Nonparametric estimation of the dependence of a spatial point process on spatial covariates

Adrian Baddeley*, Ya-Mei Chang, Yong Song and Rolf Turner

In the statistical analysis of spatial point patterns, it is often important to investigate whether the point pattern depends on spatial covariates. This paper describes non-parametric (kernel and local likelihood) methods for estimating the effect of spatial covariates on the point process intensity. Variance estimates and confidence intervals are provided in the case of a Poisson point process. Techniques are demonstrated with simulated examples and with applications to exploration geology and forest ecology.

1. INTRODUCTION

A common problem in the statistical analysis of spatial point patterns is to investigate the dependence of the point pattern on spatial covariates. Applications include spatial epidemiology (e.g. disease risk as a function of environmental exposure), spatial ecology (e.g. habitat preferences of organisms), exploration geology (e.g. prospectivity of mineral deposits predicted from survey data) and seismology.

Parametric models for this dependence, i.e. spatial point process models which include an effect due to spatial covariates, have been fitted to spatial point pattern data since the 1970's [1, 7, 15, 19, 26, 29, 60]. Formal hypothesis tests for the dependence of a point process on a spatial covariate function, under parametric assumptions, were developed in [11, 23, 48, 56, 66].

Nonparametric estimation of the effect of a spatial covariate on a spatial point process has received less attention until recently. An exception is the special case of *spatial relative risk* or *spatial residual risk* where the covariates are the Cartesian coordinates [12, 13, 25, 38, 43, 44] and/or the time coordinate [28, 57]. Nonparametric estimation is important here because simple parametric models are inappropriate, and the sample size is large.

*Corresponding author.

In this paper we consider spatial point process models that depend on one or more spatial covariates with continuous numerical values. We assume the point process intensity is a function of the covariates, and study nonparametric estimators of this function. Our estimators are rescaled versions of several existing kernel estimators for a probability density from biased sample data [30, 42] and their analogues using local likelihood density estimation [39, 49, 50]. Related kernel estimators were proposed in [33, 34].

Suppose the dataset is a finite set $\boldsymbol{y}$ of points in some d-dimensional space representing the locations and/or occurrence times of events. Additionally we have the values $\boldsymbol{X}(u)$ of a spatial covariate (real- or vector-valued) at every spatial location u. We model $\boldsymbol{y}$ as a realisation of a spatial point process $\boldsymbol{Y}$ (often, but not necessarily, assumed to be a Poisson process) with intensity function $\lambda(u)$ depending on $\boldsymbol{X}(u)$,

$$(1) \qquad \lambda(u) = \rho(\boldsymbol{X}(u))$$

where ρ is a function to be determined. This paper proposes nonparametric estimators of ρ.

In ecological applications where the points are the locations of individual organisms, ρ is a *resource selection function* [53] reflecting preference for particular environmental conditions $\boldsymbol{x}$. In geological applications where the points are the locations of valuable mineral deposits, ρ is an index of the *prospectivity* [14] or predicted frequency of undiscovered deposits as a function of geological and geochemical covariates $\boldsymbol{x}$.

The simplest and most popular parametric model for dependence of $\boldsymbol{Y}$ on $\boldsymbol{X}$ is the loglinear model

$$(2) \qquad \lambda(u) = \exp(\boldsymbol{\beta}^\top \boldsymbol{X}(u))$$

where $\boldsymbol{\beta}$ is a parameter vector. In applied literature there appears to be frequent confusion between the parametric loglinear model (2) and the nonparametric general relationship (1).

Figure 1 shows a motivating example in which the points are the locations of gold deposits observed in a geological survey of the Murchison region of Western Australia [68]. Covariate information includes the locations of geological faults (thick lines). The covariate of primary interest is $X(u)$, the distance from u to the nearest fault, shown in
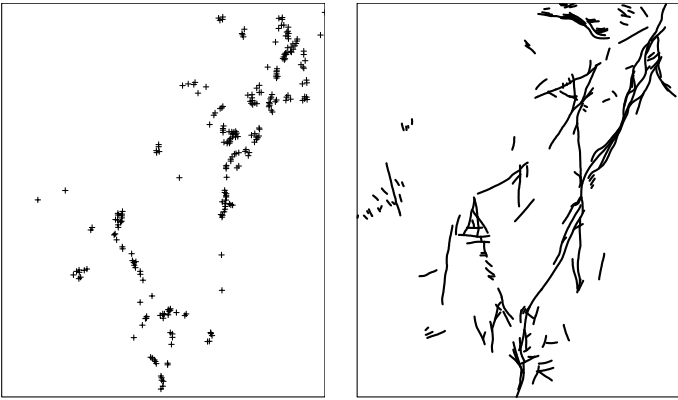
Figure 1. *Murchison data.* Left: *Points: 255 gold deposit locations in* 330×400 *km study region.* Right: *Covariate: geological faults in the same region.*
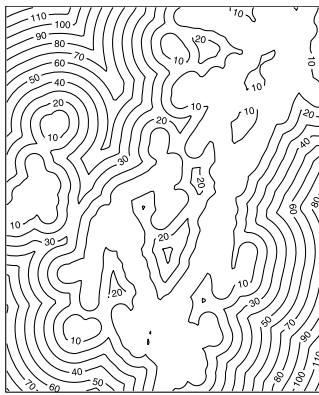

Figure 2. Contours of distance to nearest fault in the Murchison data.

Figure 2. Models of the form (1) postulate that the abundance of gold deposits depends on distance from the nearest fault. If such a model can be extrapolated to other spatial regions, then the observed pattern of faults in an exploration province can be used to identify areas of high "prospectivity" where gold deposits are more likely [11, 32, 45].

A popular technique for analysing data such as Figure 1 is pixel-based logistic regression [1]. The study region is divided into pixels; in each pixel the presence or absence of any data points is recorded; then logistic regression is used to predict the probability of the presence of a point as a function of predictor variables. Contrary to assertions in the GIS literature that logistic regression is a "nonparametric" technique (e.g. [47, p. 24]), pixel-based logistic regression is essentially equivalent to fitting a Poisson point process with intensity of loglinear form (2). See [9, 67]. True nonparametric estimation of ρ in the general form (1) is studied in this paper.

In practical applications, a nonparametric estimate of ρ may be the ultimate goal; alternatively a nonparametric es-

timate may serve as an initial guide for building parametric models (analogous to plotting p against x in binary regression [22]), or as a diagnostic for validating a parametric model.

Nonparametric estimation of ρ is closely connected to estimation of a probability density from biased sample data [30, 42]. Under regularity conditions, ρ is proportional to the ratio of two probability densities, the numerator being the density of covariate values at the points of the point process, while the denominator is the density of covariate values at random locations in space. Estimation of ρ is also related to the estimation of relative densities [37].

Estimation of a spatial point process intensity is slightly different from estimation of a probability density in several ways. First, the intensity function is not normalised (i.e. not required to integrate to unity). Second, in spatial data analysis it may be inappropriate or contentious to assume a Poisson point process (corresponding to assuming i.i.d. random variables in density estimation) so that techniques for estimation of intensity are required to remain valid for a general point process. Third, the assumption (1) may be false, and the possibility of misspecification must be considered.

Section 2 lists our technical assumptions. Section 3 derives the basic formulae for ρ. Section 4 develops kernel estimators for ρ, and confidence intervals for ρ assuming a Poisson process. Section 5 develops alternative nonparametric estimators based on local likelihood. Various extensions are discussed in Section 6. The method is tested on synthetic examples in Section 7, on the Murchison data in Section 8, and on an ecological dataset in Section 9. Section 10 describes our computational strategy and software implementation.

2. ASSUMPTIONS AND PREREQUISITES

For simplicity we assume the random points lie in d dimensional Euclidean space $\mathbb{R}^d$, where d is typically $1, 2$ or 3. However our methods apply to point processes in general spaces; the graphical methods are limited by the dimension of the values of the covariate $\boldsymbol{X}$, rather than the domain of the point process.

A spatial point pattern is a finite set $\boldsymbol{y} = \{y_1, \ldots, y_n\}$ of points observed in a sampling domain W in $\mathbb{R}^d$, where neither the number n of points nor their locations are fixed in advance. The sampling domain W is assumed to have finite positive volume. The values of the covariate function $X : W \to \mathbb{R}^m$ are assumed to be known exactly at all locations $u \in W$. In practice, the covariate values may have to be spatially interpolated from observed values at a finite set of locations u_j but we do not consider the errors arising from this step.

We assume $\boldsymbol{y}$ is a realisation of a point process $\boldsymbol{Y}$ on W with intensity function $\lambda(u)$, $u \in W$. Thus the number of random points falling in any subregion $A \subset W$ has expected value $\int_A \lambda(u) \, du$. Sometimes we may additionally assume

Y is a Poisson process. Processes with a singular intensity measure, such as earthquake epicentres concentrated along a geological fault, can be dealt with using a singular baseline measure as explained in Section 6.1.

Throughout the paper we use the following two facts. Suppose Y is *any* point process in W with intensity function $\lambda(u)$, and $h : W \to \mathbb{R}$ is a real function. Then we have *Campbell's formula*

$$(3) \qquad \mathbb{E}\left[\sum_i h(y_i)\right] = \int_W h(u)\lambda(u)\,\mathrm{d}u$$

provided $\int_W |h(u)|\lambda(u)\,\mathrm{d}u < \infty$. Additionally if Y is a *Poisson* process,

$$(4) \qquad \mathrm{var}[\sum_i h(y_i)] = \int_W h(u)^2\lambda(u)\,\mathrm{d}u$$

provided the right-hand side is finite [24, p. 188].

3. THEORY FOR A REAL COVARIATE

In this section we consider a single, real-valued covariate function X on W, i.e. a function $X : W \to \mathbb{R}$. Let

$$(5) \qquad G(x) = \frac{1}{|W|} \int_W \mathbf{1}\{X(u) \leq x\}\,\mathrm{d}u$$

be the spatial cumulative distribution function of X, where $|W|$ denotes the d-dimensional volume of the window W. Equivalently $G(x) = \mathbb{P}\{X(U) \leq x\}$ is the c.d.f. of the value $X(U)$ at a uniformly distributed random point U in W. Assume G has derivative g; a sufficient condition is that X be differentiable with nonzero gradient (see Appendix A.1). It will be convenient to consider the unnormalised versions $G^*(x) = |W|G(x)$ and $g^*(x) = |W|g(x)$.

