Biometrika (2008), **95**, 4, *pp*. 831–845
doi: 10.1093/biomet/asn045

Printed in Great Britain

A goodness-of-fit test for inhomogeneous spatial Poisson processes

BY YONGTAO GUAN

Division of Biostatistics, Yale University, New Haven, Connecticut 06520-8034, U.S.A.
yongtao.guan@yale.edu

SUMMARY

We introduce a formal testing procedure to assess the fit of an inhomogeneous spatial Poisson process model, based on a discrepancy measure function $D_c(t;\hat{\theta})$ that is constructed from residuals obtained from the fitted model. We derive the asymptotic distributional properties of $D_c(t;\hat{\theta})$ and develop a test statistic based on them. Our test statistic has a limiting standard normal distribution, so that the test can be performed by simply comparing the test statistic with readily available critical values. We perform a simulation study to assess the performance of the proposed method and apply it to a real data example.

Some key words: Goodness-of-fit test; Inhomogeneous spatial Poisson process; Residual diagnostic.

1. INTRODUCTION

Spatial Poisson processes play an important role in both statistical theory (Daley & Vere-Jones, 1988, Ch. 2) and applications (Diggle, 2003, Ch. 2). A key application for spatial Poisson processes is testing for complete spatial randomness, i.e. whether a process is homogeneous Poisson, for which a large number of testing methods have been proposed (Cressie, 1993, Ch. 8; Diggle, 2003, Ch. 2).

Many spatial point-pattern datasets include covariates, and the underlying spatial point process must be treated as inhomogeneous to model the observed spatial point pattern in terms of the observed covariates. An inhomogeneous spatial Poisson process model often appears to be appropriate, since the correlation in the data may be negligible; examples include locations of cancer patients (Diggle, 1990), of human-caused wildfires (Yang et al., 2007) and of reseeders after a fire given the locations of resprouters in a forest (Illian et al., 2008). The main modelling task is to estimate the intensity function of the process, which is often written as a parametric function of the observed covariates. In general, maximum likelihood estimation can be performed using the computationally efficient algorithm of Berman & Turner (1992). The large-sample properties of the resulting estimators for the unknown parameters were studied by Rathbun & Cressie (1994). Rathbun (1996) and Rathbun et al. (2007) considered the case in which the covariates were only partially observed.

The next step of the analysis is to assess the goodness-of-fit of the fitted model. A useful diagnostic approach, when possible, is to transform the fitted model into a homogeneous Poisson process on the real line (Ogata, 1988; Schoenberg, 2003). Available procedures for testing for complete spatial randomness can then be used. This procedure is very useful for one-dimensional data, but can also be generalized to the spatial Poisson process setting (Diggle, 1990). Other available procedures include a 'deviance residual' approach proposed by Lawson (1993) and a 'smoothed

residual field' approach proposed by Baddeley et al. (2005). For the latter, Baddeley et al. (2008) give details of the theoretical properties of the residuals. Both procedures are intended only for graphical presentations of the respective residuals and cannot be regarded as formal tests.

To obtain a formal test, it may appear reasonable simply to extend standard procedures for testing for complete spatial randomness to the inhomogeneous case, for example, comparing the theoretical and the empirical K-functions (Diggle, 2003, Ch. 7). The p-value of the test can be obtained by comparing the value of a discrepancy measure between the theoretical and the empirical K-functions, calculated from the data, with values calculated from simulations of the fitted model. However, we conjecture that this approach may have low power to detect poor fit in the inhomogeneous case, because it looks for the lack-of-fit evidence for the fitted intensity function, which models the first-order structure of the process, by evaluating the K-function, a function based on second-order structure of the process. We expect that a more powerful test could be obtained by evaluating the intensity function directly.

2. PRELIMINARY THEORETICAL RESULTS

2·1. *Definition of the discrepancy function*

Consider a spatial Poisson process N observed in a region A. Throughout the article, let $\lambda(\cdot)$ and $\lambda_c(\cdot;\theta)$ denote the true intensity function of N and a class of candidate parametric models for $\lambda(\cdot)$, respectively. Our main interest is to test the null hypothesis $H_0 : \lambda(\cdot) = \lambda_c(\cdot;\theta_0)$, for some unknown θ_0. The formulation of our problem is consistent with that of standard goodness-of-fit tests for regression problems (Kutner et al., 2004, Ch. 3).

Let $\hat{\theta}$ denote an estimator for θ. Consider a predefined shape $S \in \mathbb{R}^2$, such as a square or a circle. For any point $x \in A$, let $B(x,t)$ be the Borel set that has the same shape as S, but is inflated by the size parameter t and is centred at x. For example, for a square, t may be the length of the four sides and for a circle, t may be the diameter. Let $N(x,t)$ denote the number of events of N in $B(x,t) \cap A$; this is closely related to the scan statistic (Kulldorff, 1999), which is defined as the maximum of $N(x,t)$ over all x such that $B(x,t) \subseteq A$. For each x and t, define

$$r_c(x,t;\hat{\theta}) = N(x,t) - \int_{B(x,t)\cap A} \lambda_c(u;\hat{\theta})du; \tag{1}$$

this is a special case of the residuals defined in Baddeley et al. (2005). If the intensity function is correctly specified, then the squared value of the residual is an approximately unbiased estimator of the variance of $N(x,t)$. Furthermore, the Poisson assumption implies that $N(x,t)$ is also an unbiased estimator for the same quantity. In view of these observations, we define the discrepancy function

$$D_c(t;\hat{\theta}) = \int_A [\{r_c(x,t;\hat{\theta})\}^2 - N(x,t)]dx. \tag{2}$$

To understand better the motivation for the use of $D_c(t;\hat{\theta})$, we now comment on its properties. To simplify notation, let $\Lambda_c(x,t;\hat{\theta})$ denote the integral term in equation (1). We will also suppress the dependence of $N(x,t)$, $B(x,t)$, $r_c(x,t;\hat{\theta})$, $\Lambda_c(x,t;\hat{\theta})$ and $D_c(t;\hat{\theta})$ on t and rewrite them as $N(x)$, $B(x)$, $r_c(x;\hat{\theta})$, $\Lambda_c(x;\hat{\theta})$ and $D_c(\hat{\theta})$, respectively. Let $r(x)$ and $\Lambda(x)$ be (1) and the integral term in (1), respectively, with $\lambda_c(u;\hat{\theta})$ in (1) being replaced by the true intensity function $\lambda(\cdot)$. It

can be seen that

$$D_c(\hat{\theta}) = \int_A [\{r(x)\}^2 - N(x)]dx - 2\int_A r(x)\{\Lambda_c(x;\hat{\theta}) - \Lambda(x)\}dx + \int_A \{\Lambda_c(x;\hat{\theta}) - \Lambda(x)\}^2 dx. \tag{3}$$

The expected value of the first term on the right-hand side of equation (3) is equal to zero, since $N(x)$ is a Poisson random variable and thus its mean and variance are equal. Heuristically, if we treat $\hat{\theta}$ as fixed, then the expected value of the second term is also equal to zero. These two conclusions are true regardless of how well the fitted model fits the data. Furthermore, if $H_0 : \lambda(\cdot) = \lambda_c(\cdot;\theta_0)$ is true for some unknown θ_0 and $\hat{\theta} \doteq \theta_0$, i.e. the fitted model is a good fit, then the last term of equation (3) should also be close to zero. Thus, the expected value of $D_c(\hat{\theta})$ should be close to zero under the null hypothesis, provided that $\hat{\theta}$ estimates θ_0 well. If there is a poor fit, however, $\lambda(\cdot)$ and $\lambda_c(\cdot;\hat{\theta})$ will be very different, the last term of equation (3) will be larger than zero, and therefore the expected value of $D_c(\hat{\theta})$ will be larger than zero. To be more specific, the more $\lambda_c(\cdot;\hat{\theta})$ deviates from $\lambda(\cdot)$, the larger the expected value of $D_c(\hat{\theta})$ tends to be. To assess the goodness-of-fit of a fitted model, we thus need to assess whether or not the expected value of $D_c(\hat{\theta})$ is larger than zero. This implies that an extremely large value of $D_c(\hat{\theta})$ should be treated as an evidence for a poor fit.

