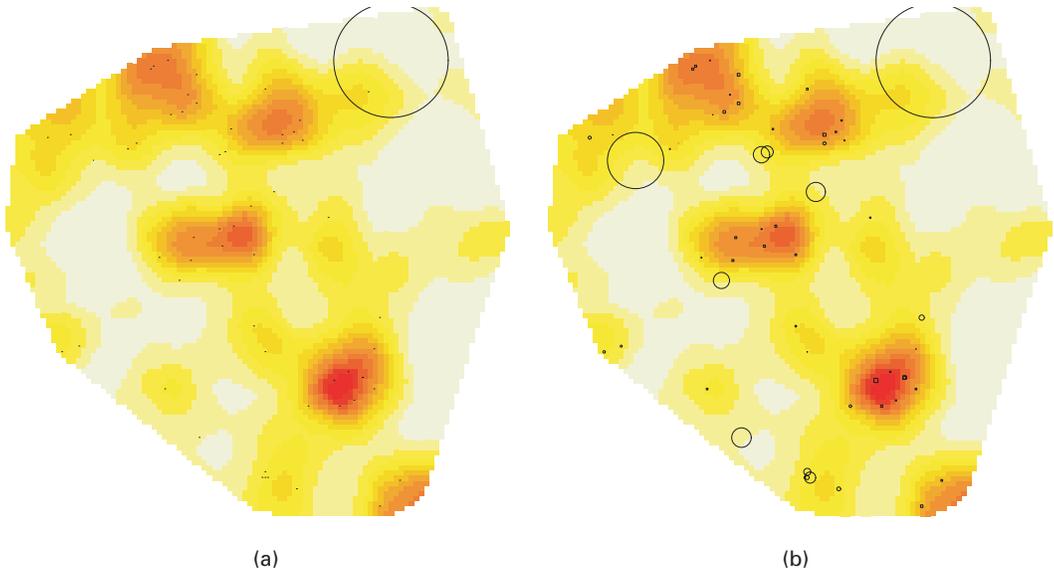
Fig. 14. (a) Mark plot of Pearson residuals for the null model fitted to the Chorley–Ribble data, showing a huge residual at location (360, 428), and (b) radii proportional to log-residuals

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 Fig. 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 Fig. 13(a). 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 GLMs, 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 $s(u)$ with the corresponding *expected* empirical quantiles for $s(u)$ under the fitted model (estimated by Monte Carlo sampling). 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 realizations from the fitted model. To each simulated data set 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 $s(u, \mathbf{x}, \hat{\theta})$ the value of the smoothed residual field at the location $u \in W$ computed for the model with fitted parameter $\hat{\theta}$ on the data set $\mathbf{x}$. Let $u_j, j = 1, \ldots, J$,

be fixed locations in W. After fitting the model to the original data set $\mathbf{x}$ we compute the values of the smoothed residual field s at these locations, $s_j = s(u_j, \mathbf{x}, \hat{\theta})$, and sort them to obtain their order statistics $s_{[1]} \leqslant s_{[2]} \leqslant \ldots \leqslant s_{[J]}$. We then generate N independent simulated realizations 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 $\hat{\theta}^{(n)}$, compute the smoothed residual field values $s_j^{(n)} = s(u_j, \mathbf{x}^{(n)}, \hat{\theta}^{(n)})$ and obtain order statistics $s_{[1]}^{(n)} \leqslant \ldots \leqslant s_{[J]}^{(n)}$. 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 Gnanadesi-kan 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 be evaluated only in the eroded window $W_{\ominus r}$ where r is the range of interpoint interaction.

12.2. Test example

Fig. 4(c) is a homogeneous clustered pattern that was generated by Geyer's saturation process. Fig. 15 shows Q–Q-plots, based on the Pearson residuals, for models fitted to this pattern. Fig. 15(a) is for a model of the correct form, the homogeneous saturation process, with the irregular parameters r and c fixed at their correct values. This suggests good agreement between the fitted model and the data. Fig. 15(b) is for the (incorrect) homogeneous Poisson model. It shows clear disagreement between the model and the data. The smoothed residual field for
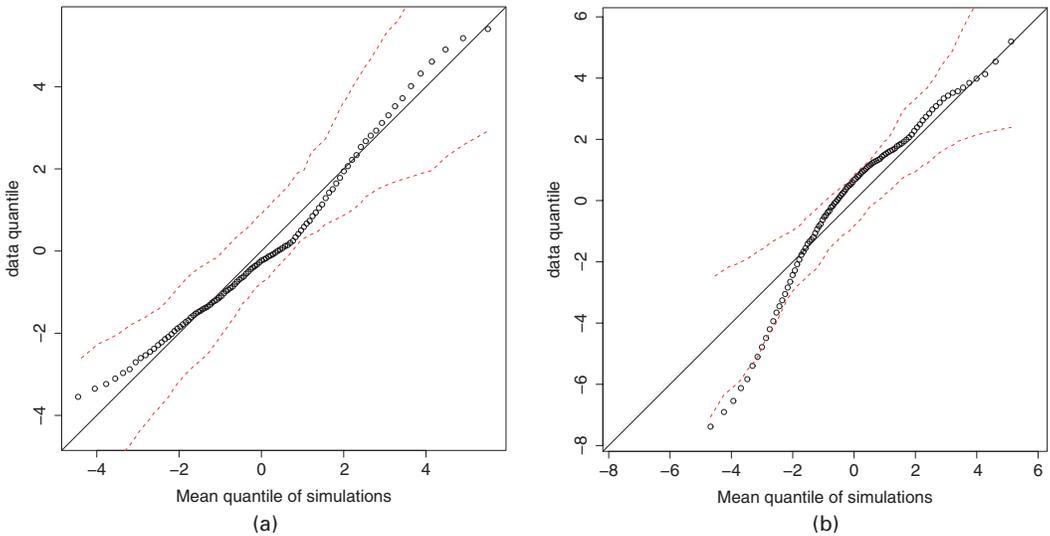
Fig. 15. *Q–Q*-plots for spatial point process models fitted to Fig. 4(c) ($\cdots\cdots$, pointwise 95% critical intervals obtained by simulation): (a) model of the correct form, homogeneous saturation process; (b) homogeneous Poisson process

the data has a heavier left-hand 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 Figs 15(a) and 15(b), and the wider prediction interval in Fig. 15(a).

For each Q–Q-plot in Fig. 15, the smoothed residual field was evaluated at a 40×40 grid of pixels, and percentiles of these values were obtained. The expected percentiles were estimated (using the 5% trimmed sample mean) from 100 simulated realizations of the fitted model. For the saturation model (Fig. 15(a)), each realization was obtained from the same simulation algorithm as the original data. The total computation time for Fig. 15 was approximately 70 min on a 2.0 GHz Linux desktop computer. For investigative work, adopting a 25×25 pixel grid, using only 40 simulated realizations and 10^4 Metropolis–Hastings iterations per realization, produces a practically useful Q–Q-plot in 1 min.

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 that is conveyed in the smoothed residual field $s(u)$ in equation (31). For heuristic purposes, suppose that the fitted model is complete spatial randomness and take the smoothing kernel k to be the uniform density on the disc of radius r centred at the origin, $k(u) = \mathbf{1}\{\|u\| \leqslant r\}/\pi r^2$. 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*, which is a well-known summary statistic that is 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

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

where $\hat{F}$ is the empirical empty space function of $\mathbf{x}$ (using the border method edge correction). $\hat{F}$ 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 $\hat{F}$. 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-hand tail. Fig. 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-hand tail. 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 Fig. 4(c) were generated by a saturation process with interaction range
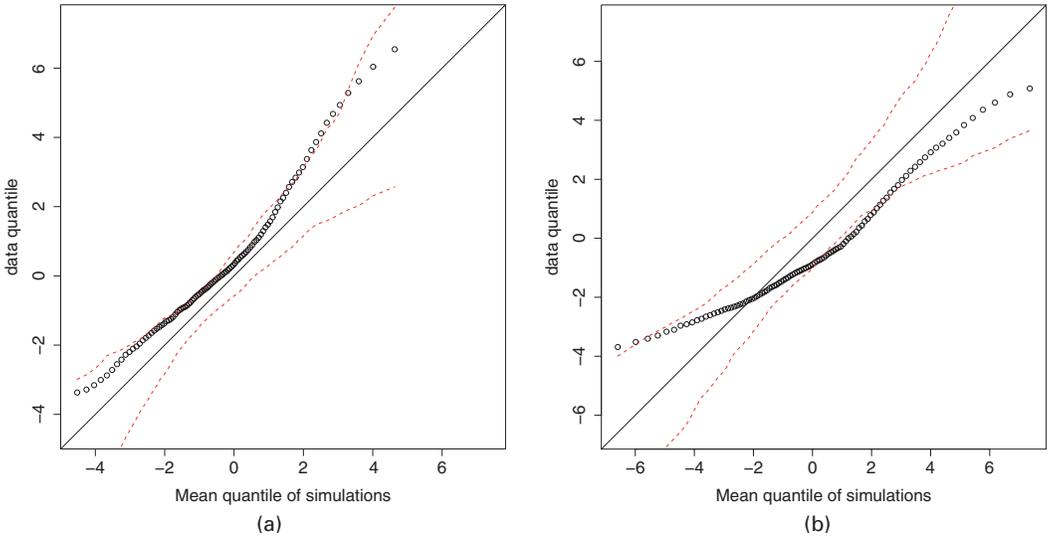
Fig. 16.   *Q–Q*-plots for fitted models in which the range of interaction has been (a) underestimated ($r = 0.04$) or (b) overestimated ($r = 0.06$) in the Geyer saturation model with $r = 0.05$: fitted to the data in Fig. 4(c); Pearson residuals

$r = 0.05$. Fig. 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 underestimation and overestimation of the interaction range *r* in the Strauss process.

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

Fig. 18 shows residual *Q–Q*-plots for two models fitted to Fig. 4(b) with the correct form of the interaction. Fig. 18(a) is for a model with a trend of the correct form and suggests that the interaction is modelled correctly. Fig. 18(b) is for a homogeneous process, i.e. $b(u)$ is constant. It suggests that the observed residuals have heavier left-hand and right-hand 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.

