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

By YONGTAO GUAN,

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

Summary

We introduce a formal testing procedure to assess the goodness-of-fit of a fitted inhomogeneous spatial Poisson process model. Our method is based on a discrepancy measure function $D_c(t;\hat{\theta})$ that is constructed by using residuals obtained from the fitted model. We derive the asymptotic distributional properties of $D_c(t;\hat{\theta})$ and then develop a test statistic based on these properties. Our test statistic has a limiting standard normal distribution so the test can be performed by simply comparing the test statistic with critical values obtained from the standard normal distribution. 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 diagnostics.

Short Title. Goodness-of-Fit Test for Poisson Processes.

1. Introduction

Spatial Poisson processes play an important role in both statistical theories (Daley & Vere-Jones, 1988, Ch. 2) and applications (Diggle, 2003, Ch. 2). A main interest for spatial Poisson processes has been concerned with testing for complete spatial randomness, i.e., if a process is homogeneous Poisson. For this a large number of testing methods have been proposed. For a survey of the results see Cressie (1993, Ch. 8) and Diggle (2003, Ch. 2).

In recent years, many carefully collected spatial point pattern data have become available where not only the spatial point pattern itself but also detailed covariates information associated with it are recorded. As a result, it becomes possible and in fact

often necessary to model the observed spatial point pattern in terms of the observed covariates. To do so the underlying spatial point process has to be treated as inhomogeneous. For many of these data examples, an inhomogeneous spatial Poisson process model appears to be appropriate since the correlation in the data may be negligible, e.g., locations of cancer patients in a region (Diggle, 1990), of human-caused wildfire in a highland (Yang et al., 2007), and of reseeders after a fire given the locations of resprouters in a forest (Illian et al., 2007). The main modeling task is to estimate the intensity function of the process, which is often written as a parametric function of the observed covariates. Maximum likelihood estimation in general can be performed with ease by using the computationally efficient algorithm of Berman and 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 further the problem when the covariates were only partially observed.

Once a model has been fitted for the intensity function, the next step of the analysis usually 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 (e.g., Ogata, 1988; Schoenberg, 2003). Available testing procedures for complete spatial randomness can then be used to assess if a homogeneous Poisson process is appropriate for the transformed data and therefore to assess if the fitted model is appropriate for the original data. This procedure is very useful for one dimensional data but can also be generalized to the spatial Poisson process setting (e.g., Diggle, 1990). Other available procedures include a "deviance residual" approach proposed by Lawson (1993) and a "smoothed residual field" approach that was recently proposed in the seminal paper by Baddeley et al. (2005). For the "smoothed residuals", Baddeley et al. (2006) gave details on the theoretical properties of the residuals. Both procedures are intended only for graphical presentations of the respective residuals but can not be regarded as formal tests. Despite their usefulness, graphical procedures have their limitations due to the arbitrariness associated with the decision process. These procedures will be more useful if supplemented with formal testing results.

To obtain a formal test, it may at first appears reasonable to just extend the standard

procedure to test for complete spatial randomness to the inhomogeneous case, say by comparing the theoretical and empirical K-functions (Diggle, 2003, Ch. 7). Under the null hypothesis of a good fit, the theoretical and empirical K-functions should be close. Otherwise we may expect to see a large discrepancy between them. The p-value of the test can be obtained by comparing a properly defined discrepancy measure between the theoretical and empirical K-functions that are calculated from the data with those calculated from simulations of the fitted model. However, we conjecture that this approach may have a low power to detect a poor fit in the inhomogeneous case. This is 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 can be obtained by evaluating the intensity function directly.

In this paper we develop a new procedure to formally test for the goodness-of-fit of a fitted spatial Poisson process model. Our method is based on a special type of the residuals defined in Baddeley et al. (2005) and therefore should also be considered as a residual analysis approach. Specifically we develop a discrepancy function, $D_c(t;\hat{\theta})$, in terms of the residuals, where t is a specified distance. As can be seen in Section 2, the proposed discrepancy function nicely reflects the discrepancy between the fitted intensity function and the data directly. We derive the variance associated with $D_c(t;\hat{\theta})$ and prove that it is asymptotically normal with mean zero under the hypothesis that the fitted model is a good fit. Based on these results, we can formally assess the goodness-of-fit by simply testing if the mean of $D_c(t;\hat{\theta})$ is larger than 0 for a preselected t. As a direct consequence of our theoretical results, pointwise confidence intervals can be obtained for $D_c(t;\hat{\theta})$ at different t values. As a graphical diagnostic procedure, we can then plot $D_c(t;\hat{\theta})$ together with the obtained pointwise intervals across a reasonable range of t values so as to facilitate diagnostics.

The rest of the article is organized as follows. In Section 2 we define the discrepancy function and study its distributional properties. We develop our testing method in Section 3, and then assess its performance through both a simulation study in Section 4 and and an application to a real data example in Section 5. We give the concluding

remarks in Section 6. Proofs of theorems are contained in the Appendix.

