

# Testing Separability in Spatial-Temporal Marked Point Processes

Frederic Paik Schoenberg

Department of Statistics, 8142 Math-Science Building, University of California,  
Los Angeles, 90095-1554, U.S.A.  
*email:* frederic@stat.ucla.edu

**SUMMARY.** Nonparametric tests for investigating the separability of a spatial-temporal marked point process are described and compared. It is shown that a Cramer–von Mises-type test is very powerful at detecting gradual departures from separability, and that a residual test based on randomly rescaling the process is powerful at detecting nonseparable clustering or inhibition of the marks. An application to Los Angeles County wildfire data is given, in which it is shown that the separability hypothesis is invalidated largely due to clustering of fires of similar sizes within periods of up to 3.9 years.

**KEY WORDS:** Nonparametric tests; Point process; Random time change; Residual analysis; Separability.

## 1. Introduction

Spatial-temporal marked point process models have been increasingly used in a wide variety of applications to represent observations of rare events such as earthquakes, wildfires, sightings of rare species, incidence of epidemics (see Ripley, 1977; Diggle, 1983; Guttorp, 1995; Schoenberg, Brillinger, and Guttorp, 2002a). The models commonly in use in such applications almost invariably have a conditional intensity that has a product form, or that, in the terminology of Cressie (1993), is separable.

The assumption of separability is quite strong: For instance, a separable model for wildfires would posit that the ratio of the risk of a fire of a certain size occurring to that of a fire of another size does not change over time. Despite the importance of this assumption, the separability of such processes is rarely rigorously scrutinized. Ogata (1988) and Schoenberg (2004) used parametric rescaling methods to observe departures from separability in the epidemic-type aftershock sequence (ETAS) model for earthquake occurrences, and Guttorp (1995) detected nonseparability in precipitation data from Middletown, Pennsylvania. Several authors have investigated spatial-temporal changes in mark distributions quite generally (e.g., Ogata and Katsura (1993) and Kagan (1999) in seismology; Johnson (1992), Keeley, Fotheringham, and Morais (1999), and Flannigan and Wotton (2001) in the case of wildfires), and separability tests have been constructed for time series and spatial autoregressive processes (e.g., Shitan and Brockwell, 1995), but other examples of tests for separability in point processes are elusive.

The purpose of the present article is to explore nonparametric methods for testing whether a marked point process is separable. The problem is defined more precisely in terms of conditional intensities in Section 2. In Section 3, direct methods of testing separability based on comparing conditional intensity estimates are explored and in Section 4 their performance is compared under various alternative hypothe-

ses. Section 5 describes a different class of tests, based on assessment of the residuals of the point process after rescaling. These tests are employed in Section 6 to detect departures from separability in wildfire data from Los Angeles County, followed by a summary and conclusions in Section 7.

## 2. Separability

A point process  $N$  is a random collection of points in some metric space  $\mathcal{X}$ . We consider the case of spatial-temporal marked point processes, where  $\mathcal{X}$  is a portion of  $\mathbf{R}^{1+n+d}$  and each point is represented by one temporal coordinate,  $n$  spatial coordinates, and a  $d$ -dimensional mark. In modeling the occurrence of wildfires or earthquakes, for instance, one may identify with each event a point  $(t, \mathbf{x}, m) \in \mathbf{R}^5$ , where  $t$  represents the time of the event's origin,  $\mathbf{x}$  the corresponding three-dimensional location, and  $m$  a real-valued measure of its size.

The construct basic to point process modeling is the Conditional Intensity (CI),  $\lambda(t, \mathbf{x}, m)$ , which may be construed as the limiting expected rate at which points of mark  $m$  accumulate around any location  $(t, \mathbf{x})$  of spacetime, conditional on the history of the process prior to time  $t$ . For background on point processes and conditional intensities, see Fishman and Snyder (1976) and Daley and Vere-Jones (2003).

Of interest here is to investigate whether the CI can be expressed as

$$\lambda(t, \mathbf{x}, m) = f(m)\lambda_1(t, \mathbf{x}), \quad (1)$$

where  $f$  is a fixed nonnegative function and  $\lambda_1$  is a nonnegative predictable process. If (1) holds we call the process separable with respect to the mark  $m$ . If the CI may be further reduced to the form

$$\lambda(t, \mathbf{x}, m) = \lambda_1(t)f_1(\mathbf{x})f_2(m), \quad (2)$$

where  $\lambda_1$  is again nonnegative and predictable and  $f_1$  and  $f_2$  are fixed nonnegative functions, then the process is *completely*

*separable*. An example of complete separability is the compound marginal temporal point process on page 386 of Rathbun (1993), designed to describe the annual germination of plants, where the event locations are modeled as completely independent of the event times: In such cases, the parameters governing each of the marginal processes may be estimated individually.

Even the most ambitious modelers would seldom claim that the temporal and spatial coordinates may be separated as in (2), so complete separability is rarely assumed. However, separability of the mark dimension is nearly always implicitly assumed in models for spatial-temporal marked point processes. The ETAS models of Ogata (1988 and 1998) which are commonly used to model the times, locations, and magnitudes of earthquake hypocenters form an important example. Note that in ETAS models, although the CI of earthquakes at time  $t$  depends on the marks of previous earthquakes, the model is nevertheless separable since the magnitude distribution at time  $t$  does not depend on  $t$  and is not influenced by prior events; see Schoenberg (2004) for further elaboration.

One nonparametric way to investigate the validity of the null hypothesis (1) is the following: obtain a nonparametric (e.g., kernel, spline, wavelet) estimate  $\bar{\lambda}_1$  of the spatial-temporal conditional intensity  $\lambda_1(t, x_1, \dots, x_n)$  and a nonparametric estimate  $\bar{f}$  of the mark density  $f$  (Vere-Jones, 1992; Brillinger, 1998). That is, in the case of  $(n + 1)$ -variate and  $d$ -variate kernel estimates for example, let

$$\bar{\lambda}_1(t, x_1, \dots, x_n) = \int_{\mathcal{X}} k_{n+1}(t - u, x_1 - y_1, \dots, x_n - y_n) dN \times (u, y_1, \dots, y_n, m) \quad (3)$$

and

$$\bar{f}(m) = \int_{\mathcal{X}} k_d(m - m') dN(t, x_1, \dots, x_n, m'), \quad (4)$$

where  $k_{n+1}$  and  $k_d$  are  $(n + 1)$ -dimensional and  $d$ -dimensional kernel densities, respectively.

Next, find a nonparametric estimate  $\hat{\lambda}$  of the full spatial-temporal-marked CI  $\lambda(t, x_1, \dots, x_n, m)$ , e.g., the kernel estimate

$$\begin{aligned} \hat{\lambda}(t, x_1, \dots, x_n, m) \\ = \int_{\mathcal{X}} k_{n+d+1}(t - u, x_1 - y_1, \dots, x_n - y_n, m - m') dN \\ \times (u, y_1, \dots, y_n, m'), \end{aligned} \quad (5)$$

where  $k_{(n+d+1)}$  is an  $n + d + 1$ -dimensional kernel density.

One may then compare the resulting CI estimates  $\hat{\lambda}(t, x_1, \dots, x_n, m)$  and  $\tilde{\lambda}(t, x_1, \dots, x_n, m) := \bar{\lambda}_1(t, x_1, \dots, x_n) \bar{f}(m) / N(\mathcal{X})$ . Thus  $\tilde{\lambda}$  is a *separable* nonparametric CI estimate satisfying (1), while  $\hat{\lambda}$  may be nonseparable.

Comparisons between  $\hat{\lambda}$  and  $\tilde{\lambda}$  apply equivalently to purely temporal marked point processes (where the spatial dimension  $n = 0$ ). Indeed, in examining the separability of the mark coordinate, the dimension of the other (spatial-temporal) coordinates of the process is quite immaterial, in the sense that changes in the mark distribution over time and over space are treated equivalently.