When the spatial point process is not Poisson, the expectation of the first term on the right-hand side of equation (3) is no longer equal to zero: it is generally larger than zero if the process is positively correlated, e.g. clustered. Thus, both the mis-specification of the intensity function model and the violation of the Poisson assumption may yield a larger-than-zero expected value for $D_c(\hat{\theta})$. For our theoretical development, we assume that the process is Poisson. The numerical properties of our proposed test when the point process is not Poisson will be investigated in the simulation study in § 4.

2·2. *Distributional properties of* $D_c(\hat{\theta})$

To study the distributional properties of $D_c(\hat{\theta})$, let D be the discrepancy measure defined in equation (2) by using the true intensity function $\lambda(\cdot)$; that is,

$$D = \int_A [\{N(x) - \Lambda(x)\}^2 - N(x)]dx, \tag{4}$$

where $\Lambda(x) = \int_{B(x)\cap A} \lambda(u)du$. We will first study the asymptotic distribution of D and then link it to that of $D_c(\hat{\theta})$.

Our asymptotic results are based on an increasing-domain framework. Consider a sequence of domains of interest, A_n. Let R_n be R obtained on A_n, where R is an arbitrary random variable/function defined on A; for example, D_n is D in equation (4) obtained on A_n. Let $|A_n|$ and $|\partial A_n|$ denote the area and the boundary length of A_n, respectively. We assume that, for some constants $0 < K_1 < K_2 < \infty$,

$$K_1 n^2 \leqslant |A_n| \leqslant K_2 n^2, \quad K_1 n \leqslant |\partial A_n| \leqslant K_2 n. \tag{5}$$

Assumption (5) requires that A_n becomes increasingly large in all directions. This is typically satisfied by the domain shapes most commonly encountered in practice, such as a sequence of square regions with side lengths of order n and circular regions with radii of order n.

We also assume that the intensity function of N is bounded from both above and below; that is, there exist constants $0 < C_1 < C_2 < \infty$ such that for all $x \in \mathbb{R}^2$,

$$C_1 \leqslant \lambda(x) \leqslant C_2. \tag{6}$$

Assumption (6) guarantees that the variance of D_n is of the same order as the area of the region on which it is defined. For any class of parametric models under consideration, this condition can be easily checked.

The following two theorems establish the asymptotic normality of D_n and $D_{c,n}(\hat{\theta}_n)$, respectively.

THEOREM 1. *Let* $\sigma_n^2 = 2\int_{A_n}\int_{A_n}\{\Lambda_n(x,y)\}^2 dxdy$, *where* $\Lambda_n(x,y)$ *is* $\Lambda_n(x)$ *obtained on* $B(x,y)$ *and* $B(x,y) = B(x) \cap B(y) \cap A_n$. *Under assumptions* (5) *and* (6),

$$\frac{D_n}{\sigma_n} \to N(0,1),$$

in distribution as $n \to \infty$.

Proof. See the Appendix. □

THEOREM 2. *Let* $\sigma_{c,n}^2(\hat{\theta}_n) = 2\int_{A_n}\int_{A_n}\{\Lambda_{c,n}(x,y;\hat{\theta}_n)\}^2 dxdy$. *Assume that equations* (5) *and* (6) *hold and that* $\lambda_c(\cdot;\theta)$ *has bounded second-order derivatives with respect to* θ. *Under* $H_0 : \lambda(\cdot) = \lambda_c(\cdot;\theta_0)$, *if* $|A_n|^{1/4}(\hat{\theta}_n - \theta_0) = o_p(1)$, *then*

$$\frac{D_{c,n}(\hat{\theta}_n)}{\sigma_{c,n}(\hat{\theta}_n)} \to N(0,1),$$

in distribution as $n \to \infty$.

Proof. See the Appendix. □

In practice, a subtle issue is to decide which type of asymptotic framework is more appropriate, such as an increasing-domain or an infill. Under the latter, the number of observations increases with n but the study region remains fixed, which appears to be more appropriate for data that are accumulated over time in a fixed region, as in the case of many public health datasets. For Poisson processes, however, the difference between these two frameworks is probably not so important because of the complete independence among events. Our main results remain true if we replace t in equation (1) by t_n, where $t_n = t/n$ for some fixed t, and replace equations (5) and (6), respectively by

$$K_1 < |A_n| < K_2, \quad K_1 \leqslant |\partial A_n| \leqslant K_2, \quad C_1 n^2 \leqslant \lambda(x) \leqslant C_2 n^2 \quad (x \in \mathbb{R}^2).$$

2·3. *Finite-sample bias correction for* $D_c(\hat{\theta})$

Theorem 2 provides the theoretical foundation for us to derive the test statistic in the next section. Unlike D_n in Theorem 1, $D_{c,n}(\hat{\theta}_n)$ is a biased estimator for zero because of the use of the random term $\hat{\theta}_n$. Although this bias is negligible in the asymptotic sense, as suggested by Theorem 2, it can be substantial for samples of modest size. Let $f^{(i)}(\theta)$ be the ith-order derivative with respect to θ for an arbitrary real function $f(\theta)$, and for two sequences of random variables a_n and b_n, write $a_n \sim b_n$ if a_n and b_n have the same limiting distribution. To investigate the bias of $D_{c,n}(\hat{\theta}_n)$, we will further assume that

$$|A_n|^{1/2}(\hat{\theta}_n - \theta_0) \sim \{V_n(\theta_0)\}^{-1} U_n \to N[0, \{V_n(\theta_0)\}^{-1}], \tag{7}$$

in distribution where

$$U_n = \frac{1}{\sqrt{|A_n|}}\left\{\sum_{y\in N\cap A_n}\frac{\lambda_c^{(1)}(y;\theta_0)}{\lambda_c(y;\theta_0)} - \int_{A_n}\lambda_c^{(1)}(x;\theta_0)dx\right\},$$

$$V_n(\theta_0) = \frac{1}{|A_n|}\int_{A_n}\frac{\lambda_c^{(1)}(x;\theta_0)\{\lambda_c^{(1)}(x;\theta_0)\}'}{\lambda_c(x;\theta_0)}dx. \tag{8}$$

Assumption (7) is typically satisfied if $\hat{\theta}_n$ is the maximum likelihood estimator of θ_0. Also define

$$W_n(\theta_0) = \frac{1}{|A_n|}\int_{A_n}\Lambda_{c,n}^{(1)}(x;\theta_0)\{\Lambda_{c,n}^{(1)}(x;\theta_0)\}'dx. \tag{9}$$

