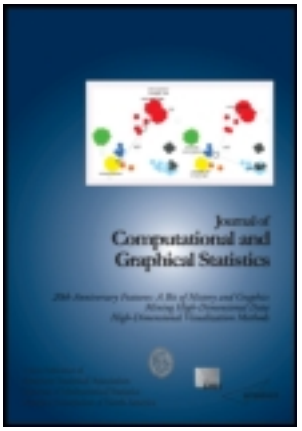

Journal of Computational and Graphical Statistics

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/ucgs20

A Monte Carlo-Adjusted Goodness-of-Fit Test for Parametric Models Describing Spatial Point Patterns

Ngoc Anh Dao & Marc G. Genton

Accepted author version posted online: 16 Jan 2013.Published online: 28 Apr 2014.

To cite this article: Ngoc Anh Dao & Marc G. Genton (2014) A Monte Carlo-Adjusted Goodness-of-Fit Test for Parametric Models Describing Spatial Point Patterns, Journal of Computational and Graphical Statistics, 23:2, 497-517, DOI: 10.1080/10618600.2012.760459

To link to this article: http://dx.doi.org/10.1080/10618600.2012.760459

Supplementary materials for this article are available online.
Please go to www.tandfonline.com/r/JCGS

A Monte Carlo-Adjusted Goodness-of-Fit Test for Parametric Models Describing Spatial Point Patterns

Ngoc Anh DAO and Marc G. GENTON

Assessing the goodness-of-fit (GOF) for intricate parametric spatial point process models is important for many application fields. When the probability density of the statistic of the GOF test is intractable, a commonly used procedure is the Monte Carlo GOF test. Additionally, if the data comprise a single dataset, a popular version of the test plugs a parameter estimate in the hypothesized parametric model to generate data for the Monte Carlo GOF test. In this case, the test is invalid because the resulting empirical level does not reach the nominal level. In this article, we propose a method consisting of nested Monte Carlo simulations which has the following advantages: the bias of the resulting empirical level of the test is eliminated, hence the empirical levels can always reach the nominal level, and information about inhomogeneity of the data can be provided. We theoretically justify our testing procedure using Taylor expansions and demonstrate that it is correctly sized through various simulation studies. In our first data application, we discover, in agreement with Illian et al., that *Phlebocarya filifolia* plants near Perth, Australia, can follow a homogeneous Poisson clustered process that provides insight into the propagation mechanism of these plants. In our second data application, we find, in contrast to Diggle, that a pairwise interaction model provides a good fit to the micro-anatomy data of amacrine cells designed for analyzing the developmental growth of immature retina cells in rabbits. This article has supplementary material online.

Key Words: Adjusted level; Distribution of p-values; Level bias; Nested Monte Carlo simulation; Spatial point process.

1. INTRODUCTION

Spatial point processes have received sustained attention in recent years in many disciplines such as ecology, biology, medicine, and material science, to name a few. For example, using spatial point processes, ecologists model the propagation mechanism of forests, biologists model the developmental growth of immature cells, and pharmacologists model the spatial structure of active agents of some drugs coupled to antigen sites (Illian et al.

Ngoc Anh Dao is Graduate Student, Department of Statistics, Texas A&M University, College Station, TX 77843 (E-mail: *dao@stat.tamu.edu*). Marc G. Genton is Professor, CEMSE Division, King Abdullah University of Science and Technology, Thuwal 23955, Saudi Arabia (E-mail: *marc.genton@kaust.edu.sa*).

Journal of Computational and Graphical Statistics, Volume 23, Number 2, Pages 497–517
DOI: 10.1080/10618600.2012.760459
Color versions of one or more of the figures in the article can be found online at *www.tandfonline.com/r/jcgs*.

2008). Inevitably, a goodness-of-fit (GOF) test is needed to assess the fit of these models. Up to now, GOF tests were primarily available for assessing the fits of simple Poisson processes. A large number of testing methods for assessing homogeneous Poisson processes, otherwise known as complete spatial randomness (CSR), are referenced and described in Cressie (1993, sec. 8) and Diggle (2003, sec. 2). Notable examples include a test based on the approximate distribution of the mean of the nearest neighbor distances (Donnelly 1978) and quadrat count tests (Diggle 2003, sec. 2.5). For sparsely sampled point patterns, Besag and Gleaves (1973) and Hines and Hines (1979) developed distance-based tests, whereas Assunção (1994) and Assunção and Reis (2000) developed an angle-based test. When the fitted model is an inhomogeneous Poisson process, Guan (2008) derived a GOF test with a statistic based on a discrepancy measure function that is constructed from residuals obtained from the fitted model. The test statistic has a limiting standard normal distribution, so the test can be performed by comparing the test statistic with available critical values. While the aforementioned methods are suitable for Poisson spatial point processes, they do not extend trivially to the fits of more complicated processes. An exploratory analysis of such data involves a diagnostic assessment of the residuals that can be handled by the graphical methods of Baddeley et al. (2005) and Baddeley, Møller, and Pakes (2008). This diagnostic approach, however, does not serve as a formal GOF test. A possible complement is the Monte Carlo GOF test (Ripley 1988; Diggle 2003), which is valid when, for example, the data consist of multiple spatial point patterns (SPPs). Analogous to training and validation datasets in regression analysis, one can obtain the parameter estimate for the null model from one, or from some, pattern(s) and apply the GOF test to the remaining patterns. This traditional Monte Carlo GOF test and its graphical version, the envelope test, are described in Section 2.2.

When the data consist of a single SPP, the parameter estimate has to come from the only pattern available, which suggests the concept of *the plug-in Monte Carlo GOF test* (PGOF), which is widely applied in the literature. However, in the general statistical context, Robins, Van der Vaart, and Ventura (2000) noted that the p-value resulting from the PGOF test does not necessarily follow the standard uniform distribution, not even in the asymptotic sense. In particular, in the context of spatial point processes, Diggle (2003, sec. 6.2) noted that "Such tests are strictly invalid, and probably conservative, if parameters have been estimated from the data. To some extent, this problem can be alleviated if we use a GOF statistic that is only loosely related to the estimation procedure." Similarly, in the general statistical context, Bayarri and Berger (2000) suggested that if the distribution of the GOF test statistic is conditioned upon a sufficient statistic, then this distribution does not depend on the true parameter and therefore the p-value calculated with respect to this conditional distribution has a standard uniform distribution. However, in the context of spatial point processes, except for the homogeneous Poisson process, one cannot find closed forms for sufficient statistics. Hence, the approach motivated by Bayarri and Berger (2000) cannot be applied. Diggle (2003, p. 10) also noted that "An inherent weakness of the Monte Carlo approach is its restriction to simple hypotheses[. . .]. Composite hypotheses can be tested if pseudo-random sampling is made conditional on the observed values of sufficient statistics for any unknown parameters, but this is seldom practicable. Note that a goodness-of-fit test which ignores the effects of estimating parameters will tend to be conservative." Diggle (2003, p. 10) noted further that this particular conservativeness does not arise with tests of CSR for mapped data,

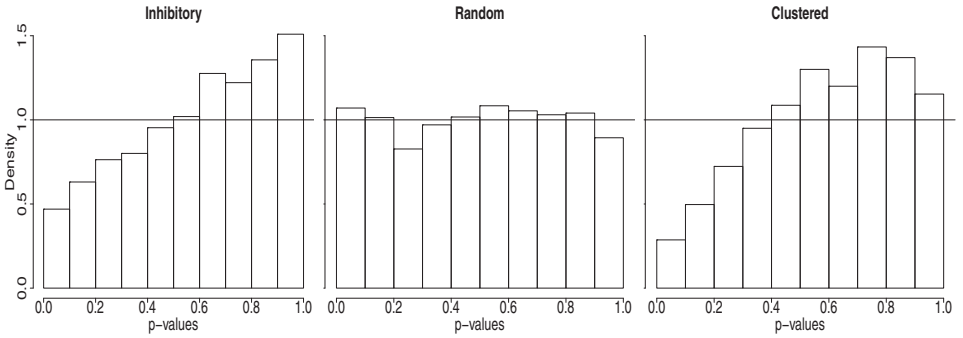

Figure 1. Distributions of the estimated p-values $\hat{p}$'s resulting from PGOF tests. The Strauss processes (left), CSR (center), and the Thomas processes (right) are defined in Section 3.1. The nearest neighbor distance distribution function, G-function, defined in Section 2.2, is used to compute the test statistic for the PGOF tests. The horizontal solid lines stand for the probability density function of the standard uniform distribution.

because the observed number of events, N, is sufficient for the intensity, λ, and, conditional on N, CSR is a simple hypothesis. But ignoring the effect of estimating parameters does affect the assessment of GOF for more general stochastic models. Figure 1 shows how far off the empirical level can be from the nominal level, $\alpha = 0.05$. The histograms show from left to right the distributions of the estimated p-values from the PGOF test assessing the fits of Strauss, CSR, and Thomas processes, which are described in Section 3.1. CSR is a simple hypothesis; hence, the distribution of the estimated p-values is standard uniform. However, by letting $\mathbb{X}$ denote the process underlying the observed SPP defined in Section 2.1, the Strauss process ($H_0 : \mathbb{X} \sim f_{\beta,\gamma,R}$, where f denotes the probability density function) and the Thomas process ($H_0 : K(\mathbb{X}) = K_{\kappa,\sigma,\mu}$, where K denotes the K-function) are composite hypotheses because there is no knowledge about the existence or the form of sufficient statistics. The distributions of the estimated p-values are not standard uniform. The PGOF test is too conservative for these two processes. It is important to correct for the bias of its empirical level because if the Type II error rate were prespecified, the PGOF test would suffer from power loss as noted by Marriott (1979): the Monte Carlo test gives a "blurred" critical region, which is not of the usual form. A conventional significance test would reject H_0 if the statistic lays in some well-defined critical region. However, for the Monte Carlo test, there is a range of values of the statistic over which there is a varying probability of rejecting H_0. This blurring of the critical region leads to a loss of power. This weakness of the PGOF test still remains when using the envelope test (Ripley 1977). If the Monte Carlo simulation is not made conditional on an observed value of a sufficient statistic, there is an undesirably high probability that the upper and lower envelopes of the null model would contain the empirical summary or second-order characteristic function of the observed SPP, which does not underlie the null model. To overcome the weaknesses with respect to the conservativeness and power loss of PGOF tests, we propose a test that adjusts for the bias of the empirical level through nested Monte Carlo simulations to reach the nominal level correctly.