Much has been written about optimally selecting kernel densities and bandwidths and correcting for edge effects. We refer the reader to Silverman (1998). We used Gaussian kernel estimates throughout, with bandwidth  $1/20$  the range in each dimension. Of concern in the present article is not the construction of suitable nonparametric conditional intensity estimates, but rather how to test for separability after such estimates have been obtained.

Note that in some circumstances nonparametric estimates such as the kernel smoothing in (5) may be viewed as estimates of the *overall* intensity or mean measure  $\mu$ , rather than the conditional intensity  $\lambda$ , and in these situations comparisons of  $\bar{\lambda}$  and  $\hat{\lambda}$ , such as those proposed in the following section, may be viewed as tests on whether the function  $\mu$  has a product form analogous to that of  $\lambda$  in (1). However, it is well known that kernel smoothings are in general poor estimates of overall intensities; for instance such estimates exhibit excessive variability when applied to stationary clustered point processes (Musmeci and Vere-Jones, 1986; Vere-Jones, 1992). Hence the tests considered here may primarily be considered tests of the separability of the CI, as expressed in (1).

### 3. Direct Tests of Separability

Under the null hypothesis (1), the two CI estimates  $\hat{\lambda}$  and  $\tilde{\lambda}$  should be similar. One way to compare the two estimates is by finding the maximum or minimum (standardized) absolute difference between the two:

$$\begin{aligned} S_1 &= \sup \left\{ |\hat{\lambda}(t, \mathbf{x}, m) - \tilde{\lambda}(t, \mathbf{x}, m)| / \sqrt{\tilde{\lambda}(t, \mathbf{x}, m)}; (t, \mathbf{x}, m) \in \mathcal{X} \right\}, \\ S_2 &= \inf \left\{ |\hat{\lambda}(t, \mathbf{x}, m) - \tilde{\lambda}(t, \mathbf{x}, m)| / \sqrt{\tilde{\lambda}(t, \mathbf{x}, m)}; (t, \mathbf{x}, m) \in \mathcal{X} \right\}. \end{aligned}$$

Other options include the Cramer-von Mises-type statistic

$$S_3 = \int_0^T \int_{\mathbf{R}^n} \int_{\mathbf{R}^d} [\hat{\lambda}(t, \mathbf{x}, m) - \tilde{\lambda}(t, \mathbf{x}, m)]^2 dm d\mathbf{x} dt,$$

or the difference of the log likelihoods under the two models  $\hat{\lambda}$  and  $\tilde{\lambda}$ :

$$\begin{aligned} S_4 &= \int_{\mathcal{X}} [\log \{\hat{\lambda}(t, \mathbf{x}, m)\} - \log \{\tilde{\lambda}(t, \mathbf{x}, m)\}] dN \\ &\quad - \int_0^T \int_{\mathbf{R}^n} \int_{\mathbf{R}^d} [\hat{\lambda}(t, \mathbf{x}, m) - \tilde{\lambda}(t, \mathbf{x}, m)] dm d\mathbf{x} dt. \end{aligned}$$

Abnormally large values of any of these test statistics indicate a departure from the separability hypothesis (1). Still other possibilities are to examine the squared differences between  $\hat{\lambda}$  and  $\tilde{\lambda}$  at the points of  $N$ , and take their mean  $S_5$  or maximum value  $S_6$  as a test statistic. Though especially simple to compute, such tests have the obvious deficiency that they cannot detect differences on portions of the space  $\mathcal{X}$  where  $N$  has no points.

Statistics similar to those reported above have been used in the detection of spatial clustering. For instance, Anderson and Titterton (1997) investigated the power of statistics such as the integrated squared distance and the maximum absolute difference between two kernel density estimates, to identify clustering in the difference between two interrelated spatial point process data sets. Zimmerman (1993) studied the integrated squared difference between the empirical distribution function of a point process and that of a corresponding

Poisson process as a test for spatial clustering, and compared variations on this statistic to more standard measures of clustering including nearest-neighbor statistics and the  $K$ - and  $L$ -functions of Ripley (1977), which are described further in Section 5. For a survey of clustering tests applied to spatial point processes, see Chapters 8.2 and 8.4 of Cressie (1993).

It is important to note, however, that clustering in spatial-temporal marked point processes is fundamentally different from separability. Indeed, a process may be strongly clustered and yet not be at all in violation of the separability condition (1). For instance, many commonly used Hawkes models, such as the ETAS model of Ogata (1998), exhibit spatial and temporal clustering, but such a model is still separable provided that despite variations in the rate of points over space and time, the mark distribution remains unchanged. When the *marks* are clustered, i.e., when spatial-temporal and mark clustering interact, then the existence of a point at  $(t, \mathbf{x}, m)$  may change the mark distribution so that in the near future the likelihood of a point having a mark near  $m$  increases, and this violates condition (1).

#### 4. Performance of Direct Tests

The performance of the tests of Section 3 may be investigated under various alternatives to (1). One such alternative is that

the interaction between  $m$  and the other variables is additive rather than multiplicative, i.e.,

$$\lambda(t, \mathbf{x}, m) = f(m) + \lambda_1(t, \mathbf{x}), \quad (6)$$

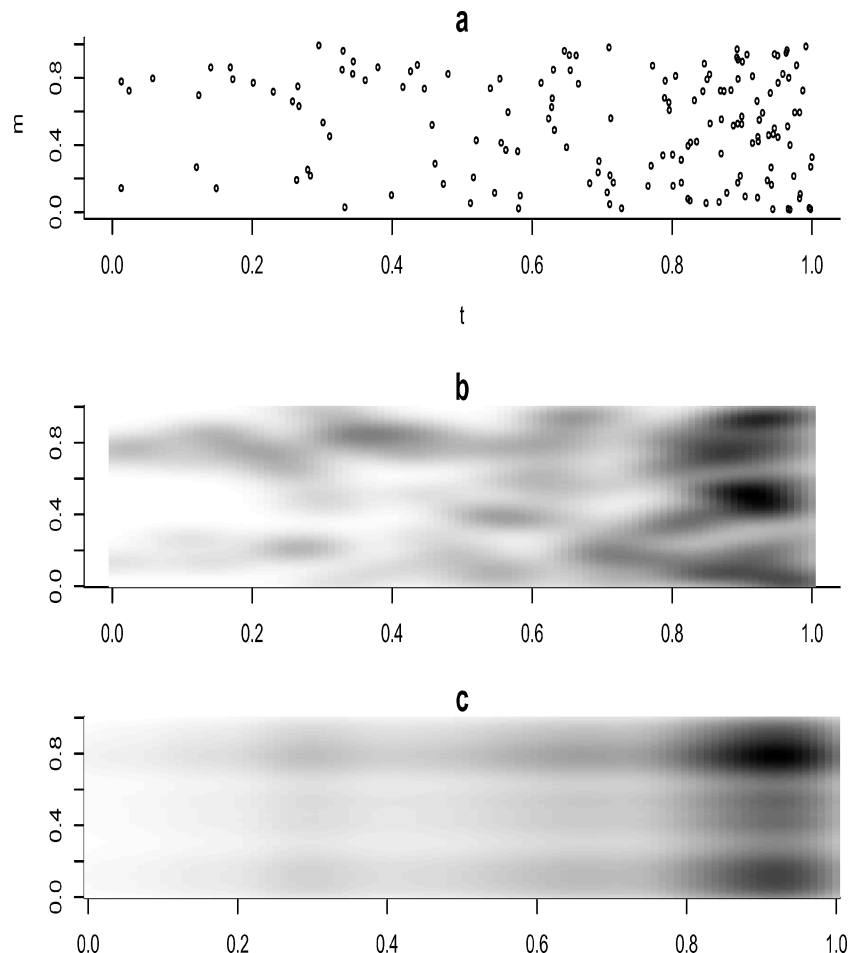
where  $\lambda_1$  is a nonnegative predictable process and  $f$  is a fixed nonnegative function.

To determine which test statistic seems most sensitive to this type of alternative, many realizations of point processes were simulated according to (6). The domain for all the simulations was set to  $[0, 1]$  for  $t$  and  $m$ , and  $[0, 1]^2$  for  $\mathbf{x}$ . For each of the simulations, kernel estimates  $\hat{\lambda}$  and  $\tilde{\lambda}$  were generated, and the Monte Carlo distribution of  $S_1-S_6$  were obtained.