We assume G^* is known exactly, or to a very high accuracy. This is true when X is a spatial coordinate, or when X values are known on a very fine pixel grid, and G^* is well approximated numerically.

Suppose Y is a point process in W with intensity function of the form (1) for some function ρ. Then [24, p. 22], [59, p. 17] the values $x_i = X(y_i)$ constitute a point process on $\mathbb{R}$ with intensity function

$$(6) \qquad f^*(x) = \rho(x)g^*(x), \quad x \in \mathbb{R}.$$

This is the pivotal relationship between the observations x_i and the target function ρ. The expected number of points in Y is

$$\mu = \int_W \lambda(u)\,\mathrm{d}u = \int_W \rho(X(u))\,\mathrm{d}u = \int_{-\infty}^{\infty} \rho(x)g^*(x)\,\mathrm{d}x.$$

Conditional on the number n of data points, the values $x_1, \ldots, x_n$ are exchangeable, with marginal probability density $f(x) = f^*(x)/\mu$. If Y is a Poisson process on W, then

the values x_i constitute a Poisson process on $\mathbb{R}$, and are conditionally i.i.d. random variables given n. Hence

$$(7) \qquad \rho(x) = \frac{f^*(x)}{g^*(x)} = \kappa\,\frac{f(x)}{g(x)}$$

where $\kappa = \mu/|W|$ is the average intensity. Apart from a scale factor, ρ is a relative probability density for the distribution of the values $x_i = X(y_i)$ relative to the distribution that would be obtained if the process Y had constant intensity on W. For practical interpretation of results, it is important that the values of $\rho(x)$ are intensities, expressed in units with dimension length^{-d}.

If (1) does *not* hold, the formulae in this section remain true when $\rho(x)$ is replaced by $\overline{\rho}(x)$, a weighted average value of $\lambda(u)$ over the contour $\{u \in W : X(u) = x\}$. See Appendix A.2. This issue of misspecification does not arise for density estimation in one-dimensional space.

4. KERNEL ESTIMATORS OF ρ

Kernel smoothing is the simplest nonparametric method in this context, with advantages that include theoretical tractability, superior computational speed and reliability, and disadvantages including bias. In large datasets, rapid computation is important and the bias of the kernel estimator is tolerable for appropriate choices of bandwidth. Kernel estimation of the intensity of a Poisson process is discussed in [46, section 6.2, pp 236–250] including the optimal rate of convergence.

Equation (6) shows that estimation of ρ is closely related to estimating a probability density from a biased sample. Jones [42] described two kernel estimators for this problem, and El Barmi and Simonoff [30] a third kernel estimator based on the probability integral transformation.

In our context the analogues of Jones' estimators of ρ from the observed values $x_i = X(y_i)$ are the "ratio" form

$$(8) \qquad \hat{\rho}_{\mathrm{R}}(x) = \frac{1}{g^*(x)} \sum_i k(x_i - x)$$

and the "reweighted" form

$$(9) \qquad \hat{\rho}_{\mathrm{W}}(x) = \sum_i \frac{1}{g^*(x_i)} k(x_i - x)$$

while the analogue of El Barmi and Simonoff's "transformation" estimator is

$$(10) \qquad \hat{\rho}_{\mathrm{T}}(x) = \frac{1}{|W|} \sum_i k(G(x_i) - G(x)).$$

Here k is a smoothing kernel on the real line, and again G is the spatial c.d.f. (5).

Guan [33] proposed a kernel estimator that is similar to the ratio form (8). It is discussed below.

The rationale for the ratio form (8) is the plug-in principle applied to (7), since

$$(11) \qquad \hat{\rho}_{\mathrm{R}}(x) = \frac{\widehat{f^*}(x)}{g^*(x)} = \hat{\kappa}\,\frac{\hat{f}(x)}{g(x)}$$

where $\hat{\kappa} = n/|W|$ is the usual unbiased estimator of κ (and the MLE if $\boldsymbol{Y}$ is Poisson),

$$(12) \qquad \hat{f}(x) = \frac{1}{n}\sum_i k(x_i - x)$$

is the usual fixed-bandwidth kernel estimator of the density f, and $\widehat{f^*}(x) = n\hat{f}(x)$ is the corresponding unnormalised kernel estimator of $f^*(x)$. The rationale for the reweighted form (9) is that the random measure with masses $1/g^*(x_i)$ at the points x_i has intensity $\rho(x)$, by an application of (3). The transformation estimator (10) is justified by the fact that the values $t_i = G(x_i)$ have intensity $q(t) = |W|\,\rho(G^{-1}(t))$ on $[0,1]$ (see Appendix A.4).

In the context of density estimation, Jones [42] showed that neither the ratio nor the reweighting estimator is uniformly optimal, but that the reweighting estimator has better performance overall. A similar statement is likely to hold in this context. The transformation estimator $\hat{\rho}_{\mathrm{T}}$ is a simple, fast and appropriate form of variable-bandwidth smoothing insofar as it depends on the covariate values but is otherwise not adaptive to features of the point process. It is likely to improve the accuracy of $\hat{\rho}(x)$ for values x where data are scarce, and may also improve bandwidth selection. The estimator $\hat{\rho}_{\mathrm{T}}$ could also be edge-corrected at the endpoints 0 and 1 using standard techniques.

Computational implementation and performance are discussed in Section 10.

Assuming (1) holds, the expectation of $\hat{\rho}_{\mathrm{R}}(x)$ is, by (3),

$$(13) \qquad \mathbb{E}[\hat{\rho}_{\mathrm{R}}(x)] = \frac{1}{g^*(x)}\int_W k(X(u)-x)\lambda(u)\,\mathrm{d}u$$
$$= \int_{-\infty}^{\infty} k(x'-x)\rho(x')\frac{g^*(x')}{g^*(x)}\,\mathrm{d}x'$$

where the last expression is obtained by a change of variables from u to $x' = X(u)$. Similarly the reweighting and transformation estimators have expectation

$$(14) \qquad \mathbb{E}[\hat{\rho}_{\mathrm{W}}(x)] = \int_{-\infty}^{\infty} k(x'-x)\rho(x')\,\mathrm{d}x'$$

$$(15) \qquad \mathbb{E}[\hat{\rho}_{\mathrm{T}}(x)] = \int_0^1 k(t-G(x))\rho(G^{-1}(t))\,\mathrm{d}t.$$

The kernel estimators are biased, as usual (e.g. [17]) with bias controlled by the kernel bandwidth and the regularity of ρ and g^* (with the exception that the regularity of g^* does not affect the bias of $\hat{\rho}_{\mathrm{W}}$). A proof for $\hat{\rho}_{\mathrm{R}}$ is given in Appendix A.3.

Guan [33] proposed a kernel estimator that is similar to the ratio form (8) except that $g^*(x)$ is replaced by $\int_W k(\boldsymbol{X}(u)-x)\,\mathrm{d}u$, a kernel smoothed counterpart of $g^*(x)$, using the same kernel k for the numerator $f^*(x)$ and denominator $g^*(x)$. However, the numerator and denominator are based on datasets of different sizes, and this strategy will typically lead to over-smoothing of the denominator $g^*(x)$. Calculations similar to those in Appendix A.3 show that Guan's estimator typically has greater bias than $\hat{\rho}_{\mathrm{R}}(x)$.

Bandwidth selection can be performed using existing methods for bandwidth selection in density estimation, because of the close connection between the two problems. Silverman's rule of thumb using the fifth root of the number of points [64, eq. (3.31), p. 48] performed well in our examples. Cross-validation methods [61, 63] tended to produce unacceptably small bandwidths.

To construct approximate confidence intervals, we assume additionally that $\boldsymbol{Y}$ is a Poisson process. Since the values x_i constitute a Poisson process with intensity f^*, the variance of $\hat{\rho}_{\mathrm{R}}(x)$ for fixed x is, by (4),

$$(16) \quad \mathsf{var}[\hat{\rho}_{\mathrm{R}}(x)] = g^*(x)^{-2}\int_W k(x-X(u))^2\lambda(u)\,\mathrm{d}u$$
$$= g^*(x)^{-2}\int_{-\infty}^{\infty} k(x-x')^2\rho(x')g^*(x')\,\mathrm{d}x'$$

Similarly

$$(17) \qquad \mathsf{var}[\hat{\rho}_{\mathrm{W}}(x)] = \int_{-\infty}^{\infty} k(x-x')^2\frac{\rho(x')}{g^*(x')}\,\mathrm{d}x'$$

$$(18) \qquad \mathsf{var}[\hat{\rho}_{\mathrm{T}}(x)] = \frac{1}{|W|}\int_0^1 k(t-G(x))^2\rho(G^{-1}(t))\,\mathrm{d}t.$$

In particular if k is the Gaussian kernel $k(v) = k_\sigma(v) = (\sigma\sqrt{2\pi})^{-1}\exp(-v^2/(2\sigma^2))$ then $k_\sigma(v)^2 = \alpha k_\tau(v)$ where $\tau = \sigma/\sqrt{2}$ and $\alpha = 1/(2\sigma\sqrt{\pi}) = 1/(2\tau\sqrt{2\pi})$, so that pointwise unbiased estimators of the variances (16)–(18) are, by (3),

$$(19) \qquad \hat{v}_R(x) = \alpha g^*(x)^{-2}\sum_i k_\tau(x_i - x)$$

$$(20) \qquad \hat{v}_W(x) = \alpha\sum_i \frac{k_\tau(x_i - x)}{g^*(x_i)^2}$$

$$(21) \qquad \hat{v}_T(x) = \alpha|W|^{-2}\sum_i k_\tau(t_i - G(x))$$

where again $x_i = X(y_i)$ and $t_i = G(x_i)$. This calculation assumes g^* (or equivalently G) is fixed and known.

For small bandwidths the variances of the three kernel estimators behave as

$$(22) \qquad \mathsf{var}[\hat{\rho}(x)] \sim \frac{\rho(x)}{g^*(x)}$$

as intuitively expected: $g^*(x)$ plays the role of the "sample size" for estimation of $\rho(x)$ at small bandwidths. Estimator

variance $\mathsf{var}[\hat{\rho}(x)]$ decreases with sample size $g^*(x)$, but increases with intensity $\rho(x)$, due to the variance properties of the Poisson process.