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 $H_0 : \lambda(\cdot) = \lambda(\cdot;\theta_0)$ for some unknown θ_0 by using the observed data. Note that the formulation of our problem is consistent with that of standard goodness-of-fit test for regression problems (e.g., Kutner et al., 2004, Ch. 3).

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

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

Note that $r_c(\mathbf{x}, t; \hat{\theta})$ defined in (1) 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 for the variance of $N(\mathbf{x}, t)$. Furthermore, the Poisson assumption implies that $N(\mathbf{x}, t)$ is also an unbiased estimator for the same quantity. In view of these observations, we thus define the following discrepancy function:

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

To better understand the motivation for the use of $D_c(t; \hat{\theta})$, we now comment on its properties. To simplify notation, let $\Lambda_c(\mathbf{x}, t; \hat{\theta})$ denote the integral term in (1). We will also suppress the dependence of $N(\mathbf{x}, t)$, $B(\mathbf{x}, t)$, $r_c(\mathbf{x}, t; \hat{\theta})$, $\Lambda_c(\mathbf{x}, t; \hat{\theta})$ and $D_c(t; \hat{\theta})$ on

t and thus rewrite them as $N(\mathbf{x})$, $B(\mathbf{x})$, $r_c(\mathbf{x};\hat{\theta})$, $\Lambda_c(\mathbf{x};\hat{\theta})$ and $D_c(\hat{\theta})$, respectively. Let $r(\mathbf{x})$ and $\Lambda(\mathbf{x})$ be (1) and the integral term in (1), respectively, with $\lambda_c(\mathbf{u};\hat{\theta})$ in (1) being replaced by the true intensity function $\lambda(\cdot)$. Straightforwardly, it can be seen that

$$\begin{aligned} D_c(\hat{\theta}) &= \int_A [\{r(\mathbf{x})\}^2 - N(\mathbf{x})]d\mathbf{x} - 2\int_A r(\mathbf{x})\{\Lambda_c(\mathbf{x};\hat{\theta}) - \Lambda(\mathbf{x})\}d\mathbf{x} \\ &+ \int_A \{\Lambda_c(\mathbf{x};\hat{\theta}) - \Lambda(\mathbf{x})\}^2 d\mathbf{x}. \end{aligned} \tag{3}$$

The expected value of the first term on the right hand side of (3) is equal to zero since $N(\mathbf{x})$ is a Poisson random variable and thus its mean and variance are equal. Heuristically if we treat $\hat{\theta}$ as a fixed value (i.e. not random), then the expected value of the second term is also equal to zero. Note that 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 $\hat{\theta} \approx \theta_0$, i.e. the fitted model is a good fit, then the last term of (3) should also be close to zero. As a result, 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 $\lambda_c(\cdot;\hat{\theta})$ then will be very different and therefore the last term of (3) will be larger than zero. As a result, the expected value of $D_c(\hat{\theta})$ will be larger than zero. More specifically, the more $\lambda_c(\cdot;\hat{\theta})$ deviates from $\lambda(\cdot)$, the larger the expected value of $D_c(\hat{\theta})$ tends to be in general. To assess the goodness-of-fit of a fitted model, we thus essentially need to assess whether 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.

We will study the distributional properties of $D_c(\hat{\theta})$ in the next section. Before we proceed, we comment on the behavior of $D_c(\hat{\theta})$ when the spatial point process is not Poisson. In this case, the expectation of the first term on the right hand side of (3) is no longer equal to zero. More specifically, it is generally larger than zero if the process is positively correlated (e.g., clustered). Thus both the misspecification 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 Section 4.

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

To formally compare $D_c(\hat{\theta})$ with zero, we need to study its distributional properties. For this let D be the discrepancy measure defined in (2) by using the true intensity function $\lambda(\cdot)$, that is,

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

where $\Lambda(\mathbf{x}) = \int_{B(\mathbf{x})\cap A} \lambda(\mathbf{u})d\mathbf{u}$. We will first study the asymptotic distribution of D and then link it to that of $D_c(\hat{\theta})$. This consideration is primarily for the ease of derivations since D has a simpler form than $D_c(\hat{\theta})$.

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

$$K_1 n^2 \leq |A_n| \leq K_2 n^2 \text{ and } K_1 n \leq |\partial A_n| \leq K_2 n. \tag{5}$$

Condition (5) essentially requires that A_n become increasingly large in all directions. This is typically satisfied by most commonly encountered domain shapes 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

$$C_1 \leq \lambda(\mathbf{x}) \leq C_2 \text{ for all } \mathbf{x} \in \mathbb{R}^2. \tag{6}$$

Condition (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 for D_n and $D_{c,n}(\hat{\theta}_n)$, respectively.