The remainder of this article is organized as follows. Section 2 describes the methodology of the proposed GOF procedure. Extensive simulation studies demonstrate in Section 3 that our method gives more accurate empirical levels than those provided by the existing PGOF

test. Section 4 provides two data applications, one to a forest dataset to study the propagation mechanism of *Phlebocarya filifolia* plants, and the other to micro-anatomy data to learn about the developmental growth of immature retina cells in rabbits. Finally, Section 5 discusses two computational techniques that make the computations more efficient.

2. MONTE CARLO GOODNESS-OF-FIT TESTS

2.1 DEFINITIONS

Hereafter, we adopt the notation given in Baddeley et al. (2005, sec. 5). An SPP is a dataset, $X = \{x_1, \ldots, x_N\}$, where the x_i's are unordered locations observed in a bounded region, W, of $\mathbb{R}^2$. We let f_θ denote the parametric model (a parametric spatial point process) fitted to X, where θ is an arbitrary finite-dimensional vector of parameters. We assume that $f_\theta(X)$ is a probability density function with respect to the unit rate Poisson process on the window W, such that f_θ satisfies the positivity condition: if $f_\theta(X) > 0$ and $Y \subset X$, then $f_\theta(Y) > 0$ for any finite point patterns, $X, Y \subset W$. Under this setup, we are interested in testing the null hypothesis:

$$H_0 : \mathbb{X} \sim f_\theta, \tag{1}$$

where $\mathbb{X}$ is the spatial point process from which the observed SPP, X, is generated. The common p-value for the hypothesis above is $p = 1 - \Pr\{t(\mathbb{X}) < t(X)\}$ for some test statistic, $t(\cdot)$, that is, this definition of the p-value is based on the assumption that a small value of $t(X)$ supports the null hypothesis, H_0. If the probability density of $t(\cdot)$ is tractable, then a Monte Carlo GOF test is not necessary because the computation of the p-value is with respect to the probability density of $t(\cdot)$, and the distribution of the p-value is exactly standard uniform for a known θ, and approximately standard uniform for an unknown θ under some conditions (Bayarri and Berger 2000). When, however, the probability density of $t(\cdot)$ is analytically unavailable, a Monte Carlo GOF test is an alternative to the classical methods to estimate the p-value to assess the fit.

2.2 A TRADITIONAL MONTE CARLO GOODNESS-OF-FIT TEST

As a basis for assessing (1), we first briefly describe the test from Diggle (2003, sec. 6), who proposed a test statistic, u, and the pseudostatistics, u_i, using the summary characteristic G-function, which will be described after introducing the estimated p-value in (5):

$$u = \int_0^\infty \{\hat{G}(h) - \overline{G}(h)\}^2 dh; \quad u_i = \int_0^\infty \{\hat{G}_i(h) - \overline{G}_i(h)\}^2 dh, \quad i = 1, \ldots, n, \tag{2}$$

where $\hat{G}(h)$ is the empirical G-function of the observed SPP, X, $\hat{G}_i(h)$ is the empirical G-function of the simulated SPP, X_i, $i = 1, \ldots, n$, under (1) with $\theta = \hat{\theta}$ *for some* $\hat{\theta}$ *not necessarily from the data*, and n is the number of Monte Carlo simulations. Also,

$$\overline{G}(h) = \frac{1}{n}\sum_{i=1}^{n} \hat{G}_i(h); \quad \overline{G}_i(h) = \frac{1}{n}\left\{\hat{G}(h) + \sum_{l=1, l\neq i}^{n} \hat{G}_l(h)\right\}, \quad i = 1, \ldots, n. \tag{3}$$

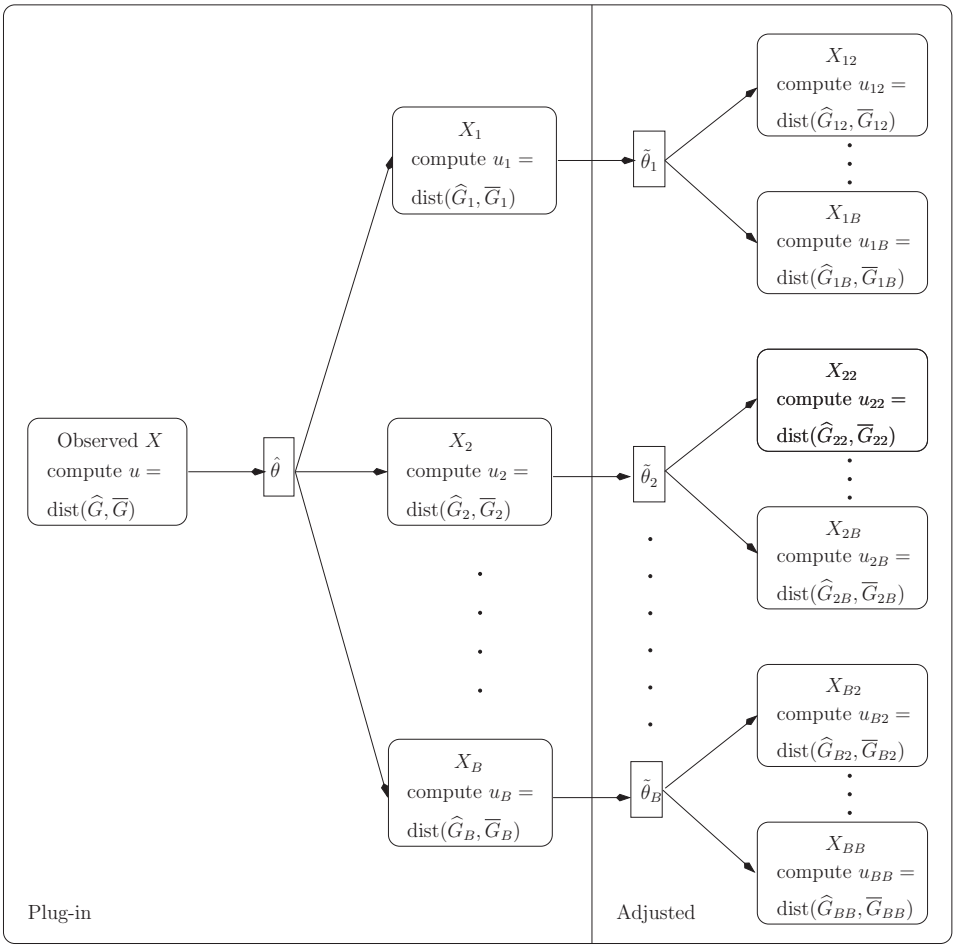

Figure 2. The test statistic, u, and its estimated p-value, $\hat{p}$, in (5) are computed using the Monte Carlo simulations represented in the "Plug-in" part. $\text{dist}(s, t) = \int_0^\infty \{s(h) - t(h)\}^2 dh$ is the squared L_2-norm of the difference between the functions s and t. The PGOF test rejects H_0 if $\hat{p} \le \alpha$, where α is the nominal level. The AGOF test would reject H_0 if $\hat{p} \le \hat{\alpha}^*$, where $\hat{\alpha}^*$ is the estimated adjusted level, which is computed in the "Adjusted" step. The Monte Carlo simulation is applied to each X_i, $i = 1, \ldots, n$, to obtain the estimated p-values $\hat{p}_1, \hat{p}_2, \ldots, \hat{p}_n$, to solve for $\hat{\alpha}^*$ in (11).

For a graphical depiction of the Monte Carlo GOF test under the plug-in fashion, see the "Plug-in" part in Figure 2. The p-value for the Monte Carlo GOF test in (1) is

$$p = 1 - \Pr(U_{\hat{\theta}} < u), \tag{4}$$

where $U_{\hat{\theta}}$ is the random variable taking u_i, $i = 1, \ldots, n$, as realizations. Practically, this p-value can be estimated from

$$\hat{p} = 1 - \frac{1 + \sum_{i=1}^{n} I(u_i < u)}{n + 1}, \tag{5}$$

where $I(\cdot)$ is the indicator function. The PGOF test would reject H_0 if $\hat{p} \leq \alpha$. Let $\hat{\mathcal{P}}$ denote the random variable corresponding to $\hat{p}$. Its distribution as noted in Section 1 is not standard uniform, except for $\mathbb{X}$ underlying a CSR.

Analogously, $u, u_i, i = 1, \ldots, n$, and $\hat{p}$ can also be computed using the other summary characteristic, the F-function, the second-order characteristic, K-, or the inhomogeneous K-function, for instance. The terminology of the summary characteristic and that of the second-order characteristic function are adopted from Illian et al. (2008). The G-function, also known as the nearest neighbor distance distribution function, is the distribution function of the distance from an arbitrary *event* in X to the nearest other event in X, and the F-function, otherwise known as the empty space function, is the distribution function of the distance from an arbitrary *point* in the chosen window, W, to the nearest event in X. Both are estimated empirically using $\widehat{G}(h) = \sum_{i=1}^{N} I(h_i \leq h)/N$ and $\hat{F}(h) = \sum_{i=1}^{m} I(h_i^* \leq h)/m$, where $h > 0$, $N \equiv N(X)$ is the number of events in X, h_i denotes the distance from the ith event in X to the nearest other event in X, and h_i^* is the distance from the ith point of a sample of m selected points in W to the nearest of N events in X. The theoretical K-function of a stationary spatial point process is defined as $K(h) = \lambda^{-1}E$ (the number of extra events within distance h of a randomly chosen event), where $h \geq 0$ and λ is the intensity of the process (Bartlett 1964; Ripley 1977). Ripley's estimator of the K-function is, for instance, in Cressie (1993, p. 640). The inhomogeneous K-function as the K-function for nonstationary processes is introduced by Baddeley, Møller, and Waagepetersen (2000).

The envelope test is a graphical view of the Monte Carlo GOF test. Using the G-function, for instance, it declares a model as a good fit if the upper envelope $U(h)$ and the lower envelope $L(h)$ envelop the $\hat{G}$-function of the observed SPP X, where $U(h) = \max_{i=1,\ldots,n} \hat{G}_i(h)$, $L(h) = \min_{i=1,\ldots,n} \hat{G}_i(h)$, with typically $n = 100$.