The kernel estimators (8)–(10) of $\rho(x)$ are asymptotically normal under large-sample conditions, e.g. [46, p. 240]. However, confidence intervals for $\rho(x)$ with good small-sample properties are difficult to construct, because the natural estimators of variance (19)–(21) have strong positive correlation with the estimators $\hat{\rho}(x)$ themselves. This parallels the well-known difficulty in constructing confidence intervals for a probability density based on kernel estimates [16, 35, 36, 40].

In density estimation, the theoretically optimal bandwidth for constructing confidence intervals is typically smaller than the optimal bandwidth for point estimation. However, this insight did not deliver any practical benefit in our experiments, because confidence bands obtained using smaller bandwidths were typically too irregular. Accordingly we propose constructing confidence intervals in the naive form $\hat{\rho}(x) \pm z\sqrt{\hat{v}(x)}$ from the estimators (8)–(10) and variance estimators (19)–(21), where z is the $100(1 - \alpha/2)\%$ critical value of the standard normal distribution.

The assumption of a Poisson point process is not essential. For a general point process, the variance of the kernel estimators can be expressed in terms of the intensity and pair correlation function of the point process, using the second order Campbell formula, e.g. [6, 33]. If the point process is regular (negatively associated) then the Poisson assumption leads to overestimates of the variance of $\hat{\rho}$, and conservative confidence intervals. Concern arises when the point process is believed to be clustered (positively associated) so that the variance is underestimated by (19)–(21). In this case, a more accurate variance estimate could be obtained by estimating the pair correlation function, provided we are willing to impose additional model assumptions, such as second-order reweighted stationarity [5] or a Cox process model [54]. However, explicit model assumptions do not sit well with the nonparametric approach, and may be difficult to verify, especially in the case of clustered patterns [10]. Alternatively, bootstrap confidence intervals might be obtained by spatial resampling [52]. Further discussion is beyond the scope of the present paper.

5. ALTERNATIVE NONPARAMETRIC ESTIMATORS

Alternatives to kernel estimation of ρ include spline smoothing, locally weighted regression [18] and local likelihood density estimation [39, 49, 50].

The appropriate counterparts of the kernel estimators (8)–(10) can be determined by expressing these estimators as multiples of the fixed-bandwidth kernel density estimator $\hat{f}(x) = (1/n) \sum_i k(x_i - x)$, and then replacing $\hat{f}$ by another nonparametric density estimator.

Let $\tilde{f}(x \mid x_1, \ldots, x_n)$ denote a nonparametric estimator of the probability density based on observations x_i, and $\tilde{f}(x \mid x_1, \ldots, x_n; w_1, \ldots, w_n)$ the corresponding estimator when observations x_i have prior weights w_i. Then the counterparts of the estimators (8)–(10) are

$$(23) \qquad \widetilde{\rho}_{\mathrm{R}}(x) = \frac{\hat{\kappa}}{g(x)} \, \widetilde{f}(x \mid x_1, \ldots, x_n)$$

$$(24) \qquad \widetilde{\rho}_{\mathrm{W}}(x) = \left(\sum_i w_i\right) \widetilde{f}(x \mid x_1, \ldots, x_n; w_1, \ldots, w_n)$$

$$(25) \qquad \widetilde{\rho}_{\mathrm{T}}(x) = \hat{\kappa} \, \widetilde{f}(G(x) \mid G(x_1), \ldots, G(x_n))$$

where $w_i = 1/g^*(x_i)$ and $\hat{\kappa} = n/|W|$.

The variance of the estimators (23)–(25) depends on the choice of nonparametric density estimator $\widetilde{f}$. Here we consider the case of local likelihood density estimation [39, 49, 50] and assume $\boldsymbol{Y}$ is a Poisson process.

The local likelihood density estimator $\widetilde{f}$ is asymptotically normal in large samples. A normal approximation to $\log \widetilde{f}$ is more accurate and more natural. Estimators of the variance of $\log \widetilde{f}$ are implemented in open-source software such as the `locfit` package [51].

Noting that all the estimators $\widetilde{\rho}$ are of the form $\widetilde{\rho}(x) = M\widetilde{f}(x)$, we have

$$\mathrm{var}\log\widetilde{\rho}(x) = \mathrm{var}\log M + \mathrm{var}\log\widetilde{f}(x) + 2\mathrm{cov}(\log M, \log\widetilde{f}(x)).$$

Using the delta method we may approximate $\mathrm{var}\log M \approx (\mathrm{var}M)/(\mathbb{E}M)^2$ and

$$\mathrm{cov}(\log M, \log\widetilde{f}(x)) \approx \frac{\mathrm{cov}(M, \widetilde{f}(x))}{\mathbb{E}M\,\mathbb{E}\widetilde{f}} \approx \frac{\mathbb{E}(\widetilde{\rho}(x) - \rho(x))}{\rho(x)},$$

the relative bias of $\widetilde{\rho}(x)$, essentially equivalent to the relative bias of $\widetilde{f}(x)$. Estimates of this quantity are available from the local likelihood procedure.

For the ratio form (23) the scale factor is $M = \hat{\kappa}/g(x) = n/g^*(x)$ so that $\mathrm{var}\log M = \mathrm{var}\log N \approx \mathbb{E}(N)/(\mathbb{E}(N))^2 = 1/\mathbb{E}(N)$.

For the reweighting estimator (24) we have $M = \sum_i 1/g^*(x_i)$ so that

$$\mathbb{E}M = \int_W \frac{1}{g^*(X(u))}\lambda(u)du$$
$$= \int \frac{1}{g^*(x)}\rho(x)g^*(x)dx = \int \rho(x)dx$$

and

$$\mathrm{var}M = \int_W \frac{1}{g^*(X(u))^2}\lambda(u)du$$
$$= \int \frac{1}{g^*(x)^2}\rho(x)g^*(x)dx = \int \frac{\rho(x)}{g^*(x)}dx.$$

For the transformation estimator (25) we have $M = \hat{\kappa}$ so again $\mathrm{var}\log M = \mathrm{var}\log N \approx \mathbb{E}(N)/(\mathbb{E}(N))^2 = 1/\mathbb{E}(N)$.

These calculations yield approximations to the variance of $\log\widetilde{\rho}(x)$, which can be used to construct asymptotically

valid confidence intervals for $\rho(x)$ based on the asymptotic normal distribution for $\log \widetilde{\rho}(x)$. The finite-sample properties of these confidence intervals are likely to be better than those of the confidence intervals for kernel estimates described in the previous section, because of the abovementioned problems with variance estimation for kernel smoothers.

6. EXTENSIONS

6.1 Relative risk model

Still assuming a real-valued covariate X, consider the generalization of (1) to

$$(26) \qquad \lambda(u) = \rho(X(u))B(u)$$

where $B(u)$ is a fixed and known function, serving as a baseline. For example, in applications to spatial epidemiology, $B(u)$ could be the spatially-varying density of the underlying population of susceptible individuals, and $\rho(X(u))$ the relative risk of the events represented by the point process, expressed as a probability or rate relative to the susceptible population, and depending on an explanatory "risk factor" X [26, 27]. Alternatively $B(u)$ may be an artificial baseline which serves to stabilise estimator variability. The theory of Section 3 can be adapted to this model, yielding the basic identity

$$(27) \qquad \rho(x) = \frac{f^*(x)}{g_B^*(x)} = \kappa_B \frac{f(x)}{g_B(x)}$$

where g_B^* is the derivative of the B-weighted spatial c.d.f. of X

$$(28) \qquad G_B^*(x) = \int_W \mathbf{1}\{X(u) \le x\}B(u)\,\mathrm{d}u$$

and $\kappa_B = \mu / \int_W B(u)\,\mathrm{d}u$ is the usual unbiased estimator (and MLE in the Poisson case) of the constant κ in the constant relative risk model $\lambda(u) = \kappa B(u)$.

The estimators proposed in Sections 4 and 5 can be applied simply by replacing g, g^*, G, G^* by their B-weighted counterparts g_B, g_B^*, G_B, G_B^* respectively. Properties of $\hat{\rho}$ described in Sections 4 and 5 and Appendix A.3 also apply *mutatis mutandis*. Thus, for example, the B-weighted analogue of (8) is

$$(29) \qquad \hat{\rho}_{\mathrm{R,B}}(x) = \frac{1}{g_B^*(x)} \sum_i k(x_i - x).$$

The practical interpretation of the relative risk (or "residual risk") term $\rho(x)$ in this Section is slightly different from that of the absolute intensity $\rho(x)$ in previous Sections. Typically the baseline function $B(u)$ is a valid intensity function. In that case, the relative risk $\rho(x)$ is dimensionless, and the constant value $\rho(x) \equiv 1$ corresponds to the baseline or null model $\lambda(u) = B(u)$.

In many applications, the baseline $B(u)$ would also be estimated from data. Variance estimation and interval estimation of $\rho(x)$ then depend on the distribution of the estimator $\hat{B}(u)$ and on the joint distribution of the observations x_i with $\hat{B}(\cdot)$. Examples of this analysis arise in case-control studies in spatial epidemiology [38, 43, 44].

Alternatively the baseline intensity function $B(u)$ may be replaced by a baseline intensity *measure*. Then (28) is replaced by an integral with respect to the baseline measure, and similarly for g_B, g_B^*, G_B. This accommodates situations where the point process intensity is singular, for example, where points are concentrated on a curve.

6.2 Vector covariate

Now consider an m-dimensional vector-valued covariate function $\boldsymbol{X}(u) = (X_1(u), \ldots, X_m(u))$. Multivariate density estimation from weighted samples was discussed in [2].

The approach described above can be applied provided the spatial distribution of $\boldsymbol{X}$ is absolutely continuous on $\mathbb{R}^m$. The counterpart of G^* is the unnormalised joint c.d.f.

$$G^*(x_1, \ldots, x_m) = \int_W \mathbf{1}\{X_1(u) \le x_1, \ldots, X_m(u) \le x_m\}\,\mathrm{d}u.$$

The ratio estimator (8) and reweighting estimator (9), and their counterparts (23)–(24) for relative/residual risk (Section 6.1), can be applied directly to an m-dimensional vector covariate function $\boldsymbol{X}$ using an m-dimensional kernel k. The variance formulae (16)–(17) generalise immediately by replacing one-dimensional by m-dimensional integration, since they are derived from (4).