By using Taylor series expansion at θ_0 for the last two terms of equation (3), we can then approximate $\{D_{c,n}(\hat{\theta}_n) - D_n\}$ by

$$|A_n|(\hat{\theta}_n - \theta_0)'W_n(\theta_0)(\hat{\theta}_n - \theta_0) - 2(\hat{\theta}_n - \theta_0)'\int_{A_n} r_n(x)\Lambda_{c,n}^{(1)}(x;\theta_0)dx. \tag{10}$$

The expected value of the first term in equation (10) can be approximated by the trace of the matrix $W_n(\theta_0)V_n(\theta_0)^{-1}$, since it can be treated as a quadratic form of the random vector $|A_n|^{1/2}(\hat{\theta}_n - \theta_0)$, which converges to a multivariate normal distribution under equation (7). Lengthy yet elementary derivations yield that the expected value of the second term can be approximated by

$$-\frac{2}{|A_n|}\int_{A_n}\{\Lambda_{c,n}^{(1)}(x;\theta_0)\}'V_n(\theta_0)^{-1}\Lambda_{c,n}^{(1)}(x;\theta_0)dx.$$

Thus, the bias of $D_{c,n}(\hat{\theta}_n)$ can be approximated by

$$-\frac{2}{|A_n|}\int_{A_n}\{\Lambda_{c,n}^{(1)}(x;\theta_0)\}'V_n(\theta_0)^{-1}\Lambda_{c,n}^{(1)}(x;\theta_0)dx + \mathrm{tr}\{W_n(\theta_0)V_n(\theta_0)^{-1}\}, \tag{11}$$

where $V_n(\theta_0)$ and $W_n(\theta_0)$ are defined in equations (8) and (9), respectively. In practice, we replace θ_0 in equation (11) by its estimator $\hat{\theta}_n$ so as to obtain an estimator for the bias of $D_{c,n}(\hat{\theta}_n)$. An alternative approach, which is much simpler in terms of programming effort, is to simulate data from the fitted model $\lambda_c(\cdot;\hat{\theta}_n)$ and then calculate $D_{c,n}(\tilde{\theta}_n)$ for each simulated realization, where $\tilde{\theta}_n$ is the version of $\hat{\theta}_n$ calculated from that realization. An estimator for the bias can be defined as the sample average of all obtained $D_{c,n}(\tilde{\theta}_n)$. This approach is used in the simulation study.

3. The proposed method

Based on the theoretical results in the last section, we develop a formal testing method to assess the goodness-of-fit of the fitted model $\lambda_c(\cdot;\hat{\theta})$. For a prespecified t, we calculate the statistic $T(\hat{\theta}) = \{D_c(\hat{\theta}) - \mathrm{bias}(\hat{\theta})\}/\sigma_c(\hat{\theta})$, where $\mathrm{bias}(\hat{\theta})$ is an estimate for the bias term in equation (11) and $\sigma_c^2(\hat{\theta}) = 2\int_A\int_A\{\Lambda_c(x,y;\hat{\theta})\}^2dxdy$. Following Theorem 2, $T(\hat{\theta})$ is approximately a standard normal random variable under $H_0 : \lambda(\cdot) = \lambda_c(\cdot;\theta_0)$. For an α-level test for H_0, we reject H_0 and thus conclude a lack of fit if $T(\hat{\theta}) > Z_\alpha$, where Z_α is the upper-tail critical value at the α level from the standard normal distribution.

To apply the proposed method it is important to select an appropriate t. Generally, a very small t will lead to a test with little power because of the insufficient sample size used to calculate $[\{r_c(x;\hat{\theta})\}^2 - N(x)]$ for each x, based on which $D_c(\hat{\theta})$ in equation (2) is defined. This is because for a small t there is often too much noise in $[\{r_c(x;\hat{\theta})\}^2 - N(x)]$, which in turn will hide any signal, indicating lack of fit. However, a large t does not necessarily lead to improved power,

since information for local lack of fit may then be smoothed out. In particular, the magnitude of $r_c(x;\hat{\theta})$ may be too small compared to $N(x)$. Furthermore, a large t-value can also cause the size of the test to deteriorate for two reasons. First, $[\{r_c(x;\hat{\theta})\}^2 - N(x)]$ is skewed to the right. If t is too large, then there are not enough replicates for $[\{r_c(x;\hat{\theta})\}^2 - N(x)]$, so that the normal approximation does not work well. Secondly, a large t will also lead to increased edge effects, in that events near the boundary are given less weight than those in the centre when forming $D_c(\hat{\theta})$. This will further reduce the effective sample size. Thus, t must be selected carefully. This is not a problem unique to our test statistic, for example, when calculating the scan statistic, one also needs to decide the size of the scanning window (Kulldorff, 1999). A simulation study in § 4 evaluates the effect of t on the performance of the test.

To obtain a data-driven approach for the selection of t, we note that t affects the test mostly through its effect on $r_c(x;\hat{\theta})$. Thus, the problem of selecting t for $D_c(\hat{\theta})$ can be treated roughly as the problem of selecting the bandwidth used to obtain the residuals $r_c(x;\hat{\theta})$, where $x \in A$. For the latter, Baddeley et al. (2005) discussed several data-driven methods. We suggest the use of one of these methods to select the bandwidth for $r_c(x;\hat{\theta})$, and thus select t for $D_c(\hat{\theta})$. From a practical point of view, our proposal is reasonable, since one may wish first to obtain and examine the 'smoothed' residual plot, i.e. the plot of $r_c(x;\hat{\theta})$. Our formal testing procedure can then be used to assess the goodness-of-fit, based on the obtained residual plot. In practice, we can also plot $D_c(\hat{\theta})$ and/or $T(\hat{\theta})$ with their respective pointwise upper confidence bounds for a range of t-values. This graphical presentation of the results will enable us to examine the evidence quickly in a more systematic way.

4. SIMULATION STUDY

4·1. *Simulation design*

We simulated both inhomogeneous Poisson processes and inhomogeneous Poisson cluster processes (Waagepetersen, 2007) on a unit square. For both types of processes, the intensity function was given by $\alpha \exp(-\beta x)$ and $\alpha \exp\{-\beta \sin(2\pi x)\}$, where x was the x-coordinate of an arbitrary point on the unit square. We refer to the first model as the linear model and the second as the sine model. To simulate the Poisson cluster process, we first simulated the parent process by using a homogeneous Poisson process with intensity equal to 50. For each parent, we then generated a Poisson number of offspring, where the location of each offspring relative to its parent was determined by a radially symmetric Gaussian dispersal variable (Diggle, 2003, p. 66). We set the standard deviation of the dispersal variable at 0·04. Finally, we thinned the offspring process independently with a thinning probability equal to $1 - \exp(-\beta x)$, as suggested by Waagepetersen (2007). We set $\beta = 1$ or 2. As β increases, so does the degree of inhomogeneity in the process. For each type of intensity function and for each β, we manipulated the value of α so that the expected number of events per realization, denoted by μ, was roughly 100 and 400. In terms of the asymptotic frameworks discussed in § 2, the setting being considered here was an infill asymptotic framework.

We used a square as the shape S for computational convenience. To study the effect of t on the performance of the test, we set the side length t equal to 0·1, 0·2 and 0·3. For each realization and each t, we applied the proposed testing method to assess the goodness-of-fit for the fitted linear and sine models. We also selected the t-value by least-squares crossvalidation (Silverman, 1999). To reduce computational time, we considered only 15 equally spaced t-values between 0·02 and 0·3. The 'optimal' t selected by least-squares crossvalidation could be larger than 0·3, but we

imposed an upper limit of 0·3 on t, since the normal approximation is not expected to be valid for large t, as mentioned in § 3.