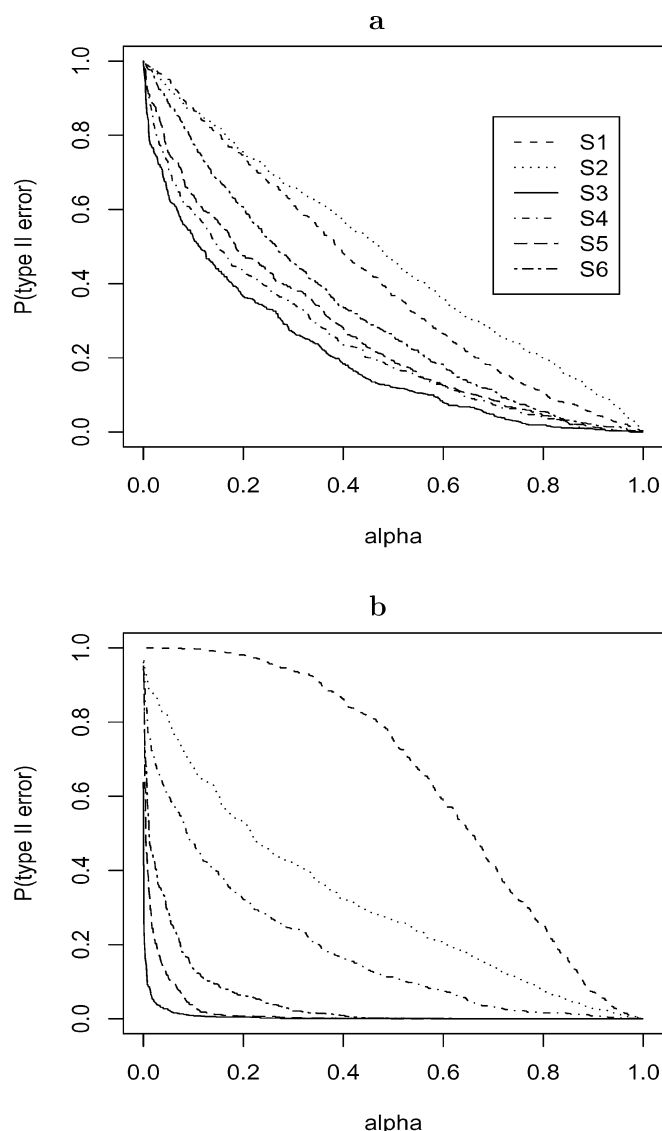
An example is shown in Figure 1. A realization of a Poisson process with nonseparable, additive CI (6) and

$$\lambda_1(t, \mathbf{x}) = \exp(a_1 + b_1 t); \quad f(m) = \exp(a_2 + b_2 m), \quad (7)$$

with  $(a_1, a_2, b_1, b_2) = (3, 3, 3, 1)$ , is depicted in Figure 1a. The model is clearly nonseparable, since near time  $t = 0$ , it is much more likely that a point has a large value of  $m$ , but near time  $t = 1$ , the distribution of  $m$  becomes somewhat closer to uniform. Figures 1b and 1c show kernel CI estimates  $\hat{\lambda}$  and  $\tilde{\lambda}$  for the simulation in Figure 1, integrated over all values of  $\mathbf{x}$ . One sees in Figure 1c the symmetry of the estimate  $\tilde{\lambda}$ ; it is



**Figure 1.** (a) Simulated Poisson process with additive, exponential CI (11 and 12); (b) nonseparable kernel conditional intensity estimate  $\hat{\lambda}$ ; and (c) separable CI estimate  $\tilde{\lambda}$ .



**Figure 2.** Power curves for  $S_1$ – $S_6$  applied to 1000 realizations of (a) the exponential additive process (11 and 12); (b) the uniform additive process (13).

precisely this symmetry that is mandated by the assumption of separability.

Figure 2a shows the power of the six test statistics,  $S_1$  through  $S_6$ , applied to simulations of the process (6 and 7). The process (6 and 7) was simulated 1000 times. For each simulation, the test statistics  $S_1, \dots, S_6$  were computed, and these statistics were compared to an empirical distribution of 1000 simulated separable processes, which was obtained as follows: For each simulation of (6 and 7), its separable CI estimate  $\hat{\lambda}$  was obtained; then separable point processes with CI equal to  $\hat{\lambda}$  were simulated; finally, for each such separable simulation, its CI was estimated via  $\tilde{\lambda}$  and  $\hat{\lambda}$ , and the statistics  $S_1, \dots, S_6$  were computed. After 1000 repetitions of this procedure, we calculated, for each statistic  $S_i$  and for values of the significance level  $\alpha \in (0, 1)$ , the fraction of the simulated nonseparable processes (7) that had values of  $S_i$  greater

than fraction  $1 - \alpha$  of the simulated separable processes. The results shown in Figure 2a suggest that the Cramer-von Mises test  $S_3$  is considerably more powerful than the alternatives. For instance, for  $\alpha = 0.05$ , we found that for 287 of the 1000 simulated nonseparable processes (7) was its test statistic  $S_4$  greater than that of 95% of the simulated separable processes. Hence our estimate of the probability of a type II error for test  $S_4$  at significance level  $\alpha = 0.05$  is  $1 - 28.7\%$ , or 71.3%. By contrast, the estimated probability of a type II error for the test  $S_3$  at  $\alpha = 0.05$  is just 63.8%. For small  $\alpha$ , the weakest of the six tests was  $S_1$ , which essentially could not detect any nonseparability in the process (6 and 7): for  $\alpha = 0.05$  the estimated probability of a type II error for  $S_1$  was 95.1%.

Other nonseparable Poisson models were simulated and the results were similar to that above. The Cramer-von Mises-type statistic  $S_3$  was most powerful at detecting nonseparability, particularly for small values of  $\alpha$ . For instance, for a model such as

$$\lambda(t, \mathbf{x}, m) = 200 (\mathbf{1}_{\{t \in [0.4, 0.6]\}} + \mathbf{1}_{\{m \leq 0.3\}}), \quad (8)$$

the change in the distribution of the  $m$ -values over time is readily apparent, since for  $t$  between 0.4 and 0.6 marks greater than 0.3 are feasible, whereas for all other times such marks are excluded. Figure 2b displays the power of the six test statistics for 1000 simulations of the model (6, 8). Here again the statistic  $S_3$  is substantially more sensitive than the others to the departure from separability in the additive model. Note that the Kolmogorov-Smirnov-type statistic  $S_1$  appears to have surprisingly low power. The reason for this is that for a typical realization of the model (6, 8), differences between  $\tilde{\lambda}$  and  $\hat{\lambda}$  are moderately sized across the entire domain, but there are no times and marks in particular where these differences are especially large, and  $S_1$  fails to detect gradual departures from separability. By contrast, the statistic  $S_3$ , which integrates the squared difference between the two CI estimates over all times and locations, appears to be a very powerful test statistic under Poisson alternatives.