Absolute continuity of the distribution of $\boldsymbol{X}$ implies, in particular, that the image $\boldsymbol{X}(W) = \{\boldsymbol{X}(u) : u \in W\}$ must not be contained in any lower-dimensional subset of $\mathbb{R}^m$. This excludes algebraically dependent covariates such as polynomials (i.e. where $X_i(u) = Z(u)^i$ is the ith term in a polynomial in a real-valued covariate Z).

Polynomials are a special case of the separable model

$$(30) \qquad \lambda(u) = \rho_1(X_1(u)) \ldots \rho_m(X_m(u))A(u)$$

where ρ_j are functions to be estimated, and $A(u)$ is a known function serving as a baseline. When attention is focused on one of the functions ρ_j, holding others fixed, (30) collapses to a model of the form (26). The technique of section 6.1 applies, and ρ_j can be estimated by the analogue of (29). The appropriate variance calculations depend on the method used to estimate the reference function B.

If covariates are high-dimensional and if prior information is lacking, it becomes important to reduce dimensionality. Dimension reduction techniques for spatial point processes were proposed in [34].

7. SIMULATED EXAMPLES

Figure 3 shows simulated realisations of two Poisson point processes, with intensities

$$\lambda(u) = \exp(\beta_0 + \beta_1 d_R(u)) \tag{31}$$

$$\lambda(u) = \exp(\beta_0 + \beta_2 d_R(u)^2) \tag{32}$$

respectively, where $(\beta_0, \beta_1) = (5, -3)$ for the left panel and $(\beta_0, \beta_2) = (5, -4)$ for the right panel. Here $d_R(u)$ denotes the shortest distance from a given location u to the set of lines arranged in the shape of the letter "R" shown in the Figure. Both models (31) and (32) are of the general form (1) with covariate $\boldsymbol{X} = d_R$ and covariate effects $\rho(x) = \exp(\beta_0 + \beta_1 x)$ and $\rho(x) = \exp(\beta_0 + \beta_1 x^2)$ respectively.
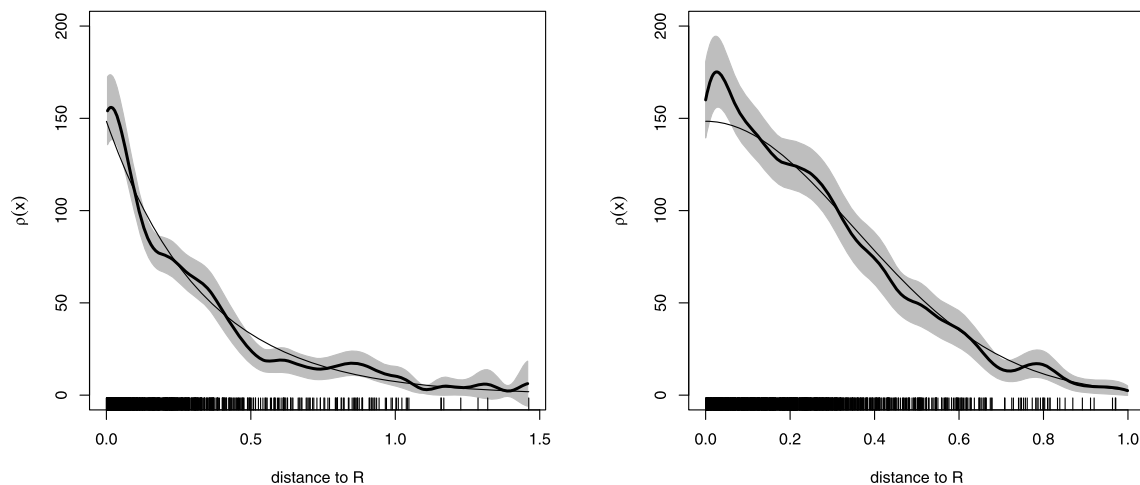
Figure 4 shows the ratio-form kernel estimates $\hat{\rho}_R$ as a function of the covariate $\boldsymbol{X}(u) = d_R(u)$ for these two datasets, calculated by (8), with ± 2 standard deviation (nominally 95% pointwise confidence) interval calculated from the variance estimate (19). The true functions ρ are also shown; the estimates of ρ appear to be quite accurate. Rug plots [65] show the values x_i.

The estimated standard deviation in Figure 4 is roughly constant as a function of distance x (apart from erratic behaviour at large x due to the paucity of data). This is accidental; in the present example, ρ and g are both decreasing functions of x with roughly similar shape, so that $\mathsf{var}[\hat{\rho}(x)]$ does not vary greatly, by (22).

Figure 5 shows all six estimates (8)–(10) and (23)–(25) for the data in the right panel of Figure 3. The ratio and reweighting estimates are very similar here, while the transformation estimate is slightly different. For small x values, the kernel estimates (left column) are overestimates of ρ, while the local likelihood estimates (right column) are generally more accurate. Note that the confidence bands for the kernel estimates (left column) are pointwise 95% confidence limits, but the confidence bands for the local likelihood estimates (right column) are *simultaneous* 95% confidence limits, due to a technical limitation of the `locfit` package.

For comparison, Figure 6 shows the maximum likelihood estimate of ρ for a model of the correct parametric form (32) fitted to the data in the right panel of Figure 3, together with the asymptotic (likelihood based) pointwise 95% confidence bands.

In other experiments we have found that the results obtained using the ratio, reweighting and transformation methods are generally similar, except that the ratio method is more sensitive to smoothing of g.

Figure 3. *Simulated data.* Left: *intensity a loglinear function of distance to letter R.* Right: *intensity a log-quadratic function of distance to letter R.*

Figure 4. Kernel estimates of ρ (thick solid lines) for the simulated point pattern datasets in Figure 3 using the ratio form (8), with pointwise two-standard-deviation confidence limits (grey shading). Correct function ρ also shown (thin solid lines). Left: *left panel of Figure 3.* Right: *right panel of Figure 3.*
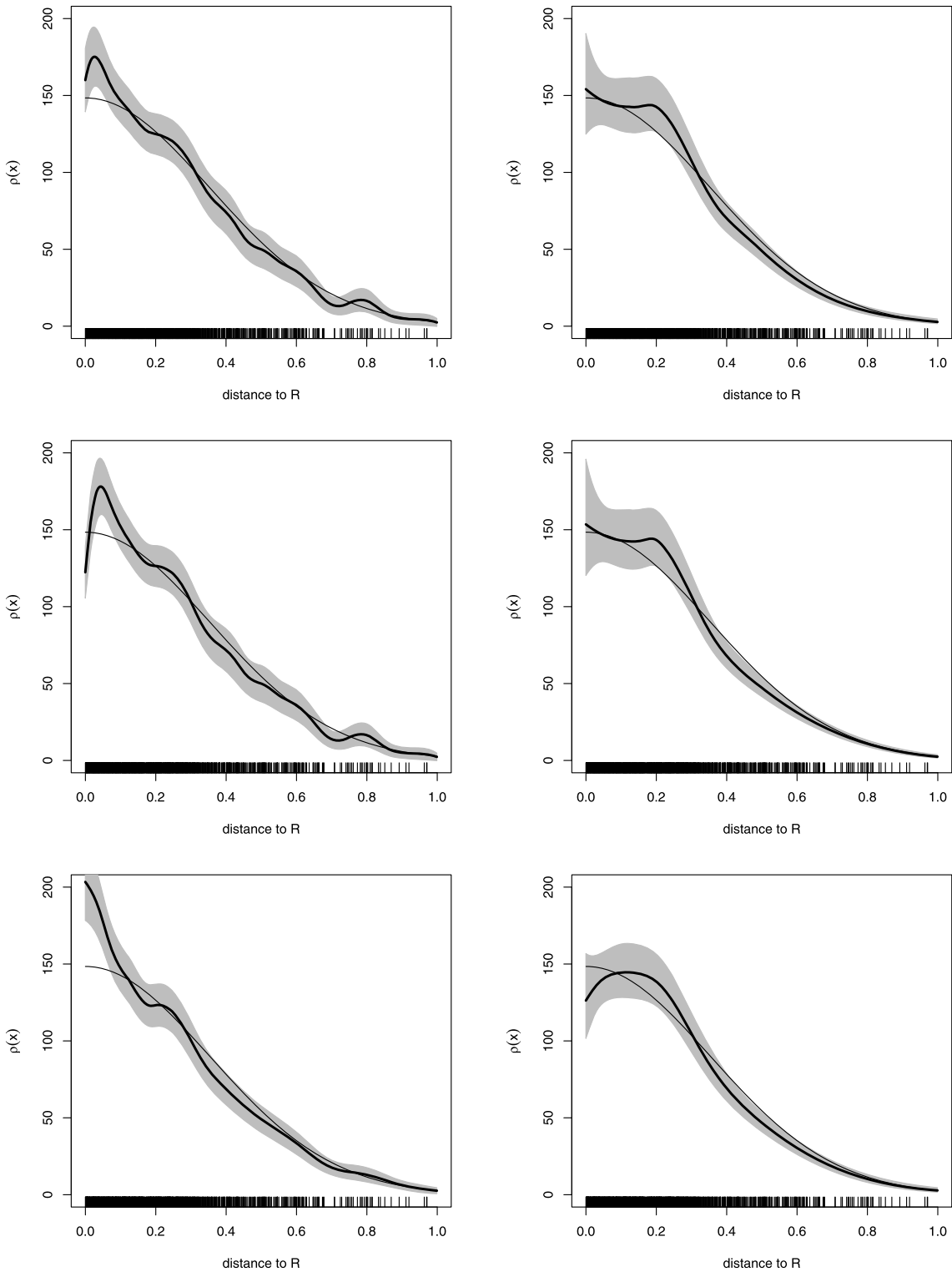
Figure 5. *Six estimates of ρ for the simulated point pattern dataset in the right panel of Figure 3.* Left column: *kernel estimates.* Right column: *local likelihood estimates.* Top row: *ratio method.* Middle row: *reweighting method.* Bottom row: *transformation method. Estimate ρ̂ (thick solid lines), pointwise or simultaneous 95% confidence limits (grey shading), correct function ρ (thin solid lines).*
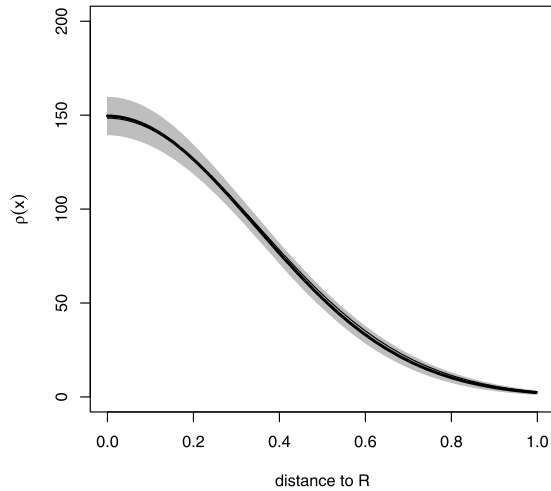
Figure 6. Maximum likelihood estimate of ρ for the right panel of Figure 3 fitted using the correct parametric form. Estimate ρ̂ (thick solid lines), pointwise two-standard-deviation confidence limits (grey shading), correct function ρ (thin solid lines).