To compare with existing methods, we also applied two other competing tests. The first was a simulation-based approach that compared the empirical and theoretical K-functions. We will refer to this approach as the K-function approach. The second approach was based on the idea that a spatial Poisson process could be transformed to be a homogeneous Poisson process on [0,1] by using an appropriate transformation. Standard goodness-of-fit tests for uniformity were then applied to the transformed data. We will refer to the second approach as the transformation approach.

For the K-function approach, let $\hat{K}(r)$ denote the nonstationary version of the empirical K-function at lag r as defined in Baddeley et al. (2000). We used the following popular discrepancy measures between $\hat{K}(r)$ and $K(r)$ (Ho & Chiu, 2007):

$$\mathrm{dk}_1 = \sup_{r\in[0,r_0]} |\hat{K}(r)^{1/2} - K(r)^{1/2}|, \quad \mathrm{dk}_2 = \int_0^{r_0} \{\hat{K}(r)^{1/2} - K(r)^{1/2}\}^2 dr,$$

where $r_0 = 0{\cdot}125, 0{\cdot}062$ for $\mu = 100, 400$, respectively, and $K(r) = \pi r^2$. The choices of r_0 were based on the recommendation of Ripley (1979). The square-root transformation of $\hat{K}(r)$ was suggested by Besag (1977) as a variance stabilizer. To perform the test, we simulated 39 realizations from the fitted model and obtained dk_1 and dk_2 for each realization. We rejected H_0 if dk_1, respectively dk_2, from the original realization ranked in the top 10% of the pooled dk_1, respectively dk_2. This led to tests with a nominal size equal to 10%. In our simulation, the test based on dk_2 was slightly more powerful than that based on dk_1. We therefore present only the results for the former.

For the transformation approach, let $I(\cdot)$ denote an indicator function. A referee suggested transforming any given event of the process, say x, to be $\mu(-\infty, f(x)]/\mu(-\infty, \infty)$, where $f(x)$ was a known, continuous function of location and

$$\mu(-\infty, t] = \int_A I\{f(u) \leqslant t\}\lambda_c(u;\hat{\theta})du.$$

In the above expression, $\lambda_c(\cdot)$ was either the linear model or the sine model. For the function $f(\cdot)$, we used $f(x) = x$ for the linear model and $f(x) = \sin(2\pi x)$ for the sine model. We then calculated the Kolmogorov–Smirnov test statistic by comparing the empirical distribution of the transformed process with a uniform distribution on [0, 1]. To perform the test, we simulated 39 realizations from the fitted model and obtained the same test statistic for each realization. We rejected H_0 if the calculated statistic from the original realization ranked in the top 10% of the pooled statistics. This in turn led to a test with a nominal size equal to 10%.

4·2. *Simulation results*

Table 1 lists percentages of rejections at the 10% nominal level from 500 simulations in the Poisson process case. When the true intensity model was used, the resulting rejection percentages were all close to the nominal size. When a wrong intensity model was used, the resulting rejection percentages, i.e. powers, increased as the expected number of events increased; a larger sample size should lead to a more powerful test. When t increased, the power first increased and then decreased, agreeing with our general comment regarding the effect of t on the test. Simulation results not included also suggested that too large a t-value, $t = 0{\cdot}4$ say, not only caused the size of the test to deteriorate, but also lowered the power significantly.

Heuristically, we may treat $|A|/|B(\cdot)|$ as the number of independent replicates for a region of the same size as $B(\cdot)$. In our simulation, $|A| = 1$ and $|B(\cdot)| = t^2$. Thus $t = 0{\cdot}3$ roughly corresponded

Table 1. *Simulated relative frequencies of rejection at the 10% nominal level in the Poisson process case. Test* 1 *and Test* 2 *are the proposed tests with t fixed and selected by least-squares crossvalidation, respectively. Test* 3 *is the test based on the K-functions. Test* 4 *is the test based on transformation to a homogeneous Poisson process*

				Test 1 with different t			Test 2	Test 3	Test 4
	Model	μ	β	0·1	0·2	0·3			
Size	Linear	100	1	0·096	0·096	0·084	0·094	0·110	0·064
			2	0·130	0·130	0·104	0·110	0·072	0·070
		400	1	0·096	0·094	0·082	0·090	0·114	0·080
			2	0·112	0·098	0·078	0·086	0·090	0·090
	Sine	100	1	0·100	0·100	0·076	0·094	0·080	0·052
			2	0·098	0·108	0·080	0·096	0·094	0·060
		400	1	0·112	0·082	0·062	0·074	0·092	0·066
			2	0·126	0·114	0·088	0·110	0·078	0·078
Power	Linear	100	1	0·162	0·200	0·162	0·166	0·112	0·124
			2	0·400	0·482	0·436	0·464	0·164	0·284
		400	1	0·354	0·466	0·424	0·442	0·136	0·252
			2	0·944	0·994	0·990	0·988	0·630	0·840
	Sine	100	1	0·550	0·728	0·712	0·730	0·192	0·902
			2	0·996	1·000	1·000	1·000	0·206	1·000
		400	1	1·000	1·000	1·000	1·000	0·628	1·000
			2	1·000	1·000	1·000	1·000	0·758	1·000

to 11 independent replicates. The actual sample size might be slightly bigger, since $D(\hat{\theta})$ integrated $[\{r_c(x;t;\hat{\theta})\}^2 - N(x,t)]$ over all $x \in A$, and not just over a set of x_i $(i = 1, \ldots, |A|/|B(\cdot)|)$ such that $\cup_i B(x_i) = A$ and $\cap_i B(x_i) = \emptyset$. Nevertheless, it does not seem appealing to use a t that is even larger. One possibility in practice is to set an upper limit, 0·3 say, for t and then use least-squares crossvalidation to select t. From Table 1 we see that this approach worked well in our simulations.

Figure 1 plots the true intensity models and the average of the fitted incorrect intensity models obtained from the simulations for each scenario. For the models being considered here, the difference between the true and the incorrect models was higher when the true model was the sine for each fixed β, and increased as β increased for each fixed model. The powers in Table 1 indicate that there was a much higher power to detect a lack of fit when the sine model was the true model and when $\beta = 2$.