However, while the test statistic  $S_3$  is quite powerful at detecting broad, gradual variations in the conditional distribution of  $m$ , its sensitivity to nonseparability in the form of clustering or inhibition of the marks is not optimal. Figure 3 shows the power of the test statistics  $S_1 - S_6$  for 1000 realizations of temporal marked Hawkes cluster and inhibition processes, simulated using the Lewis-Shedler-Ogata thinning method (Lewis and Shedler, 1979; Ogata, 1981). The Hawkes cluster process corresponding to Figure 3a has CI

$$\lambda(t, m) = \alpha + \beta \sum_{t' < t} \exp\{\gamma(t' - t) - \phi|m - m'|\}, \quad (9)$$

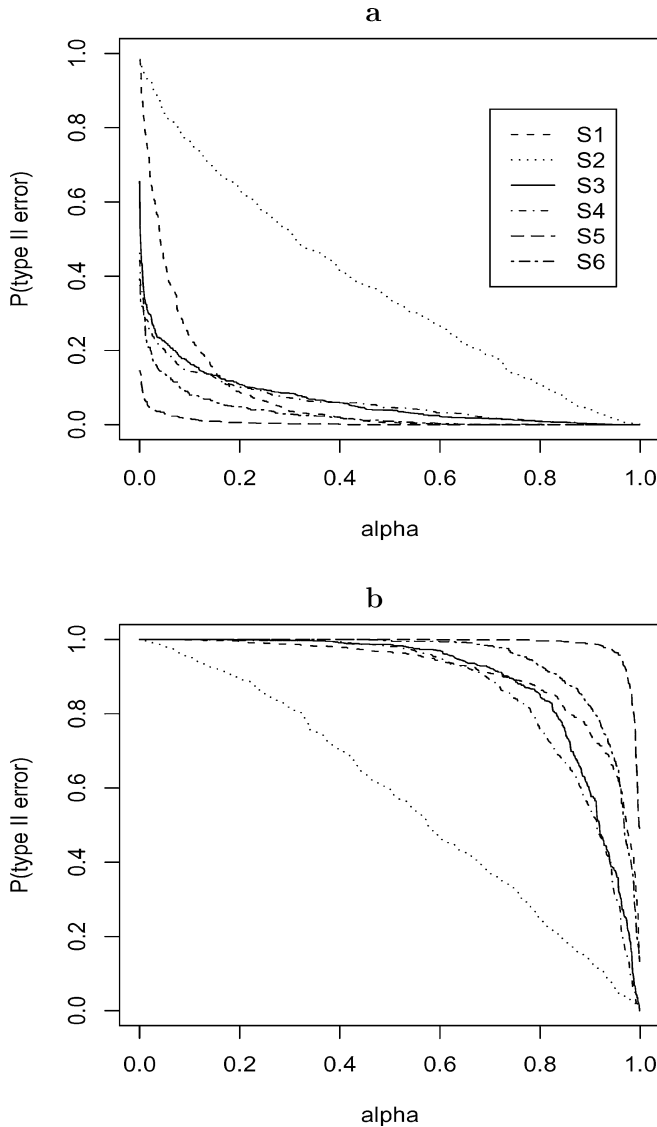
with  $(\alpha, \beta, \gamma, \phi) = (20, 800, 50, 50)$  so the distribution of the marks changes dramatically over time in violation of (1). Figure 3b corresponds to an inhibitory analog of (9) with CI

$$\lambda(t, m) = \left[ \alpha - \beta \sum_{t' < t} \exp\{\gamma(t' - t) - \delta|m - m'|\} \right]^+, \quad (10)$$

with  $\alpha = 100$  and the other parameters the same as in the clustering case. Here a point at time  $t$  and mark  $m$  decreases the CI for points with similar marks at times shortly thereafter,

and the positivity restriction is imposed merely to ensure that the CI is nonnegative. The resulting processes are both highly nonseparable, because for the cluster (resp., inhibition) process the occurrence of a point with mark  $m$  greatly increases (resp., decreases) the likelihood that a point occurring shortly thereafter will have a mark near  $m$ . Thus the mark distribution at any time  $t$  depends on previous points, in violation of (1). However, in these cases the nonseparability is localized, and statistics such as  $S_3$  have difficulty detecting these types of nonseparability.

Figure 3a shows that the test statistic  $S_3$  is somewhat sensitive at detecting nonseparability in the form of clustered marks, though not optimally so.  $S_3$  is a bit more powerful than statistics  $S_1$  and  $S_2$ , but less powerful than the other three statistics. For the inhibitory process (10), all of the six test statistics shown in Figure 3b have extremely low power:



**Figure 3.** Power curves for  $S_1$ – $S_6$  applied to 1000 realizations of (a) a Hawkes process with CI (14); (b) its inhibitory analog with CI (15).

$S_3$  is a bit more powerful than the other statistics with the exception of  $S_2$ , but even  $S_2$  appears to be quite insensitive at detecting this very strong inhibition between the marks. One might suspect that statistics such as  $S_3$  would take on unusually *small* values in the presence of inhibition and could thus be used to detect inhibition of the marks, but unfortunately this is not the case. Indeed, typical values of  $S_3$ , for instance, are well within the range of most of the simulated values for separable processes, and the medians of the two distributions are quite close. The reason for the lack of power of these statistics is not that the statistics take abnormally low values in the presence of inhibition, but rather that their inhibition prevents them from ever taking very large values, whereas separable processes occasionally do.

We applied the six tests listed above to several other different cluster and inhibition processes, and in each case the results were similar to those of the Hawkes cluster process (9) and inhibition process (10) shown in Figure 3. Some results are summarized in Table 1, which shows, for each of the models, the proportion of times each statistic fails to reject the null hypothesis of separability when implemented at significance level  $\alpha = 0.05$ . For each entry in Table 1, the statistics were compared to an empirical distribution of 1000 simulated separable processes exactly as with Figure 2; for each simulation of each model, the estimate  $\hat{\lambda}$  was obtained, then separable point processes with CI equal to  $\hat{\lambda}$  were simulated, and then these separable processes' CIs were in turn estimated via  $\hat{\lambda}$  and  $\hat{\lambda}$ , and the statistics  $S_1, \dots, S_6$  computed.

The temporal marked point processes of Table 1, which were simulated for times  $t$  and marks  $m$  in the unit interval, may be described briefly as follows. A Neyman–Scott cluster process is constructed by generating clusters of points whose centers are governed by a homogeneous Poisson process of rate  $\alpha$ . For the particular example which is simply abbreviated “N-S” in Table 1, the clusters each consist of a fixed integer  $k$  of points which are positioned uniformly and independently within a ball of radius  $r$  around each cluster’s center.

**Table 1**

*Estimated probabilities of type II error at significance level  $\alpha = 0.05$ , for various nonseparable models. All probabilities are expressed in percentages. “Hawkes” denotes the Hawkes cluster model of (9), and “Inhib” refers to its inhibitory analog in (10). “N-S” means Neyman–Scott, with parameters  $\alpha = 20$ ,  $k = 3$ ,  $r = 0.06$ . “C-M” means Cox–Matern, with  $\alpha = 20$ ,  $\beta = 3$ ,  $r = 0.06$ . For the Thomas process,  $\alpha = 20$ ,  $\beta = 3$ ,  $\sigma = 0.03$ . “Mat1” means Matern(I), with  $\alpha = 70$ ,  $r = 0.03$ . “Mat2” means Matern(II), with  $\alpha = 50$ ,  $r = 0.03$ . For the SSI process,  $k = 20$ ,  $r = 0.03$ .*

	$S_1$	$S_2$	$S_3$	$S_4$	$S_5$	$S_6$
Hawkes	43.8	83.9	22.3	20.2	2.9	13.9
N-S	34.7	82.8	14.0	9.1	2.4	31.4
C-M	21.8	81.3	10.0	8.3	0.3	14.2
Thomas	28.5	78.8	8.3	6.6	1.1	13.0
Inhib	100.0	98.1	100.0	100.0	100.0	100.0
Mat1	97.4	94.4	97.7	85.1	99.7	98.5
Mat2	93.9	91.7	96.0	83.3	98.3	98.0
SSI	95.4	92.0	98.5	80.4	98.3	97.5

The Cox–Matern cluster process and Thomas process are two other examples of Neyman–Scott processes, where the cluster sizes are not fixed but instead are i.i.d. Poisson random variables with mean  $\beta$ . For the Cox–Matern process, the points in each cluster are again distributed uniformly and independently within a ball of radius  $r$  around each cluster center, while for the Thomas process these points are distributed independently and isotropically according to a Gaussian distribution with standard deviation  $\sigma$ . The Matern (I) inhibition process is constructed by initially generating a homogeneous Poisson process of rate  $\alpha$ , and then deleting all pairs of points  $(t_i, m_i)$  and  $(t_j, m_j)$  with  $[(t_i - t_j)^2 + (m_i - m_j)^2]^{(1/2)} < r$ . For the Matern (II) model, one generates points according to a homogeneous Poisson process of rate  $\alpha$ , indexes the points  $i = 1, 2, \dots, n$ , at random, and then successively deletes any point  $i$  if it is within a distance  $r$  from any retained point with smaller index. The simple sequential inhibition (SSI) model is constructed similarly, except that candidate points continue to be simulated until ultimately exactly  $k$  points are retained, where  $k$  is some fixed nonrandom integer. For further descriptions see Chapter 8.5 of Cressie (1993).