8. MURCHISON DATA

The Murchison data were introduced in Figure 1. The covariate of primary interest is $d(u) = $ distance from u to the nearest geological fault, shown in the right panel of Figure 1.

To deal with edge effects, we restrict attention to the subset $W_\ominus = \{u \in W : d(u) \leq b(u)\}$ of the study region W, where $b(u)$ denotes the shortest distance from u to the boundary of W. This subset is shown in Figure 7. By the "local knowledge principle" [62, pp. 11,49,62], [3] $W_\ominus$ is the largest set on which the value of $d(u)$ is computable from observations inside W.

Figure 8 shows the six estimates $\hat\rho$ for the Murchison gold data of Figure 1 (in the restricted set $W_\ominus$) as a function of distance from the nearest fault. The estimated standard deviation of $\hat\rho$ is a generally decreasing function of distance, because the density g of the spatial distribution of the distance covariate is decreasing.

Figure 8 suggests ρ is approximately an exponential function, except for a possible bump at distance $x = 5$ km. This impression is confirmed when the y axis is plotted on a log scale, shown in the left panel of Figure 9. If nonparametric estimation is used as a preliminary step before choosing an appropriate parametric model, then Figure 8 would support the initial choice of a loglinear model

$$(33) \qquad \lambda(u) = \exp(\alpha + \beta d(u))$$

where α, β are parameters to be fitted. This is equivalent to pixel-based logistic regression of the probability of the presence of a gold deposit against distance to nearest fault [1, 9, 67]. We estimated the parameters α, β by maximum likelihood [9].
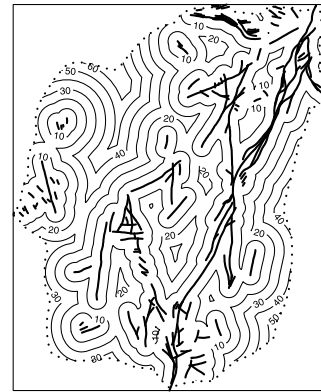

Figure 7. Modification of covariate to avoid edge effects. Contours of distance to nearest fault (thin solid lines) shown within the restricted domain $W_\ominus$ (dotted lines).

Next we apply the techniques for vector-valued covariates given in Section 6.2. Relative to the loglinear model (33), Figure 10 shows the relative risk estimate $\hat\rho$ as described in Section 6.2, using the reweighting method, for the vector covariate $(Z_1(u), Z_2(u))$, where $Z_1(u) = X(u)$ is the distance to the nearest fault, and $Z_2(u)$ is the geographic northing (Cartesian y-coordinate). This suggests that the fitted loglinear relationship (33) gives an underestimate of the true abundance for small distances, and an overestimate for large distances.

Figure 11 shows the spatial relative risk estimate $\hat\rho$ where Z_1, Z_2 are the geographic easting and northing, corresponding to Cartesian x and y-coordinates respectively. This is an estimate of the familiar spatially-varying residual risk relative to the fitted model (33). This shows a clear departure from the loglinear model (33). In the northeast corner, there is an overdensity of gold relative to the loglinear model (i.e. the model underestimates the true abundance of gold deposits) while the northwest and southeast corners show an underdensity (the model overestimates the true abundance). This strongly suggests that other spatial covariates need to be taken into account: these might be geostructural or geochemical observations, or information about the survey effort.

A perfunctory analysis of the Murchison data using pixel-based logistic regression, equivalently fitting a Poisson process with loglinear intensity (33), would probably have failed to identify key features of these data that were detected using nonparametric estimation.

9. BEILSCHMIEDIA DATA

The top panel of Figure 12 shows the positions of 3605 trees of the species *Beilschmiedia pendula* (Lauraceae) in a $1,000 \times 500$ metre rectangular sampling region in a tropical rainforest. The other panels show the terrain elevation
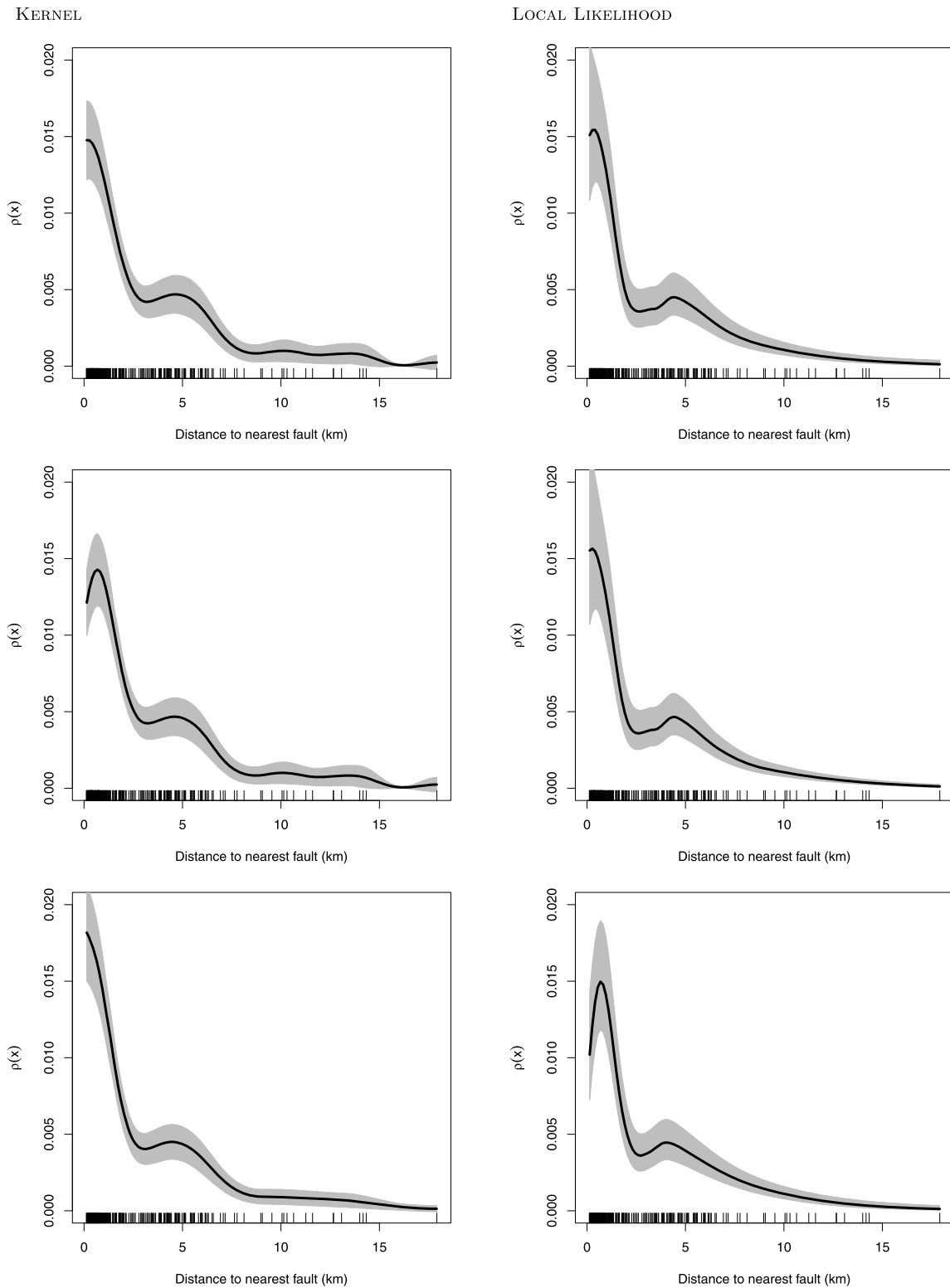
Figure 8. *Estimates of ρ for Murchison gold deposits (Figure 1) as a function of distance to nearest fault.* Left column: *kernel estimates;* Right column: *local likelihood estimates.* Top: *ratio estimator;* Middle: *weighted estimator;* Bottom: *transformation estimator. Solid lines are estimates of ρ. Grey shading indicates* ±2 *standard deviation (nominally 95% pointwise confidence) intervals.*
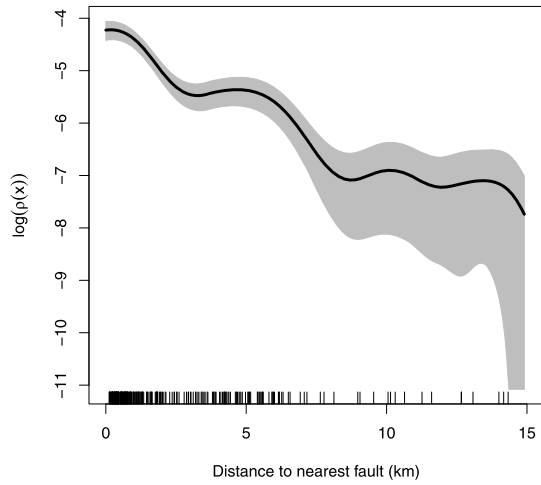
Figure 9. Log-transformed kernel estimate $\log \hat{\rho}$ *for Murchison gold deposits (Figure 1) as a function of distance to the nearest fault. Equivalent to top left panel of Figure 8.*
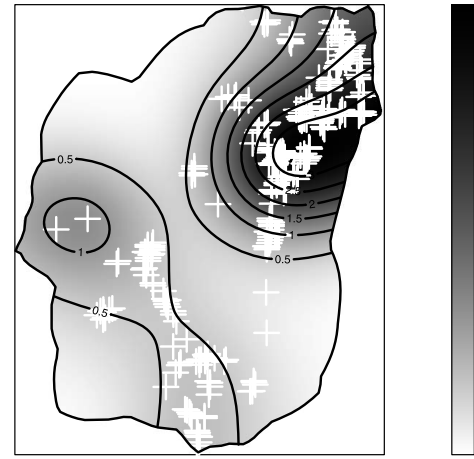