When compared with the K-function approach, our test was much more powerful in all cases. The improvement was often quite substantial. For example, when $\mu = 400$ was used and the true model was the sine model, the proposed test rejected the incorrect model 100% of the time, whereas the K-function approach rejected it only 62·8%, for $\beta = 1$, and 75·8%, for $\beta = 2$, of the time. This difference was probably due to the fundamental difference when deriving the test statistics for these two methods. Our approach first calculated the local discrepancy between the data and the fitted intensity function directly and then combined them to derive a global measure for the overall discrepancy. The K-function approach, on the other hand, first calculated a number of 'indirect' global discrepancy measures; $\hat{K}(r)$, $r \in [0, r_0]$ is a global measure, since each $\hat{K}(r)$ was obtained by pooling all data together. It then combined these global measures to derive another global measure for the overall discrepancy. As a result, valuable information related to local discrepancies may have been diluted, if not lost completely. This in turn led to a poor power for the K-function approach. When compared with the transformation approach, our

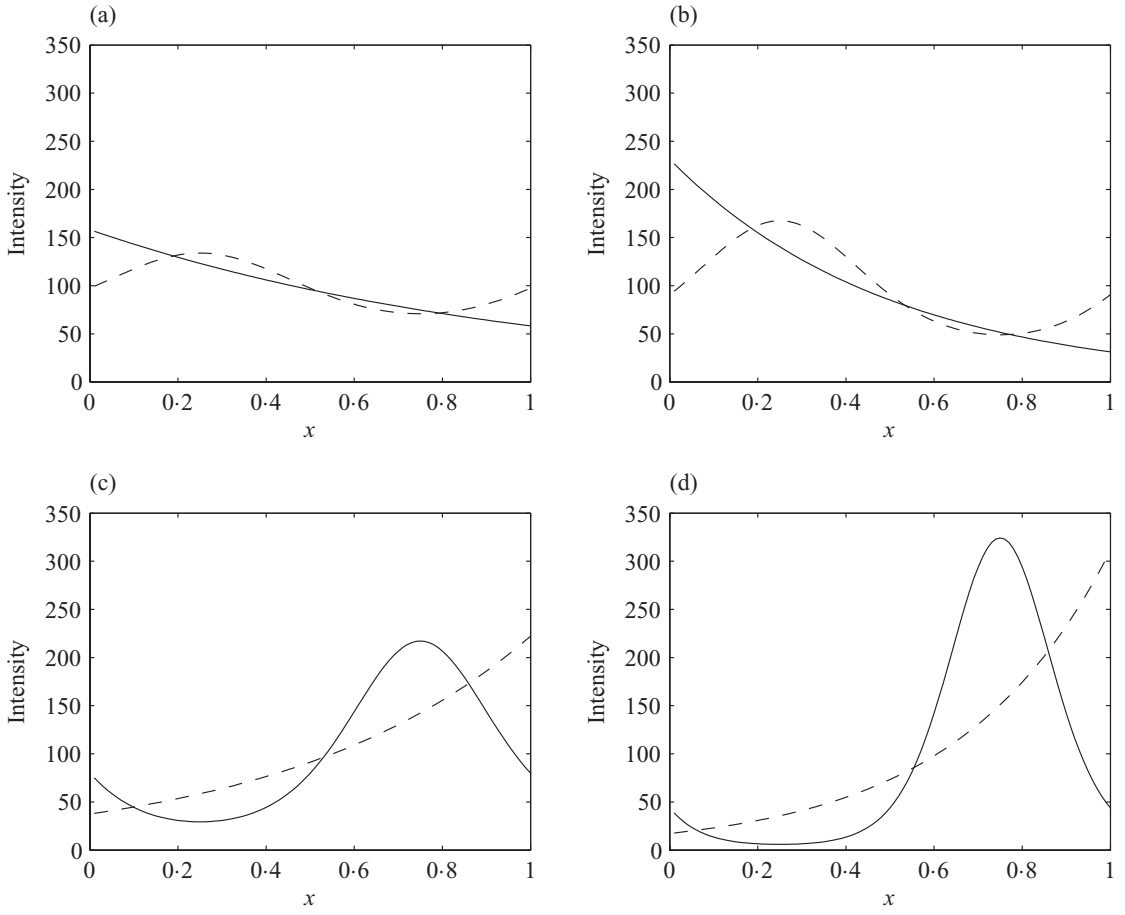

Fig. 1. Simulation study. Plots of the true intensity function model (solid line) and the average of the fitted incorrect models (dashed line) versus x. The x-axes in the plots are the x coordinate values on a unit square and the y-axes are the true and estimated intensities. Panel (a) corresponds to the linear model with $\beta = 1$; panel (b) corresponds to the linear model with $\beta = 2$; panel (c) corresponds to the sine model with $\beta = 1$; and panel (d) corresponds to the sine model with $\beta = 2$.

test was more powerful in the linear model case, and was comparable in the sine model case. The K-function approach was consistently less powerful than the transformation approach.

Table 2 lists percentages of rejections at the 10% nominal level from 500 simulations in the Poisson cluster process case. Only the results in the linear model case were included. The main findings in the sine model case were similar, and thus were omitted. Regardless of which intensity model was fitted, all three tests concluded that there was lack of fit at a rate much higher than the nominal 10% level. In particular, the percentages of rejections were either equal to or close to 100% for both the proposed approach and the K-function approach. This indicated the high power of these approaches to detect a violation of the Poisson assumption. When $\mu = 100$, our test still slightly outperformed the K-function approach, but the improvement became much smaller. When $\mu = 400$, both approaches always rejected H_0. However, the transformation approach rejected H_0 much less frequently, especially when $\mu = 100$. The transformation approach was defined in terms of the intensity function only, so it was not surprising to see its low power to detect a violation of the Poisson assumption, since this is related to the higher order structures of the process.

Table 2. *Simulated relative frequencies of rejection at the* 10% *nominal level in the Poisson cluster process case. The true intensity model is the linear model. See Table* 1 *for details of the tests*

			Test 1 with different t			Test 2	Test 3	Test 4
Model	μ	β	0·1	0·2	0·3			
Linear	100	1	0·988	0·972	0·892	0·978	0·944	0·382
		2	0·994	0·980	0·886	0·978	0·834	0·434
	400	1	1·000	1·000	1·000	1·000	1·000	0·826
		2	1·000	1·000	1·000	1·000	1·000	0·918
Sine	100	1	0·988	0·978	0·904	0·974	0·964	0·294
		2	0·998	0·988	0·956	0·990	0·968	0·466
	400	1	1·000	1·000	1·000	1·000	1·000	0·732
		2	1·000	1·000	1·000	1·000	1·000	0·806