One sees from Table 1 that the statistics  $S_1$  through  $S_6$  are quite powerless to detect inhibitory behavior in the marks. Regarding detection of clustering of the marks, the results are somewhat mixed. The six statistics have some sensitivity to violations of (1) due to clustering of the marks, but their power is hardly overwhelming. Comparison of the power of these statistics for different processes is complicated by the different distributions of the total numbers of points per realization for each process, but the results generally suggest gross lack of power among the statistics for the inhibitive processes and moderate power at detecting clustering of the marks. An alternative test that is more sensitive to nonseparable inhibition and clustering is described in Section 5.

## 5. Residual Test of Separability

An alternative way to test a point process for separability is to inspect the residuals of the point process after rescaling in such a way that the rescaled points, under the null hypothesis of separability, should be approximately homogeneous Poisson. The source of the rescaling method dates back to Meyer (1971) who showed that for a simple temporal marked point process with countable mark space  $\{m_1, m_2, \dots\}$ , if one transforms each point  $(t, m_i)$  by moving it to  $(\int_0^t \lambda(u, m_i) du, m_i)$ , what results is a sequence of independent Poisson processes of unit rate. Meyer's theorem generalizes readily to multidimensional point processes: one focuses on one nontemporal coordinate  $m$  in  $\mathbf{R}_+$ , and each point  $(t, \mathbf{x}, m)$  is shifted to  $(t, \mathbf{x}, \int_0^m \lambda(t, \mathbf{x}, m') dm')$ . Under quite general conditions, the resulting process is Poisson with unit rate (Schoenberg, 1999).

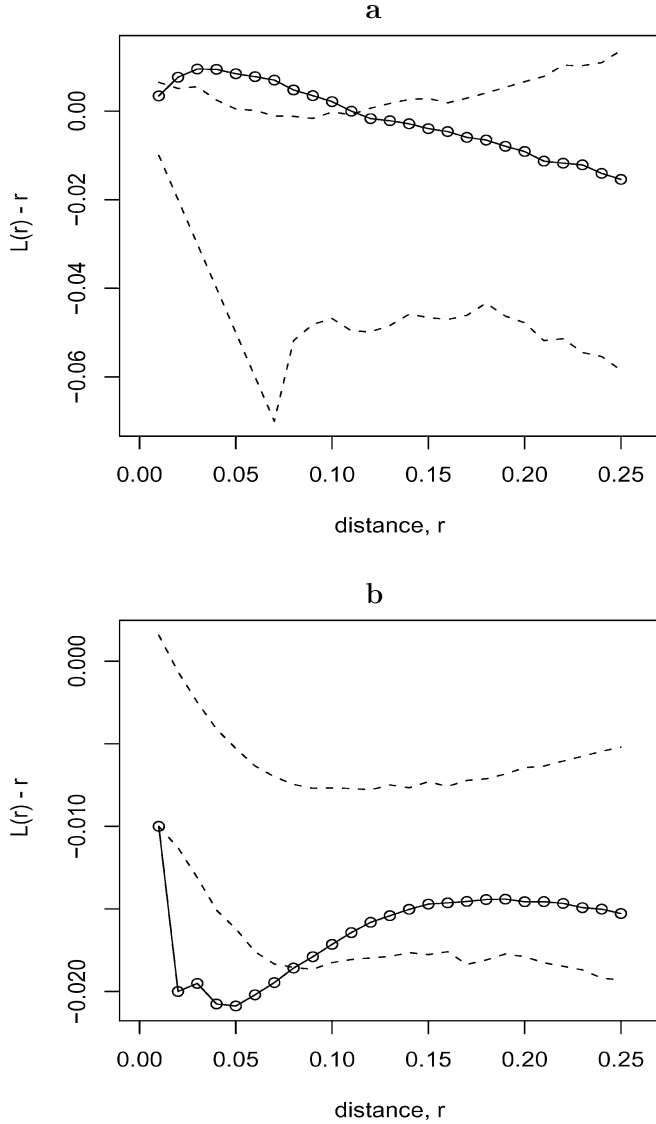
This method of rescaling may be used in conjunction with nonparametric (e.g., kernel) CI estimation to construct a nonparametric test for separability, as follows. Given a nonparametric separable CI estimate such as  $\tilde{\lambda}$  described in Section 2, one may rescale the process, moving each point  $(t, \mathbf{x}, m)$  to

$$\left(t, \mathbf{x}, \int_0^m \tilde{\lambda}(t, \mathbf{x}, m') dm'\right); \quad (11)$$

the resulting points form the *rescaled* or *residual* process. If the process is indeed separable, then  $\tilde{\lambda}$  should closely approximate  $\lambda$  and thus the residual process should closely resemble a Poisson process of unit rate. One may then apply any of a multitude of tests to examine whether this is the case. For instance, section 8.2 of Cressie (1993) lists 17 different tests based on nearest-neighbors, 6 statistics based on quadrat counts, and several others. Perhaps the most used among these, particularly if one is interested in detecting clustering among the residual points, is the estimated  $L$ -function (Ripley, 1977; Diggle, 1983; Baddeley, Møller, and Waagepetersen, 2000). The  $L$ -function, which is a normalized version of  $\hat{K}(u)$ , the (boundary-corrected) mean number of points within a distance  $u$  of any given point, indicates the amount of clustering or inhibition in the process. For instance, in  $\mathbf{R}^2$ , using Euclidean distance,  $\hat{L}(u)$  is defined as  $(\hat{K}(u)/\pi)^{1/2}$ . Positive and negative values of  $\hat{L}(u)$  indicate more or less clustering, respectively, at scale  $u$  than one would expect of a homogeneous Poisson process. Note that clustering of points in the original process  $N$  is quite different from separability, and such clustering will not be present in the residuals if the original process is separable. The reason is that such clusters will concentrate in locations, times, and marks of higher estimated CI, so that the difference between integrals of the form  $\int_0^m \tilde{\lambda}(t, \mathbf{x}, m') dm'$  in (11) will be especially large where such clusters occur and the rescaled mark coordinates of the points will hence be well-dispersed. Thus, clustering in the residuals does not indicate clustering in the original process, but rather is an indication of nonseparable clustering of the marks in violation of the separability hypothesis (1). Note also that the  $L$ -function cannot distinguish nonstationary and clustering alternatives, as is well known, and that the  $L$ -function of the residuals is unlikely to detect departures from separability in the form of gradual changes in rate among the residuals, and hence should be viewed as a complement to, rather than a replacement for, the statistics discussed in Sections 3 and 4.

Figures 4a and 4b summarize the  $L$ -function estimates for the simulated nonseparable Hawkes clustered and inhibitory processes corresponding to Figure 3a and 3b, respectively. The  $L$ -functions are estimated for the residual points (11) for the two processes, after adjusting the axes of the residual processes so that the scales of each dimension are commensurate. The solid curve shows the median centered  $L$ -function for these nonseparable processes, and the dotted curves show upper and lower 95 percentile bounds for their corresponding simulated separable processes (i.e., the simulated processes with separable CI  $\tilde{\lambda}$ , whose CI is in turn estimated via  $\tilde{\lambda}$ ; the points are then rescaled and the  $K$ -function estimated, as with the nonseparable processes). From these bounds, one sees the negative bias in the estimated  $L$ -function applied to the simulated separable processes. This is due to the fact that the CIs of the processes are estimated, causing the residual points to be slightly more well-dispersed than a homogeneous Poisson process (Solow, 1992; Schoenberg, 2002).