Figure 11. Spatial relative/residual risk estimate $\hat{\rho}(x, y)$ *for the Murchison data relative to the fitted loglinear intensity model (33). Crosses mark the gold locations* y_i.
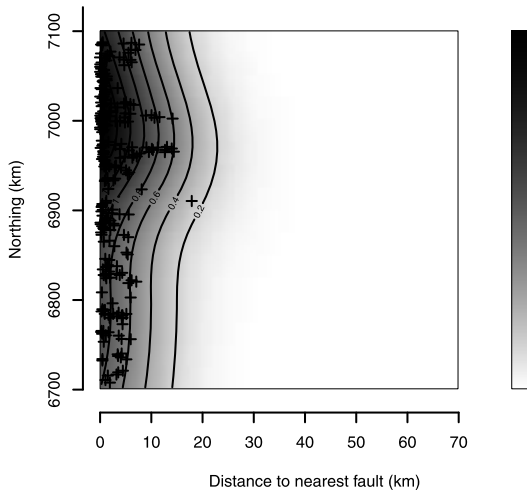

Figure 10. Bivariate relative/residual risk estimate $\hat{\rho}(Z_1, Z_2)$ *for the Murchison data relative to the fitted loglinear intensity model (33).* Z_1 *is distance to the nearest fault,* Z_2 *is geographic northing. Crosses mark the observed values* $(Z_1(y_i), Z_2(y_i))$ *at the gold locations* y_i.

Figure 12. Beilschmiedia data. Top: *positions of 3,605 trees.* Middle: *terrain elevation.* Bottom: *terrain slope. Reproduced by kind permission of S.P. Hubble, R.B. Foster and R. Condit. Data kindly supplied by R.P. Waagepetersen.*

and terrain slope for the same sampling region. This survey forms part of a much larger study of forest dynamics on Barro Colorado Island in the Panama Canal [20, 21, 41]. We thank the study authors for releasing this portion of the data. The data in Figure 12 have been analysed in [55].

The left panel of Figure 13 shows a kernel estimate of intensity as a function of terrain elevation alone. More precisely, writing $e(u)$ and $s(u)$ for the terrain elevation and slope respectively, the left and right panels of Figure 13 show
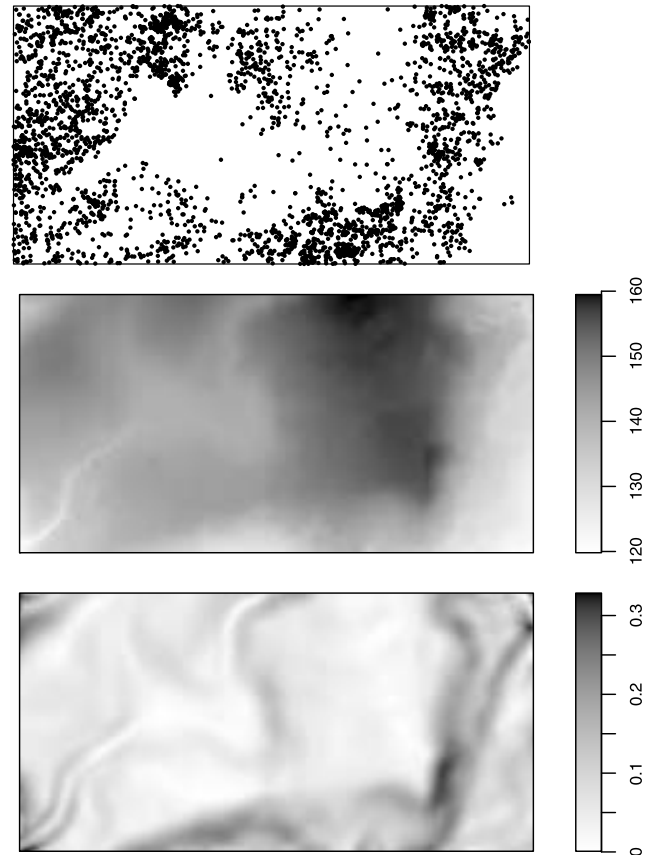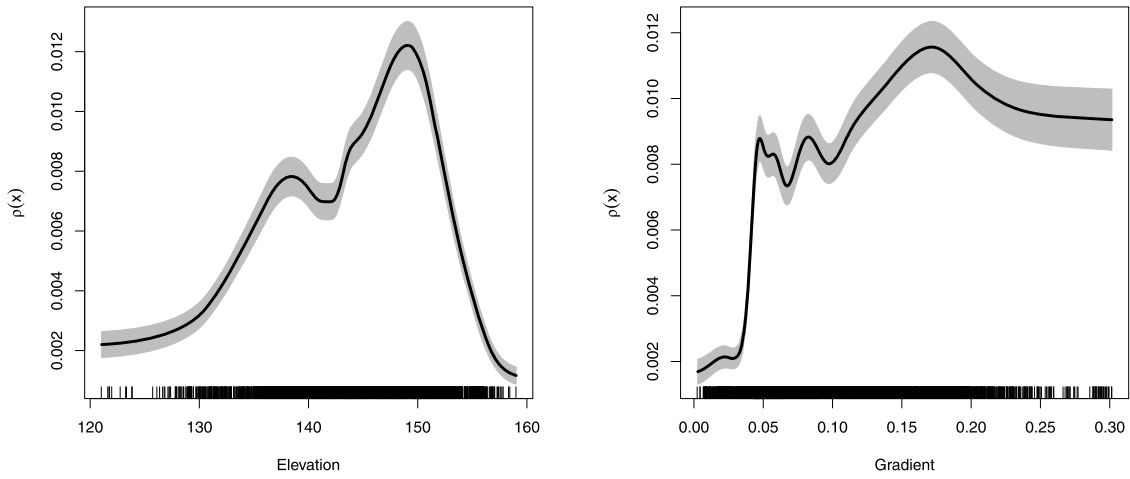
Figure 13. *Estimates of* B. pendula *intensity as a function of terrain elevation alone (*Left*) and as a function of terrain slope alone (*Right*).*
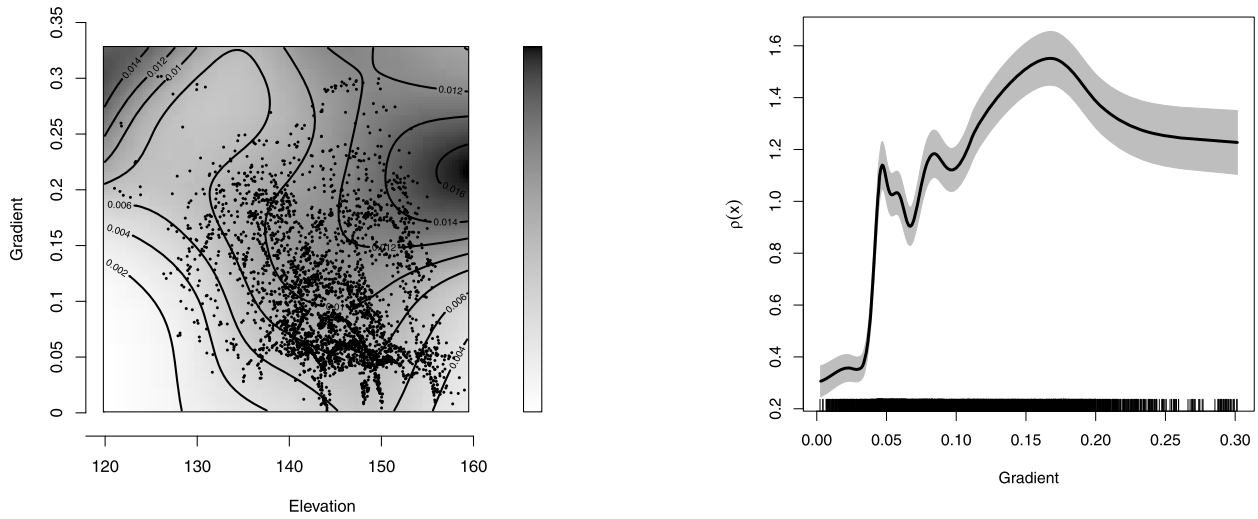

Figure 14. *Bivariate kernel estimate of intensity as a function of terrain elevation and terrain slope. Dots show observed values of elevation and slope at each data point.*

Figure 15. *Relative/residual risk estimate of* Beilschmiedia *intensity as a function of terrain slope, relative to fitted function of terrain elevation in left panel of Figure 13.*

estimates of the functions ρ_E, ρ_S respectively, in the models

$$\lambda(u) = \rho_E(e(u)) \tag{34}$$

$$\lambda(u) = \rho_S(s(u)). \tag{35}$$

The estimate of ρ_E suggests a preference for higher elevations (up to 150 metres), while $\hat{\rho}_S$ suggests a tendency to avoid very flat terrain but otherwise little preference for particular slopes.

Figure 14 shows the bivariate kernel estimate of intensity as a function jointly of terrain elevation and terrain slope. This broadly reinforces the interpretation of the previous Figure.

Figure 15 shows a kernel estimate of relative or residual risk of *B. pendula* as a function of terrain slope, relative to the estimated function of terrain elevation shown in the left panel of Figure 13. That is, the "separable" model

$$\lambda(u) = \rho_{S|E}(s(u))\rho_E(e(u)) \tag{36}$$

was fitted by first estimating ρ_E as in (34) using (8), then taking $B(u) = \hat{\rho}_E(e(u))$ and computing the relative or residual risk estimate (29). Figure 15 shows the estimate $\hat{\rho}_{S|E}$. This is very similar in shape to the right-hand panel of Figure 13, indicating strong support for the separable model.