12.6.   *Applications*
12.6.1.   *Japanese pines data*
Fig. 19(a) shows a *Q–Q*-plot for the Ogata–Tanemura model, based on the Pearson residuals. It suggests that the soft core interaction is a good fit. Fig. 19(b) 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.
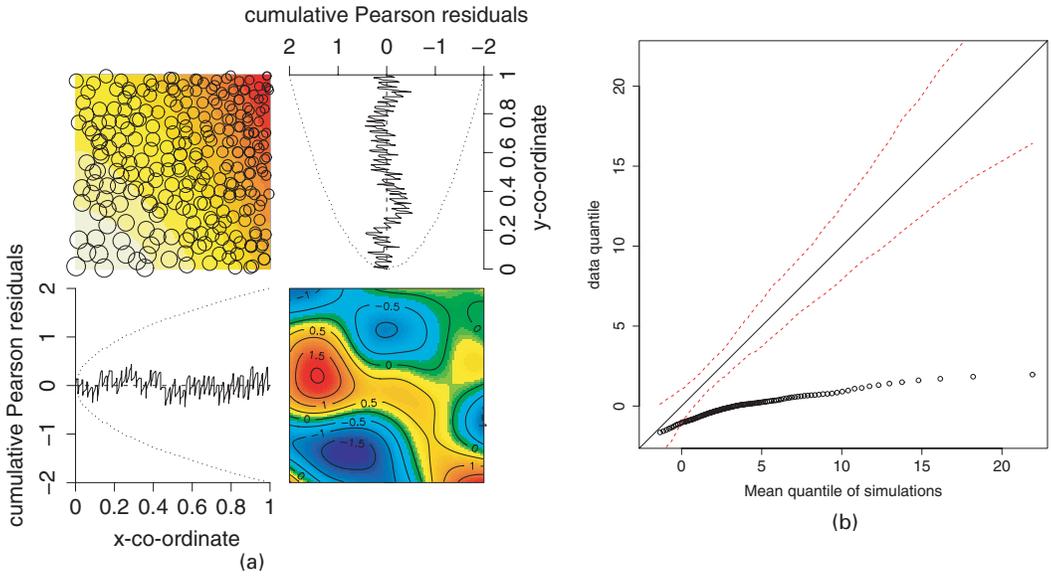
Fig. 17. Diagnostics for a Poisson process with log-quadratic intensity fitted to the data in Fig. 4(b)
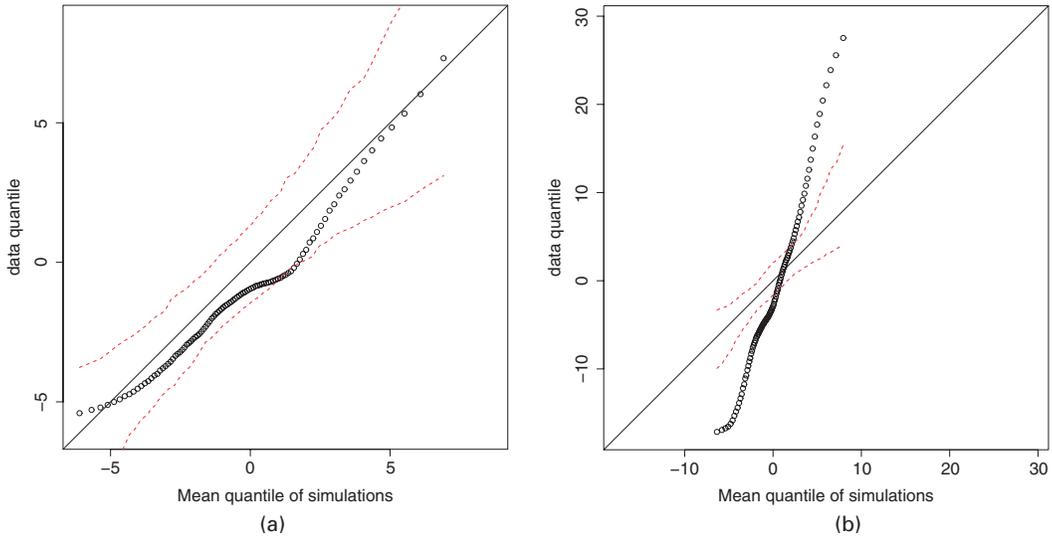
Fig. 18. *Q–Q*-plots for models with the correct interaction (Strauss) fitted to the data in Fig. 4(b): (a) trend of the correct form (log-quadratic); (b) trend of the incorrect form (homogeneous)

12.6.2. *Queensland copper data*

Although the lurking variable plots in Fig. 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 by 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 pseudo-likelihood (Baddeley and Turner, 2000) to obtain $\hat{r} = 1.19$ km, $\hat{c} = 2$ and $\hat{\gamma} = 2.86$, reflecting quite
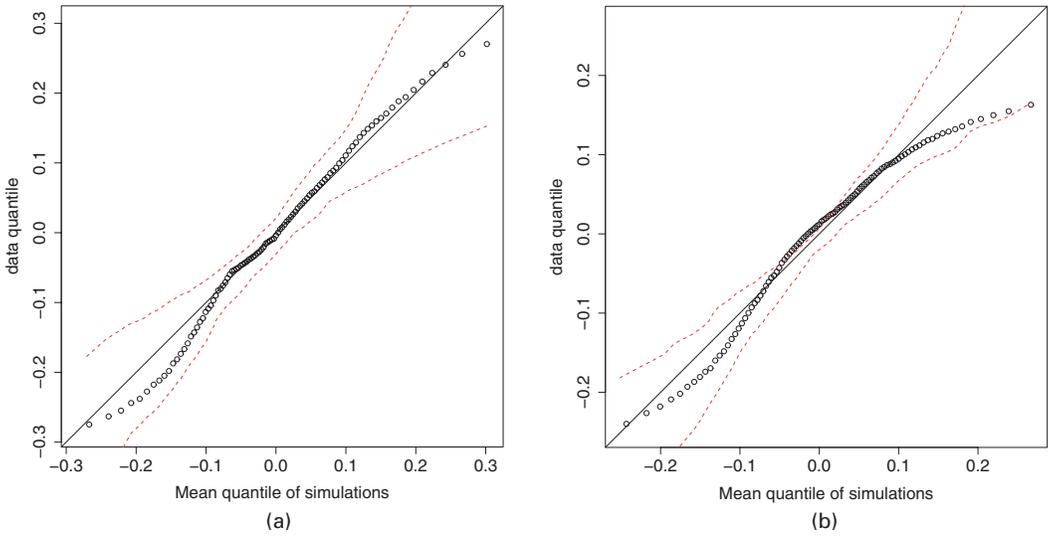
Fig. 19. *Q–Q*-plots for soft core models fitted to the Japanese pines data of Fig. 1: (a) Ogata–Tanemura model, log-cubic trend; (b) homogeneous trend
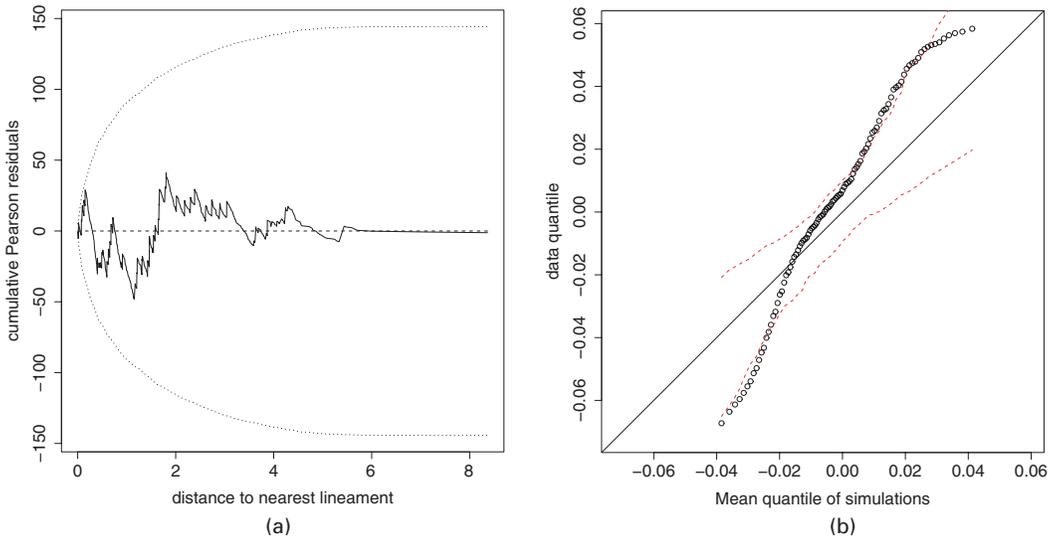
Fig. 20. Diagnostics for a homogeneous saturation model fitted to the Queensland copper data: (a) lurking variable plot against distance to nearest lineament; (b) *Q–Q*-plot; Pearson residuals

strong clustering. Diagnostics for this model are shown in Fig. 20. The *Q–Q*-plot suggests that the data may be even more clustered than the fitted model, assuming that 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).

12.6.3. Cells data
An extreme example of interaction is the 'biological cells' data set (Ripley, 1977) which is shown in Fig. 21(a). It is often modelled by a hard core process (Ripley, 1981; Diggle, 2003) with hard
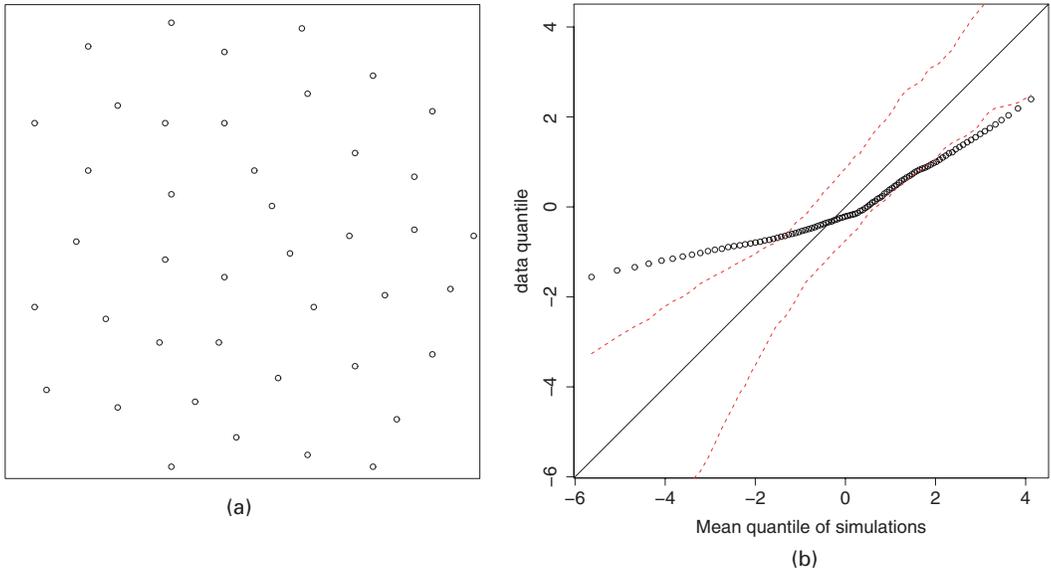
Fig. 21.   (a) Cells data and (b) *Q–Q*-plot for the fitted Strauss process model (Pearson residuals)

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 $\hat{\beta} = 201, 142$ with and without edge correction. Four-panel plots of the two models (which are not shown) indicate clearly that $\hat{\beta} = 142$ is an underestimate, whereas $\hat{\beta} = 201$ appears reasonable.

The *Q–Q*-plot in Fig. 21(b) is for the model with $\hat{\beta} = 201$. The empirical quantiles have a lighter left-hand 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 (see Mase *et al.* (2001)).

In practice, a hard core model is often fitted by 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 λ residuals. Since the residual density

$$\hat{h}(u, \mathbf{x})\, \hat{\lambda}(u, \mathbf{x}) = \mathbf{1}\{\hat{\lambda}(u, \mathbf{x}) > 0\}$$

is a discontinuous function of $\hat{\lambda}(u, \mathbf{x})$, the residuals are unstable to numerical error when $\hat{\lambda}(u, \mathbf{x}) \approx 0$, i.e. when $\hat{\gamma} \approx 0$. The remedy is to constrain γ to be exactly 0.

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.   *Discretized residual measure*
Some discretization 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 discretization device to compute the residual measure.

In the Berman–Turner device, the observed point pattern **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 $\Sigma_j \mathbf{1}\{u_j \in B\} w_j g(u_j)$. The residual measure $R(B, h, \hat{\theta})$ will be approximated by

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

where $\hat{h}_j = h_{\hat{\theta}}(u_j, \mathbf{x} \backslash \{u_j\})$, $\hat{\lambda}_j = \lambda_{\hat{\theta}}(u_j, \mathbf{x})$ 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 log-linear regression. An individual summand in equation (39),

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

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

Note that the residuals that are 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 that are defined in Section 7.2 attach unit mass to each data point x_i, whereas the discretized raw residual $r_j^1 = z_j - \hat{\lambda}_j w_j$ is not equal to 1 for a data point; rather it approximates the residual sum over a region that is associated with u_j.

In other statistical modelling contexts where a weighted likelihood is used, the Pearson residuals would usually be defined as $r_j^* = \sqrt{w_j}(y_j/\hat{\lambda}_j^{1/2} - \hat{\lambda}_j^{1/2})$ 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 equation (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 minimize the mean-square error (Diggle (1985), Berman and Diggle (1989) and Stoyan and Stoyan (1995), pages 237–238). Alternatively there are rules of thumb for bandwidth selection for the pair correlation function (Stoyan and Stoyan (1995), page 285).