Theorem 1. *Let* $\sigma_n^2 = 2\int_{A_n}\int_{A_n}\{\Lambda_n(\mathbf{x},\mathbf{y})\}^2 d\mathbf{x}d\mathbf{y}$*, where* $\Lambda_n(\mathbf{x},\mathbf{y})$ *is* $\Lambda_n(\mathbf{x})$ *obtained on* $B(\mathbf{x},\mathbf{y})$ *and* $B(\mathbf{x},\mathbf{y}) = B(\mathbf{x}) \cap B(\mathbf{y}) \cap A_n$*. Assume that conditions (5) and (6) hold. Then*

$$D_n/\sigma_n \to N(0,1) \text{ in distribution as } n \to \infty.$$

Proof. See the Appendix. □

Theorem 2. *Let* $\sigma_{c,n}^2(\hat{\theta}_n) = 2\int_{A_n}\int_{A_n}\{\Lambda_{c,n}(\mathbf{x},\mathbf{y};\hat{\theta}_n)\}^2 d\mathbf{x}d\mathbf{y}$*. Assume that conditions (5) and (6) hold and* $\lambda_c(\cdot;\theta)$ *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*

$$D_{c,n}(\hat{\theta}_n)/\sigma_{c,n}(\hat{\theta}_n) \to N(0,1) \text{ in distribution as } n \to \infty.$$

Proof. See the Appendix. □

A subtle issue in practice is to decide which type of asymptotic framework is more appropriate, e.g., increasing-domain or infill? The latter says that the number of observations increases with n but the study region remains fixed. This type of framework appears to be more appropriate for data that are accumulated over time in a fixed region (e.g., many public health data). For Poisson processes, however, the difference between these two frameworks is probably not so important due to the complete independence among events. We note here that our main results continue to be true if we replace t in (1) by t_n where $t_n = t/n$ for some fixed t and replace (5) and (6) respectively by

$$K_1 < |A_n| < K_2 \text{ and } K_1 \le |\partial A_n| \le K_2,$$

and

$$C_1 n^2 \le \lambda(\mathbf{x}) \le C_2 n^2 \text{ for all } \mathbf{x} \in \mathbb{R}^2.$$

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. Note that unlike D_n in Theorem 1, $D_{c,n}(\hat{\theta}_n)$ is a biased estimator for zero due to 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 data with a modest sample

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, denote $a_n \sim b_n$ 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

$$|A_n|^{1/2}(\hat{\theta}_n - \theta_0) \sim \{\mathbf{V}_n(\theta_0)\}^{-1}\mathbf{U}_n \to N(\mathbf{0}, \{\mathbf{V}_n(\theta_0)\}^{-1}) \text{ in distribution,} \tag{7}$$

where

$$\mathbf{U}_n = \frac{1}{\sqrt{|A_n|}}\Big\{ \sum_{\mathbf{y}\in N\cap A_n} \frac{\lambda_c^{(1)}(\mathbf{y};\theta_0)}{\lambda_c(\mathbf{y};\theta_0)} - \int_{A_n} \lambda_c^{(1)}(\mathbf{x};\theta_0)d\mathbf{x}\Big\}, \tag{8}$$

$$\mathbf{V}_n(\theta_0) = \frac{1}{|A_n|}\int_{A_n} \frac{\lambda_c^{(1)}(\mathbf{x};\theta_0)\{\lambda_c^{(1)}(\mathbf{x};\theta_0)\}'}{\lambda_c(\mathbf{x};\theta_0)}d\mathbf{x}. \tag{9}$$

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

$$\mathbf{W}_n(\theta_0) = \frac{1}{|A_n|}\int_{A_n} \Lambda_{c,n}^{(1)}(\mathbf{x};\theta_0)\{\Lambda_{c,n}^{(1)}(\mathbf{x};\theta_0)\}'d\mathbf{x}, \tag{10}$$

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

$$|A_n|(\hat{\theta}_n - \theta_0)'W_n(\theta_0)(\hat{\theta}_n - \theta_0) - 2(\hat{\theta}_n - \theta_0)'\int_{A_n} r_n(\mathbf{x})\Lambda_{c,n}^{(1)}(\mathbf{x};\theta_0)d\mathbf{x}. \tag{11}$$

The expected value of the first term in (11) can be approximated by the trace of the matrix $\mathbf{W}_n(\theta_0)\mathbf{V}_n(\theta_0)^{-1}$ 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 due to condition (7). For the second term, lengthy yet elementary derivations yield that its expected value can be approximated by

$$-\frac{2}{|A_n|}\int_{A_n} \{\Lambda_{c,n}^{(1)}(\mathbf{x};\theta_0)\}'\mathbf{V}_n(\theta_0)^{-1}\Lambda_{c,n}^{(1)}(\mathbf{x};\theta_0)d\mathbf{x}.$$