To overcome the weaknesses with respect to the conservativeness and the power loss of PGOF tests, we propose a test that adjusts for the bias of its empirical level through nested Monte Carlo simulations to reach the nominal level correctly.

2.3 THE MONTE CARLO-ADJUSTED GOODNESS-OF-FIT TEST

The motivation for our proposed test arises as follows. To make the PGOF test *correctly sized*, that is, its empirical level, $\hat{\alpha}$, reaches the nominal level, α, our goal is to find $\hat{\alpha}^*$ such that

$$\Pr\{\hat{\mathcal{P}} < \hat{\alpha}^*\} = \alpha, \tag{6}$$

where $\hat{\alpha}^*$ is an estimate of α^*, which is referred to as *the adjusted level*. Usually α^* is unknown, and $\alpha^* = \alpha$ only in case of CSR. The decision rule of rejecting H_0 if $\hat{p} \leq \hat{\alpha}^*$ yields a correctly sized test because Pr{Reject (1) at level $\hat{\alpha}^*$| (1) is true} $= \alpha$ by (6).

Our proposed Monte Carlo-adjusted GOF (AGOF) test consists of three steps. The first step is to determine $\hat{p}$. For that $u, u_1, \ldots, u_n$ are computed, based, for instance, on the G-function, as given in (2) with

$$\overline{G}(h) = \frac{1}{n-1}\sum_{j=2}^{n} \hat{G}_j(h); \quad \overline{G}_i(h) = \frac{1}{n-1}\sum_{j=1, j\neq i}^{n} \hat{G}_j(h), \quad i = 1, \ldots, n. \tag{7}$$

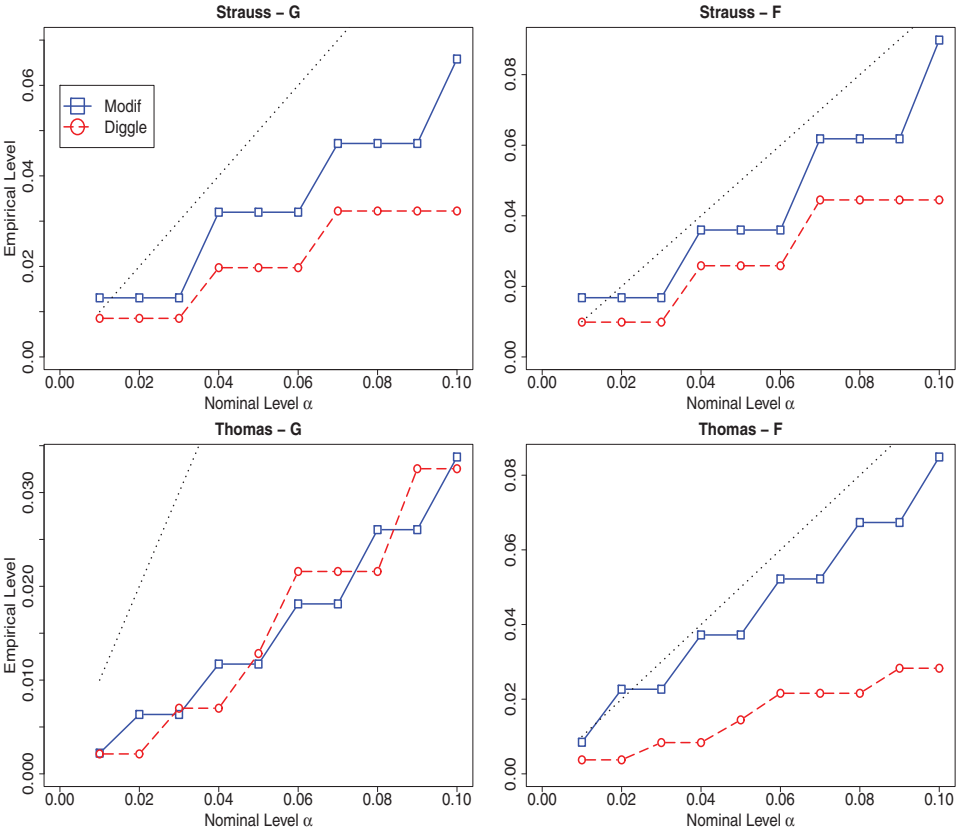

Figure 3. Empirical level curves resulting from the traditional Monte Carlo GOF tests in Section 2.2, represented by the dashed lines going through circles labeled as "Diggle," and their modified version through (7), represented by the solid lines going through squares labeled as "Modif," are shown; the dotted line is the reference 45° line. The tests use $n = 50$ based on $M = 3000$ generated patterns of Thomas and Strauss processes.

The above $\overline{G}(h)$ and $\overline{G}_i(h)$, $i = 1, \ldots, n$, deliberately do not include $\hat{G}(h)$ like in (3). Thus, the $\overline{G}_i(h)$'s are not contaminated with the observed information through $\hat{G}(h)$ but are computed purely based on patterns generated under H_0. Also, we choose the above $\overline{G}(h) = \sum_{j=2}^{n} \hat{G}_j(h)/(n-1)$ as opposed to $\sum_{j=1}^{n} \hat{G}_i(h)/n$ because the $\overline{G}_i(h)$'s are normalized with the constant $(n-1)$. Simulations using Strauss and Thomas processes with $n = 50$ in Figure 3 demonstrate that, except for one case, the empirical levels resulting from the modification in (7) are higher than those resulting from the traditional GOF test as proposed in (3). Only for the case of the Thomas process using the GOF-G test, the empirical levels resulting from (7) are not higher than that from (3). Overall, this modification makes the traditional GOF test become slightly more sensitive, which can be advantageous for small n. Eventually, the advantage of this modification will diminish as n increases, in particular for $n \geq 100$.

The second step is to find $\hat{\alpha}^*$. To achieve that, we have to reconstruct the distribution of $\hat{\mathcal{P}}$ via pseudorandom sampling using a Monte Carlo technique. We claim that the following procedure can provide the pseudovalues, $\hat{p}_1, \ldots, \hat{p}_n$ for $\hat{\mathcal{P}}$. They are estimates of the

unknown plug-in p-values, $p_1, \ldots, p_n$, of the following hypotheses

$$H_{i,0} : \mathbb{X}_i \sim f_\theta, \tag{8}$$

with $\mathbb{X}_i, i = 1, \ldots, n$, as the underlying spatial point processes corresponding to the simulated $X_i, i = 1, \ldots, n$, under (1). In Figure 2, the AGOF test is illustrated as the extension of the PGOF test: we apply the PGOF test to each X_i, that is, we simulate $X_{i,j}$, $j = 2, \ldots, n$, under (8) with $\theta = \widetilde{\theta}_i$, where $\widetilde{\theta}_i$ is the parameter estimate obtained from X_i, which is also $X_{i,1}$ in our notation. To compute $\hat{p}_1, \ldots, \hat{p}_n$, we define

$$u_{i,j} = \int_0^\infty \{\hat{G}_{i,j}(h) - \overline{G}_{i,j}(h)\}^2 dh, \quad i, j = 1, \ldots, n,$$

where

$$\overline{G}_{i,1}(h) = \frac{1}{n-2} \sum_{j=3}^{n} \hat{G}_{i,j}(h); \quad \overline{G}_{i,j}(h) = \frac{1}{n-2} \sum_{l=2, l\neq j}^{n} \hat{G}_{i,l}(h), \quad j = 2, \ldots, n.$$

Then, the plug-in p-value,

$$p_i = 1 - \Pr(U_{\widetilde{\theta}_i} < u_{i,1}), \tag{9}$$

for (8) with $U_{\widetilde{\theta}_i}$ as the random variable corresponding to $u_{i,2}, \ldots, u_{i,n}$ can be estimated by

$$\hat{p}_i = 1 - \frac{1 + \sum_{j=2}^{n} I(u_{i,j} \leq u_{i,1})}{n}. \tag{10}$$

We claim that the distribution of $\hat{\mathcal{P}}$ is herewith reconstructed via the pseudovalues, $\hat{p}_1, \ldots, \hat{p}_n$. Now, we find $\hat{\alpha}^*$ such that

$$\sum_{i=1}^{n} I\{\hat{p}_i < \hat{\alpha}^*\} = \alpha n. \tag{11}$$

Via the quantile function, $\hat{\alpha}^*$ can be obtained from the α-quantile of $\hat{p}_1, \ldots, \hat{p}_n$. The third, and also the last, step is making the conclusion about the fit, that is, reject H_0 if $\hat{p} \leq \hat{\alpha}^*$. The validity of the AGOF test follows with the proof and pseudocode given in the online supplementary material.

Proposition 1. Let $\hat{\mathcal{P}}$ denote the random variable corresponding to the estimated plug-in p-value, $\hat{p}$, resulting from (5), and let $\hat{\alpha}^*$ be that found in (11). Then, we have (6). In particular when the null model is true, the probability of rejecting the null model to the significance level of $\hat{\alpha}^*$ is α, that is, $\Pr\{\hat{\mathcal{P}} < \hat{\alpha}^*\} = \Pr\{\text{Reject } H_0 \text{ at level } \hat{\alpha}^* | \ H_0 \text{ is true}\} = \alpha$.

Remark 1. The "*traditional*" Monte Carlo GOF test means that the Monte Carlo GOF test is applied to: (i) data consisting of at least two datasets; or (ii) data consisting of a single dataset with the parameter estimate depending only "loosely" on the estimation procedure. The PGOF test is the Monte Carlo GOF test with a plug-in, that is, when the test is applied to the data consisting of a single dataset, and the parameter estimate can depend on the estimation procedure. The AGOF test is the Monte Carlo-adjusted GOF test, which is an extension of the PGOF test.

3. SIMULATION STUDIES

From a process of interest, we generate $M = 3000$ patterns to compute the size of a PGOF or AGOF test. As in Section 2, n gives the number of Monte Carlo simulations within a single Monte Carlo GOF test. M, however, sets the number of replications of the Monte Carlo GOF tests. We use various processes to represent four main types of patterns, which are random (independent), inhibitory (regular), clustered (aggregated), and inhomogeneous. To compare fairly our simulation results with those in Guan (2008), Illian et al. (2008), and Diggle (2003), we generate processes with patterns having a comparable number of events as described below. However, other simulations confirm that the performance of the PGOF and AGOF tests are not sensitive to the choice of specific parameter values. Computational work is done using the `spatstat` package (Baddeley and Turner 2005) in R (R Development Core Team 2012).