Departure from separability is readily detected from Figure 4a, which shows that for a typical realization of this Hawkes clustered process, its residuals remain significantly clustered at small scales after rescaling according to  $\tilde{\lambda}$  as in (11). Similarly, from Figure 4b one discerns that the residuals of a typical realization of the process (10) are significantly



**Figure 4.** Estimated centered  $L$ -functions applied to the rescaled residuals of (a) a Hawkes process with CI (14); (b) its inhibitory analog with CI (15). The solid curves show the median value of  $\hat{L}(r) - r$  over 1000 realizations of each process. The dashed curves represent empirical 95% bounds for the separable Monte Carlo distributions, i.e., for 1000 simulated separable processes whose true conditional intensities are given by the separable CI estimate  $\tilde{\lambda}$  applied to each Hawkes and inhibitory process.

less clustered than one would expect if the process were indeed separable. Although the median  $L$ -function estimates shown in Figure 4a and 4b do not appear to be very far from the 95% bounds, they are nevertheless highly significant. Note that the linear decrease in the lower bound is due to the fact that, for a substantial percentage of the simulated separable processes, the rescaled residuals are so well dispersed that there are no pairs of points at all within these small distances of one another. For the other clustered and inhibitive processes listed in Table 1, the results were quite similar to Figure 5a and 5b,

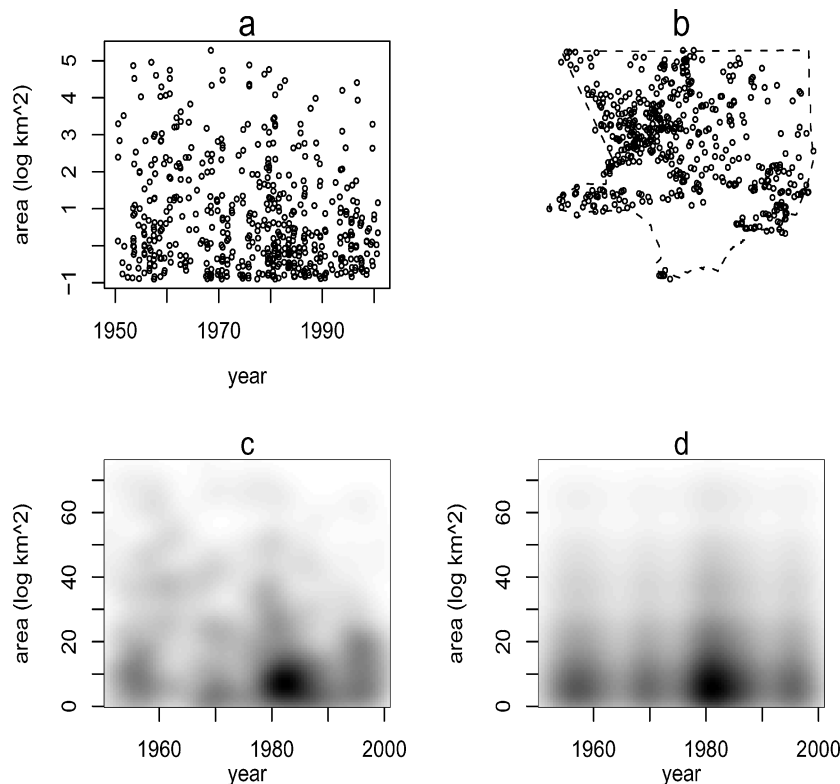
respectively. The results suggest that the  $\hat{L}$ -test applied to the rescaled residuals can be a quite powerful test for nonseparable clustering and inhibition in violation of assumption (1), compared with  $S_3$  or the other tests described in Section 3.

## 6. Application: Los Angeles County Wildfires

Wildfires in Los Angeles County, California, are important public safety concerns, often causing significant ecological upheaval, millions of dollars in property damage, and occasionally loss of human lives (Whelan, 1995; Pyne, Andrews, and Laren, 1996). The hot, dry climate, the warm, seasonal Santa Ana winds, and the fact that the predominant vegetation is highly flammable chaparral combine to make Southern California one of the most fire-prone areas in the world (Yool, Eckhardt, and Cosentino, 1985; Keeley and Keeley, 1988). Data on wildfires have been systematically recorded in Los Angeles County since the late 19th century; the records on fires burning greater than 100 acres are believed to be nearly complete dating back to 1950. Figures 5a and 5b show the times, centroid locations, and burn areas of such wildfires, from January 1950 to January 2001, recorded by the Los Angeles County Fire Department (LACFD) and Los Angeles County Department of Public Works (LACDPW). A few of the wildfires have centroids outside the County border; all such fires are included provided they burned at least 100 acres of area in Los Angeles County. Numerous variables on each fire have been analyzed (see, e.g., Schoenberg et al., 2002b), but for this example we focus exclusively on wildfire sizes as marks. Point process models for such data sets, as well as other forestry data, are typically separable (Stoyan and Pettinen, 2000), though this assumption is generally not checked.

Nonseparable and separable kernel CI estimates of the LACFD data are shown in Figure 5c and 5d. Though some clustering appears to be present, no other obvious departures are immediately observable at a glance. The distribution of wildfire sizes does not appear to change drastically over time. The statistic  $S_3$  applied to the wildfire data has a  $p$ -value of 14.9%, when compared with its Monte Carlo distribution based on simulating 1000 separable processes each with CI  $\tilde{\lambda}$  shown in Figure 5d. As  $S_3$  is quite powerful at detecting nonseparability in the form of broad regions where the distribution of wildfire sizes changes, the fact that the observed value is not significant suggests that such departures from separability in this data set are not excessive.

On the other hand, there does seem to be highly significant clustering of wildfire sizes, in violation of the separability hypothesis, as demonstrated in Figure 6. Figure 6a shows the times and rescaled sizes of the rescaled residual points, using the separable space-time-mark CI estimate  $\tilde{\lambda}$ . The high volatility of the boundary is a result of the fact that  $\tilde{\lambda}$  varies dramatically across locations, resulting in large changes of the integral  $\int_0^m \tilde{\lambda}(t, \mathbf{x}, m') dm'$  from point to point. One can more easily discern how the wildfire sizes vary over time from Figure 6b, which shows the rescaled residual points after integrating  $\tilde{\lambda}$  over all spatial locations,  $\mathbf{x}$ . From Figure 6b one may detect some clustering of the marks in the residual points directly. This is confirmed from Figure 6c, which shows the estimated centered  $L$ -function of these rescaled points, along with 99%-confidence bounds based on 1000 Monte Carlo



**Figure 5.** (a) Times and burn areas greater than 100 acres from January 1950 to January 2001 in Los Angeles County, according to LACFD wildfire data; (b) centroid locations of each such fire burn area; (c) nonseparable kernel CI estimate  $\tilde{\lambda}$ ; and (d) separable kernel CI estimate  $\hat{\lambda}$ .

simulations: For each of the 1000 simulations of a separable process with CI  $\tilde{\lambda}$  shown in Figure 5d, we reestimated its CI via kernel smoothing, constructing a new separable conditional intensity estimate  $\tilde{\lambda}'$ , and then rescaled this simulated process according to  $\tilde{\lambda}'$  and calculated the resulting  $L$ -function estimate. From Figure 6c one sees that the clustering in the residuals is highly significant for rescaled distances up to 0.077, which correspond to differences between points in the original data set of up to 3.9 years.