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

$$-\frac{2}{|A_n|}\int_{A_n} \{\Lambda_{c,n}^{(1)}(\mathbf{x};\theta_0)\}'\mathbf{V}_n(\theta_0)^{-1}\Lambda_{c,n}^{(1)}(\mathbf{x};\theta_0)d\mathbf{x} + trace\{\mathbf{W}_n(\theta_0)\mathbf{V}_n(\theta_0)^{-1}\}, \tag{12}$$

where $\mathbf{V}_n(\theta_0)$ and $W_n(\theta_0)$ are defined in (9) and (10), respectively. In practice we replace θ_0 in (12) by its estimates $\hat{\theta}_n$ so as to obtain an estimate 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 $D_{c,n}(\hat{\hat{\theta}}_n)$ on each simulated realization. An estimate for the bias can be defined as the sample average of all obtained $D_{c,n}(\hat{\hat{\theta}}_n)$. This approach is being 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 for the fitted model $\lambda_c(\cdot;\hat{\theta})$. Specifically for a prespecified t, we calculate the statistic $T(\hat{\theta}) = \{D_c(\hat{\theta}) + bias(\hat{\theta})\}/\sigma_c(\hat{\theta})$, where $bias(\hat{\theta})$ is an estimate for the bias term in (12) and $\sigma_c^2(\hat{\theta}) = 2\int_A\int_A\{\Lambda_c(\mathbf{x},\mathbf{y};\hat{\theta})\}^2 d\mathbf{x}d\mathbf{y}$. 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 due to the insufficient sample size used to calculate $[\{r_c(\mathbf{x};\hat{\theta})\}^2 - N(\mathbf{x})]$ for each $\mathbf{x}$, based on which $D_c(\hat{\theta})$ in (2) is defined. This is because for a small t there is often too much noise in $[\{r_c(\mathbf{x};\hat{\theta})\}^2 - N(\mathbf{x})]$, which in turn will hide any signal for the 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 $r_c(\mathbf{x};\hat{\theta})$ may be too small compared to $N(\mathbf{x})$. Furthermore, a large t value can also deteriorate the size of the test due to two reasons. Firstly, note the fact that $[\{r_c(\mathbf{x};\hat{\theta})\}^2 - N(\mathbf{x})]$ is skewed to the right. If t is too large, then we won't have enough replicates for $[\{r_c(\mathbf{x};\hat{\theta})\}^2 - N(\mathbf{x})]$ so the normal approximation won't work well. Secondly, a large t will also lead to increased edge effects, where edge effects here refer to that events near the boundary are given less weights than those in the center when forming $D_c(\hat{\theta})$. This will further reduce the effective sample size. Thus t must be selected carefully. Note that this is not a problem unique to our test statistic. When calculating the scan statistic, for example, one also needs to decide the size of the scanning window (e.g., Kulldorff, 1999). We will perform a simulation study in the next

section to evaluate the effect of t on the performance of the test.

It will be desirable to have a data-driven approach for the selection of t. To do so, we note that t affects the test mostly through its effect on $r_c(\mathbf{x};\hat{\theta})$. Thus the problem to select t for $D_c(\hat{\theta})$ can be roughly treated as the problem to select the bandwidth used to obtain the residuals $r_c(\mathbf{x};\hat{\theta})$, where $\mathbf{x} \in A$. For the latter Baddeley et al. (2005) discussed several possible data-driven methods. We suggest to use one of these methods to select the bandwidth for $r_c(\mathbf{x};\hat{\theta})$ and thus to select t for $D_c(\hat{\theta})$. From a practical point of view, our proposal is reasonable since one may wish to first obtain and examine the "smoothed" residual plot (i.e. the plot of $r_c(\mathbf{x};\hat{\theta})$). Our formal testing procedure can then be used to formally 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 quickly examine the evidence 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 value of an arbitrary point on the unit square. Throughout this section, we will 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 (e.g., Diggle, 2003, P. 66). We set the standard deviation of the dispersal variable at .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 $\beta = 1, 2$. Note that the larger β was, the more inhomogeneous the process was. For each type of intensity function and for each given β, we manipulated the value of α such that the expected

number of events per realization (denoted by μ) was roughly 100 and 400. In terms of the asymptotic frameworks discussed in Section 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 .1, .2 and .3. For each realization and each given 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 cross-validation (Silverman, 1999). To reduce computational time, we considered only fifteen equally spaced t values between .02 and .3. The "optimal" t selected by least-squares cross-validation could be larger than .3. We nevertheless imposed an upper limit .3 for t since it did not appear wise to use a larger t for a unit square. Recall from the discussion in Section 3 that we needed t to be small enough so that the normal approximation could work.

To compare with existing methods, we also applied two other competing tests. The first was a simulation-based approach by comparing 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 a proper transformation. The test was then done by using standard goodness-of-fit tests to test for uniformity for 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 non-stationary 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 $K(r)$ (e.g., Ho & Chiu, 2007):

$$DK_1 = \sup_{r\in[0,r_0]} |\hat{K}(r)^{1/2} - K(r)^{1/2}|,$$

$$DK_2 = \int_0^{r_0} \{\hat{K}(r)^{1/2} - K(r)^{1/2}\}^2 dr,$$

where $r_0 = .125, .062$ 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 (or DK_2) from the original realization ranked in the top 10% of the pooled DK_1 (or 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 thus will present only the results for the former.