Figure 16 shows the estimated bivariate relative or residual risk (for the elevation and gradient) relative to the sep-
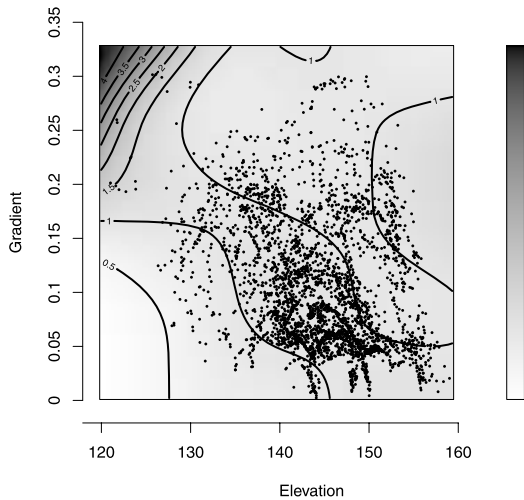
Figure 16. Validation of separable model. Bivariate (elevation and slope) relative/residual risk estimate of excess intensity relative to separable model. Crosses show observed values of elevation and slope at each data point.

arable model (36). The relative risk is close to 1 except at the upper left of the Figure, corresponding to elevation values below 130 metres and slopes greater than 0.15. The data satisfying these constraints are found at the lower left corner of the study plot in Figure 12 which contains a distinct linear feature such as a gully or river. The terrain slope values were derived from the terrain elevation by an image processing algorithm, so the very high values of terrain slope near the corner may be physically real or may be an artefact of the computation. In either case it would be prudent to exclude this corner from analysis, and when this is done, the analogue of Figure 16 shows that the separable model (36) is adequate.

A final caveat is that clustering is ever-present in forest data, and this may invalidate the confidence intervals shown here, which were based on the Poisson assumption.

10. COMPUTATION

All algorithms were implemented in the R language [58] using the spatial statistics package spatstat [8] to handle point pattern data, and the locfit package [51] for local likelihood density estimation. Fixed-bandwidth kernel estimators were computed using the Fast Fourier Transform (provided by the R base library function fft), while local likelihood density estimators require an iterative optimization algorithm (provided by locfit). The spatial pdf g^* was estimated by kernel or local likelihood smoothing of the covariate values on a fine pixel grid. Then G^* was estimated by numerical integration.

Kernel smoothing bandwidths were selected by Silverman's rule of thumb [64, eq. (3.31), p. 48] applied separately to each dataset. That is, Silverman's rule was applied first to the pixel grid of covariate values to obtain g^*, and then separately applied to the point pattern data yielding a different bandwidth for each of (8)–(10). Local likelihood density estimates were locally quadratic approximations of the log density, weighted by the tricubic kernel $k(t) = (1 - |t|^3)^3_+$. Due to an apparent bug in locfit, pointwise confidence intervals were not available, and simultaneous confidence bands have been provided.

On a 2.5 GHz laptop, computation of each panel of Figure 5 took about 0.4 seconds, using a 100×100 pixel grid, while computation of each panel in Figure 8 took about 8 seconds using a 512×512 pixel grid.

Software for computing these diagnostics, along with the data used in the paper, is available in the authors' R package spatstat [4, 8].

APPENDIX A. MISCELLANEOUS PROOFS

A.1 Differentiability of the spatial cdf

Assume that the covariate function $X : W \to \mathbb{R}$ is differentiable with nonzero gradient $\nabla X(u)$ at every $u \in W$. Then by the coarea formula [31, Thm 3.2.22],

$$
\begin{aligned}
G^*(x) &= \int_W \mathbf{1}\{X(u) \le x\} \, \mathrm{d}u \\
&= \int_\mathbb{R} \int_{X^{-1}\{y\}} w(u) \, \mathbf{1}\{X(u) \le x\} \, \mathrm{d}\mathcal{H}^{d-1}(u) \, \mathrm{d}y \\
&= \int_{-\infty}^x \int_{X^{-1}\{y\}} w(u) \, \mathrm{d}\mathcal{H}^{d-1}(u) \, \mathrm{d}y
\end{aligned}
$$

where $X^{-1}\{y\} = \{u \in W : X(u) = y\}$ is the level set where the covariate takes the value y, and $\mathcal{H}^{d-1}$ is $d-1$ dimensional Hausdorff measure. Here $w(u) = (\|\nabla X(u)\|)^{-1}$, where $\|\nabla X(u)\|$ is the absolute gradient or 1-dimensional Jacobian of X at u. The integral over $X^{-1}\{y\}$ is finite for almost all y. Hence the cdf $G(x)$ is differentiable almost everywhere, with density

$$
g^*(x) = \int_{X^{-1}\{x\}} w(u) \, \mathrm{d}\mathcal{H}^{d-1}(u).
$$

A.2 Intensity of X values

For any point process $\boldsymbol{Y}$ with any intensity function $\lambda(u)$, again using the coarea formula, the values $x_i = X(y_i)$ have intensity

$$
f^*(x) = \int_{X^{-1}\{x\}} \lambda(u) w(u) \, \mathrm{d}\mathcal{H}^{d-1}(u).
$$

If the model $\lambda(u) = \rho(X(u))$ holds, then $\lambda(u) = \rho(x)$ for all $u \in X^{-1}(x)$, so

$$
f^*(x) = \rho(x) g^*(x).
$$

However, if the model $\lambda(u) = \rho(X(u))$ is *not* true, then we still have

$$f^*(x) = \overline{\rho}(x)g^*(x)$$

where

$$\overline{\rho}(x) = f^*(x)/g^*(x) = \frac{\int_{X^{-1}\{x\}} \lambda(u)w(u)\,\mathrm{d}\mathcal{H}^{d-1}(u)}{\int_{X^{-1}\{x\}} w(u)\,\mathrm{d}\mathcal{H}^{d-1}(u)}$$

is the weighted average intensity over the level set $\{u : X(u) = x\}$ with weight $w(u)$ inversely proportional to the absolute gradient of X.

A.3 Bound on bias of $\hat{\rho}$

The fixed-bandwidth kernel estimator $\hat{\rho}$ in (8) has relative bias, by (13),

$$(37) \qquad \mathbb{E}\left[\frac{\hat{\rho}(x) - \rho(x)}{\rho(x)}\right] = \int_{\mathbb{R}} k(t-x)B(t,x)\,\mathrm{d}t$$

where

$$B(t,x) = \frac{\rho(t)}{\rho(x)}\left[\frac{g^*(t)}{g^*(x)} - 1\right] + \left[\frac{\rho(t)}{\rho(x)} - 1\right].$$

Consider a kernel k of the standard form $k(x) = b^{-1}k_1(x/b)$ where b is the bandwidth. Assume the kernel has compact support, say $k_1(x) = 0$ for $|x| > 1$. Define the logarithmic modulus of continuity of a function h by $\epsilon_h(x,\delta) = \inf\{|t - x| : \|h(t)/h(x) - 1\| \geq \delta\}$. Assume the functions $h = \rho$ and $h = g^*$ both satisfy $\epsilon_h(x,\delta) < \infty$ for all x, δ, so that $|t - x| < \epsilon_h(x,\delta)$ implies $|h(t)/h(x) - 1| < \delta$. Then it is easy to show that, for a given $0 < \delta < 1$, the relative bias (37) is smaller than δ in absolute value when $b < \epsilon$, where $\epsilon = \min\{\epsilon_\rho(x,\delta/2), \epsilon_{g^*}(x,\delta^{1/2}/4)\}$.

A.4 Probability integral transformation for $\hat{\rho}$

Under the model (1), and assuming G has a continuous inverse function G^{-1}, the values t_i constitute a point process Ψ on $[0,1]$ with intensity function $q(t) = |W|\,\rho(G^{-1}(t))$. If $\mathbf{Y}$ is Poisson then $\Psi = \{t_i\}$ is Poisson.

To prove this, note that the expected number of values t_i satisfying $t_i \leq t$ is

$$\begin{aligned}
\mathbb{E}[\#\{t_i \leq t\}] &= \int_W \mathbf{1}\{G(X(u)) \leq t\}\lambda(u)\,\mathrm{d}u \\
&= \int_W \mathbf{1}\{G(X(u)) \leq t\}\rho(X(u))\,\mathrm{d}u \\
&= \int_{-\infty}^{\infty} \mathbf{1}\{G(x) \leq t\}\rho(x)g^*(x)\,\mathrm{d}x \\
&= \int_0^1 \mathbf{1}\{s \leq t\}\rho(G^{-1}(s))\frac{g^*(G^{-1}(s))}{g(G^{-1}(s))}\,\mathrm{d}s \\
&= |W|\int_0^t \rho(G^{-1}(s))\,\mathrm{d}s
\end{aligned}$$

by changing variables from $u \in W$ to $x = X(u) \in \mathbb{R}$ and to $s = G(x) \in [0,1]$. Hence the intensity of $\{t_i\}$ is $q(t) = |W|\,\rho(G^{-1}(t))$.

REFERENCES

[1] AGTERBERG, F. P. (1974). Automatic contouring of geological maps to detect target areas for mineral exploration. *Journal of the International Association for Mathematical Geology* **6** 373–395.

[2] AHMAD, I. A. (1995). On multivariate kernel estimation for samples from weighted distributions. *Statistics and Probability Letters* **22** 121–129. MR1327737

[3] BADDELEY, A. J. (1999). Spatial sampling and censoring. In: *Stochastic Geometry: Likelihood and Computation* (O. E. Barndorff-Nielsen, W. S. Kendall and M. N. M. van Lieshout, eds.) 2 37–78. Chapman and Hall, London. MR1673114

[4] BADDELEY, A. (2010). Analysing spatial point patterns in R. Technical Report, CSIRO. Version 4. Available at www.csiro.au/resources/pf16h.html.

[5] BADDELEY, A., MØLLER, J. and WAAGEPETERSEN, R. (2000). Non- and semiparametric estimation of interaction in inhomogeneous point patterns. *Statistica Neerlandica* **54** 329–350. MR1804002

[6] BADDELEY, A., MØLLER, J. and PAKES, A. G. (2008). Properties of residuals for spatial point processes. *Annals of the Institute of Statistical Mathematics* **60** 627–649. MR2434415

[7] BADDELEY, A. and TURNER, R. (2000). Practical maximum pseudolikelihood for spatial point patterns (with discussion). *Australian and New Zealand Journal of Statistics* **42** 283–322. MR1794056

[8] BADDELEY, A. and TURNER, R. (2005). Spatstat: an R package for analyzing spatial point patterns. *Journal of Statistical Software* **12** 1–42. URL: www.jstatsoft.org, ISSN: 1548-7660.

[9] BADDELEY, A., BERMAN, M., FISHER, N. I., HARDEGEN, A., MILNE, R. K., SCHUHMACHER, D., SHAH, R. and TURNER, R. (2010). Spatial logistic regression and change-of-support for Poisson point processes. *Electronic Journal of Statistics* **4** 1151–1201. doi: 10.1214/10-EJS581. MR2735883

[10] BARTLETT, M. S. (1964). A note on spatial pattern. *Biometrics* **20** 891–892.