3.1 GENERATING PATTERNS

In our simulation studies, the observation window of all point patterns is the unit square. In the first three types of patterns, the average number of events of each pattern is 100. First, for the random pattern, we choose the CSR process with intensity $\lambda = 100$. Second, for the inhibitory pattern, we choose a Strauss process with $(\beta, \gamma, R) = (200, 0.25, 0.05)$, where β is the chemical activity parameter, γ is the interaction parameter, and R is the interaction radius, and the probability density function $f(\mathbb{X}) = \nu\beta^n\gamma^s$, with ν as the normalizing constant and s as the number of distinct unordered pairs of events closer than R units apart (Møller and Waagepetersen 2003). Third, for the clustered pattern, we consider a Thomas process as representative of this group. Here, a CSR with intensity κ is generated to obtain the so-called "parent" points. Each parent point is replaced by a random cluster of "children" points that are Poisson distributed with intensity μ. The positions of the children points are distributed about the parent location according to a bivariate Gaussian distribution with covariance $\sigma^2 I_2$, where I_2 is the 2×2 identity matrix (Møller and Waagepetersen 2003). Here the likelihood is intractable, but the theoretical K-function is available, $K(h) = \pi h^2 + \kappa^{-1}[1 - \exp\{-h^2/(4\sigma^2)\}]$. For our purposes, we chose a Thomas process with $(\kappa, \sigma, \mu) = (20, 0.05, 5)$. Finally, for the inhomogeneous Poisson processes, we adopt the models in the simulation studies in Guan (2008). That is, we consider the following intensity functions:

$$\lambda(x) = \exp(\beta_0 + \beta_1 x), \tag{12}$$

$$\lambda(x) = \exp\{\beta_0 + \beta_1 \sin(2\pi x)\}, \tag{13}$$

where β_0 is the normalizing constant and β_1 specifies the inhomogeneity in the data. For each intensity function, we consider the cases $\beta_1 = -1$ and $\beta_1 = -2$. The bigger the absolute value of β_1, the more inhomogeneous the pattern. Here, β_0 is manipulated to obtain $\mu = 100$ and $\mu = 400$ number of events for each combination of β_1 and the intensity functions. We refer to the process with the intensity function (12) as linear and (13) as sine models.

3.2 PARAMETER ESTIMATION AND PERFORMANCE OF GOF TESTS

We use various methods to obtain the estimates. For the CSR, the number of events is the estimate of the intensity parameter. For the Strauss process, we use the maximum profile pseudolikelihood (MPPL) estimator (Baddeley and Turner 2000) to obtain the estimate of R via `profilepl` and the maximum pseudolikelihood (MPL) estimators (Besag 1978; Berman and Turner 1992; Baddeley and Turner 2000) to obtain the estimates of β and γ via `ppm`. For the Thomas process, we use a minimum contrast method using the K-function (MCM-K) (Diggle 2003, sec. 6) via `thomas.estK`. For the inhomogeneous Poisson processes, we obtain the maximum likelihood estimate for β_0 and β_1 via `ppm`.

In the following, GOF-G, -F, -K, or -K^{in} tests denote GOF tests using a G-, F-, K-, or K^{in}-function to compute the test statistics. If the K-function is used in the estimation procedure, Diggle (2003, sec. 6.2) recommended not using it again in computation of the statistic and the pseudostatistics in (2) and (7). Hence, we have just GOF-G and -F tests for the Thomas process, and also have the GOF-K test for the Strauss and CSR processes. Additionally, for the inhomogeneous Poisson processes, we also employ the GOF-K^{in} test, where the K^{in}-function denotes the *inhomogeneous* K-function, which does not assume stationarity as the G-, F-, and K-functions do. With respect to the significance level, we choose the common nominal levels $\alpha = 0.05$, and $\alpha = 0.10$, due to Guan (2008). A GOF test is correctly sized if its empirical level, $\hat{\alpha}$, reaches the nominal level, α, correctly according to the usual definition, that is, the interval $(\hat{\alpha} - 2se, \hat{\alpha} + 2se)$, where $se = \sqrt{\hat{\alpha}(1-\hat{\alpha})/M}$, contains α.

Table 1 presents two facts. First, the PGOF test is correctly sized only for the CSR. In contrast, our proposed AGOF test is correctly sized for all processes, except for the

Table 1. For nominal level α, $\widehat{\alpha}_{\text{PGOF}}$ and $\widehat{\alpha}_{\text{AGOF}}$ are empirical levels from the PGOF and AGOF tests using $n = 100$, $M = 3000$ replicates. $\overline{\alpha}^* = \sum_{k=1}^{M} \hat{\alpha}_k^*/M$. Notations -$G$, -$F$, and -$K$ label computations of the GOF test statistics using the G-, F-, K-functions, respectively. The CSR, Strauss, and Thomas processes are defined in Section 3.1. Estimated standard errors multiplied by 100 are in parentheses

			$\alpha = 0.05$			$\alpha = 0.10$		
Model	No. of events	Pattern	$\widehat{\alpha}_{\text{PGOF}}$	$\overline{\alpha}^*$	$\widehat{\alpha}_{\text{AGOF}}$	$\widehat{\alpha}_{\text{PGOF}}$	$\overline{\alpha}^*$	$\widehat{\alpha}_{\text{AGOF}}$
CSR-G	100	Random	0.053 (0.410)	0.049	0.054 (0.414)	0.096 (0.538)	0.106	0.105 (0.560)
CSR-F		Random	0.049 (0.396)	0.049	0.048 (0.389)	0.095 (0.536)	0.107	0.099 (0.545)
CSR-K		Random	0.051 (0.401)	0.048	0.050 (0.398)	0.093 (0.529)	0.099	0.091 (0.524)
Strauss-G	100	Inhibitory	0.026 (0.292)	0.141	0.068 (0.459)	0.044 (0.374)	0.233	0.125 (0.603)
Strauss-F		Inhibitory	0.020 (0.226)	0.106	0.047 (0.385)	0.043 (0.371)	0.184	0.090 (0.522)
Strauss-K		Inhibitory	0.040 (0.356)	0.062	0.053 (0.410)	0.078 (0.488)	0.124	0.100 (0.548)
Thomas-G	100	Clustered	0.008 (0.153)	0.153	0.050 (0.374)	0.027 (0.277)	0.241	0.107 (0.517)
Thomas-F		Clustered	0.037 (0.322)	0.073	0.055 (0.388)	0.074 (0.446)	0.140	0.100 (0.512)

AGOF-G test applied in the Strauss case. Second, the estimated adjusted level, $\hat{\alpha}^*$, can serve as an indication for some patterns, as opposed to randomness, in the data. The AGOF test would provide $\hat{\alpha}^*$ strongly deviating from α if the SPP comes from a process that is very different from CSR.

For CSR, as explained in Section 1, the AGOF test is not needed. Here, the PGOF test coincides with the traditional Monte Carlo GOF test; thus, $\alpha^* = \alpha$. Nevertheless, the simulations in Table 1 validate the correctness of our method as $\widehat{\alpha}_{\text{AGOF}}$ conforms with $\widehat{\alpha}_{\text{PGOF}}$, $\overline{\alpha}^* = \sum_{k=1}^{M} \hat{\alpha}_k / M$, and α. For the Strauss process, the MPPL $\hat{R}$ of the irregular parameter, R, is a poor estimator, which strongly affects the estimation of the other two regular parameters, β and γ, which can be obtained via MPL or approximate maximum likelihood (AML) estimation (Huang and Ogata 1999). Due to the high computational complexity of the latter method, we use the MPL estimators of β and γ. Simulations show that $\hat{\gamma}$ tends to be greater than the true $\gamma = 0.25$. In fact, the median of $\hat{\gamma}$ is 0.277 and the mean is 0.289. The difficulty in parameter estimation carries on in the estimation of α^*. However, increasing n from 100 to 150 makes the conclusion of the AGOF test more accurate. We also conjecture that using the AML instead of the MPL estimators of β and γ would lead to a correctly sized AGOF-G test for the Strauss process.

Table 2 shows that the PGOF-G tests are correctly sized in two cases of the model (12) when $\beta_1 = -1$ with $\mu = 100$ and $\mu = 400$. This seems to contradict our observation made in Section 1 and the finding in the previous paragraph that the PGOF test is correctly sized only for CSR. Our other simulations show that if the pattern and its H_0 are not too inhomogeneous as in the case of (12) with $\beta_1 = -1$, the PGOF test can, but does not have to, be correctly sized as the G-function of a CSR and of those models can be similar. However, the PGOF-F and -K^{in} tests remain incorrectly sized. Table 2 shows that the PGOF tests are overall incorrectly sized and their empirical levels deviate away from the nominal levels in two situations: (i) when the inhomogeneity factor, $|\beta_1|$, increases and (ii) when the intensity function changes from linear, (12), to nonlinear, (13). In contrast, the AGOF tests are correctly sized throughout. Overall, the AGOF tests clearly outperform the PGOF tests and do not underperform the test by Guan (2008). Here, we did not implement the test of Guan (2008), which was described in Section 1. The empirical levels, $\hat{\alpha}_{\text{Guan}}$, are extracted from $t = 0.2$ in Guan (2008, Tables 1 and 2).