The bandwidths in this paper were chosen by cross-validation (Wand and Jones, 1995) of the discretized residuals. Let $r_j = r_j^h$ denote one of the discretized residuals that were 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 minimize $C(b) = \Sigma_i (s_i - \hat{s}_i)^2$ where $\hat{s}_i = \Sigma_{j \neq i} a_{ij} s_j / \Sigma_{j \neq i} a_{ij}$ 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 **X** driven by a random intensity function $\Lambda(u)$, $u \in W$ (i.e. **X** is conditionally a Poisson process with intensity function Λ), the conditional intensity is

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

The right-hand side can be estimated by Markov chain Monte Carlo 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 Markov chain Monte Carlo 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

References

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

Anselin, L. (1995) Local indicators of spatial association—LISA. *Geogr. Anal.*, **27**, 93–115.

Atkinson, A. (1985) *Plots, Transformations and Regression*. Oxford: Oxford University Press.

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

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

Baddeley, A., Møller, J. and Waagepetersen, R. (2006) Residual versions of summary functions for spatial point patterns. To be published.

Baddeley, A. and Turner, R. (2000) Practical maximum pseudolikelihood for spatial point patterns (with discussion). *Aust. New Zeal. J. Statist.*, **42**, 283–322.

Baddeley, A. and Turner, R. (2005a) Modelling spatial point patterns in R. *Lect. Notes Statist.*, **185**, in the press.

Baddeley, A. and Turner, R. (2005b) Spatstat: an R package for analyzing spatial point patterns. *J. Statist. Softwr.*, **12**, no. 6, 1–42.

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

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

Berman, M. and Diggle, P. (1989) Estimating weighted integrals of the second-order intensity of a spatial point process. *J. R. Statist. Soc.* B, **51**, 81–92.

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

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

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

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

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

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

Brillinger, D. (1994) Time series, point processes, and hybrids. *Can. J. Statist.*, **22**, 177–206.

Brillinger, D. and Preisler, H. (1986) Two examples of quantal data analysis: a) multivariate point process, b) pure death process in an experimental design. In *Proc. 13th Int. Biometric Conf., Seattle*. Seattle: International Biometric Society.

Clyde, M. and Strauss, D. (1991) Logistic regression for spatial pair-potential models. In *Spatial Statistics and Imaging* (ed. A. Possolo), ch. II, pp. 14–30. Hayward: Institute of Mathematical Statistics.

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

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

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

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

Cressie, N. and Collins, L. (2001a) Analysis of spatial point patterns using bundles of product density LISA functions. *J. Agric. Biol. Environ. Statist.*, **6**, 118–135.

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

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

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

Diggle, P. (1978) On parameter estimation for spatial point processes. *J. R. Statist. Soc.* B, **40**, 178–181.

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

Diggle, P. J. (1990) A point process modelling approach to raised incidence of a rare phenomenon in the vicinity of a prespecified point. *J. R. Statist. Soc.* A, **153**, 349–362.

Diggle, P. (2003) *Statistical Analysis of Spatial Point Patterns*, 2nd edn. London: Arnold.

Diggle, P., Chetwynd, A., Haggkvist, R. and Morris, S. (1995) Second order analysis of space-time clustering. *Statist. Meth. Med. Res.*, **4**, 124–136.

Diggle, P. J. and Rowlingson, B. S. (1994) A conditional approach to point process modelling of elevated risk. *J. R. Statist. Soc.* A, **157**, 433–440.

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

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

Georgii, H.-O. (1976) Canonical and grand canonical Gibbs states for continuum systems. *Communs Math. Phys.*, **48**, 31–51.

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

Geyer, C. (1999) Likelihood inference for spatial point processes. In *Stochastic Geometry: Likelihood and Computation* (eds O. Barndorff-Nielsen, W. Kendall and M. van Lieshout), ch. 3, pp. 79–140. Boca Raton: Chapman and Hall–CRC.

Geyer, C. and Møller, J. (1994) Simulation procedures and likelihood inference for spatial point processes. *Scand. J. Statist.*, **21**, 359–373.

Gnanadesikan, R. B. and Wilk, M. (1970) A probability plotting procedure for general analysis of variance. *J. R. Statist. Soc.* B, **32**, 88–101.

Horvitz, D. and Thompson, D. (1952) A generalization of sampling without replacement from a finite universe. *J. Am. Statist. Ass.*, **47**, 663–685.

Jensen, J. and Møller, J. (1991) Pseudolikelihood for exponential family models of spatial point processes. *Ann. Appl. Probab.*, **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 *Recent Advances on Scan Statistics* (eds J. Glaz and N. Balakrishnan), pp. 303–322. Boston: Birkhauser.

Landwehr, J., Pregibon, D. and Shoemaker, A. (1984) Graphical methods for assessing logistic regression models. *J. Am. Statist. Ass.*, **79**, 61–83.

Lawson, A. (1992) On fitting non-stationary Markov point process models on GLIM. In *Computational Statistics* (eds Y. Dodge and J. Whittaker), vol. 1, pp. 35–40. Heidelberg: Physica.

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 *Stochastic Point Processes* (ed. P. Lewis), pp. 1–54. New York: Wiley.

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

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

Lindsey, J. K. (1995) Fitting parametric counting processes by using log-linear models. *Appl. Statist.*, **44**, 201–212.

Mase, S., Møller, J., Stoyan, D., Waagepetersen, R. and Döge, G. (2001) Packing densities and simulated tempering for hard core Gibbs point processes. *Ann. Inst. Statist. Math.*, **53**, 661–680.

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

Møller, J. (2003) Shot noise Cox processes. *Adv. Appl. Probab.*, **35**, 614–640.

Møller, J., Syversveen, A.-R. and Waagepetersen, R. (1998) Log Gaussian Cox processes. *Scand. J. Statist.*, **25**, 451–482.

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

Møller, J. and Waagepetersen, R. P. (2003b) An introduction to simulation-based inference for spatial point processes. *Lect. Notes Statist.*, **173**, 143–198.

Nair, M. (1990) Random space change for multiparameter point processes. *Ann. Probab.*, **18**, 1222–1231.

Nguyen, X. and Zessin, H. (1979) Integral and differential characterizations of Gibbs processes. *Math. Nacht.*, **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). *Bull. Choshi Mar. Lab.*, no. 6, 27–37.

Ogata, Y. (1988) Statistical models for earthquake occurrences and residual analysis for point processes. *J. Am. Statist. Ass.*, **83**, 9–27.

Ogata, Y. and Tanemura, M. (1981) Estimation of interaction potentials of spatial point patterns through the maximum likelihood procedure. *Ann. Inst. Statist. Math.* B, **33**, 315–338.

Ogata, Y. and Tanemura, M. (1986) Likelihood estimation of interaction potentials and external fields of inhomogeneous spatial point patterns. In *Proc. Pacific Statist. Congr.* (eds I. Francis, B. Manly and F. Lam), pp. 150–154. Amsterdam: Elsevier.

Papangelou, F. (1974) The conditional intensity of general point processes and an application to line processes. *Z. Wahrsch. Ver. Geb.*, **28**, 207–226.

Pregibon, D. (1981) Logistic regression diagnostics. *Ann. Statist.*, **9**, 705–724.

R Development Core Team (2004) *R: a Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Ripley, B. (1977) Modelling spatial patterns (with discussion). *J. R. Statist. Soc.* B, **39**, 172–212.

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

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

Ripley, B. and Kelly, F. (1977) Markov point processes. *J. Lond. Math. Soc.*, **15**, 188–192.

Särkkä, A. (1993) *Pseudo-likelihood Approach for Pair Potential Estimation of Gibbs Processes*. Jyväskylä: University of Jyväskylä.

Schoenberg, F. (1999) Transforming spatial point processes into Poisson processes. *Stochast. Process. Appl.*, **81**, 155–164.

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

Stoyan, D., Kendall, W. and Mecke, J. (1995) *Stochastic Geometry and Its Applications*, 2nd edn. Chichester: Wiley.

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

Venables, W. and Ripley, B. (1997) *Modern Applied Statistics with S-Plus*, 2nd edn. New York: Springer.

Vere-Jones, D. (1970) Stochastic models for earthquake occurrence (with discussion). *J. R. Statist. Soc.* B, **32**, 1–62.

Vere-Jones, D. and Schoenberg, F. P. (2004) Rescaling marked point processes. *Aust. New Zeal. J. Statist.*, **46**, 133–143.

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

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

Zhuang, J., Ogata, Y. and Vere-Jones, D. (2005) Diagnostic analysis of space-time branching processes for earthquakes. *Lect. Notes Statist.*, **185**, in the press.

Discussion on the paper by Baddeley, Turner, Møller and Hazelton

Wilfrid S. Kendall (*University of Warwick, Coventry*)

We owe thanks to Baddeley, Turner, Møller and Hazelton for a most interesting and stimulating paper, for the clarity with which it has been written and for the pleasure of exploring their attractive synthesis of spatial pattern, open source computing and statistical theory.

It has long been notorious that it is hard for the untutored eye to assess the statistical significance of perceived deviations from models of spatial pattern, most famously in the case of the humble Poisson planar point process. Statistics and diagnostic plots (such as Ripley's celebrated K-function) have been developed over a long period of time to assist in more reliable assessment. This paper builds on the tradition, and particularly the Stoyan–Grabarnik diagnostic based on the Papangelou conditional intensity, while linking

organically to the theory of generalized linear models. (I note, in passing, that I believe the impetus for the definition of conditional intensity arose from the strictly theoretical motivation of attacking the then unresolved Davidson conjecture: it is delightful to find it now at the centre of a topic in applied statistics!) Neglecting *all* technicalities, important or not, conditional intensity

$$\lambda(u; \mathbf{x}) = \frac{f(\mathbf{x} \cup \{u\})}{f(\mathbf{x} \setminus \{u\})}$$

tells us how much we should expect to see a point at a specified location u if we know the point pattern everywhere else ($\mathbf{x} \setminus \{u\}$). The residuals of a model are 'observed' minus 'expected', so in this particular case we must have a residual *random measure*, placing an atom of unit mass on each observed point and compensating by subtracting a density formed from the (fitted) conditional intensity. And then of course we should consider the possibility of reweighting or scaling the residuals, just as in the case of generalized linear models.

Consideration of residuals leads inevitably to attractive and even informative supplementary graphics. Here the challenge is to deliver useful graphics when the subject of enquiry is not a sequence of residual differences but the realization of a random measure. The authors propose a four-panel presentation:

(a) a *mark plot* of the measure representing atoms by circles and density by a grey scale plot,
(b), (c) plots of residuals against x-and y-co-ordinates and
(d) a smoothed version of (a) presented as a contour plot.

As shown by the examples, (a) is pretty but not easy to interpret, plots (b) and (c) are straightforward and often informative, and (d) carries most of the useful information to be derived from (a); the further computational burden of Q–Q-plots must be incurred to gain information about interactions.

Here are some questions which the paper raises for me. The proposed residual diagnostic follows the pattern of classical regression residuals and is clearly the right thing to do; however, in classical theory residuals can be masked when more than one observation deviates from the pattern, and we might expect this to be even more of a hazard in the point process case. It would be informative to see a simple salutary point process example, after the fashion of the Baddeley and Silverman (1984) cautionary example on second-order methods.

In Section 9.4 the authors describe the effect on raw residuals of the spatial Markov property when the underlying model is a finite range Markov point process (and make a conjecture about asymptotics, the strong law of large numbers and the central limit theorem for residuals; perhaps a proof would follow by adapting the methods of Penrose and Yukich (2001) and Penrose (2003)). I cannot resist the temptation to ask, does the effect of the spatial Markov property carry through for nearest neighbour Markov point processes (Baddeley and Møller, 1989)?

I find myself undecided about a detail of the mark plot: might it be more appropriate to represent pattern points by discs of radius the square root of (weighted) residual mass, to calibrate their visual effect? But this is a rather minor issue.