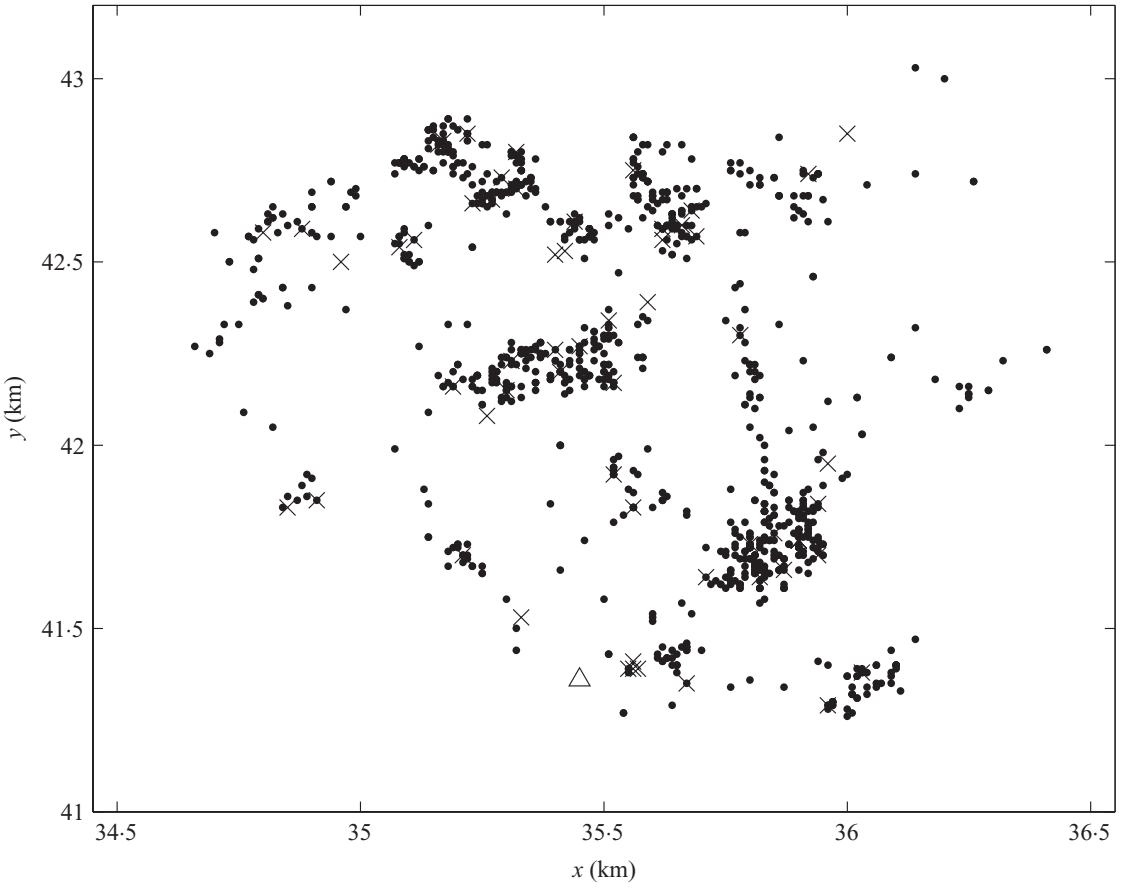

Fig. 2. Locations of larynx (×) and lung (·) cancers and the incinerator (△).

5. AN APPLICATION

Figure 2 plots the locations of 58 cases of larynx cancer and 978 cases of lung cancer in the Chorley and South Ribble Health Authority of Lancashire during 1974–1983 (Diggle, 1990). The main interest is to model the locations of larynx cancer cases in relation to the location of an industrial incinerator. To do so, Diggle (1990) fitted the following inhomogeneous Poisson process model to the data, with the distance from each larynx cancer case to the location of the

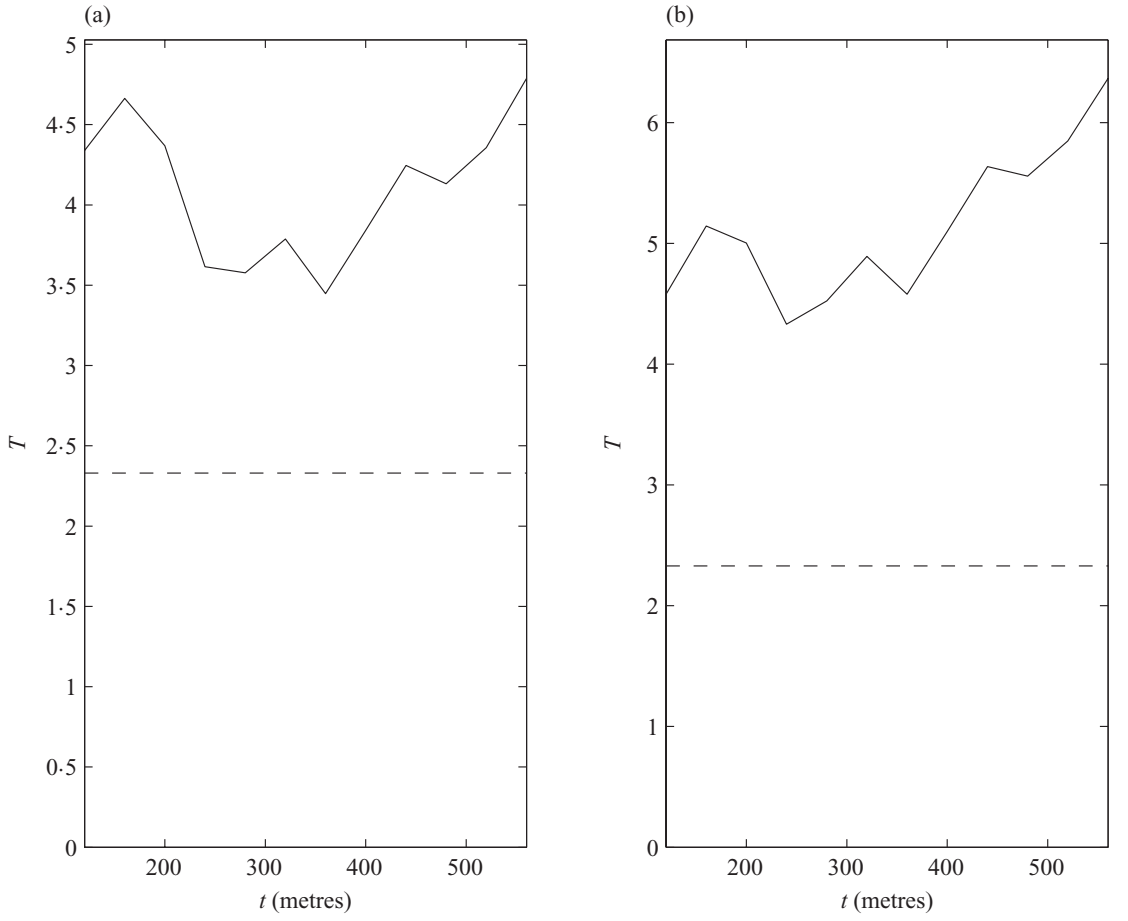

Fig. 3. Residual plots for the larynx cancer data, showing the test statistic $T(\hat{\theta})$ calculated at different values of t. Panel (a) is for the model with $(\hat{\alpha}, \hat{\beta}) = (23{\cdot}67, 0{\cdot}91)$ and panel (b) is for that with $(\hat{\alpha}, \hat{\beta}) = (33{\cdot}69, 1{\cdot}11)$. The solid lines are the test statistics and the dashed lines are the 99% confidence limits under each model.

incinerator as a covariate,

$$\lambda(x) = \rho\lambda_0(x)\{1 + \alpha \exp(-\beta||x - x_0||^2)\}.$$

Here, ρ is the overall number of events per unit area, $\lambda_0(\cdot)$ is the spatial intensity of the population at risk and x_0 is the location of the incinerator. To estimate $\lambda_0(\cdot)$, the lung cancer cases were treated as a surrogate for the susceptible population. Diggle (1990) estimated $\lambda_0(\cdot)$ by kernel smoothing the lung cancer cases, using an isotropic Gaussian kernel with standard deviation $\sigma = 0{\cdot}15$ km. Diggle (1990) obtained the estimates $(\hat{\alpha}, \hat{\beta}) = (23{\cdot}67, 0{\cdot}91)$ using a maximum likelihood approach, whereas Diggle & Rowlingson (1994) obtained the estimates $(\hat{\alpha}, \hat{\beta}) = (33{\cdot}69, 1{\cdot}11)$ using a conditional approach. Both analyses indicated raised incidence of larynx cancers near the incinerator.