3.3 EFFECTIVE SIMULATION SIZE AND COMPUTATIONAL TIME

Marriott (1979) proposed to consider $m/n = \alpha$, where m is chosen in a way that if u is among the m largest values of $u_1, \ldots, u_n$, H_0 is rejected. Besag and Diggle (1977) suggested that $m = 5$ might be a suitable value for the traditional Monte Carlo tests for spatial patterns. Consequently, $n = 100$ should be used for $\alpha = 0.05$ and $n = 500$ for $\alpha = 0.01$. Hence, the smaller the α, the bigger the n. Our simulation studies shown in Tables 1 and 2 are run using $n = 100$ to compare fairly the performance of the PGOF and AGOF tests at the nominal level $\alpha = 0.05$. However, Table 3 shows that except for AGOF-G and -F tests for the Strauss process, the AGOF tests need only $n = 20$ to be correctly sized at the $\alpha = 0.05$ level, and $n = 100$ to be correctly sized at the $\alpha = 0.01$ level. That is, the effective simulation size is reduced by a factor of 5 in the setting of the AGOF test, except for the AGOF-G and -F tests for processes having parameters being estimated by the MPPL method,

Table 2. For nominal level α, $\widehat{\alpha}_{\mathrm{PGOF}}$ and $\widehat{\alpha}_{\mathrm{AGOF}}$ are empirical levels from the PGOF and AGOF tests using $n = 100$, $M = 3000$ replicates. $\overline{\alpha}^* = \sum_{k=1}^{M} \hat{\alpha}_k^*/M$. Linear and sine models and β_0 and β_1 are defined in (12) and (13) in Section 3.1. Notations -G, -F, and -K^{in} label computations of the GOF test statistics using the G-, F-functions, and the *inhomogeneous* K-function, respectively. From Guan (2008, Tables 1 and 2), $t = 0.2$, $\hat{\alpha}_{\mathrm{Guan}}$ is extracted. Estimated standard errors multiplied by 100 are in parentheses

				$\alpha = 0.05$			$\alpha = 0.10$			
Model	No. of events	β_0	$-\beta_1$	$\widehat{\alpha}_{\mathrm{PGOF}}$	$\overline{\alpha}^*$	$\widehat{\alpha}_{\mathrm{AGOF}}$	$\widehat{\alpha}_{\mathrm{PGOF}}$	$\overline{\alpha}^*$	$\widehat{\alpha}_{\mathrm{AGOF}}$	$\hat{\alpha}_{\mathrm{Guan}}$
Linear-G	100	5.06	1	0.048	0.054	0.052	0.095	0.107	0.104	0.096
				(0.390)		(0.404)	(0.535)		(0.557)	
Linear-F	100	5.06	1	0.032	0.077	0.055	0.072	0.139	0.102	0.096
				(0.323)		(0.417)	(0.473)		(0.552)	
Linear-K^{in}	100	5.06	1	0.034	0.072	0.063	0.077	0.138	0.109	0.096
				(0.331)		(0.443)	(0.486)		(0.570)	
Linear-G	100	5.45	2	0.039	0.062	0.052	0.075	0.124	0.099	0.130
				(0.353)		(0.405)	(0.480)		(0.546)	
Linear-F	100	5.45	2	0.015	0.128	0.051	0.035	0.213	0.101	0.130
				(0.112)		(0.345)	(0.155)		(0.469)	
Linear-K^{in}	100	5.45	2	0.035	0.069	0.046	0.074	0.132	0.105	0.130
				(0.333)		(0.385)	(0.478)		(0.560)	
Linear-G	400	6.45	1	0.051	0.049	0.052	0.094	0.107	0.099	0.094
				(0.349)		(0.357)	(0.478)		(0.497)	
Linear-F	400	6.45	1	0.033	0.087	0.052	0.059	0.161	0.102	0.094
				(0.327)		(0.408)	(0.431)		(0.552)	
Linear-K^{in}	400	6.45	1	0.043	0.063	0.053	0.083	0.128	0.107	0.094
				(0.359)		(0.409)	(0.502)		(0.565)	
Linear-G	400	6.85	2	0.041	0.061	0.052	0.083	0.124	0.108	0.098
				(0.365)		(0.405)	(0.503)		(0.567)	
Linear-F	400	6.85	2	0.012	0.146	0.052	0.027	0.238	0.101	0.098
				(0.201)		(0.405)	(0.297)		(0.550)	
Linear-K^{in}	400	6.85	2	0.025	0.086	0.049	0.053	0.160	0.093	0.098
				(0.283)		(0.395)	(0.408)		(0.531)	
Sine-G	100	4.37	1	0.026	0.081	0.044	0.054	0.149	0.090	0.100
				(0.288)		(0.374)	(0.412)		(0.523)	
Sine-F	100	4.37	1	0.004	0.176	0.045	0.012	0.272	0.100	0.100
				(0.115)		(0.378)	(0.204)		(0.548)	
Sine-K^{in}	100	4.37	1	0.023	0.076	0.044	0.056	0.145	0.096	0.100
				(0.275)		(0.373)	(0.420)		(0.537)	
Sine-G	100	3.80	2	0.019	0.096	0.046	0.048	0.169	0.093	0.108
				(0.253)		(0.382)	(0.393)		(0.532)	
Sine-F	100	3.80	2	0.004	0.195	0.053	0.011	0.293	0.108	0.108
				(0.115)		(0.409)	(0.193)		(0.567)	
Sine-K^{in}	100	3.80	2	0.048	0.048	0.048	0.093	0.109	0.091	0.108
				(0.393)		(0.393)	(0.532)		(0.527)	
Sine-G	400	5.80	1	0.029	0.076	0.049	0.058	0.142	0.094	0.082
				(0.329)		(0.412)	(0.447)		(0.558)	
Sine-F	400	5.80	1	0.007	0.178	0.046	0.015	0.278	0.097	0.082
				(0.162)		(0.400)	(0.237)		(0.566)	
Sine-K^{in}	400	5.80	1	0.037	0.075	0.054	0.064	0.143	0.100	0.082
				(0.361)		(0.432)	(0.465)		(0.573)	
Sine-G	400	5.15	2	0.028	0.093	0.057	0.063	0.166	0.106	0.114
				(0.325)		(0.452)	(0.476)		(0.602)	
Sine-F	400	5.15	2	0.004	0.189	0.056	0.014	0.288	0.104	0.114
				(0.125)		(0.428)	(0.230)		(0.596)	
Sine-K^{in}	400	5.15	2	0.042	0.064	0.052	0.072	0.128	0.101	0.114
				(0.394)		(0.435)	(0.506)		(0.589)	

Table 3. The empirical levels $\widehat{\alpha}_{\text{AGOF}}$ result from AGOF tests using $n = 20, 50, 100$ to assess the effective simulation size n to nominal levels $\alpha = 0.01, 0.05, 0.10$. Sine, Strauss, and Thomas processes are defined in Section 3.1

	$\alpha = 0.01$			$\alpha = 0.05$			$\alpha = 0.10$		
Model	20	50	100	20	50	100	20	50	100
CSR-G	0.049	0.024	0.012	0.049	0.057	0.054	0.110	0.111	0.105
CSR-F	0.054	0.025	0.008	0.054	0.058	0.048	0.112	0.111	0.099
CSR-K	0.056	0.023	0.009	0.056	0.058	0.049	0.109	0.111	0.091
Sine-G	0.054	0.024	0.011	0.054	0.051	0.051	0.105	0.105	0.095
Sine-F	0.051	0.022	0.010	0.051	0.050	0.052	0.101	0.100	0.100
Sine-K	0.048	0.020	0.011	0.048	0.051	0.049	0.100	0.105	0.094
Strauss-G	0.076	0.026	0.015	0.076	0.065	0.068	0.101	0.123	0.125
Strauss-F	0.069	0.036	0.010	0.069	0.079	0.047	0.133	0.133	0.090
Strauss-K	0.051	0.017	0.010	0.051	0.051	0.053	0.109	0.101	0.100
Thomas-G	0.053	0.024	0.012	0.053	0.046	0.050	0.099	0.100	0.092
Thomas-F	0.053	0.023	0.011	0.053	0.057	0.049	0.106	0.110	0.100

including Strauss processes. For CSR, while Table 3 shows that the AGOF tests just need $n = 20$, other simulations show that the PGOF tests need $n \geq 100$ to be correctly sized at $\alpha = 0.05$.

Table 4 shows our investigation of the computational time of the PGOF and AGOF tests. The computation was done on compute nodes that have 8 CPU cores, 32 GB of RAM, and the CPU processors clocked at 2.4 GHz or faster. For each of the CSR, sine, Strauss, and Thomas processes defined in Section 3.1, $M = 3000$ patterns are generated to provide the average computational time given in seconds. The time is to obtain $\hat{p}$ from a PGOF test, $\hat{p}$ and $\hat{\alpha}^*$ from an AGOF test using G-, F-, K- or K^{in}-, and pc- or pc^{in}-functions altogether. The pc-function denotes the pair correlation function $pc(h) = K'(h)/(2\pi h)$ (Illian, Møller, and Waagepetersen 2009). The pc^{in}-function results analogously from the K^{in}-function.

Correctly sizing the AGOF test does come with a cost as the computational time is always longer than that of the PGOF test. While the Monte Carlo simulations are of the size n for the PGOF test, they are of the size n^2 for the AGOF test as illustrated in Figure 2, which also shows that the AGOF test is an extension of the PGOF test.

Table 4. For each of CSR, sine, Strauss, and Thomas processes defined in Section 3.1, $M = 3000$ patterns are generated to provide the average computational time given in seconds. The time in seconds is to obtain results from a GOF test using G-, F-, K- or K^{in}-, and pc- or pc^{in}-functions altogether

	$n = 20$		$n = 50$		$n = 100$	
Model	PGOF	AGOF	PGOF	AGOF	PGOF	AGOF
CSR	2	42	7	227	9	835
Sine	7	87	19	541	36	1897
Strauss	269	5138	642	31,907	1288	121,671
Thomas	2	52	5	313	9	1076

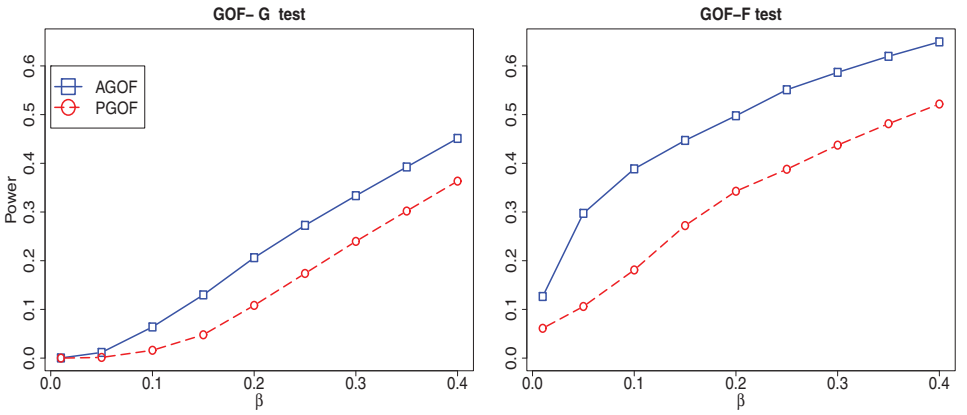

Figure 4. Power curves of AGOF-G, -F and PGOF-G, -F tests, using $n = 100$, are evaluated based on $M = 3000$ simulated patterns of the sine model in (13) when wrongly fitting a Thomas process. The power curves of the AGOF tests are the solid curves going through nine squares, and of the PGOF tests are the dashed curves going through nine circles. β denotes the Type II error rate at 0.01, 0.05, 0.10, 0.15, 0.20, 0.25, 0.30, 0.35, 0.40.