Finally, this graphical apparatus of point process residual diagnostics is all about presentation of a realization of a particular random measure. May we look forward to future developments of `spatstat`, integrating this apparatus into some kind of extension from the `ppp` point process class to a general random measure class `prm`? For example, for data sets such as the copper data it would be natural to use a Hough transform (see for example the mention in Ohser and Mücklich (2000) and also Baddeley and van Lieshout (1992)) to convert the point pattern into a fibre process on line space, and to examine smoothings of induced random measures. My preliminary (and computationally inefficient) experiments suggest that no structure is revealed by this specific procedure in the copper data case, but such an extension would have many possible uses.

Expressing these hopes for future developments, it therefore gives me great pleasure to propose the vote of thanks for this paper.

Eric Renshaw (*University of Strathclyde, Glasgow*)
Although the study of spatial processes has seen massive development through the past four decades, there has been remarkably little discussion concerning model fit. This paper therefore provides a valuable contribution through the construction of residuals and residual plots for models fitted to spatial point patterns, using a natural generalization from temporal point processes. Replacing the hazard rate of the lifetime process by the Papangelou conditional intensity is a particularly neat approach, and the time is ripe for applying these ideas to marked point processes, which may be regarded as a specialized version

of a general point structure. This situation is particularly complex, since mark size and point position will not only be interdependent in many real life situations, but will also often develop through time.

Spatial–temporal modelling has witnessed a surge of interest fuelled by large amounts of data on pollution and global climate monitoring, through geographical information systems, remote sensing platforms, monitoring networks and computer simulation models (Särkkä and Renshaw, 2005). Multivariate time series methods involve the construction of specific space–time dynamic models. Geostatistical models are used to study continuous spatial processes that are observed at a finite set of locations, though the joint space–time covariance structure can be difficult to specify and implement. Space–time dynamic models combine both approaches; estimation can be recursive and produces a powerful modelling strategy called the kriged Kalman filter. Stochastic, integrodifference equations allow the redistribution kernel to vary with space and/or time, whilst extensive use can also be made of Bayesian frameworks and general additive mixed models. Disease mapping often involves hierarchical spatial–temporal models, whilst descriptive statistical methods in climatology include empirical orthogonal functions, principal oscillation pattern analysis, canonical correlation analysis and space–time spectral analysis. However, such studies generally involve sampling at fixed locations and so do not easily relate to physical processes which develop through continuous space and time.

Often there is a genuine dependence between point locations and their associated mark variables, e.g. the heights, diameter at breast height and volume of trees, which are directly affected by competition for light and nutrient, and also their locations which relate to the ability of offspring and immigrants to survive under varying levels of competition. Cressie (1993), for example, examines the locations and diameters of 584 long leaf pines in a region $200 \, \text{m} \times 200 \, \text{m}$. Spectral analysis (Renshaw, 2002) shows strong evidence of clustering with intercluster distances that are in agreement with Cressie's nested block quadrat analysis, and Stoyan and Stoyan (1995) provide a compelling space domain analysis assuming that trees exist in clusters. Since only rarely will we know the exact process-generating mechanism, we need a general, flexible and computationally fast stochastic model as a realistic surrogate (Renshaw and Särkkä, 2001).

Injecting stochasticity through the simple immigration–death process generates useful and ubiquitous paradigms in many practical applications. The ith immigrant is allocated a random mark $m_i(t)$ and location x_i, and, for computational speed, in $(t, t + \text{d}t)$ $m_i(t)$ undergoes the *deterministic* change

$$m_i(t + \text{d}t) = m_i(t) + f\{m_i(t)\} \, \text{d}t + \sum_{j \neq i} h\{m_i(t), m_j(t); \|x_i - x_j\|\} \, \text{d}t.$$

This generic process caters both for competition (repulsion) and attraction, is extremely wide ranging and can be made totally general across any appropriate system involving stochastic arrival–death and deterministic growth interaction. Here $f(\cdot)$ is an individual growth function in the absence of spatial interaction, and $h(\cdot)$ is a spatial interaction function taken over all points $j \neq i$. If $m_i(t + \text{d}t) \leqslant 0$ then the individual has died 'interactively' and point i is deleted. Choosing $f(\cdot)$ equal to $\lambda m_i(t)\{1 - m_i(t)/K\}$ (logistic) or $\lambda\{1 - m_i(t)/K\}^2$ (quadratic) ensures that the population size remains bounded, whilst the linear form $\lambda\{1 - m_i(t)/K\}$ generates genuine 'upper' and 'lower' canopy structures. Two forms for $h(\cdot)$ that produce particularly informative patterns are

$$h_1(\cdot) = -b \, I[\|x_i - x_j\| < r\{m_i(t) + m_j(t)\}],$$
$$h_2(\cdot) = -b \, \text{area}[D\{x_i, r \, m_i(t)\} \cap D\{x_j, r \, m_j(t)\}]/\pi r^2 \, m_i^2(t).$$

The former corresponds to hard-edge interaction; as soon as two discs overlap then competitive interaction takes place with force b. As this function is symmetric (the larger and smaller of an interactive pair are affected equally), it may not be appropriate if marks are substantially different. The latter corresponds to soft-edge interaction, is asymmetric (the smaller of two interacting marks is affected more than the larger), and is relevant to many ecological situations. For the competitive force on a mark is proportional to the extent to which its influence zone, $D(\cdot)$, is overlapped by those of neighbouring marks.

Renshaw and Särkkä (2001) develop a pseudolikelihood approach to construct parameter estimates at successive time points $t_1, t_2, \ldots$. However, if single estimates are to be constructed from data across the full range $t_1, \ldots, t_n$, then this approach is no longer computationally feasible. Fortunately, a parallel least squares procedure can be developed, and study of simple stochastic models shows that this is as powerful as likelihood-based procedures, yet is mathematically and computationally simpler to employ (Särkkä and Renshaw, 2005).

Until now 'hypothesis testing' has been based purely on summary statistics such as mark histograms, age plots and interpoint distances, combined with visual comparison of real and simulated patterns, and

qualititative assessment of key features. For example, linear growth is needed to effect an established lower tier, whereas soft-edge interaction produces more realistic forest patterns than hard-edge interaction. Applying the general approach that is developed in this paper specifically to marked point processes would clearly represent a major advance, especially since growth and interaction functions can be changed to suit any given area of application. Comas (2005), for example, analyses management strategies for thinning, planting, regeneration, extraction, mixed stands, etc. to maximize forest yield. As model choice is made on a fairly subjective basis, a more formal development of the production and testing of realistic forest model structures would be of great practical use. Other techniques which show considerable promise include splitting the interdependence between marks and points through a harmonic decomposition of the mark frequencies (Renshaw *et al.*, 2005), and the development and application of spatial wavelet analysis (Saura *et al.*, 2006).

This stimulating, comprehensive and timely paper is not only a milestone in the advance of spatial processes, but also there is clearly huge scope for expanding the approach into many fields of application. I am delighted to second the vote of thanks.

The vote of thanks was passed by acclamation.

Andrew Lawson (*University of South Carolina, Columbia*)
With regard to spatial epidemiology (SE) (see Fig. 3), Diggle (1990) is a very early example of work in this subject. Indeed much has happened in the field since then (see for example Elliott *et al.* (2000) and Lawson (2001) for reviews). An example is the emphasis on *cumulative distance residuals* based on transformations. It is now well established that air pollution hazards have a bivariate exposure distribution where both distance and angle and functions of both are involved in the description of risk around sources.

It is reasonable for a range of point processes of interest in health studies that we assume a flexible Bayesian model where effectively correlation is handled within a higher level of the hierarchy and then the points are assumed to be conditionally independent given higher level parameters. Hence, conditional on $\lambda(u, x)$, the case events follow a heterogeneous Poisson process. A stochastic intensity model can accommodate many features of interest in SE including random effects and hidden random object processes as well as covariate effects.

An example would be

$$\{x_i\} \sim \text{PP}\{\lambda(x_i; \theta)\},$$

$$\lambda(x_i; \theta) = \exp(F_i' \alpha + S_i + \nu_i),$$

$$\{S_i\} \sim \text{Gau}(\mathbf{0}, \Sigma),$$

$$\{\alpha_j\} \sim N(0, \kappa_\alpha),$$

$$\{\nu_i\} \sim N(0, \kappa_\nu).$$

This has been successfully employed in SE in a variety of examples (see for example Lawson (2001)). Indeed residuals based on a saturated estimate of $\lambda(u, x)$ compared with a parameterized estimate of $\lambda(u, x)$ (Section 8.2) were proposed by Lawson (1993) and applied in SE examples (Lawson and Harrington, 1996).

The choice of B is a major concern. The authors do not discuss this and simply suggest that a smoothing approach is better than a quadrat count approach. Both approaches are smoothing approaches, however. The residuals that are derived will be highly dependent on this choice (i.e. quadrat size, shape and location) as will the smoothed residual field ($s(u)$). This is worrying as it is known that estimation with modulated Poisson process models where the background ($\rho(u)$) is estimated is extremely sensitive to the smoothing parameter choice (Lawson and Williams, 1994).

Cluster modelling has been extensively developed for SE and sophisticated Bayesian models for cluster detection have been developed for case event data. The Geyer saturation process model is a general clustering model that does not model cluster locations (see for example Lawson and Denison (2002)).

J. Mateu and F. Saura (*University Jaume I, Castellón*)
We enjoyed reading this stimulating and timely paper, which is both certainly interesting and useful for the practice of spatial point process modelling.

It is surely crucial that models should be chosen not only for their mathematical convenience but also because they reflect the scientist's insight into the nature of the phenomena observed. Particularly welcome in this account is the possibility of checking and analysing the residuals coming from general inhomoge-

neous point process models where the intensity may depend on general covariates, such as an underlying continuous random field. This, in turn, sheds light on the exciting analysis of spatial point and geostatistical process connections.

The first point that we would like to focus on is in a parallel way to evaluate the residuals of a lurking variable through wavelet analysis. Wavelet theory has been applied to many fields, but its use for point processes has not yet been extended (Brillinger, 1997). A possible methodology is based on computing a transformation $f_h(t)$ of a suitable estimator of the intensity function (Diggle, 1985) of a one-dimensional point pattern (e.g. one of the residuals shown in this paper) $\hat{\lambda}_h(t)$, with h the bandwidth. Then, its discrete wavelet decomposition can be computed providing the detail coefficients to detect the significant features (Saura *et al.*, 2006). Wavelet analysis applied to the residuals can show a deeper insight into the appropriateness of the fitted model. For example, it can show more evidence of significance when an incorrect model is fitted to the data. Additionally, wavelet analysis can be used to discriminate between possible residual types.

Our second point is why not consider a more general expression for the definition of the particular residuals by considering

$$h(u, \mathbf{x}) = \lambda(u, \mathbf{x})^{-\alpha},$$

where α is a parameter that must be estimated from the data and fitted model? Note that when $\alpha = 1$ and $\alpha = 0.5$ we obtain the inverse and Pearson residuals respectively. This is like asking the data to indicate which is the most adequate residual form to use. Thus, this seems a more objective selection of the type of residuals, although certainly more complicated.

The final point to consider is the particular or specialized case of marked point processes. Applying the general residual approach to the marked case would clearly represent a major advance. For example, when the marks are modelled through a (quite general) harmonic decomposition, we could define a residual measure, called *discrepancy* in Renshaw *et al.* (2005), to evaluate the goodness of fit. The question we pose here is which kind of relationship has the discrepancy function with any of the tools presented in this paper?

E. F. Harding (*Little Ouse*)
I would like to reinforce the compliments already paid. This deep and unifying paper is very clearly written. Already, at the end of Section 1, I felt that I knew well what to expect and could then enjoy watching the rest unfold in detail.

The paper closes a few loops for me. Adrian Baddeley and I were in the Statistical Laboratory together more then 20 years ago; I am very happy at the closing of this loop!

The use of the Papangelou conditional intensity, leading to residuals defined everywhere, is the key insight.

Wilfrid Kendall's reference to Papangelou's work in the early 1970s reminded me of Papangelou's (1974) related paper on Palm probabilities in the Rollo Davidson memorial volume *Stochastic Geometry*, which I co-edited with David Kendall.