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

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

In the above, $\lambda_c(\cdot)$ was either the linear model or the sine model. For the function $f(\cdot)$, we used $f(\mathbf{x}) = x$ for the linear model and $f(\mathbf{x}) = \sin(2\pi x)$ 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. Note that when the true intensity model was used, the resulting percentages of rejections (i.e., sizes) were all close to the nominal size. This provided evidence that our asymptotic results were appropriate for these models and data. When a wrong intensity model was used, the resulting percentages of rejections (i.e., powers) increased as the expected number of events increased. This was in accordance with the general knowledge that a larger sample size should lead to a more powerful test. When t increased, the power first increased and then decreased. This agreed with our general comment regarding the effect of t on the test. Simulation results not included also suggested that a too large t value (say $t = .4$) not only deteriorated the size of the test 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 $t = .3$ roughly corresponded to 11 independent replicates. Note that the actual sample size might be slightly bigger since $D(\hat{\theta})$ integrated $[\{r_c(\mathbf{x}; t; \hat{\theta})\}^2 - N(\mathbf{x}, t)]$ over all $\mathbf{x} \in A$, but not over just a set of $\mathbf{x}_i$, $i = 1, \cdots, |A|/|B(\cdot)|$, such that $\cup_i B(\mathbf{x}_i) = A$ and $\cap_i B(\mathbf{x}_i) = \emptyset$. Nevertheless it does not appear appealing to use an even larger t. One possibility in practice is to set an upper limit (say .3) for t and then use least-squares cross-validation to select t. From Table 1 we see this approach worked well in our simulation.

Figure 1 plots the true intensity model (linear or sine) and the average of the fitted incorrect intensity models obtained from the simulation for each situation. For the models being considered here, the difference between the true and the incorrect models was higher when the true model was the sine model for each fixed β, and increased as β increased for each fixed model. Reflected from the powers in Table 1, we see 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, we see our test was much more powerful in all cases. The improvement often was 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 likely 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 (i.e. $\hat{K}(r)$, $r \in [0, r_0]$, this is a global measure since each $\hat{K}(r)$ was obtained by pooling all data together) and then combined these global measures to derive another global measure for the overall discrepancy. As a result, valuable information related to local discrepancies might 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, we see our test was more powerful in the

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

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. Note that regardless of which intensity model was fitted, all three tests concluded a 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 one 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. However, the improvement became much smaller. When $\mu = 400$, both approaches rejected H_0 all the time. The transformation approach, however, rejected H_0 much less frequently especially when $\mu = 100$. Note that the transformation approach was defined in terms of the intensity function only. Thus 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.

5. An Application

We applied our testing method to a real data example from an epidemiological study. 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. The data were given and analyzed in Diggle (1990). The main interest of the study is to model the locations of larynx cancer cases in relation to the location of an industrial incinerator. To do so, Diggle (1990) fitted an inhomogeneous Poisson process model to the data where the distance from each larynx cancer case to the location of the incinerator served as a covariate. Specifically, he used the following model

$$\lambda(\mathbf{x}) = \rho\lambda_0(\mathbf{x})\{1 + \alpha \exp(-\beta||\mathbf{x} - \mathbf{x}_0||^2)\}.$$

In the above, ρ is the overall number of events per unit area, $\lambda_0(\cdot)$ is the spatial intensity of the population at risk and $\mathbf{x}_0$ 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 $\sigma = .15$ km. Diggle (1990) obtained the estimates $(\hat{\alpha}, \hat{\beta}) = (23.67, 0.91)$ using a maximum likelihood approach, whereas Diggle & Rowlinson (1994) obtained the estimates $(\hat{\alpha}, \hat{\beta}) = (33.69, 1.11)$ 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 in an increasing order in terms of their distances to $\mathbf{x}_0$, where $\mathbf{x}_0$ is the location of the incinerator. Let E_i denote the disc with center $\mathbf{x}_0$ and radius equal to the ith ordered distance. The following quantity was then defined

$$T_i = \int_{E_i} \lambda_0(\mathbf{x})\{1 + \hat{\alpha}\exp(-\hat{\beta}||\mathbf{x} - \mathbf{x}_0||^2)\}, \quad i = 1, \cdots, 58.$$

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 goodness-of-fit for the second model by using a residual analysis approach. Specifically, they considered the residuals in (1) where $\mathbf{x}$ therein was equal to $\mathbf{x}_0$ 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) concluded 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 (i.e., t) for the squares being used, and the corresponding 99% confidence limits. There was a striking evidence that none of the two fitted models appeared to be a good fit since all $T(\hat{\theta})$ were above the 99% confidence limits. Note that our conclusion is contradictory to that of Diggle (1990) in which the author acknowledged the limit of the approach being used therein. Furthermore, our analysis formally confirmed the lack-of-fit for the second model as detected by Baddeley et al. (2005). In a personal communication, Peter Diggle suggested that the lack-of-fit for the first model was likely 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 $(\hat{\alpha}, \hat{\beta}) = (33.69, 1.11)$ were obtained by using the conditional approach in Diggle & Rowlinson (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 a lack-of-fit. It should be noted that an estimate of $\lambda_0(\cdot)$ was required in order to calculate our test statistics for both fitted models. This in turn affected the calculated test statistics. Thus our analysis is only an illustrating example but not a thorough analysis.