3.4 STATISTICAL POWER

In the following, we show that the power of the AGOF test is at least as good as the power of the PGOF test via a simulation study. We simulate $M = 3000$ patterns under the sine model in (13), and fit wrongly a Thomas process (null model) to each of the patterns. In our opinion, this example is realistic because patterns underlying both (13) and a Thomas process contain clusters. In a real scenario, there might be a pattern underlying (13) that is incorrectly categorized as a Thomas process. After fitting a Thomas process to these datasets, we obtained 3000 $\hat{p}$'s from the PGOF-G and -F tests. The Type II error rate, β, was set at 0.01, 0.05, 0.10, 0.15, 0.20, 0.25, 0.30, 0.35, 0.40. The power was the rejection rate based on α for the PGOF tests and based on $\hat{\alpha}^*$ for the AGOF tests. In Figure 4, the power curve of the AGOF test is the solid curve going through nine squares and the one of the PGOF test is the dashed curve going through nine circles. The power curve of the AGOF test is clearly higher than the one of the PGOF at all given β.

Remark 2. Our simulations show that there can be contradictory conclusions for the GOF-G, - F, -K, and -K^{in} tests based on one spatial pattern although fitting the correct model. Indeed, each of the G-, F-, K-, and K^{in}-functions measures completely different things. Thus, disagreement among the AGOF tests can occur, but their performances can still be correctly sized.

4. DATA APPLICATIONS

4.1 *PHLEBOCARYA FILIFOLIA* PLANTS

Figure 5 (left) displays, in a 22 m × 22 m window, positions of 207 *Phlebocarya filifolia* plants that are typically located in scattered positions throughout Western Australia. Illian

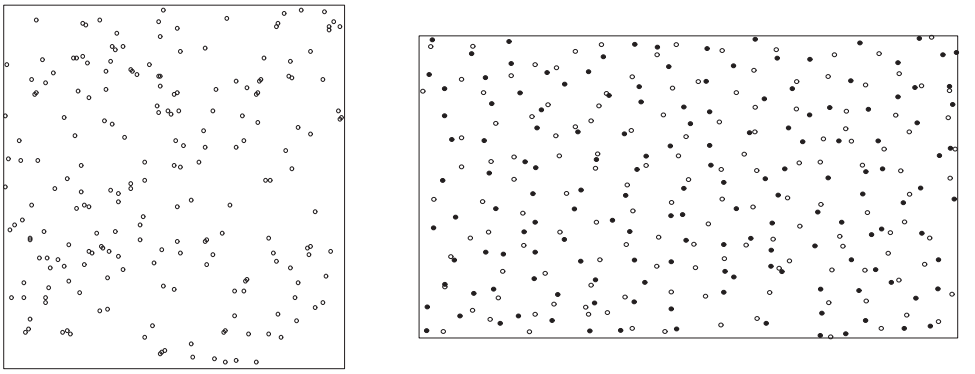

Figure 5. Left: positions of 207 *Phlebocarya filifolia* plants in a 22 m × 22 m window at Cooljarlo near Perth, Australia. Right: positions of 294 displaced amacrine cells in the retina of a rabbit. Solid and open circles represent on and off cells, respectively. The observation window for the data is (0, 1.62) × (0, 1).

et al. (2008, Example 4.19) used this dataset as an example of an homogeneous clustered pattern by demonstrating the occurrence of aggregation via a large value of the Clark–Evans index and a small value of the mean-direction index. Illian et al. (2008, Example 7.2) concluded that the Matérn cluster process provides an acceptable fit via the envelope test with a plug-in as the envelopes using the G- and F-functions with $n = 100$ envelop the empirical G- and F-functions, respectively. We would like to reevaluate the GOF of a process with an homogeneous clustered pattern with our AGOF-G and -F tests.

As clustering is not immediately obvious, we first used the PGOF test to assess the fit of CSR to detect whether actual modeling is necessary. The PGOF-G, -F, and -K tests yielded estimated p-values of 0, 0.515, and 0.01, respectively, suggesting that the fit of CSR is disputable. Figure 6 shows the empirical G-, F-, and K-function (dash-dotted curves); other references are the functions of a CSR (solid curve) representing a random pattern, of a Thomas process (dashed curve) representing a clustered pattern, and of a Strauss process (dotted curve) representing an inhibitory pattern. These processes have approximately 200 events. While the left column of plots shows these functions on their complete domains, the right column shows these functions on a much smaller domain to scrutinize their behavior for distances shorter than 0.01. While the empirical F-function aligning with the one of the CSR can suggest random behavior in agreement with the conclusion of the PGOF-F test, the empirical G-function lying above the one of the CSR indicates clustering, except for distances shorter than 0.0044. The empirical K-function significantly signals slight clustering as it lies completely above the one of the CSR, except for distances shorter than 0.0047. At this distance, the behavior of this dataset and that of the reference processes seem to be similar because the biggest discrepancy is less than 10^{-4}. While the G- and F-functions are believed to be "short-sighted" as they describe the distances to nearest neighbors of reference events or points and say little or nothing about the spatial dependence of the events beyond the nearest neighbor (Illian et al. 2008, sec. 4.3.1), the K-function can be an effective summary of spatial dependence over a wide range of scales (Cressie 1993, sec. 8.4.3). Thus, we ascribe clustering to this dataset due to the plots of the K-function. We consider a Matérn cluster process (Illian et al. 2008) and a Thomas process as

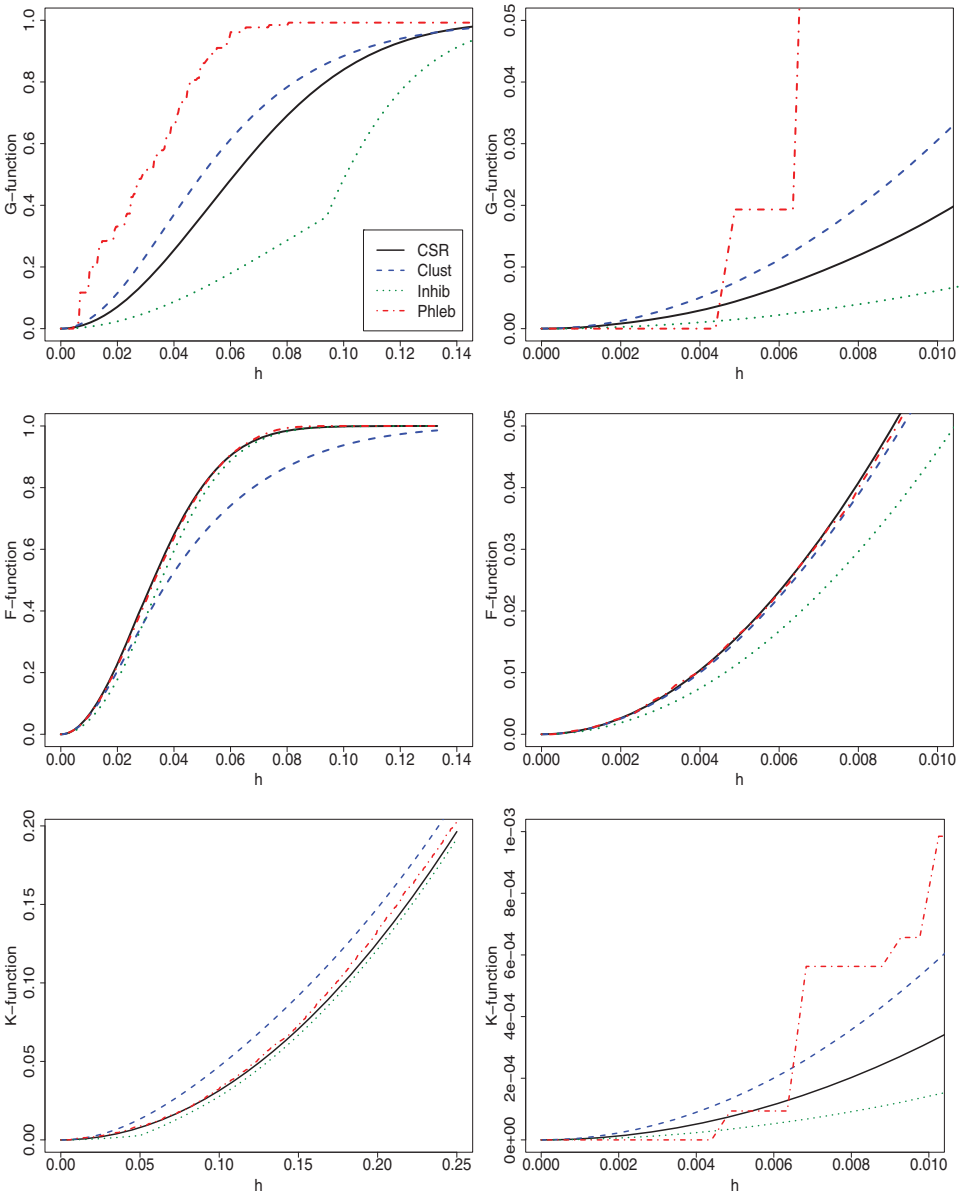

Figure 6. Study of the summary characteristics based on the G- and F-functions and second-order characteristic based on the K-function of the *Phlebocarya filifolia* plants. Left column: complete domain; right column: smaller domain. The empirical K-function in dashed curve indicates slight clustering.

reasonable models because they are representatives of processes with homogeneous clustered patterns.