I have comments on a couple of issues (anything else must wait until the rest has properly sunk in).

I share Wilfrid Kendall's ambivalence towards the 'mark plot' in Fig. 5(a), for instance. The varying radii at first suggest inadequate fit, even though it is the true model. However, what they really exhibit is that occurrence of a point is increasingly anomalous as intensity decreases; the radii measure this anomaly. You obtain the same in an ordinary histogram, where observations are less to be expected in the tails of the fitted distribution. Years ago, Tukey (1972) devised the brilliantly simple 'hanging rootogram': take the square root of everything, and hang the histogram bars off the theoretical curve; then the bottoms of the bars exhibit informative residuals. This links naturally with Wilfrid Kendall's square-root suggestion.

The other issue arises explicitly in the Chorley–south Ribble data, where apparent inhomogeneity is strongly influenced by the spatial distribution of susceptible individuals. But it is implicit also for the seedlings and saplings in a natural forest stand of Japanese black pine. The forest includes mature specimens, whose presence will influence the propensity for seedlings and saplings (positively, by seeding; negatively, by competition). This suggests a theoretical question. Inhomogeneity due to 'intrinsic' phenomena (e.g. interaction or clustering models) is in general unidentifiable relative to inhomogeneity of an unobserved underlying 'susceptibility'; and both may be realizations of spatial processes. However, such models theoretically imply observable features enabling identification relative to an underlying 'susceptibility' influenced by a different kind of model (e.g. spatial variation of resources). Do the authors have general views on possible theoretical approaches to such identification?

Dietrich Stoyan (*Technische Universität Bergakademie Freiberg*) **and Aila Särkkä** (*Chalmers University of Technology, Gothenburg*)
The authors have defined residuals for point process models and introduced several diagnostic plots based on them. We think that this is a very important piece of work which will frequently be used in the future in point process statistics and we congratulate them for their valuable contribution. However, a weak point of their approach is perhaps that models with hard core are excluded. In practice many point patterns have at least small hard cores. Therefore we have developed an alternative method of residual analysis which is perhaps still closer to the classical idea of residuals in regression analysis, where residuals are defined as differences between observed responses and predictions (Särkkä and Stoyan, 2005).

When considering spatial point patterns, the observed responses are the measured locations of the points. Therefore, a natural counterpart of the classic residual is to define the residual of a point to be the difference between the observed location and some predicted location given the model. We suggest computing the 'predicted location' by using the Papangelou conditional intensity.

Let us assume that we have fitted a model with conditional intensity $\lambda(u, \mathbf{x})$ to the data. Let us then consider a point $x_i \in \mathbf{x}$ and a disc $b(x_i, r)$ of radius r centred at x_i where $r > 0$ is a fixed radius. Then, we search for the location u in $b(x_i, r)$ where the point x_i should be placed to maximize the conditional intensity $\lambda(u, \mathbf{x})$. Let $\hat{x}_i$ be the predicted location. Then we define the residual of the point x_i as $x_i - \hat{x}_i$. If there are different possibilities to choose for $\hat{x}_i$, then we take the point that is closest to x_i. If the fitted model is good, the length of the residual is expected to be small, i.e. either $\hat{x}_i = x_i$ or $\hat{x}_i$ is very close to x_i. Note that the residuals depend on the radius r and that the choice of r can be crucial.

The residuals that we suggest turned out to be reasonable tools for non-trivial Gibbs point process models with or without a hard core. Also, they could be used to detect outliers in point process data. However, our residual analysis is not appropriate when checking for example the Poisson assumption since in this case the residuals would always be zero vectors independently of the point pattern.

Jiancang Zhuang (*Institute of Statistical Mathematics, Tokyo*)
I congratulate the authors on presenting us with a most excellent and interesting paper. This paper systematically develops the theory of residual analysis for spatial point processes in great detail together with applications to several typical data sets showing plausible statistical analysis techniques. It has overstimulated me. In particular, it helped me in finding the solutions to problems in the point process inferences to earthquake data (Zhuang *et al.*, 2004; Zhuang, 2005).

I would next like to comment briefly on two points. The first is that the residuals that are discussed in this paper are first order. Similarly to second-order residuals for temporal processes (Zhuang, 2005), we can also define second-order residuals for spatial processes based on the formula

$$\mathbf{E}\left[\int\int_{D \setminus \text{diag}(D)} h(u, v, \mathbf{X} \setminus \{u, v\})\, \mathbf{X}(\mathrm{d}u)\, \mathbf{X}(\mathrm{d}v) - \int_D h(u, v, \mathbf{X})\, \lambda(u, v, \mathbf{X})\, \mu_2(\mathrm{d}u\, \mathrm{d}v) \right] = 0, \qquad (42)$$

where $\lambda(u, v, \mathbf{X})$ is the two-point conditional intensity defined in Section 9.3, $\mu_2 = \mu \times \mu$, μ denotes the Lebesgue measure on W and $\text{diag}(D) = \{(x, y) : (x, y) \in D \wedge x = y\}$, $D \subseteq W^2$. More generally, higher order residuals can be defined based on

$$\mathbf{E}\left[\int \ldots \int_{\text{nd}(D^{(k)})} h(u_1, u_2, \ldots, u_k, \mathbf{X} \setminus \{u_1, u_2, \ldots, u_k\})\, \mathbf{X}(\mathrm{d}u_1)\, \mathbf{X}(\mathrm{d}u_2) \ldots \mathbf{X}(\mathrm{d}u_k) \right]$$
$$= \mathbf{E}\left[\int_{D^k} h(u_1, u_2, \ldots, u_k, \mathbf{X})\, \lambda(u_1, u_2, \ldots, u_k, \mathbf{X})\, \mu_k(\mathrm{d}u_1\, \mathrm{d}u_2 \ldots, \mathrm{d}u_k) \right], \qquad (43)$$

where

$$\text{nd}(D^{(k)}) = \{(x_1, x_2, \ldots, x_k) : (x_1, x_2, \ldots, x_k) \in D^{(k)} \wedge \forall i\, \forall j\, (i \neq j \Rightarrow x_i \neq x_j)\},$$

$D^{(k)} \subseteq W^k$ and

$$\lambda(u_1, u_2, \ldots, u_k, \mathbf{X}) = \frac{f(u_1, u_2, \ldots, u_k, \mathbf{X} \cup \{u_1, u_2, \ldots, u_k\})}{f(u_1, u_2, \ldots, u_k, \mathbf{X} \setminus \{u_1, u_2, \ldots, u_k\})}. \qquad (44)$$

The second point is on Gibbs processes. Let the conditional intensity take the form

$$\lambda(u, \mathbf{X}) = b(u) \prod_{i:x_i \in \mathbf{X} \setminus \{u\}} c(\|u - x_i\|). \tag{45}$$

Except the example residuals for this kind of processes that are discussed in the paper, the following two kinds of residuals are possibly useful for statistical inferences. Let

$$h(u, \mathbf{X} \setminus \{u\}) = b(u) / \lambda(u, \mathbf{X}); \tag{46}$$

then, for any $B \subseteq W$,

$$\mathbf{E}\left[\sum_{x_i \in \mathbf{X} \cap B} h(x_i, \mathbf{X} \setminus \{x_i\})\right] = \mathbf{E}\left[\int_B h(u, \mathbf{X}) \, \lambda(u, \mathbf{X}) \, \mu(\mathrm{d}u)\right] = \int_B b(u) \, \mu(\mathrm{d}u). \tag{47}$$

If we choose B to be a set of small volume around point x, then the above quantity is approximately $b(x) \mu(B)$, leading to a reconstruction of the function $b(u)$. In other words,

$$R(B, h, \lambda) = \sum_{x_i \in \mathbf{X} \cap B} h(x_i, \mathbf{X}) - \int_B b(u) \, \mu(\mathrm{d}u) \tag{48}$$

could be called the background residual in some sense.

Similarly, taking

$$h(u, v, \mathbf{X} \setminus \{u, v\}) = \frac{1}{\lambda(u, \mathbf{X} \setminus \{u, v\}) \, \lambda(\nu, \mathbf{X} \setminus \{u, v\})}, \tag{49}$$

by equation (42),

$$\mathbf{E}\left[\sum_{(x_i, x_j) \in (\mathbf{X} \times \mathbf{X}) \cap (D \setminus \mathrm{diag}(D))} h(x_i, x_j, \mathbf{X} \setminus \{x_i, x_j\})\right] = \mathbf{E}\left[\int_D h(u, v, \mathbf{X}) \, \lambda(u, v, \mathbf{X}) \, \mu(\mathrm{d}u \, \mathrm{d}v)\right]$$

$$= \int_D c(\|u - v\|) \, \mu_2(\mathrm{d}u \, \mathrm{d}v). \tag{50}$$

Again, if we choose a bounded

$$D \subset \{(u, v) : x - \delta \leqslant \|u - v\| \leqslant x + \delta\} \cap W,$$

the above integral can be approximately estimated by $c(x) \mu_2(D)$, which can be used to reconstruct $c(x)$. This is an example of the usage of second-order residuals.

These are among the many points that I would like to raise. I have enjoyed reading this paper very much.

Julian Besag (*University of Washington, Seattle*)
I welcome this interesting and elegant paper, which is the culminatation of a long-standing crusade, particularly by Adrian Baddeley, to develop and popularize frequentist methods for the analysis of spatial point processes that have amenable full conditional (Papangelou) distributions. However, I am disappointed by the examples. Fig. 1 seems a bad start in a 21st-century paper. One has to ask why anyone would want to model the locations of 204 seedlings and saplings in a $10 \, \mathrm{m} \times 10 \, \mathrm{m}$ square by a log-cubic trend with some pairwise interactions thrown in. More generally, modern data acquisition in ecology can generate large scale data sets where the scientific questions involve concomitant environmental variables and/or interactions between multitype point processes. The epidemiologic example comes closer to this but it is not obvious that anything much is gained by reformulating the model in terms of its conditional intensities; and I think the same would hold for the application in Diggle *et al.* (2005), say. I hope the authors can provide better examples in the future.

I find the division between residual analysis and goodness of fit less clear cut than the authors, particularly given the free choice of test statistic with exact (Markov chain) Monte Carlo methods (Besag and Clifford, 1989, 1991) that can be extended to some point process models, though perhaps not to those of most practical interest.

I am surprised that my pseudolikelihood estimation is still recommended. Is this justified 20 years after the introduction of Markov chain Monte Carlo maximum likelihood (Penttinen, 1984) about which we

know much more these days (e.g. Geyer (1999))? Finally, the authors mention similarities with diagnostics for logistic regression and I wonder whether these can be formalized using the construction of pairwise interaction point processes as limits of autologistic binary lattice schemes (Besag *et al.*, 1982).

The following contributions were received in writing after the meeting.

Peter Diggle (*Lancaster University*)
As with residual analysis in other areas of statistics, the real test of the tools proposed in the paper will be their ability to reveal, in substantive applications, insights which would have been missed by existing diagnostic methods. The inclusion of the authors' proposals within the spatstat package should greatly help this process.

Although accepting that the examples in the paper are largely illustrative, I would like to comment on the lung and larynx cancer example, which raises general questions about the role of spatial point process methods in epidemiology. I have argued elsewhere (Diggle (2003), chapter 9) that, in this context, point process modelling can help to formalize an epidemiological question in statistical terms, but that we should not necessarily rely on the point process model for inference.

In Diggle (1990), I modelled the two types of cancers as independent Poisson processes with respective intensities $\rho(x)$ and

$$\lambda(x) = \beta \, \rho(x) \, f\{d(x)\}$$

where β corresponds to the rarity of larynx relative to lung cancer, $\rho(x)$ is proportional to the control population density, $d(x)$ denotes distance from the incinerator's former location and

$$f(d) = 1 + \alpha \, \exp\{-(d/\phi)^2\}$$

describes how excess risk for larynx cancer varies with distance.

The most contentious part of this formulation is $\rho(x)$. Even if we admit the concept of a continuously varying $\rho(x)$, its estimation is extremely difficult because it varies over orders of magnitude between urban and rural areas. But the scientific focus is on $f(d)$, not on $\rho(x)$.