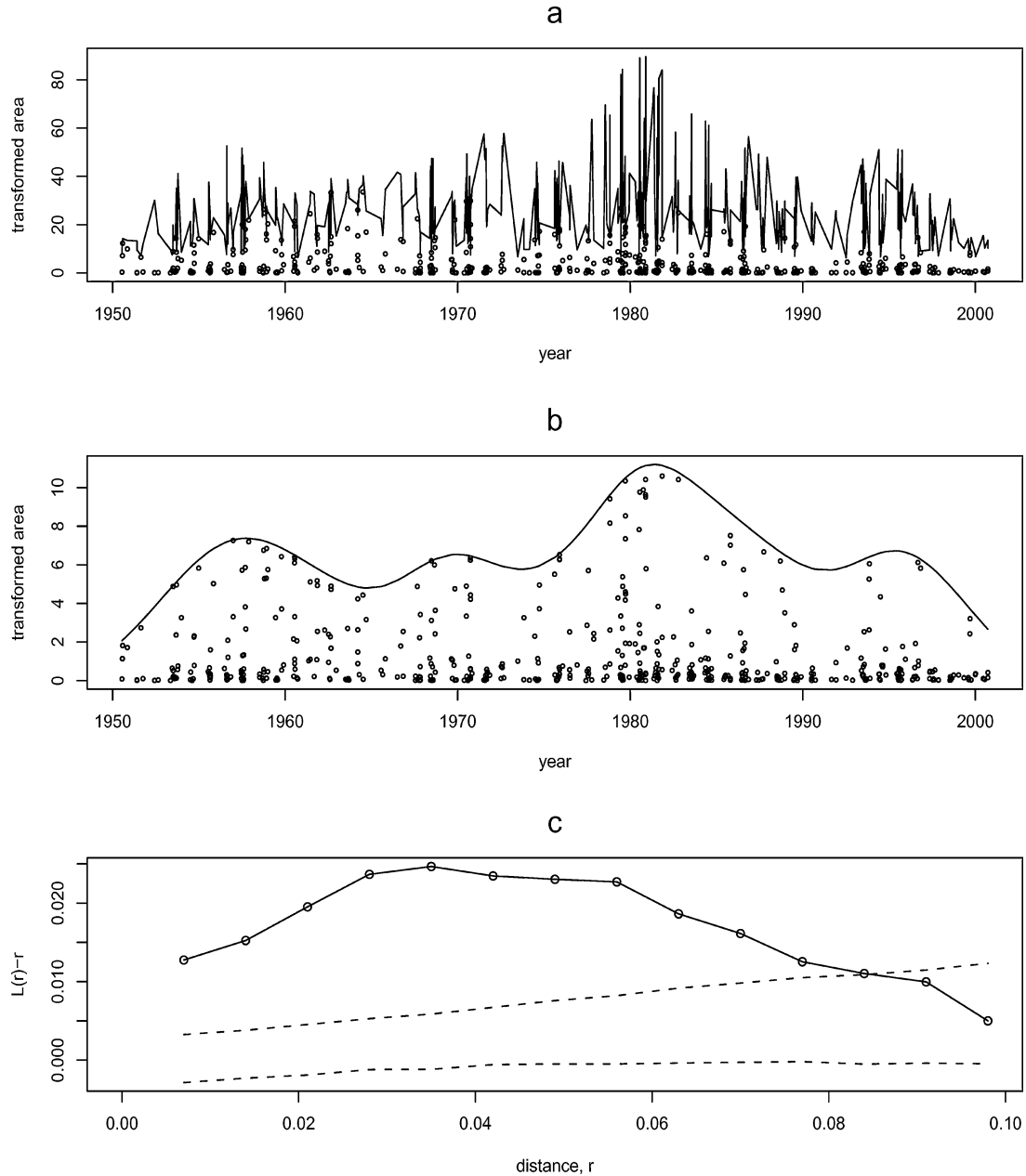
Further inspection of Figures 5a and 6b can illuminate some of the main sources of clustering of the marks in the residuals. In rescaling, the mark axis is stretched where  $\tilde{\lambda}$  is large and compressed where  $\tilde{\lambda}$  is small. Hence clusters of fires for which  $\tilde{\lambda}$  is small, i.e., those of large area in times when fewer fires occurred, are moved even closer together, and a large fraction of the clustering in the residuals is attributable to such clusters. For instance, consider the two fires in Figure 5a occurring in the year 1970 and with areas of 4.5 and 4.7 log km<sup>2</sup>. In Figure 5a these two points do not appear particularly close together. However, fires of that size are rare, so  $\tilde{\lambda}$  is small in that portion of the space. Therefore the residuals corresponding to these two points are bunched very closely together: Their  $y$ -coordinates in Figure 6b are both approximately 6.3. Another example is the cluster of four fires occurring in 1997, of sizes 0.54 to 0.75 log km<sup>2</sup>. Because 1997 was a year with relatively few fires,  $\tilde{\lambda}$  is rather low in this year, so these fires are clumped together in the residual plot of Figure 6b, with

all four residual points having  $y$ -coordinates of very nearly 0.5.

## 7. Discussion

While the lack of a very significant gradual change in the wildfire area distribution over time is not surprising, the significant, nonseparable small-scale clustering observed in the Los Angeles County wildfire dataset may seem curious. Note that most of the fires within any given year occur at very disparate spatial locations, and the notion that they are causally related to one another, i.e., that certain fires are causing other fires of similar size to occur shortly thereafter, seems highly implausible. The clustering may not be attributed to boundary effects or bias due to the estimation of  $\tilde{\lambda}$ , since the  $L$ -function estimates were boundary-corrected and the simulated separable processes used for the confidence bounds in Figure 6c had rescaled processes with similar boundaries and had their CIs estimated using the identical method as that used on the wildfire data. Nor can the apparent nonseparability reasonably be attributed to errors or rounding in the data set: Wildfire areas were recorded by LACFD officials using digitized wildfire maps which are believed to be accurate to approximately 10–20 meters (Schoenberg et al., 2002b). The apparent clustering in the data set also cannot be due to the insuitability of a model, because the methods used in its detection were nonparametric.





**Figure 6.** (a) Rescaling of LACFD wildfire data using the separable space-time-magnitude CI estimate  $\tilde{\lambda}(t, \mathbf{x}, m)$ ; (b) rescaling of LACFD wildfire data using the separable CI estimate  $\tilde{\lambda}(t, m)$ , i.e., the integral of  $\tilde{\lambda}(t, \mathbf{x}, m)$  over all locations; (c) estimated  $L$ -function (solid curve) applied to rescaled residuals in (b), along with 99% bounds (dashed curves) for  $L$ -function estimates applied to 1000 Monte Carlo simulations of separable processes with CI  $\tilde{\lambda}(t, m)$ , after estimating the CI of each simulated process and rescaling it accordingly.

One possible explanation is that the clustering of sizes in the wildfire data set may be partly due to climatic and temporal variations. Not only does fire incidence vary dramatically by season in Los Angeles, largely due to the presence or absence of warm Santa Ana winds, but also average temperatures in Los Angeles County increased quite dramatically throughout the 20th century while precipitation and winds exhibited significant variations (Moritz, 1997; Keeley and Fotheringham, 2001; Schoenberg et al., 2002b),

and fires of certain sizes may be especially likely under particular weather conditions. Clustering of marks may similarly result from variations in human response or ignition patterns, which may cause more fires of a given size to occur in some periods rather than others (Kauffman, 1993). In addition, clustering of the type observed here may be a natural feature of the mosaic of fire patterns in Los Angeles County. For instance, in view of the often-noted cyclic or renewal-type behavior of wildfires due to slow regeneration of fuel in general

and chaparral in particular (Hanes, 1971; Johnson and Gutsell, 1994), clusters of fires of similar size that happen to occur at one time may tend to repeat, resulting in multiple clusters throughout the data set. However, perhaps the most plausible explanation is that separability is simply a very strong condition. There are many ways in which the size distribution can change over time, so it should not be surprising to observe some significant nonseparability in a data set; rather it would be quite surprising if the separability condition were met. This last conclusion has serious implications for multidimensional point process modeling, in which at present separable models are regularly assumed without testing whether the assumption of separability appears to be reasonable.

The Cramer-von Mises type statistic  $S_3$  and the  $L$ -function applied to rescaled residuals appear to be quite powerful tests for separability in multidimensional point process models. The statistic  $S_3$  seems most sensitive to gradual or global changes in the distribution of the marks, while the  $L$ -function on the residuals appears to capture local nonseparability in the form of clustering and inhibition of the marks quite well. This conclusion closely parallels that of Zimmerman (1993), whose study of tests for detecting inhomogeneity in spatial point processes suggests that Cramer-von Mises type tests are powerful against broad, global departures from spatial homogeneity, but weaker than distance-based methods such as the  $L$ -function in the face of clustered and inhibitory alternatives. The application of separability tests to other wildfire data sets in other areas, and to other point process data in general, should be performed in the future.