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

[12] BITHELL, J. F. (1990). An application of density estimation to geographical epidemiology. *Statistics in Medicine* **9** 691–701.

[13] BITHELL, J. F. (1991). Estimation of relative risk functions. *Statistics in Medicine* **10** 1745–1751.

[14] BONHAM-CARTER, G. (1995). *Geographic Information Systems for geoscientists: modelling with GIS. Computer Methods in the Geosciences* 13. Pergamon Press/Elsevier, Kidlington, Oxford, UK.

[15] BRILLINGER, D. R. (1978). Comparative aspects of the study of ordinary time series and of point processes. In *Developments in Statistics* (P. R. Krishnaiah, ed.) 33–133. Academic Press. MR0501668

[16] Chen, S. X. (1996). Empirical likelihood confidence intervals for nonparametric density estimation. *Biometrika* **83** 329–341. MR1439787

[17] Chu, C. K. and Marron, J. S. (1991). Choosing a kernel regression estimator. *Statistical Science* **6** 404-436. MR1146907

[18] Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association* **74** 829–836. MR0556476

[19] Clyde, M. and Strauss, D. (1991). Logistic regression for spatial pair-potential models. In *Spatial Statistics and Imaging*, (A. Possolo, ed.). *Lecture Notes – Monograph Series* **20** II 14–30. Institute of Mathematical Statistics ISBN 0-940600-27-7. MR1195558

[20] Condit, R. (1998). *Tropical Forest Census Plots*. Springer Verlag.

[21] Condit, R., Hubbell, S. P. and Foster, R. B. (1996). Changes in tree species abundance in a neotropical forest: impact of climate change. *Journal of Tropical Ecology* **12** 231–256.

[22] Copas, J. B. (1983). Plotting *p* against *x*. *Applied Statistics* **32** 25–31. MR0713965

[23] Cox, D. R. (1972). The statistical analysis of dependencies in point processes. In *Stochastic Point Processes* (P. A. W. Lewis, ed.) 55–66. Wiley, New York. MR0375705

[24] Daley, D. J. and Vere-Jones, D. (1988). *An Introduction to the Theory of Point Processes*. Springer Verlag, New York. MR0950166

[25] Diggle, P. J. (1985). A kernel method for smoothing point process data. *Journal of the Royal Statistical Society, Series C (Applied Statistics)* **34** 138–147.

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

[27] Diggle, P. J. and Rowlingson, B. (1994). A conditional approach to point process modelling of elevated risk. *Journal of the Royal Statistical Society, Series A (Statistics in Society)* **157** 433–440.

[28] Diggle, P. J., Rowlingson, B. and Su, T. L. (2005). Point process methodology for on-line spatio-temporal disease surveillance. *Environmetrics* **16** 423–434. MR2147534

[29] Diggle, P., Morris, S., Elliott, P. and Shaddick, G. (1997). Regression modelling of disease risk in relation to point sources. *Journal of the Royal Statistical Society, Series A* **160** 491–505.

[30] El Barmi, H. and Simonoff, J. S. (2000). Transformation based density estimation for weighted distributions. *Journal of Nonparametric Statistics* **12** 861–878. MR1802580

[31] Federer, H. (1969). *Geometric Measure Theory*. Springer Verlag, Heidelberg. MR0257325

[32] Groves, D. I., Goldfarb, R. J., Knox-Robinson, C. M., Ojala, J., Gardoll, S., Yun, G. Y. and Holyland, P. (2000). Late-kinematic timing of orogenic gold deposits and significance for computer-based exploration techniques with emphasis on the Yilgarn Block, Western Australia. *Ore Geology Reviews* **17** 1–38.

[33] Guan, Y. (2008). On consistent nonparametric intensity estimation for inhomogeneous spatial point processes. *Journal of the American Statistical Association* **103** 1238–1247. MR2528839

[34] Guan, Y. and Wang, H. (2010). Sufficient dimension reduction for spatial point processes directed by Gaussian random fields. *Journal of the Royal Statistical Society, Series B* **72** 367–387. MR2758117

[35] Hall, P. (1992). The effect of bias estimation on coverage accuracy of bootstrap confidence intervals for a probability density. *Annals of Statistics* **20** 675–694. MR1165587

[36] Hall, P. (2002). *The bootstrap and Edgeworth expansion*. Springer.

[37] Handcock, M. S. and Morris, M. (1999). *Relative Distribution Methods in the Social Sciences*. Springer-Verlag, New York. MR2000e:91112 MR1705294

[38] Hazelton, M. L. and Davies, T. M. (2009). Inference based on kernel estimates of the relative risk function in geographical epidemiology. *Biometrical Journal* **51** 98–109. MR2667514

[39] Hjort, N. L. and Jones, M. C. (1996). Locally parametric density estimation. *Ann. Statist.* **24** 1619–1649. MR1416653

[40] Horowitz, J. L. (2001). The bootstrap. In *Handbook of Econometrics*, (J. J. Heckman and E. Leamer, eds.) **5** 3159–3228. North-Holland, Amsterdam.

[41] Hubbell, S. P. and Foster, R. B. (1983). Diversity of canopy trees in a neotropical forest and implications for conservation. In *Tropical Rain Forest: Ecology and Management* (S. L. Sutton, T. C. Whitmore and A. C. Chadwick, eds.) 25–41. Blackwell Scientific Publications, Oxford.

[42] Jones, M. C. (1991). Kernel density estimation for length-biased data. *Biometrika* **78** 511–519. MR1130919

[43] Kelsall, J. E. and Diggle, P. J. (1995a). Kernel estimation of relative risk. *Bernoulli* **1** 3–16. MR1354453

[44] Kelsall, J. E. and Diggle, P. J. (1995b). Non-parametric estimation of spatial variation in relative risk. *Statistics in Medicine* **14** 2335–2342.

[45] Knox-Robinson, C. M. and Groves, D. I. (1997). Gold prospectivity mapping using a geographic information system (GIS), with examples from the Yilgarn Block of Western Australia. *Chronique de la Recherche Minière* **529** 127–138.

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

[47] Kvamme, K. L. (2006). There and back again: revisiting archeological locational modeling. In *GIS and Archaeological Site Modelling* (M. W. Mehrer and K. L. Wescott, eds.) 3–40. CRC Press.

[48] Lawson, A. B. (1993). On the analysis of mortality events around a prespecified fixed point. *Journal of the Royal Statistical Society, Series A* **156** 363–377.

[49] Loader, C. (1996). Local likelihood and density estimation. *Ann. Statist.* **24** 1602–1618. MR1416652

[50] Loader, C. (1999). *Local Regression and Likelihood*. Springer, New York. MR1704236

[51] Loader, C. (2010). locfit: Local Regression, Likelihood and Density Estimation. R package version 1.5-6.

[52] Loh, J. M. (2008). A fast and valid spatial bootstrap for correlation functions. *Astrophysical Journal* **681** 726–734.

[53] Manly, B. J. F., McDonald, L. L. and Thomas, D. L. (1993). *Resource Selection by Animals: Statistical Design and Analysis for Field Studies*. Chapman and Hall, London.

[54] Møller, J. and Waagepetersen, R. P. (2004). *Statistical Inference and Simulation for Spatial Point Processes*. Chapman and Hall/CRC, Boca Raton. MR2004226

[55] Møller, J. and Waagepetersen, R. P. (2007). Modern spatial point process modelling and inference (with discussion). *Scandinavian Journal of Statistics* **34** 643–711. MR2396935

[56] Morton-Jones, A. J., Diggle, P. J. and Elliott, P. (1999). Investigation of excess environmental risk around putative sources: Stone's test with covariate adjustment. *Statistics in Medicine* **18** 189–197.

[57] Ogata, Y. (2001). Increased probability of large earthquakes near aftershock regions with relative quiescence. *Journal of Geophysical Research* **106** 8729–8744. MR1915525

[58] R Development Core Team, (2009). R: A language and environment for statistical computing R Foundation for Statistical Computing, Vienna, Austria ISBN 3-900051-07-0.

[59] Reiss, R. D. (1993). *A Course on Point Processes*. Springer. MR1199815

[60] Ripley, B. D. (1981). *Spatial Statistics*. John Wiley and Sons, New York. MR0624436

[61] Scott, D. W. (1992). *Multivariate Density Estimation. Theory, Practice and Visualization*. Wiley, New York. MR1191168

[62] Serra, J. (1982). *Image Analysis and Mathematical Morphology*. Academic Press, London. MR0753649

[63] Sheather, S. J. and Jones, M. C. (1991). A reliable data-based method for kernel density estimation. *Journal of the Royal Statistical Society, Series B* **53** 683–690. MR1125725

[64] Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London. MR0848134

[65] Tufte, E. R. (1983). *The Visual Display of Quantitative Information*, First ed. Graphics Press.

[66] Waller, L., Turnbull, B., Clark, L. C. and Nasca, P. (1992). Chronic Disease Surveillance and testing of clustering of disease and exposure: Application to leukaemia incidence and TCE-contaminated dumpsites in upstate New York. *Environmetrics* **3** 281–300.

[67] Warton, D. I. and Shepherd, L. C. (2010). Poisson point process models solve the "pseudo-absence problem" for presence-only data in ecology. *Annals of Applied Statistics* **4** 1383–1402. MR2758333

[68] Watkins, K. P. and Hickman, A. H. (1990). Geological evolution and mineralization of the Murchison Province, Western Australia Bulletin report No. 137, Geological Survey of Western Australia. Published by Department of Mines, Western Australia, 1990. Available online from Department of Industry and Resources, State Government of Western Australia, www.doir.wa.gov.au.

Adrian Baddeley
CSIRO Mathematics
Informatics and Statistics
Floreat, Perth
Western Australia

School of Mathematics & Statistics
University of Western Australia
Australia
E-mail address: Adrian.Baddeley@csiro.au

Ya-Mei Chang
Department of Statistics
Tamkang University
Taiwan
E-mail address: yamei628@gmail.com

Yong Song
CSIRO Land and Water Highett
Melbourne
Australia
E-mail address: Yong.Song@csiro.au

Rolf Turner
Department of Statistics
University of Auckland
Auckland
New Zealand
E-mail address: r.turner@auckland.ac.nz