6. Concluding remarks

In this paper we have introduced a formal testing method to assess the goodness-of-fit of a fitted inhomogeneous spatial Poisson process model. The test is based on a discrepancy measure function $D_c(t; \hat{\theta})$ that is constructed in terms of the residuals from the fitted model. We have theoretically justified the validity of the testing approach and compared through simulations its performance with the traditional K-function approach and an approach by transforming the data into a homogeneous Poisson process. In our simulations, the proposed test is consistently more powerful than the K-function approach, and performed competitively with, if not better than, the transformation approach to detect a lack-of-fit due to a misspecified intensity function alone. If the process is not Poisson, our test has a slightly higher (or similar) power to detect a lack-of-fit when compared to the K-function approach, and a much higher power when compared to the transformation approach. Based on evidence from our simulation study, we recommend selecting t within a reasonable predefined range by least-squares cross-validation.

When a lack-of-fit is detected, we need to decide whether the lack-of-fit is due to the use of an incorrect intensity function model or the existence of correlation, i.e. the process is not Poisson. This is often a difficult issue as heterogeneity in intensity and correlation especially clustering can lead to point patterns with similar characteristics (e.g., 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 on the intensity function. One sensible approach in practice may be to first 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 modeling 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 National Science Foundation. The author thanks Peter Diggle and Adrian Baddeley for their helpful comments, Andrew Lawson for pointing out an error in the reference list, and the Editor and two referees for their constructive comments that have greatly improved the paper.

Appendix

Proofs

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

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

The last equality is due to the fact that for any Poisson random variable X with an expected value μ,

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

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 $A'_n = \cup_i A_l^i$. Condition (5) guarantees that $|A'_n|/|A_n| \to 1$ as $n \to \infty$. Let D_l^i be D calculated on A_l^i. Define $D'_n = \sum_i^{k_n} D_l^i$ and $(\sigma'_n)^2 = var(D'_n)$.

We first want to show that

$$\frac{D_n}{\sigma_n} - \frac{D'_n}{\sigma'_n} \to 0 \text{ in probability.} \tag{13}$$

To show the above, we only need to show that

$$\frac{cov(D_n, D'_n)}{\sigma_n \sigma'_n} \to 1.$$

This follows from some lengthy yet elementary algebra due to the fact that $|A'_n|/|A_n| \to 1$ and condition (6). Then (13) holds from Chebyshev's inequality.

We then want to show that

$$\frac{D'_n}{\sigma'_n} \to N(0,1) \text{ in distribution.} \tag{14}$$

To prove (14), we first verify the following

$$\sup_n E(|D_l|^4) < cl^4. \tag{15}$$

Note that $(\sigma'_n)^2$ is of order n^2 due to condition (6) and the way in which the subblocks were constructed. (14) then follows trivially from (15) by the application of the Lyapunov's theorem since D_l^i are independent.

In what follows, let $\int$ stand for $\int_{D_l}$ unless specified otherwise. Define $B^{-1}(\mathbf{x}) = \{\mathbf{s} : \mathbf{x} \in B(\mathbf{s})\} \cap A_l$ and $B^{-1}(\mathbf{x}, \mathbf{y}) = B^{-1}(\mathbf{x}) \cap B^{-1}(\mathbf{y})$. To prove (15), we first rewrite D_l as

$$\Big[\sum_{\mathbf{x}\neq\mathbf{y}}\sum |B^{-1}(\mathbf{x},\mathbf{y})| - \int \{\Lambda(\mathbf{s})\}^2 d\mathbf{s}\Big] - 2\Big[\sum_{\mathbf{x}} \int_{B^{-1}(\mathbf{x})} \Lambda(\mathbf{s}) d\mathbf{s} - \int \{\Lambda(\mathbf{s})\}^2 d\mathbf{s}\Big].$$

Denote the two terms in the big brackets by F_l and G_l, respectively. A sufficient condition for (15) to hold is that

$$\sup_n E(|F_l|^4) < cl^4 \text{ and } \sup_n E(|G_l|^4) < cl^4.$$

Let $g(\mathbf{x}) = \int_{B^{-1}(\mathbf{x})} \Lambda(\mathbf{s}) d\mathbf{s}$. Tedious yet elementary algebra shows

$$E\{(G_l)^4\} = 6\Big[\int \{g(\mathbf{x})\}^2 \lambda(\mathbf{x}) d\mathbf{x}\Big]^2 + \int \{g(\mathbf{x})\}^4 \lambda(\mathbf{x}) d\mathbf{x}.$$

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

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

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

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

Lastly, it can be shown that the seven-fold and the eight-fold integrals are both equal to zero. Thus (15) holds.