#### ACKNOWLEDGEMENTS

This material is based upon work supported by the National Science Foundation under grant 9978318. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation. The author is especially grateful to Jean Jacod and l'Université Paris VI, where much of this research was performed. Thanks also to James Woods, the LACDPW, and LACFD (especially Mike Takeshita and Frank Vidales) for their generosity in sharing data, to Adrian Baddeley and Rolf Turner for their very useful *spatstat* R package, and to Roger Peng, Emmanuel Roy, James Woods, the associate editor, and two anonymous referees for helpful suggestions.

#### RÉSUMÉ

Des tests non paramétriques pour étudier la séparabilité d'un processus spatiaux temporels de points marqués sont décrits et comparés. Il est montré qu'un test de type Cramer-von Mises est très puissant dans la détection de déviations graduelles de la séparabilité, et qu'un test résiduel basé sur la re-gradation aléatoire du processus est puissante dans la détection de groupes non séparables ou d'inhibition des marques. Une application aux données de traînée de poudre du comté de Los Angeles est réalisée, on montre que l'hypothèse de séparabilité n'est pas validée en raison de feux de groupes de tailles semblables pendant les périodes et jusqu'à 3.9 ans.

#### REFERENCES

- Anderson, N. H. and Titterton, D. M. (1997). Some methods for investigating spatial clustering, with epidemiological applications. *Journal of Royal Statistical Society A* **160**, 7–105.
- Baddeley, A. J., Møller, J., and Waagepetersen, R. (2000). Non- and semi-parametric estimation of interaction in inhomogeneous point patterns. *Statistica Neerlandica* **54**, 329–350.
- Brillinger, D. R. (1998). Some wavelet analyses of point process data. In *31st Asilomar Conference on Signals, Systems and Computers*, 1087–1091. IEEE Computer Society. California: Los Alamitos.
- Cressie, N. A. (1993). *Statistics for Spatial Data, Revised edition*. New York: Wiley.
- Daley, D. and Vere-Jones, D. (2003). *An Introduction to the Theory of Point Processes*, 2nd edition. New York: Springer.
- Diggle, P. (1983). *Statistical Analysis of Spatial Point Patterns*. London: Academic Press.
- Fishman, P. M. and Snyder, D. L. (1976). The statistical analysis of space-time point processes. *IEEE Transactions on Information Theory*, **IT-22**, 257–274.
- Flannigan, M. and Wotton, B. (2001). Climate, weather, and area burned. In *Forest Fires: Behavior and Ecological Effects* E. Johnson and K. Miyanishi (eds), 351–373. San Diego: Academic Press.
- Guttorp, P. (1995). *Stochastic Modeling of Scientific Data*. London: Chapman and Hall.
- Hanes, T. (1971). Succession after fire in the chaparral of Southern California. *Ecological Monographs* **41**, 27–52.
- Johnson, E. (1992). *Fire and Vegetation Dynamics: Studies from the North American Boreal Forest*. Cambridge: Cambridge University Press.
- Johnson, E. and Gutsell, S. (1994). Fire frequency models, methods and interpretations. *Advances in Ecological Research* **25**, 239–287.
- Kagan, Y. (1999). Universality of the seismic moment frequency relation. *Pure Applied Geophysics* **155**, 233–258.
- Kauffman, J. (1993). Group report: The role of humans in shaping fire regime and ecosystem properties. In *Fire in the Environment: The Ecological, Atmospheric and Climatic Importance of Vegetation Fires*. Chichester, England: Wiley.
- Keeley, J. E. and Keeley, S. (1988). Chaparral. In *North American Terrestrial Vegetation*, M. G. Barbour and D. W. Billings (eds), 165–207. Cambridge: Cambridge University Press.
- Keeley, J. E. and Fotheringham, C. J. (2001). History and management of crown-fire ecosystems: A summary and response. *Conservation Biology* **15**, 1561–1567.
- Keeley, J., Fotheringham, C., and Morais, M. (1999). Re-examining fire suppression impacts on brushland fire regimes. *Science* **284**, 1829–1832.
- Lewis, P. and Shedler, G. (1979). Simulation of nonhomogeneous Poisson processes by thinning. *Naval Research Logistics Quarterly* **26**, 403–413.
- Meyer, P. (1971). Démonstration simplifiée d'un théorème de Knight. *Lecture Notes in Mathematics* **191**, 191–195.

- Moritz, M. (1997). Analyzing extreme disturbance events: Fire in Los Padres National Forest. *Ecological Applications* **7**, 1252–1262.
- Musmeci, F. and Vere-Jones, D. (1986). A variable-grid algorithm for smoothing clustered data. *Biometrics* **42**, 483–494.
- Ogata, Y. (1981). On Lewis' simulation method for point processes. *IEEE Transactions on Information Theory* **27**, 23–31.
- Ogata, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of American Statistical Association* **83**, 9–27.
- Ogata, Y. (1998). Space-time point process models for earthquake occurrences. *Ann. Inst. Statist. Math.* **50**, 379–402.
- Ogata, Y. and Katsura, K. (1993). Analysis of temporal and spatial heterogeneity of magnitude frequency distribution inferred from earthquake catalogs. *Geophysical Journal International* **113**, 727–738.
- Pyne, S., Andrews, P., and Laren, R. (1996). *Introduction to Wildland Fire*. New York: Wiley.
- Rathbun, S. L. (1993). Modeling marked spatio-temporal point patterns. *Bulletin of International Statistical Institute* **55**, 379–396.
- Ripley, B. (1977). Modelling spatial patterns. *JRSS-B* **39**, 172–192.
- Schoenberg, F. P. (1999). Transforming spatial point processes into poisson processes. *Stochastic Processes and their Applications* **81**, 155–164.
- Schoenberg, F. P. (2002). On rescaled Poisson processes and the Brownian bridge. *Annales of International Statistical Mathematics* **54**, 445–457.
- Schoenberg, F. P. (2004). Multi-dimensional residual analysis of point process models for earthquake occurrences. *Journal of American Statistical Association* (in press).
- Schoenberg, F. P., Brillinger, D. R., and Guttorp, P. M. (2002a). Point processes, spatial-temporal. In *Encyclopedia of Environmetrics*, A. El-Shaarawi and W. Piegorsch (eds), Volume 3, 1573–1577. New York: Wiley.
- Schoenberg, F. P., Peng, R., Huang, Z., and Rundel, P. (2002b). Detection of nonlinearities in the dependence of burn area on fuel age and climatic variables. *International Journal of Wildland Fire* **12**, 1–10.
- Shitan, M. and Brockwell, P. J. (1995). An asymptotic test for separability of a spatial autoregressive model. *Communications Statistics Theory Mathematics* **24**, 2027–2040.
- Silverman, B. W. (1998). *Density Estimation for Statistics and Data Analysis*, 2nd ed. Boca Raton: Chapman and Hall.
- Solow, A. (1992). Model-checking in non-stationary Poisson processes. *Applied Stochastic Models and Data Analysis* **8**, 129–132.
- Stoyan, D. and Penttinen, A. (2000). Recent applications of point process methods in forestry statistics. *Statistical Science* **15**, 61–78.
- Vere-Jones, D. (1992). Statistical methods for the description and display of earthquake catalogs. In *Statistics in Environmental and Earth Sciences*, A. Walden and P. Guttorp (eds), 220–236. London: Edward Arnold.
- Whelan, R. (1995). *The Ecology of Fire*. Cambridge: Cambridge University Press.
- Yool, S., Eckhardt, D., and Cosentino, M. (1985). Describing the brushfire hazard in Southern California. *Analysis of Association of American Geography* **75**, 431–442.
- Zimmerman, D. L. (1993). A bivariate Cramér-von Mises type of test for spatial randomness. *Applied Statistics* **42**, 43–54.

Received April 2003. Revised November 2003.

Accepted November 2003.

## Query

**Q1** Author: Please update Schoenberg (2004).