Assuming that the pattern comes from a Matérn cluster process (Møller and Waagepetersen 2003), an homogeneous Poisson point process with intensity κ is generated to obtain the "parent" points. Each parent point is replaced by a random cluster of

"children" points that are Poisson distributed with intensity μ. The positions of the children points are placed independently and uniformly inside a disk of radius R centered at the parent point. Neither likelihood nor pseudolikelihood is available, but the K-function is $K(h) = \pi h^2 + s(h)\{h/(2R)\}/\kappa$, where $s(h) = 2 + (1/\pi)\{(8h^2 - 4)\arccos(h) - 2\arcsin(h) + 4h\sqrt{(1-h^2)^3} - 6h\sqrt{1-h^2}\}$ for $h \leq 1$, and $s(h) = 1$ for $h > 1$. The MCM-K provides the estimate $(\hat{\kappa}, \hat{R}, \hat{\mu}) = (132.996/\{22\text{ m} \times 22\text{ m}\}, 3.482\text{ m}, 1.556)$. The AGOF-$G$ and -F tests give the $\hat{p}$'s of 0.054, 0.634 and $\hat{\alpha}^*$'s of 0.123, 0.449, respectively. The AGOF-G test does not support the fit as $\hat{p} < \hat{\alpha}^*$ $(0.023 < 0.123)$. However, the AGOF-F test supports the good fit of the Matérn cluster model because $\hat{p} > \hat{\alpha}^*$ $(0.634 > 0.449)$.

Now, suppose the pattern comes from a Thomas process, where the MCM-K provides the estimate $(\hat{\kappa}, \hat{\sigma}, \hat{\mu}) = (111.455/\{22\text{ m} \times 22\text{ m}\}, 2.099\text{ m}, 1.857)$. The AGOF-$G$ and -F tests give $\hat{p}$'s of 0.044, 0.761 and $\hat{\alpha}^*$'s of 0.115, 0.620, respectively. The AGOF-G test does not support the fit as $\hat{p} < \hat{\alpha}^*$ $(0.044 < 0.115)$. However, the AGOF-F supports the fit of the Thomas process because $\hat{p} > \hat{\alpha}^*$ $(0.761 > 0.620)$.

We think that the AGOF-G tests reject the fits of the Matérn cluster and the Thomas processes because the G-function in Figure 6 indicates, on the most part, clustering, but it suggests inhibition on distances, h, less than 0.0044. Thus, we suspect that a fit of any clustered or inhibitory model would unlikely gain support by the AGOF-G test. Overall, we would support either the fit of the Matérn cluster or the Thomas process because the AGOF-F test in each case supports the fit and the plot of the empirical K-function significantly indicates slight clustering. This finding is in agreement with the conclusion made by Illian et al. (2008, Example 7.2). One might understand the propagation mechanism of *Phlebocarya filifolia* plants better: there is indeed clustering, although just slightly. One might also prefer a Thomas to a Matérn cluster process as the Thomas process models more "children" points at closer distances to parent points; this characteristic might be more natural for forest data.

4.2 AMACRINE CELLS

The data displayed in Figure 5 (right) show the bivariate pattern of amacrine cells within the retina of a rabbit. Interest lies in distinguishing between two developmental hypotheses in studying the retinas of rabbits. The observation window for the data is the rectangle $(0, 1.62) \times (0, 1)$ according to Diggle (2003). The two types of cells are responses to a light being switched *on* and *off*. The *separate layer* hypothesis is that the on and off cells are initially formed in two separate layers that later fuse to form a mature retina. The *single layer* hypothesis is that the two types of cells are initially undifferentiated in a single layer and acquire their separate functions at a later stage. Via a Monte Carlo test of independence, Diggle (2003) failed to reject the on and off cells being independent processes, supporting the separate layer hypothesis (Diggle 2003, sec. 4.7). To answer whether they would fuse to form a mature retina, one might consider fitting a statistical model to one and run a GOF test on the other dataset.

Diggle (2003, sec. 4.7) showed that these processes have very similar second-order properties including inhibitory behaviors at small distances. Subsequently, a model is fitted to the on cells only and the off cells are reserved for a GOF assessment. A pairwise interaction point process, that is, a Markov point process, was chosen due to the inhibitory behavior

at small distances, together with the absence of any obvious longer-range heterogeneity or other form of aggregation. Ripley (1977) formalized the class of pairwise interaction processes, $f(X) = \nu\beta^n \prod_i \prod_{j\neq i} d\{\|x_i - x_j\|\}$, where ν is the normalizing constant, β reflects the intensity of the process, $d(h)$ is nonnegative for all $h > 0$, and the product is over all pairs of distinct points in the SPP, X. The parametric model with the interaction function for amacrine cells is taken from Diggle and Gratton (1984):

$$d(h) = \begin{cases} 0, & h < \delta, \\ \{(h-\delta)/(\rho-\delta)\}^\kappa, & \delta \le h \le \rho, \\ 1, & h > \rho. \end{cases}$$

The distance from δ to ρ can be interpreted as an interaction distance. The parameters δ, ρ, and κ are nonnegative, and $\delta < \rho$. The strength of inhibition increases with κ. For $\kappa = 0$, the model is a hard-core process with radius δ, whereas for $\kappa = \infty$ it has radius ρ.

Diggle (2003, sec. 7.2) obtained $(\hat{\delta}, \hat{\rho}, \hat{\kappa}) = (0.020, 0.12, 4.90)$ from the on cells via the MPL estimation. Then, to assess the GOF to the off cells, a traditional GOF-G test and a traditional GOF-F test were employed. The p-values were reported to be 0.01 and 0.37, respectively. Since these two tests give opposite conclusions with respect to any commonly used nominal level, Diggle (2003) was then motivated to obtain different estimates $(\hat{\delta}, \hat{\rho}, \hat{\kappa}) = (0.016, 0.12, 1.96)$ via AML estimation. Diggle (2003, fig. 7.7) showed that there is agreement between the F-function of the off cells and its corresponding envelopes. However, the GOF-G test still shows a poor fit of the pairwise interaction point process to the off cells. Diggle (2003) did not give a clear evaluation of the GOF for the amacrine cells data.

We fit the pairwise interaction process to the off cells dataset (training set) and obtained estimates of δ, ρ via the MPPL, and of κ via the AML or the MPL estimation. Then we ran AGOF-G, -F, and -K tests. If they concluded a good fit, these parameter estimates were used for the PGOF-G, -F, -K tests applied to the on cells data (validation set).

Via MPPL and AML estimation, we obtained the estimates $(0.015, 0.11, 3.435)$ from the off cells dataset, and $(0.026, 0.28, 1.083)$ from the on cells dataset for (δ, ρ, κ). Due to a very high computational intensity, we used $n = 50$ for the AGOF-G, -F, and -K tests, which provided p-values of 0.868, 0.647, 0.092 and $\hat{\alpha}^* = 0.022, 0, 0.091$, respectively. All AGOF tests indicated a good fit of the pairwise interaction process to the off cells dataset because all $\hat{p}$'s are greater than the corresponding $\hat{\alpha}^*$'s. Then we used the estimate of the off cells dataset as the parameter for the pairwise interaction process to fit to the on cells dataset. The PGOF-G, -F, and -K tests provided p-values of 0.200, 0.519, and 0.340, which indicated a good fit because the $\hat{p}$'s were greater than $\alpha = 0.05$. The AGOF tests drew the same conclusions as the PGOF tests as the $\hat{p}$'s were greater than the corresponding $\hat{\alpha}^* = 0.032, 0.061, 0.191$.

Via MPPL and MPL estimation, we obtained the estimates $(0.015, 0.11, 2.44)$ from the off cells dataset, and $(0.026, 0.28, 1.062)$ from the on cells dataset for (δ, ρ, κ). The AGOF-G, -F, and -K tests provided $\hat{p}$'s of 0.203, 0.011, 0.478 and $\hat{\alpha}^* = 0.015, 0.052, 0.052$, respectively. Only the AGOF-G and AGOF-K tests indicated a good fit of the pairwise interaction process to the off cells dataset because only their $\hat{p}$'s were greater than the corresponding $\hat{\alpha}^*$'s. We could have stopped here, but for curiosity we proceeded to run PGOF tests using the estimate of the off cells dataset as the parameter for the pairwise

interaction process to fit to the on cells dataset. The PGOF-G, -F, and -K tests provided p-values of 0.685, 0.092, and 0.606, which indicated a good fit because the $\hat{p}$'s were greater than $\alpha = 0.05$. The AGOF tests drew the same conclusions as the PGOF tests as the $\hat{p}$'s were greater than the corresponding $\hat{\alpha}^* = 0.050, 0.050, 0.174$.

The conclusions of the AGOF-F tests for the fit of the off cells dataset differed, but the conclusions of the PGOF and AGOF tests for the fit of the on cells dataset agreed. Since the AML estimators had better statistical properties (Huang and Ogata 1999), we relied on the finding using AML estimates and concluded that the separate layer hypothesis holds (Diggle 2003, sec. 4.7), that is, these two independent processes later fuse to form a mature retina.

5. DISCUSSION

This article discussed a method using nested Monte Carlo simulations to obtain the estimated adjusted level, $\hat{\alpha}^*$, corresponding to a prespecified nominal level, α. We rejected the null model if $\hat{p} \leq \hat{\alpha}^*$, where $\hat{p}$ is the estimated p-value. The nested Monte Carlo method was, however, computationally intensive, and we now discuss two techniques to improve the computational complexity, the first of which also estimates α^* more efficiently.

The first technique uses the idea of interpolation. Due to the computational intensity, we reduce the number of replications of Monte Carlo simulations, n, and compensate that by estimating the new u_i and $u_{i,j}$ with interpolation techniques, such as kernel density estimation. From the Monte Carlo simulations in Section 2, we obtain two sets, $S = \{u_1, \ldots, u_n\}$ and $S^* = \{u_{i,j}, i, j = 1, \ldots, n\}$. Via kernel density estimation, we estimate the new u_i's augmenting S to $S_{\text{aug}} = \{u_1, u_2, \ldots, u_n, \hat{u}_1, \hat{u}_2, \ldots, \hat{u}_\tau\}$ to obtain a better $\hat{p}$. Analogously, we estimate u_{ij}'s augmenting S^* to $S^*_{\text{aug}} = \{u_{1,2}, \ldots, u_{1,n}, \hat{u}_{1,1}, \ldots, \hat{u}_{1,\tau}; u_{2,2}, \ldots, u_{2,n}, \hat{u}_{2,1}, \ldots, \hat{u}_{2,\tau}; \ldots; u_{n,2}, \ldots, u_{n,n}, \hat{u}_{n,1}, \ldots, \hat{u}_{n,\tau}\}$ to obtain better $\hat{p}_1, \ldots, \hat{p}_n$, which leads to estimating a better $\hat{\alpha}^*$. The reason is that $\hat{\alpha}^*$ is the α-quantile of $\hat{p}_1, \ldots, \hat{p}_n$ as formalized in (11). Here, the quantities $\hat{u}_i$'s and $\hat{u}_{i,j}$'s denote the new estimated pseudostatistics, and τ^{-1} specifies the interpolation factor. Using Taylor expansions, we can show that the new $\hat{p}$ is similar to the one in Section 2 minus a term of order, $O(\tau^{-1})$. Since τ^{-1} can be chosen arbitrarily small, this term is negligible and the new $\hat{p}$ possesses the same properties as discussed in Section 2. The simulation studies use CSR, sine, Strauss, and Thomas processes defined in Section 3.1 and $n = 20$ and $\tau = 100$. Table 5 demonstrates that interpolation techniques can make the AGOF tests more accurate as the mean squared errors are smaller in 15 out of 22 studied cases. The strength of interpolation might be increased or optimized by a different choice of τ.