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

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

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,

$$\Lambda_{c,n}(\mathbf{x};\hat{\theta}_n) - \Lambda_{c,n}(\mathbf{x};\theta_0) = (\hat{\theta}_n - \theta_0)'\Lambda^{(1)}_{c,n}(\mathbf{x};\theta^*_n),$$

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

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

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

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

$$\{D_{c,n}(\hat{\theta}_n) - D_n\}/|A_n|^{1/2} \to 0 \text{ in probability,}$$

and

$$\sigma_{c,n}(\hat{\theta}_n)/\sigma_n \to 1 \text{ in probability.}$$

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

REFERENCES

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

Baddeley, A. J., Møller, J. & Pakes, A. G. (2006). Properties of residuals for spatial point processes. *Annals of the Institute of Statistical Mathematics*, to appear.

Baddeley, A. J., Turner, R., Møller, J. & Hazelton, M. (2005). Residual analysis for spatial point processes (with discussion). *Journal of the Royal Statistical Society, Series B*, **67**, 617–666.

Besag, J. (1977). Contribution to the discussion of Dr. Ripley's paper. *Journal of the Royal Statistical Society, Series B*, **39**, 193–195.

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

Brix, A., Senoussi, R., Couteron, P. & Chaduf, J. (2001). Assessing goodness of fit of spatially inhomogeneous Poisson processes. *Biometrika*, **88**, 487–497.

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-Verlag.

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

Diggle, P. J. & Rowlinson, B. S. (1994). A conditional approach to point process modeling of elevated risk. *Journal of the Royal Statistical Society, Series A*, **157**, 433–440.

Diggle, P. J. (2003). *Statistical analysis of spatial point patterns*. New York: Oxford University Press.

Ho, L. P. and Chiu, S. N. (2007). Testing uniformity of a spatial point pattern. *Journal of Computational and Graphical Statistics*, **16**, 378–398.

Illian, J. B., Møller, J. & Waagepetersen, R. (2007). Spatial point process analysis for a plant community with high biodiversity. *Submitted*.

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

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

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

Ogata, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical Association*, **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. *Advances in Applied Probability*, **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**, 153–165.

Ripley, B. D. (1979). Tests of 'randomness' for spatial point patterns. *Journal of the Royal Statistical Society, Series B*, **41**, 368–374.

Schoenberg, F. P. (2003). Multidimensional Residual analysis of point process models for earthquake occurrences. *Journal of the American Statistical Association*, **98**, 789–795.

Silverman, B. W. (1999). *Density estimation for statistics and data analysis*. New York: Chapman & Hall.

Waagepetersen, R. (2007). An estimating function approach to inference for inhomogeneous Neyman-Scott processes. *Biometrics*, **63**, 252–258.

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 Science*, **53**, 1–15.

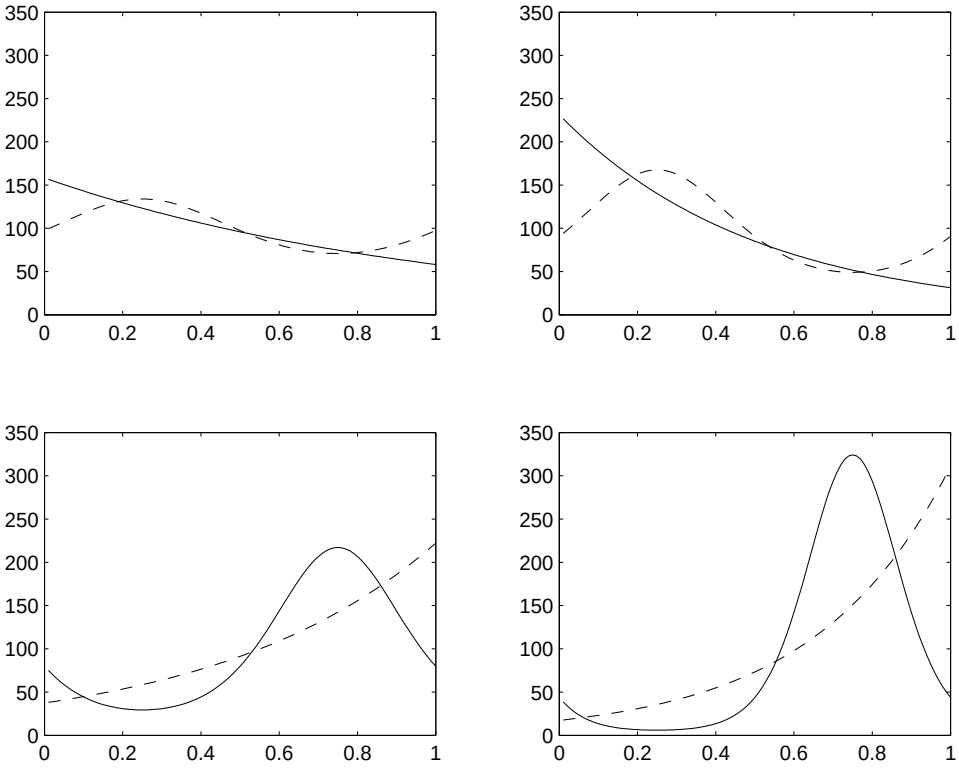

Figure 1: Plots of the true intensity function model (solid line) and the average of the fitted incorrect models (dashed line) versus x. The x axes in the plots are the x coordinate values on a unit square and the y axes are the true and estimated intensity function values. The top two plots are for the linear model and the bottom plots are for the sine model, where $\beta = 1, 2$ from the left to right.