Diggle and Rowlingson (1994) noted that under the assumed Poisson process model, and conditional on the observed lung and larynx cancer locations, the binary labels of the locations form a set of independent Bernoulli trials with 'success' (larynx) probabilities

$$p(x) = \frac{\beta \, f\{d(x)\}}{1 + \beta \, f\{d(x)\}}.$$

The independence assumption implicit in the Poisson process formulation is reasonable for cancers. The assumed form of $f(\cdot)$ may or may not be reasonable, although its derivation from a point process model ensures that it is qualitatively sensible: $f(d) \approx 1$ for large d; hence β is the constant relative risk at locations remote from the incinerator; α is the elevation in relative risk at the incinerator. Perhaps more importantly, the reformulation as a binary regression problem allows inference on $f(\cdot)$ while eliminating $\rho(\cdot)$. The residuals from this binary regression suggest nothing untoward about the point highlighted in Fig. 14—like every other point remote from the incinerator, it has $\hat{f}\{d(x)\} \approx 1$ and hence approximate fitted value $\hat{\beta}$. I conclude that, as the authors surmise, their 'huge residual' is a consequence of a poorly estimated $\rho(\cdot)$—which is fair comment, but of limited scientific interest for these data. Of greater concern is the high leverage exerted by a tight 'cluster' of four larynx cases close to the incinerator.

Paul Fearnhead (*Lancaster University*)
I have a minor comment relating to the lurking variable plots (Fig. 7), and in particular the approximate confidence limits that were used.

By estimating the variance of the residuals by using equation (27), the authors are ignoring the effect that fitting a Poisson process ensures that the residuals (as defined by equation (34)) evaluated at $x = 1$ will always be 0. Thus the estimate of the variance of the residuals for values of x close to 1 will be a large overestimate of the truth.

This can lead to an inconsistency, whereby the conclusions from a lurking variable plot will depend on the direction in which we consider the spatial covariate.
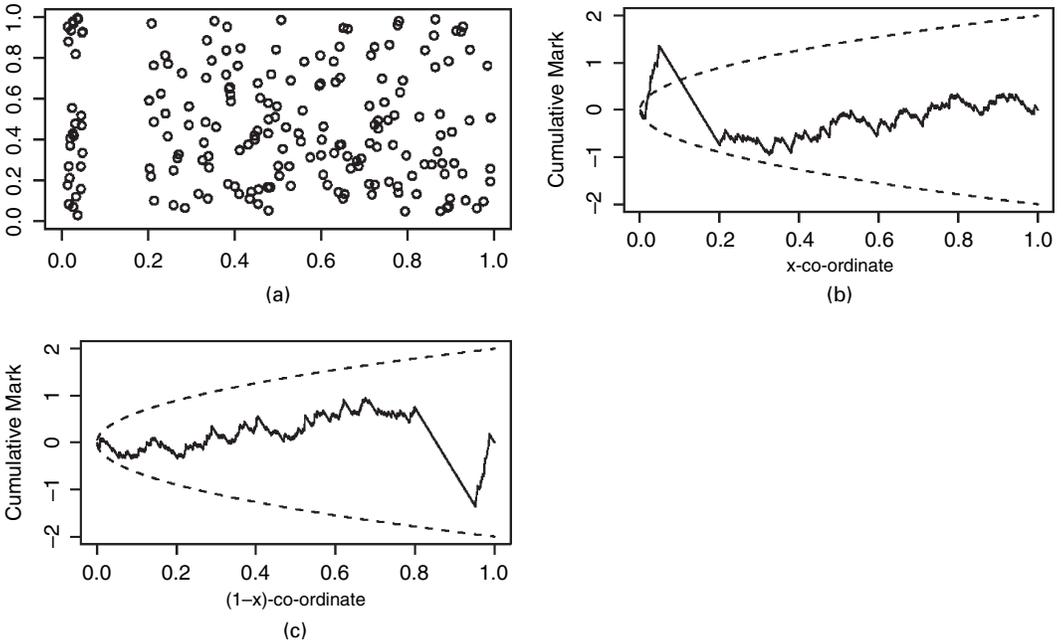
Fig. 22. (a) Simulated data from an inhomogeneous Poisson process, and lurking variable plots for the x-co-ordinate based on sublevel sets (b) W_1 and (c) W_2 (------, confidence limits calculated by using 2σ-limits based on equation (27))

As a trivial, and admittedly somewhat contrived, example, Fig. 22(a) shows data simulated from a Poisson process with intensity

$$\lambda(x, y) = \begin{cases} 800 & x < 0.05, \\ 0 & 0.05 \leqslant x < 0.2, \\ 200 & x \geqslant 0.2. \end{cases}$$

Define W to be the unit square, and for $u \in W$ define $X(u)$ as the x-co-ordinate of a point u. Finally define two sublevel sets:

$$W_1(z) = \{u \in W : X(u) < z\},$$
$$W_2(z) = \{u \in W : 1 - X(u) > z\}.$$

These sublevel sets differ solely in whether they define intervals of the x-axis as $[0, z]$ or $[1 - z, 1]$. Lurking variable plots for each of these sublevel sets is a suitable diagnostic plot for testing whether
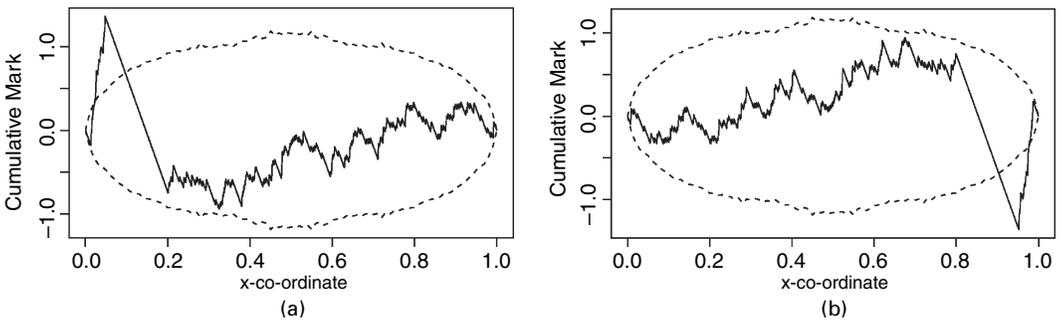
Fig. 23. Lurking variable plots for the x-co-ordinate based on sublevel sets (a) W_1 and (b) W_2 (------, confidence limits calculated by using the parametric bootstrap)

the x-co-ordinate is a covariate that the data depend on. One should hope that the conclusions from each of these sublevel plots would be the same, but as is shown in Figs 22(b) and 22(c) this is not so, at least if we compare the residuals with the plotted 2σ-limits. (For such a simple example, it is obvious from the data, and arguably from the resulting shape of the residual plot, that a homogeneous Poisson process is an inappropriate model—though it is easy to imagine less simple examples for which these two different but equivalent lurking variable plots lead to different conclusions.)

A simple way to avoid this problem, and which will also give appropriate confidence limits in more general cases, is to use simulation to construct the confidence limits. Fig. 23 shows the resulting lurking variable plots where the confidence limits have been constructed by using a parametric bootstrap. These are now symmetric, and both plots lead to the same conclusion that the homogeneous Poisson process is inappropriate.

Charles J. Geyer (*University of Minnesota, Minneapolis*)
The authors have written a most enlightening paper. I find compelling their argument for basing diagnostics for spatial point processes on the Papangelou conditional intensity. Proposition 3.3 in Geyer (1999) says that the Markov chain Monte Carlo samplers that are used to simulate these processes (Geyer and Møller, 1994) are geometrically ergodic if the Papangelou conditional intensity is uniformly bounded. Another important application of this conditional intensity seems natural.

Baddeley and his colleagues present many examples where their diagnostics work, but none where they fail. But we know from the much simpler and much better understood area of regression diagnostics that diagnostics are a fairly weak methodology for finding problems with models. How much heteroscedasticity does there need to be for a residuals *versus* fitted values plot to reveal it clearly? Simple experiments, which anyone can do, say that for $n = 100$ there needs to be a factor of 3 in the error standard deviation from one side of the plot to the other. So diagnostics cannot replace hypothesis tests for model comparison. Of course, in regression, no one suggests that they should.

Baddeley and his colleagues also do not claim that diagnostics can replace hypothesis tests, in particular likelihood ratio tests (Geyer (1999), section 3.17); they simply ignore hypothesis tests. I wonder what their diagnostics would show about whether the triplets process (Geyer (1999), section 3.9.1), what one obtains by truncating the Gibbs expansion (equation (4) in the paper) at three terms, fits the data of Fig. 4(c). Presumably, as in Geyer (1999), section 3.17, the likelihood ratio test would show that the triplets process does not fit. These two processes are so close in distribution that humans cannot distinguish them (Geyer (1999), Figs 3.4 and 3.5). I conjecture that diagnostics would fail to find what the likelihood ratio test clearly reveals.

I should also like to pose a question. One method of dealing with edge effects is to treat the process outside the observation window as missing data (Geyer (l999), section 3.15). In such a situation should we still only apply diagnostics to data in a subwindow away from the edges or should we average the diagnostics over the missing data outside the observation window?

Pavel Grabarnik (*Russian Academy of Sciences, Pushchino*)
This paper is of fundamental importance since it puts on a firm basis earlier attempts at residual analysis for Gibbs point processes (Stoyan and Grabarnik, 1991; Särkkä, 1993). My comments concern a generalization of forms of spatial residuals and a possible application of the presented methodology to a special case of point patterns.

In the paper the Georgii–Nguyen–Zessin formula (8) is a source of various forms of residuals obtained by h-weighting. More generally, any estimating function can be a candidate for the h-weighted innovation process. For instance, the time invariance estimating functions (Baddeley, 2000) is one option. The second-order pseudoscore (Goulard *et al.*, 1996) or other weightings of the innovation based on the Georgii–Nguyen–Zessin formula for the two-point conditional intensity is another possibility. Perhaps there is not an optimal choice, and the best form of residuals depends on the type of data and the model fitted; therefore it would be helpful for practitioners to have some guidelines for such a choice.

A potential application of the methodology proposed is modelling a spatial mixture of regular and clustered point patterns. This type of point patterns occurs in mixed age forests, when locations of old trees form a regular pattern as a result of a self-thinning mortality process, whereas young trees grow in clusters that are associated with gaps in the canopy created by falls of dead trees. Difficulties with goodness-of-fit testing by means of standard summary functions, e.g. Ripley's K-function, arise here due to their inability to detect clustering and regularity simultaneously at the same scale. In a case of the complete spatial

randomness testing a solution was proposed in Grabarnik and Chiu (2002). It seems that a diagnostic based on spatial residuals can help to validate fitted models that are capable of producing the mixture of clustered and regular patterns (e.g. the 'bipattern' model in Grabarnik and Särkkä (2001)). In addition to lurking variable plots, a useful diagnostic tool for the mixture patterns could be a calculation of the spatial autocorrelation function.