To evaluate the goodness-of-fit for a fitted model, Diggle (1990) ordered the larynx cancer cases according to their distances from x_0, where x_0 is the location of the incinerator. Let E_i denote the disc with centre x_0 and radius equal to the ith-order distance, and let

$$T_i = \int_{E_i} \lambda_0(x)\{1 + \hat{\alpha} \exp(-\hat{\beta}||x - x_0||^2)\} \quad (i = 1, \ldots, 58).$$

Under the fitted model, T_i can be roughly treated as a realization from a homogeneous one-dimensional Poisson process. The goodness-of-fit of the fitted model can then be evaluated by testing if T_i are from a homogeneous Poisson process. A satisfactory fit was concluded for the first set of estimates under this approach (Diggle, 1990). Recently, Baddeley et al. (2005) assessed the fit of the second model by considering the residuals in equation (1), where x therein was equal to x_0 and the Borel set B was a circle with radius t. By comparing the obtained residuals and their respective 2σ-limits, Baddeley et al. (2005) found a slight lack of fit near $t = 0$ for the second model.

We performed our proposed goodness-of-fit test for both models. As in the simulation, we used a square for the shape S. Figure 3 plots the test statistic values for each model, at various side lengths, t, for the squares being used, and the corresponding 99% confidence limits. There is striking evidence that neither of the two fitted models is a good fit, since all $T(\hat{\theta})$ were above the 99% confidence limits. Our conclusion contradicts that of Diggle (1990), who acknowledged the limitations of his approach. Furthermore, our analysis formally confirmed the lack of fit of the second model as detected by Baddeley et al. (2005). In a personal communication, Peter Diggle suggested that the lack of fit of the first model was probably due to the biased estimate of $\lambda_0(\cdot)$ that was produced by the kernel smoothing method. The reliability of these estimates may be in question, since $\lambda_0(\cdot)$ was used in the estimation of the parameters. On the other hand, the estimates $(\hat{\alpha}, \hat{\beta}) = (33{\cdot}69, 1{\cdot}11)$ were obtained by using the conditional approach in Diggle & Rowlingson (1994) which eliminated the need to estimate $\lambda_0(\cdot)$. As a result, we believe that these results are more reliable, although our testing method still indicated lack of fit. An estimate of $\lambda_0(\cdot)$ was required in order to calculate our test statistics for both fitted models which affected the calculated test statistics. Thus our analysis is only illustrative, rather than being a thorough analysis.

6. DISCUSSION

When lack of fit is detected, we need to decide whether the lack of fit is caused by the use of an incorrect intensity function model or the existence of correlation, i.e. the process is not Poisson. This is often difficult, as heterogeneity in intensity and correlation, especially clustering, can lead to point patterns with similar characteristics (Diggle, 2003, Ch. 9). Brix et al. (2001) proposed a method to test if an observed spatial point pattern could be treated as a realization from an inhomogeneous spatial Poisson process. Their approach did not require any specific parametric form for the intensity function. One sensible approach in practice may be first to apply their method to evaluate if the Poisson assumption is reasonable. If the assumption is not rejected, then the focus of the analysis should be on modelling the intensity function alone. Otherwise, it is necessary to consider alternative processes that allow correlation in the data.

ACKNOWLEDGEMENT

This work was supported by the U.S. National Science Foundation. The author thanks Peter Diggle and Adrian Baddeley for helpful comments, Andrew Lawson for pointing out an error in the reference list, and Professor D. M. Titterington and two referees for constructive comments that greatly improved the paper.

APPENDIX

Proofs

Derivation of the variance of D. Let $N(x, y)$ and $r(x, y)$ be $N(x)$ and $r(x)$ defined on $B(x, y)$, respectively. We have

$$\begin{aligned}\text{var}(D) &= \int_A\int_A E([\{r(x)\}^2 - N(x)][\{r(y)\}^2 - N(y)])dxdy \\ &= \int_A\int_A E([\{r(x, y)\}^2 - N(x, y)]^2)dxdy \\ &= 2\int_A\int_A \{\Lambda(x, y)\}^2 dxdy.\end{aligned}$$

The last equality follows because, for any Poisson random variable X with an expected value μ,

$$E[\{(X - \mu)^2 - X\}^2] = 2\mu^2. \qquad \square$$

Proof of Theorem 1. We shall denote several constants by the same letter, c. To prove Theorem 1, we use k_n subsquares to approximate A_n. Each of the subsquares has a side length l, where $l = cn^\alpha$ for some $0 < \alpha < 1$. Let A_l^i be the ith subsquare and $A_n' = \cup_i A_l^i$. Assumption (5) guarantees that $|A_n'|/|A_n| \to 1$ as $n \to \infty$. Let D_l^i be D, calculated on A_l^i and define $D_n' = \sum_i^{k_n} D_l^i$ and $(\sigma_n')^2 = \text{var}(D_n')$.

We first want to show that

$$\frac{D_n}{\sigma_n} - \frac{D_n'}{\sigma_n'} \to 0, \tag{A1}$$

in probability. To show this, we only need to show that

$$\frac{\text{cov}(D_n, D_n')}{\sigma_n\sigma_n'} \to 1.$$

This follows after some lengthy yet elementary algebra from the fact that $|A_n'|/|A_n| \to 1$ and assumption (6). Then equation (A1) holds from Chebyshev's inequality.

We then want to show that

$$\frac{D_n'}{\sigma_n'} \to N(0, 1), \tag{A2}$$

in distribution. To prove this, we first verify that

$$\sup_n \text{E}(|D_l|^4) < cl^4. \tag{A3}$$

Note that $(\sigma_n')^2$ is of order n^2 because of condition (6) and the way in which the sub-blocks were constructed. Result (A2) then follows trivially from equation (A3) by the application of the Lyapunov theorem, since the D_l^i are independent.

In what follows, let $\int$ stand for $\int_{D_l}$ unless specified otherwise. Define $B^{-1}(x) = \{s : x \in B(s)\} \cap A_l$ and $B^{-1}(x, y) = B^{-1}(x) \cap B^{-1}(y)$. To prove equation (A3), we first rewrite D_l as

$$\left[\sum\sum_{x \neq y} |B^{-1}(x, y)| - \int \{\Lambda(s)\}^2 ds\right] - 2\left[\sum_x \int_{B^{-1}(x)} \Lambda(s)ds - \int \{\Lambda(s)\}^2 ds\right].$$

Denote the two terms in the square brackets by F_l and G_l, respectively. A sufficient condition for equation (A3) to hold is that

$$\sup_n E\left(|F_l|^4\right) < cl^4, \quad \sup_n E\left(|G_l|^4\right) < cl^4.$$

Let $g(x) = \int_{B^{-1}(x)} \Lambda(s)ds$. Elementary algebra shows that

$$E\{(G_l)^4\} = 6\left[\int \{g(x)\}^2\lambda(x)dx\right]^2 + \int \{g(x)\}^4\lambda(x)dx.$$

Clearly, $E\{(G_l)^4\} < cl^4$ for some c because of condition (6). Furthermore,

$$\begin{aligned}(F_l)^4 = &\left\{\sum\sum_{x\neq y}|B^{-1}(x,y)|\right\}^4 + \left\{\iint |B^{-1}(u,v)|\lambda(u)\lambda(v)dudv\right\}^4\\ &-4\left\{\sum\sum_{x\neq y}|B^{-1}(x,y)|\right\}^3\iint |B^{-1}(u,v)|\lambda(u)\lambda(v)dudv\\ &+6\left\{\sum\sum_{x\neq y}|B^{-1}(x,y)|\right\}^2\left\{\iint |B^{-1}(u,v)|\lambda(u)\lambda(v)dudv\right\}^2\\ &-4\sum\sum_{x\neq y}|B^{-1}(x,y)|\left\{\iint |B^{-1}(u,v)|\lambda(u)\lambda(v)dudv\right\}^3.\end{aligned}$$