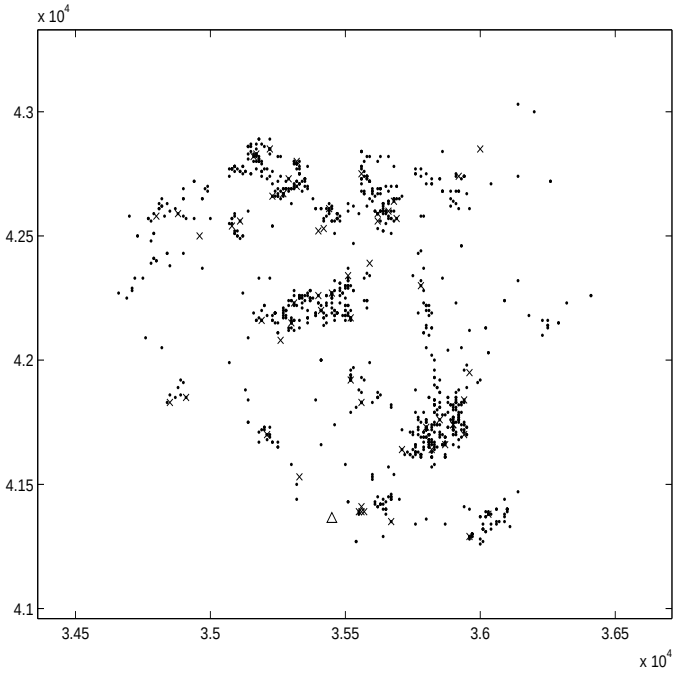

Figure 2: Locations of larynx and lung cancers. $\cdot$ stands for lung cancer, $\times$ stands for larynx cancer and $\triangle$ stands for the incinerator.

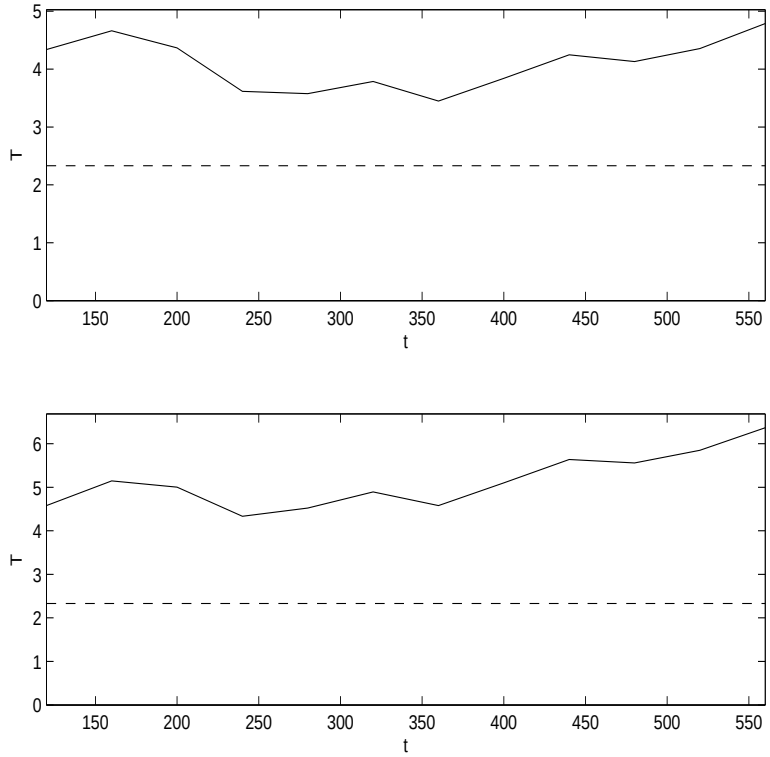

Figure 3: Residual plots for the Larynx cancer data. The y axis label T stands for the test statistic $T(\hat{\theta})$, calculated at different t values, where the t values are defined in meters. The top plot is for the model with $(\hat{\alpha}, \hat{\beta}) = (23.67, 0.91)$ and the bottom plot is for that with $(\hat{\alpha}, \hat{\beta}) = (33.69, 1.11)$. The solid lines are the test statistics and the dashed lines are the 99% confidence limits under each model.

Table 1: Percentages of rejections at the 10% nominal level in the Poisson process case. Test 1 and Test 2 are the proposed test with t fixed and selected by least-squares cross-validation, 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

Table 2: Percentages of rejections at the 10% nominal level in the Poisson cluster process case. The true intensity model is the linear model. Test 1 and Test 2 are the proposed test with t fixed and selected by least-squares cross-validation, 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			
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