Yoshihiko Ogata (*Institute of Statistical Mathematics, Tokyo*)
I congratulate the authors for their illuminating paper on the diagnostic analysis of spatial point processes. Residual analysis is indeed an exciting part of statistical analysis, leading to many important discoveries from the data. I have been using a hierarchical Bayesian space–time point process model (Ogata *et al.*, 2003; Ogata, 2004) for the precise prediction of the earthquake occurrence rate $\lambda(t, x, y)$ at any time and location, the conditional intensity depending on the history of the earthquakes. The parameters of the model also depend on the location (x, y) and quantify the regional characteristics of the seismic activity depending on the geology and tectonics. For example, one of the parameters indicating aftershock frequency takes low and high values in and around a number of regions on the plate boundary surface respectively, where the friction coefficient is high, called asperities. I then carry out a 'Bayesian diagnostic analysis', where a flexible parametric function $\xi(t, x, y|\theta)$ is estimated to rectify the estimated model $\hat{\lambda}(t, x, y)$ in such a way that the intensity

$$\eta(t, x, y|\theta) = \hat{\lambda}(t, x, y)\, \xi(t, x, y|\theta)$$

of the doubly stochastic Poisson process is optimally adapted to the data. In fact, the dimension of θ is as large as the number of earthquakes, and we consider the likelihood of θ subject to the smoothness prior on the ξ-function. Eventually, although a good fit of the estimated model $\hat{\lambda}$ is indicated in most of the entire space–time volume where $\xi \approx 1$ holds, anomalies (activation and quiescence) relative to the modelled (or predicted by) $\hat{\lambda}$ are revealed in the periods and regions where $\xi > 1$ or $\xi < 1$ hold respectively. Taking account of predominant angles of fault mechanisms of the earthquakes in each region, such anomalies appear to be related to changes in failure stress within the crust from a far field rupture or silent slip. My summary results so far support the fact that even a small exogenous stress increment of the order of millibars can trigger such seismicity anomalies. Thus, I expect that an anomaly is sufficiently sensitive to detect slight changes in stress which even the geodetic records from the global positioning system network can barely detect in the time series of displacement records.

Frederic Paik Schoenberg (*University of California, Los Angeles*)
Congratulations go to the authors for their outstanding summary and implementation of residual methods for spatial point processes, for unifying several different treatments of point process residuals and for putting these techniques in context by comparing them with residual methods for ordinary regression data. The authors' emphasis on graphical methods for goodness-of-fit assessment is also appreciated. However, I hope that readers are not left with the impression that the problem of how to assess the goodness of fit for point processes has been completely solved. Existing methods all seem to have substantial drawbacks, and there is still much important work to be done in this area.

The authors' decision to smooth the residuals seems a little curious, and one may wonder whether some power is lost as a result. If an analogous smoothing procedure is routinely done with regression residuals, I am unaware of it. One may also question the effectiveness of the smoothed residual plots in discriminating between point process models. The authors state that this technique can be used to identify departures from Poisson processes. However, in most of the examples that are provided, the smoothed residuals identify cases where the trends, rather than the second-order properties, are incorrectly specified.

The residual Q–Q-plots seem more powerful at detecting departures from modelled interaction effects but do not identify locations where models fit poorly; nor do they readily suggest ways of improving models. Only the cumulative distribution is inspected and, as the authors note, the Q–Q-plots are primarily useful for detecting deviations in the tails. Further, situations may arise where there are more residuals in a certain interval than would be expected, and many fewer than expected in the next interval, so that the cumulative effect is negligible.

Another graphical method for model checking would be to thin the point process randomly, keeping any point x_i with probability $b/\lambda(x_i)$, where b is the infimum of λ over the observed space. The resulting process is Poisson in the space–time context when the conditional intensity λ is used (Schoenberg, 2003). Under general conditions this should also result in a Poisson process for purely spatial point processes, with λ the Papangelou conditional intensity; a proof of this would be welcome.

Rasmus Waagepetersen (*Aalborg University*)
To me, this great paper is a proof that statistics for spatial point processes has grown up into an adult branch of modern statistics. Now we can not only fit complex models to spatial point patterns but also assess the fitted models with tools that are analogous to those used for model assessment in the major fields of generalized linear models and survival analysis. My specific comments pertain to the lurking variable plots, residuals for Cox processes and the K-function.

The lurking variable plot seems to some extent superfluous given that fitting a model including the variable is now routine by using the eminent R library `spatstat`. Another potential use of a lurking variable type of plot, however, would be to identify the correct scoring of a covariate. In survival analysis, for example, it has been proposed to use martingale residuals to identify the proper scoring of a covariate.

For many types of Cox and cluster processes including inhomogeneous log-Gaussian Cox processes and certain Neyman–Scott processes, the intensity function is known in closed form whereas the conditional intensity is not. Hence, for such models it is more natural to construct residuals by using the intensity, i.e., letting $\lambda(\cdot)$ denote the intensity function, we might consider $n(X \cap B) - \int_B \lambda(u) \, \mathrm{d}u$ or $\Sigma_{x_i \in X \cap B} 1/\lambda(x_i) - |B|$ by analogy with the raw or inverse λ innovation measures. Moreover, the second-order properties of the inverse intensity residuals are determined by the K-function as adapted to inhomogeneous point processes in Baddeley *et al.* (2000). Problems regarding the interpretation of the inhomogeneous case K-function occur when the intensity function is estimated nonparametrically. Waagepetersen (2005) in contrast considers a parameterized log-linear intensity function for a certain class of inhomogeneous Neyman–Scott processes. In Waagepetersen (2005) I first obtain asymptotically normal estimates of the regression parameters and secondly use the consistently estimated intensity function to obtain a useful estimate of the K-function.

Returning to models that are specified in terms of a conditional intensity we might, inspired by Baddeley *et al.* (2000), define a K-type function by

$$\mathbb{E}\left[\frac{1}{|B|} \sum_{x_i \in X \cap B, x_j \in X} \frac{\mathbf{1}[0 < \|x_i - x_j\| < t]}{\lambda(x_i, X) \, \lambda(x_j, X)} \right].$$

This would characterize the second-order moments of the inverse λ innovation measure; for a pairwise interaction process with translation invariant interaction function c it has known expectation $\int_{\|v\| \leqslant t} c(v) \, \mathrm{d}v$; see also equation (28).

The **authors** replied in later, in writing, as follows.

We thank the discussants and all contributors for their enthusiastic and stimulating comments.

Wilfrid Kendall has raised several intriguing and challenging questions for further research. This paper is clearly only a beginning, in which we have identified the right notion of 'residuals' for a spatial point process. Related concepts, such as leverage and influence, should now be developed from this standpoint. The right way to plot residuals also needs investigation, as Kendall and Harding have both diplomatically pointed out.

To study limit behaviour and independence properties, we may exploit the fact that, for a Gibbs point process X on $\mathbb{R}^d$ with conditional intensity $\lambda(u, X)$ (see section 6.4 in Møller and Waagepetersen (2003)), the innovation

$$I(B|X_{B^c}) = n(X \cap B) - \int_B \lambda(u, X_B \cup X_{B^c}) \, \mathrm{d}u$$

is a martingale with respect to an increasing sequence of bounded Borel sets $B = B_1, B_2, \ldots \subset \mathbb{R}^d$. For a nearest neighbour Markov point process, $\lambda(u, X_B \cup X_{B^c})$ depends on X_{B^c} only through its splitting boundary (Kendall, 1990).

Eric Renshaw encourages us to develop residuals for marked point processes. The basic theory for such residuals is covered in the present paper, since a marked point process is 'just' a special type of point process. However, the practicalities, including diagnostic plots, may be quite different, and we shall pursue them.

Renshaw and Ogata have discussed models for space–time point processes. There is clearly scope for adapting our results to space–time (though much of this is already covered by the work of Vere-Jones, Ogata, Renshaw and others). Residuals in space–time are simpler to handle than residuals for spatial point processes because they are Markov in time.

Harding rightly warns that clustering and inhomogeneity are indistinguishable in some contexts. See Bartlett (1964). Unidentifiability reduces the sensitivity of residuals and diagnostics to departures from

the fitted model (and indeed reduces the power of hypothesis tests) but does not affect their fundamental validity. It does complicate the interpretation of diagnostic plots.

For Besag, Diggle and Lawson, the bone of greatest contention seems to be our choice of examples in Section 2 and their subsequent analysis. We concede that the examples can be impeached on various grounds. They were selected because they are accessible, illustrative, sufficiently small to exhibit sampling variability and have already been analysed and discussed in the literature, where indeed the issue of model checking was raised for each of them. Like Fisher's irises, any data which have been subjected to several analyses have inevitably lost some of their scientific freshness.

Note especially that we do not advocate or endorse the particular *models* that were fitted to the example data. Our paper is concerned with the validation of any point process model, rather than with finding the correct model for each example. In each case we used the models that were *fitted by the original researchers*; indeed the examples were selected with a view to the potential for weaknesses in these fitted models. The 'huge residual' that was highlighted in Fig. 14 results from our residual analysis of the model that was fitted by Diggle (1990) and surely satisfies Diggle's criterion that our residuals be able to find new features in existing analyses.

The method that is used to fit models is also largely irrelevant to our message. We certainly do not 'recommend' maximum pseudolikelihood estimation. For the synthetic examples, we fitted models by maximum pseudolikelihood simply because a fast software implementation was available. We are unaware of any software that is capable of fitting a wide range of (non-stationary; non-Poisson) spatial point process models by maximum likelihood estimation. We have recently implemented the Huang–Ogata Monte Carlo approximation to the maximum likelihood estimate for a wide class of models, so that residual analysis can now be performed for approximate maximum likelihood estimate fits. For further explanation see Baddeley and Turner (2005).

Diggle and Lawson rightly observe that a point process model may not be the appropriate vehicle for inference from spatial epidemiological data, in so far as the main interest lies in the relative risk rather than the absolute frequency of cases. To put it another way, inference ought to be performed conditionally on the domicile locations, as advocated by Diggle and Rowlinson (1994), rather than unconditionally, as demonstrated by Diggle (1990). For example, Diggle's discussion contribution confirms that the observation causing the huge residual in Fig. 14 does not cause a lack of fit for the *conditional* model. Our residuals have a counterpart in the conditional setting, which we shall elaborate in Baddeley and Møller (2005).

Another key issue is the role of the Poisson point process. Lawson and Diggle say that a Poisson model is reasonable for human cancer cases, conditional on any random effects. Surely this is a crucial assumption and should be open to critique. Indeed Lawson and Harrington (1996) argued that it is inappropriate, because the data often do not allow us to condition on random effects. Diggle (1990) noted that the raised incidence of laryngeal cancer near the incinerator in the Chorley–Ribble data could be attributable to clustering. In the hierarchical model that was described by Lawson, if we do not condition on the random effects, the point process is a Cox process. It is amenable to our residuals, but not to established model checking techniques for the Poisson process.

In spatial statistics, the Poisson process has enjoyed popularity out of proportion to its realism, simply because it is tractable. Now that we have some practical techniques for validating non-Poisson models, there is at least an opportunity to countenance alternative models.

Geyer comments that diagnostics are a weak way of checking models. This is a general issue on which we share the viewpoint of Cox and Snell (1981). Residuals are a weak but general tool whereas formal hypothesis testing is a stronger but more specific tool. They are used for complementary purposes in statistical modelling. Typically, if the null hypothesis is rejected, we inspect residuals to guess the type of departure from the null model. Formal methods require more restrictive assumptions and may fail if those assumptions are wrong.

Standard text-books on spatial statistics place heavy emphasis on hypothesis testing. Consequently, many users of spatial statistics in applied science will carry out a formal test of the uniform Poisson model as their first step in any analysis. This is not good statistical practice in our eyes. We are endeavouring to restore the balance, by explaining how to inspect residuals as *part of* a mature statistical analysis.

Geyer suggests that a good test of the diagnostics for interpoint interaction would be to fit a saturation model to his 'triplets' process. We shall implement this and report elsewhere.

Stoyan and Särkkä, Lawson and Schoenberg favour different definitions of residuals for spatial point processes. What convinces us that our definition is the right one? Firstly, it has a sound theoretical basis. The Georgii–Nguyen–Zessin formula (8) is essentially the only identity that is applicable to point process

models of very general kind, and hence 'must' have a key role in defining residuals. The conditional intensity has a canonical role, as also pointed out by Geyer. Secondly, our residuals can be applied to a very wide class of point process models, rather than being derived from properties of a particular model. The residuals of Lawson (1993) and Schoenberg (2003) essentially apply only to Poisson processes. Cumulative distance residuals and similar transformation techniques depend on particular properties of the model and have other weaknesses that are canvassed in Section 3. Thirdly, there is a very strong analogy between our residuals and standard residuals for generalized linear models. The new residuals of Stoyan and Särkkä are certainly inventive, but they are hardly closer to classical residuals for the linear model than our residuals, and we see no statistical theory behind them.