Some simple algebra shows that $E\{(F_l)^4\}$ can be written as the sum of seven integrals, ranging from a two-fold integral to an eight-fold integral. Note that $B^{-1}(x,y) \neq \emptyset$ only on a bounded set for each fixed x. This fact leads all the two- to five-fold integrals being of order no higher than l^4. The six-fold integral, with a multiplicative constant ignored, can be shown to satisfy

$$\left[\iiint |B^{-1}(x,y)||B^{-1}(x,z)|\{\lambda(x)\}^2\lambda(y)\lambda(z)dxdydz\right]^2 < cl^4.$$

Lastly, it can be shown that the 7-fold and the 8-fold integrals are both equal to zero. Thus equation (A3) holds. □

Proof of Theorem 2. Under $H_0 : \lambda(\cdot) = \lambda_c(\cdot;\theta_0)$ for some θ_0, we have

$$D_{c,n}(\hat{\theta}_n) = D_n - 2\int_{A_n} r_n(x)\{\Lambda_{c,n}(x;\hat{\theta}_n) - \Lambda_{c,n}(x;\theta_0)\}dx + \int_{A_n}\{\Lambda_{c,n}(x;\hat{\theta}_n) - \Lambda_{c,n}(x;\theta_0)\}^2dx.$$

Let $\Lambda^{(i)}_{c,n}(\cdot;\theta)$ be the ith-order derivative of $\Lambda_{c,n}(\cdot;\theta)$ with respect to θ. By Taylor series expansions, we obtain

$$\begin{aligned}&\Lambda_{c,n}(x;\hat{\theta}_n) - \Lambda_{c,n}(x;\theta_0) = (\hat{\theta}_n - \theta_0)'\Lambda^{(1)}_{c,n}(x;\theta^*_n),\\ &\Lambda_{c,n}(x;\hat{\theta}_n) - \Lambda_{c,n}(x;\theta_0) = (\hat{\theta}_n - \theta_0)'\Lambda^{(1)}_{c,n}(x;\theta_0) + (\hat{\theta}_n - \theta_0)'\Lambda^{(2)}_{c,n}(x;\theta^{**}_n)(\hat{\theta}_n - \theta_0),\end{aligned}$$

where both θ^*_n and θ^{**}_n lie between $\hat{\theta}_n$ and θ_0. Thus,

$$\begin{aligned}D_{c,n}(\hat{\theta}_n) - D_n = &(\hat{\theta}_n - \theta_0)'\left[\int_{A_n}\Lambda^{(1)}_{c,n}(x;\theta^*_n)\{\Lambda^{(1)}_{c,n}(x;\theta^*_n)\}'dx\right](\hat{\theta}_n - \theta_0)\\ &-2(\hat{\theta}_n - \theta_0)'\left\{\int_{A_n} r_n(x)\Lambda^{(2)}_{c,n}(x;\theta^{**}_n)dx\right\}(\hat{\theta}_n - \theta_0)\\ &-2(\hat{\theta}_n - \theta_0)'\int_{A_n} r_n(x)\Lambda^{(1)}_{c,n}(x;\theta_0)dx.\end{aligned}$$

Note that σ_n^2 is of order $|A_n|^{1/2}$ because of assumption (6). Thus, to prove Theorem 2, we only need to show that, in probability,

$$\{D_{c,n}(\hat{\theta}_n) - D_n\}/|A_n|^{1/2} \to 0, \quad \sigma_{c,n}(\hat{\theta}_n)/\sigma_n \to 1.$$

The first is true if both $\Lambda^{(1)}(x;\theta)$ and $\Lambda^{(1)}(x;\theta)$ are bounded in a small neighbourhood of θ_0 and $\hat{\theta}_n - \theta_0 = o_p(1/|A_n|^{1/4})$. The second is true because of condition (6). Thus Theorem 2 is proved. □

References

Baddeley, A. J., Møller, J. & Pakes, A. G. (2008). Properties of residuals for spatial point processes. *Ann. Inst. Statist. Math.* **60**, 629–49.

Baddeley, A. J., Møller, J. & Waagepetersen, R. (2000). Non- and semi-parametric estimation of interaction in inhomogeneous point patterns. *Statist. Neerlandica* **54**, 329–50.

Baddeley, A. J., Turner, R., Møller, J. & Hazelton, M. (2005). Residual analysis for spatial point processes (with Discussion). *J. R. Statist. Soc.* B **67**, 617–66.

Berman, M. & Turner, R. T. (1992). Approximating point process likelihoods with GLIM. *Appl. Stat.* **41**, 31–8.

Besag, J. (1977). Contribution to the discussion of a paper by B. D. Ripley. *J. R. Statist. Soc.* B **39**, 193–5.

Brix, A., Senoussi, R., Couteron, P. & Chaduf, J. (2001). Assessing goodness-of-fit of spatially inhomogeneous Poisson processes. *Biometrika* **88**, 487–97.

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

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

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**, 349–62.

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

Diggle, P. J. & Rowlingson, B. S. (1994). A conditional approach to point process modeling of elevated risk. *J. R. Statist. Soc.* A **157**, 433–40.

Ho, L. P. & Chiu, S. N. (2007). Testing uniformity of a spatial point pattern. *J. Comput. Graph. Statist.* **16**, 378–98.

Illian, J. B., Møller, J. & Waagepetersen, R. (2008). Hierarchical spatial point process analysis for a plant community with high biodiversity. *Environ. Ecol. Stat.*, doi: 10.1007/s10651-007-0070-8.

Kulldorff, M. (1999). Spatial scan statistics: Models, calculations, and applications. In *Scan Statistics and Applications*, Ed. J. Glaz and N. Balakrishnan, pp. 303–22. Boston, MA: Birkhäuser.

Kutner, M. H., Nachtsheim, C. J. & Neter, J. (2004). *Applied Linear Regression Models*. Boston, MA: McGraw-Hill.

Lawson, A. B. (1993). A deviance residual for heterogeneous spatial Poisson processes. *Biometrics* **49**, 889–97.

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

Rathbun, S. L. (1996). Estimation of Poisson intensity using partially observed concomitant variables. *Biometrics* **52**, 226–42.

Rathbun, S. L. & Cressie, N. A. C. (1994). Asymptotic properties of estimators for the parameters of spatial inhomogeneous Poisson point processes. *Adv. in Appl. Probab.* **26**, 122–54.

Rathbun, S. L., Shiffman, S. & Gwaltney, C. J. (2007). Modelling the effects of partially observed covariates on Poisson process intensity. *Biometrika* **94**, 153–65.

Ripley, B. D. (1979). Tests of 'randomness' for spatial point patterns. *J. R. Statist. Soc.* B **41**, 368–74.

Schoenberg, F. P. (2003). Multidimensional residual analysis of point process models for earthquake occurrences. *J. Am. Statist. Assoc.* **98**, 789–95.

Silverman, B. W. (1999). *Density Estimation for Statistics and Data Analysis*. New York: Chapman and Hall.

Waagepetersen, R. (2007). An estimating function approach to inference for inhomogeneous Neyman–Scott processes. *Biometrics* **63**, 252–8.

Yang, J., He, H. S., Shifley, S. R. & Gustafson, E. J. (2007). Spatial patterns of modern period human-caused fire occurrence in the Missouri Ozark Highlands. *Forest Sci.* **53**, 1–15.

[*Received April* 2007. *Revised March* 2008]