The second technique considers the characteristics of sequential Monte Carlo p-values. Besag and Clifford (1991) introduced several ways of calculating exact Monte Carlo p-values by sequential sampling. Instead of fixing the sample size, n, sampling is continued until a prespecified number, H, of values larger than the value u of the test statistic, U, has been observed. Let l be the value of the random sample size, L, at termination. Besag and Clifford (1991) showed that a p-value equal to H/l has properties similar to those obtained by sampling with a fixed sample size, n. With this method, Monte Carlo simulations may become less computationally intensive as they can terminate much earlier.

Table 5. To nominal level α, the empirical level $\hat{\alpha} = \widehat{\alpha}_{\text{AGOF}}$ using $\tau = 100$, $n = 20$, and $M = 5000$ is shown. $\overline{\alpha}^* = \sum_{k=1}^{M} \hat{\alpha}_k^*/M$. MSE is the mean squared error. The "Standard" columns display empirical levels from the AGOF tests from Section 2.3. The "Interpolation" columns display empirical levels from the AGOF tests incorporating interpolation as described in Section 5. Notations -*G*, -*F*, and -*K* label computations of the GOF test statistics using the *G*-, *F*-, *K*-functions, respectively. The CSR, sine, Strauss, Thomas processes are defined in Section 3.1

	$\alpha = 0.05$						$\alpha = 0.10$					
	Standard			Interpolation			Standard			Interpolation		
	$\overline{\alpha}^*$	$\hat{\alpha}$	MSE	$\overline{\alpha}^*$	$\hat{\alpha}$	MSE	$\overline{\alpha}^*$	$\hat{\alpha}$	MSE	$\overline{\alpha}^*$	$\hat{\alpha}$	MSE
CSR-*G*	0.029	0.051	0.103	0.025	0.046	0.254	0.077	0.110	1.117	0.72	0.098	0.210
CSR-*F*	0.027	0.055	0.334	0.023	0.046	0.233	0.072	0.115	2.576	0.069	0.103	0.263
CSR-*K*	0.028	0.057	0.656	0.023	0.049	0.100	0.074	0.115	2.394	0.069	0.099	0.183
Sine-*G*	0.058	0.064	2.024	0.052	0.052	0.279	0.119	0.116	2.765	0.121	0.103	0.287
Sine-*F*	0.109	0.049	0.103	0.116	0.044	0.444	0.192	0.102	0.223	0.224	0.095	0.463
Sine-*K*	0.031	0.056	0.419	0.027	0.048	0.124	0.079	0.119	3.974	0.078	0.103	0.287
Strauss-*G*	0.088	0.076	7.005	0.086	0.066	2.683	0.169	0.134	12.066	0.182	0.122	5.323
Strauss-*F*	0.062	0.062	1.605	0.060	0.054	0.246	0.127	0.112	1.591	0.136	0.099	0.197
Strauss-*K*	0.036	0.053	0.217	0.031	0.043	0.519	0.088	0.107	0.626	0.086	0.092	0.839
Thomas-*G*	0.096	0.045	0.317	0.100	0.044	0.493	0.172	0.091	1.048	0.196	0.092	0.745
Thomas-*F*	0.046	0.051	0.107	0.044	0.043	0.545	0.106	0.105	0.399	0.115	0.094	0.485

SUPPLEMENTARY MATERIALS

Supplements: R-code for goodness-of-fit (GOF) tests described in the article (GOF.R).

***Phlebocarya filifolia* plants data:** Positions of 207 plants described in the article (Phlebocarya.txt).

Amacrine cells: Positions of 294 on/off cells in the retina of a rabbit (amacrineson.txt and amacrinesoff.txt).

Proof of Proposition 1 and pseudocode of the Monte Carlo-adjusted GOF test.

All files can be found in a single zip file (AGOF.zip).

ACKNOWLEDGMENTS

This publication is based in part on work supported by award no. KUS-C1-016-04 made by King Abdullah University of Science and Technology (KAUST) and by NSF grants DMS-1007504 and DMS-1106494. The authors thank the editor, the associate editor, and a referee for their helpful comments and suggestions. The first author thanks Dr. Naisyin Wang for suggesting the topic of this article. The authors also acknowledge the Texas A&M University Brazos HPC cluster that contributed to the research reported here.

[Received December 2011. Revised December 2012.]

REFERENCES

Assunção, R. M. (1994), "Testing Spatial Randomness by Means of Angles," *Biometrics*, 50, 531–537. [498]

Assunção, R. M., and Reis, I. A. (2000), "Testing Spatial Randomness: A Comparison Between T^2 Methods and Modifications of the Angle Test," *Brazilian Journal of Probability and Statistics*, 14, 71–86. [498]

Baddeley, A. J., Møller, J., and Pakes, A. G. (2008), "Properties of Residuals for Spatial Point Processes," *Annals of the Institute of Statistical Mathematics*, 60, 629–649. [498]

Baddeley, A. J., Møller, J., and Waagepetersen, R. P. (2000), "Non- and Semi-Parametric Estimation of Interaction in Inhomogeneous Point Patterns," *Statistica Neerlandica*, 54, 329–350. [502]

Baddeley, A. J., and Turner, R. (2000), "Practical Maximum Pseudolikelihood for Spatial Point Patterns," *Australian & New Zealand Journal of Statistics*, 42, 283–322. [506]

——— (2005), "Spatstat: An R Package for Analyzing Spatial Point Patterns," *Journal of Statistical Software*, 12, 1–42. [505]

Baddeley, A. J., Turner, R., Møller, J., and Hazelton, M. (2005), "Residual Analysis for Spatial Point Processes" (with discussion), *Journal of the Royal Statistical Society,* Series B, 67, 617–666. [498,500]

Bartlett, M. S. (1964), "The Spectral Analysis of Point Processes," *Biometrika*, 51, 299–311. [502]

Bayarri, M. J., and Berger, J. O. (2000), "*P* Values in Composite Null Models," *Journal of the American Statistical Association*, 95, 1127–1142. [498,500]

Berman, M., and Turner, T. R. (1992), "Approximating Point Process Likelihoods With `GLIM`," *Applied Statistics*, 41, 31–38. [506]

Besag, J. E. (1978), "Some Methods of Statistical Analysis of Spatial Data," *Bulletin of the International Statistical Institute*, 47, 77–92. [506]

Besag, J. E., and Clifford, P. (1991), "Sequential Monte Carlo *p*-Values," *Biometrika*, 78, 301–304. [515]

Besag, J. E., and Diggle, P. J. (1977), "Simple Monte Carlo Tests for Spatial Patterns," *Applied Statistics*, 26, 327–333. [507]

Besag, J. E., and Gleaves, J. T. (1973), "On the Detection of Spatial Pattern in Plant Communities," *Bulletin of the International Statistical Institute*, 45, 153–158. [498]

Cressie, N. A. C. (1993), *Statistics for Spatial Data* (revised ed.), New York: Wiley. [498,502,511]

Diggle, P. J. (2003), *Statistical Analysis of Spatial Point Patterns* (2nd ed.), London: Arnold. [498,500,505,506,513,514,515]

Diggle, P. J., and Gratton, R. J. (1984), "Monte Carlo Methods of Inference for Implicit Statistical Models" (with discussion), *Journal of the Royal Statistical Society,* Series B, 46, 193–227. [514]

Donnelly, K. (1978), "Simulations to Determine the Variance and Edge-Effect of Total Nearest Neighbor Distance," in *Simulation Methods in Archaeology*, ed. I. Hoddler, London: Cambridge University Press, pp. 91–95. [498]

Guan, Y. (2008), "A Goodness-of-Fit Test for Inhomogeneous Spatial Poisson Processes," *Biometrika*, 95, 831–845. [498,505,506,507]

Hines, W. G. S., and Hines, R. J. O. (1979), "The Eberhardt Index and the Detection of Non-randomness of Spatial Point Distributions," *Biometrika*, 66, 73–80. [498]

Huang, F., and Ogata, Y. (1999), "Improvements of the Maximum Pseudo-Likelihood Estimators in Various Spatial Statistical Models," *Journal of Computational and Graphical Statistics*, 8, 510–530. [507,515]

Illian, J., Penttinen, A., Stoyan, H., and Stoyan, D. (2008), *Statistical Analysis and Modelling of Spatial Point Patterns* (Statistics in Practice), West Sussex, England: Wiley. [498,502,505,511,513]

Illian, J. B., Møller, J., and Waagepetersen, R. P. (2009), "Hierarchical Spatial Point Process Analysis for a Plant Community With High Biodiversity," *Environmental and Ecological Statistics*, 16, 389–405. [509]

Marriott, F. H. C. (1979), "Monte Carlo Tests: How Many Simulations," *Applied Statistics*, 28, 75–77. [499,507]

Møller, J., and Waagepetersen, R. P. (2003), *Statistical Inference and Simulation for Spatial Point Processes*, London: Chapman and Hall/CRC Press. [505,512]

R Development Core Team (2012), *R: A Language and Environment for Statistical Computing*, Vienna, Austria: R Foundation for Statistical Computing. [505]

Ripley, B. D. (1977), "Modelling Spatial Patterns" (with discussion), *Journal of the Royal Statistical Society,* Series B, 39, 172–212. [499,502,514]

——— (1988), *Statistical Inference for Spatial Processes*, Australia: Cambridge University Press. [498]

Robins, J. M., Van der Vaart, A., and Ventura, V. (2000), "Asymptotic Distribution of *P* Values in Composite Null Models," *Journal of the American Statistical Association*, 95, 1143–1156. [498]