Stoyan and Särkkä suggest that our method does not apply to hard core processes. This is simply false; our residuals do apply to hard core processes (see Sections 6, 8.2 and 12.6.3).

Schoenberg conjectures that random thinning techniques apply to non-Poisson processes. Unfortunately they do not: the thinned process **Y** is not Poisson. If the numerator b is a constant upper bound on the conditional intensity, then using the first- and second-order versions of the Georgii–Nguyen–Zessin formula it can be shown that the pair correlation function of **Y** is

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

The expression in brackets is identically equal to 1 if and only if **X** is a Poisson process. Unless **X** is Poisson, we have $g \neq 1$ in general, and hence **Y** is not Poisson.

Waagepetersen suggests that residuals be defined by using the intensity function rather than the conditional intensity, since the intensity is simpler to evaluate for many Cox and Poisson cluster processes. This is feasible and might be valuable for assessing spatial trend. However, parameter estimates for such models are often found by the method of minimum contrast (section 10.1 in Møller and Waagepetersen (2003)); then the intensity function would be used both for parameter estimation and for checking the fitted model, and Waagepetersen's proposed residuals might be insensitive to lack of fit. We agree completely with Waagepetersen's other comments.

Zhuang describes further possibilities for application of the residuals. The apparent generalizations are in fact applications of the original definition, which are explored in Baddeley *et al.* (2006).

In response to Besag's question, our residuals are indeed the limits, under increasingly fine discretization of space, of residuals for discrete autologistic regression models. In fact, residuals can be defined in a very general point process setting, including Markov random fields, which we study in Baddeley and Møller (2005).

Lawson worries about the 'choice' of the set B in the definition of our residuals. This is a misunderstanding of the mathematical formalities. In our paper, B denotes a free variable; the residual measure R is a quantity $R(B)$ defined for *every* set B. The residual measure consists of atoms (concentrated mass) at the data points, together with a mass density in the background region. No 'choice of B' is required for this definition. Nor does the definition of the residuals involve smoothing. Fig. 6 and the top left panel of Fig. 8 are visual representations of the residuals without any smoothing nor any choice of B.

Lawson notes that a lurking variable plot (against distance from the incinerator) may not reveal a complex effect that depends both on distance and on direction. We emphasize that the lurking variable plot is a generic tool that can be applied to any spatial covariate, and not just to Euclidean distance or a Cartesian co-ordinate. For example, to investigate the effects that are mooted by Lawson, lurking variable plots may be based on polar co-ordinates, or based on an anisotropic measure of distance (e.g. where the contours of equidistance are ellipses or cones) or confined to an arbitrary subregion of the data.

Lawson and Schoenberg object to smoothing the residuals. We emphasize that smoothing is not inherent in our definition of the residuals; it is just one of several techniques proposed for *displaying* the residuals to gain insight. Smoothing does not necessarily lead to a loss of 'power'; in general, smoothing involves a trade-off between a loss of information and a reduction in variance. An analogy with smoothing the residuals of a generalized linear model is misplaced; our residuals are strongly correlated to begin with; the data points are not observations on separate experimental units.

This should not be confused with the separate issue of smoothing in the terms in the model. Our analysis of Diggle's (1990) model fitted to the Chorley–Ribble data suggests that the background intensity $\rho(u)$ has been underestimated by smoothing, as also pointed out by Diggle.

Fearnhead's point is well taken. For example, for a uniform Poisson process fitted to uniform Poisson data, the exact variance of the residual $R(B)$ is given in the three equations at the end of Section 9.3, rather than by equation (25). The correct formula produces envelopes with a shape similar to those in Fearnhead's

Fig. 23. Baddeley *et al.* (2005) gives exact formulae for the variance of residuals (i.e. including the effect of parameter estimation) both for the lurking variable plot and for its derivative (equation (35)). We hope to avoid the need for extensive simulation.

Mateu and Saura suggest tuning the residual weights, scaling them by a factor $\lambda(u, x)^{\alpha}$ where α is data dependent. This seems rather complicated; presumably the mean will not be equal to 0, and it is unclear how α should be estimated.

We agree with all Grabarnik's comments. We should like to conclude by acknowledging the crucial contributions of Grabarnik and Stoyan to this topic.

References in the discussion

Baddeley, A. J. (2000) Time-invariance estimating equations. *Bernoulli*, **6**, 783–808.

Baddeley, A. and van Lieshout, M. (1992) Object recognition using Markov spatial processes. In *Proc. 11th International Association for Pattern Recognition Int. Conf. Pattern Recognition, Los Alamitos*, pp. B136–B139. Institute of Electrical and Electronic Engineers Computer Society Press.

Baddeley, A. and Møller, J. (1989) Nearest-neighbour Markov point processes and random sets. *Int. Statist. Rev.*, **57**, 89–121.

Baddeley, A. and Møller, J. (2005) Residuals for Markov random fields. To be published.

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

Baddeley, A. J., Møller, J. and Waagepetersen, R. (2000) Non- and semi-parametric estimation of interaction in inhomogeneous point patterns. *Statist. Neerland.*, **54**, 329–350.

Baddeley, A., Møller, J. and Waagepetersen, R. (2006) Residual versions of summary functions for spatial point patterns. To be published.

Baddeley, A. J. and Silverman, B. W. (1984) A cautionary example on the use of second-order methods for analyzing point patterns. *Biometrics*, **40**, 1089–1093.

Baddeley, A. and Turner, R. (2005) Spatstat: an R package for analyzing spatial point patterns. *J. Statist. Softwr.*, **12**, 1–42.

Bartlett, M. (1964) A note on spatial pattern. *Biometrics*, **20**, 891–892.

Besag, J. E. and Clifford, P. (1989) Generalized Monte Carlo significance tests. *Biometrika*, **76**, 633–642.

Besag, J. E. and Clifford, P. (1991) Sequential Monte Carlo *p*-values. *Biometrika*, **78**, 301–304.

Besag, J. E., Milne, R. and Zachary, S. (1982) Point process limits of lattice processes. *J. Appl. Probab.*, **19**, 210–216.

Brillinger, D. R. (1997) Some wavelet analyses of point process data. *Technical Report*. University of California, Berkeley.

Comas, C. (2005) Modelling forest dynamics through the development of spatial and temporal marked point processes. *PhD Thesis*. University of Strathclyde, Glasgow.

Cox, D. and Snell, E. (1981) *Applied Statistics: Principles and Examples*. London: Chapman and Hall.

Cressie, N. A. C. (1993) *Statistics for Spatial Data*, revised edn. New York: Wiley.

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

Diggle, P. J. (1990) A point process modelling approach to raised incidence of a rare phenomenon in the vicinity of a prespecified point. *J. R. Statist. Soc.* A, **153**, 349–362.

Diggle, P. (2003) *Statistical Analysis of Spatial Point Patterns*, 2nd edn. London: Arnold.

Diggle, P. J. and Rowlingson, B. S. (1994) A conditional approach to point process modelling of elevated risk. *J. R. Statist. Soc.* A, **157**, 433–440.

Diggle, P., Zheng, P. and Durr, P. (2005) Nonparametric estimation of spatial segregation in a multivanate point process: bovine tuberculosis in Cornwall, UK. *Appl. Statist.*, **54**, 645–658.

Elliott, P., Wakefield, J., Best, N. and Briggs, D. (2000) *Spatial Epidemiology; Methods and Applications*. Oxford: Oxford University Press.

Geyer, C. (1999) Likelihood inference for spatial point processes. In *Stochastic Geometry: Likelihood and Computation* (eds O. Barndorff-Nielsen, W. Kendall and M. van Lieshout), ch. 3, pp. 79–140. Boca Raton: Chapman and Hall–CRC.

Geyer, C. and Møller, J. (1994) Simulation procedures and likelihood inference for spatial point processes. *Scand. J. Statist.*, **21**, 359–373.

Goulard, M., Särkkä, A. and Grabarnik, P. (1996) Parameter estimation for marked Gibbs point processes through the maximum pseudo-likelihood method. *Scand. J. Statist.*, **23**, 365–379.

Grabarnik, P. and Chiu, S. N. (2002) Goodness-of-fit test for complete spatial randomness against mixtures of regular and clustered spatial point processes. *Biometrika*, **89**, 411–421.

Grabarnik, P. and Särkkä, A. (2001) Interacting neighbour point processes: models for clustering. *J. Statist. Computn Simuln*, **68**, 113–126.

Kendall, W. (1990) A spatial Markov property for nearest-neighbour Markov point processes. *J. Appl. Probab.*, **28**, 767–778.

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

Lawson, A. B. (2001) *Statistical Methods in Spatial Epidemiology*. New York: Wiley.

Lawson, A. and Denison, D. (eds) (2002) *Spatial Cluster Modeling*. Boca Raton: CRC Press.

Lawson, A. and Harrington, N. (1996) The analysis of putative environmental pollution gradients in spatially correlated epidemiological data. *J. Appl. Statist.*, **23**, 301–310.

Lawson, A. B. and Williams, F. L. R. (1994) Armadale: a case-study in environmental epidemiology. *J. R. Statist. Soc.* A, **157**, 285–298.

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

Ogata, Y. (2004) Space-time model for regional seismicity and detection of crustal stress changes. *J. Geophys. Res.*, **109**, B3, B03308.

Ogata, Y., Katsura, K. and Tanemura, M. (2003) Modelling heterogeneous space–time occurrences of earthquakes and its residual analysis. *Appl. Statist.*, **52**, 499–509.

Ohser, J. and Mücklich, F. (2000) *Statistical Analysis of Microstructures in Materials Science*. New York: Wiley.

Papangelou, F. (1974) On the Palm probabilities of processes of points and processes of lines. In *Stochastic Geometry* (eds E. F. Harding and D. G. Kendall), pp. 114–147. London: Wiley.

Penrose, M. D. (2003) *Random Geometric Graphs*. Oxford: Oxford University Press.

Penrose, M. D. and Yukich, J. E. (2001) Central limit theorems for some graphs in computational geometry. *Ann. Appl. Probab.*, **11**, 1005–1041.

Penttinen, A. (1984) Modelling interaction in spatial point patterns: parameter estimation by the maximum likelihood method. *Jyväsk. Stud. Comput. Sci. Econ. Statist.*, **7**.

Renshaw, E. (2002) Two-dimensional spectral analysis for marked point processes. *Biometr. J.*, **44**, 718–745.

Renshaw, E., Mateu, J. and Saura, F. (2005) Disentangling mark/point interaction in marked point processes. To be published.

Renshaw, E. and Särkkä, A. (2001) Gibbs point processes for studying the development of spatial-temporal stochastic processes. *Computnl Statist. Data Anal.*, **36**, 85–105.

Särkkä, A. (1993) *Pseudo-likelihood Approach for Pair Potential Estimation of Gibbs Processes*. Jyväskylä: University of Jyväskylä.

Särkkä, A. and Renshaw, E. (2005) The analysis of marked point patterns evolving through space and time. To be published.

Särkkä, A. and Stoyan, D. (2005) Some residual analysis for Gibbs point process models. To be published.

Saura, F., Mateu, J. and Renshaw, E. (2006) Structure detection for one-dimensional patterns through wavelets. To be published.

Schoenberg, F. P. (2003) Multi-dimensional residual analysis of point process models for earthquake occurrences. *J. Am. Statist. Ass.*, **98**, 789–795.

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

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

Tukey, J. W. (1972) Some graphic and semigraphic displays. In *Statistical Papers in Honor of George W. Snedecor* (ed. T. A. Bancroft), pp. 293–316. Ames: Iowa State University Press.

Waagepetersen R. (2005) An estimating equation approach to inference for inhomogeneous Neyman-Scott processes. To be published.

Zhuang, J. (2005) Multi-dimensional second-order residual analysis of space-time point processes and its applications in modelling earthquake data. Submitted to *J. R. Statist. Soc.* B.

Zhuang, J., Ogata, Y. and Vere-Jones, D. (2004) Analyzing earthquake clustering features by using stochastic reconstruction. *J. Geophys. Res.*, **109**, B5, B05301.