Title
Testing Separability in Multi-dimensional Point Processes

Author
Schoenberg, Frederic

Publication Date
2003

Testing separability in multi-dimensional point processes

Frederic Paik Schoenberg[1], *University of California, Los Angeles*

Abstract

Nonparametric tests for investigating the separability of a multi-dimensional 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 non-separable clustering or inhibition. 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 one to two years.

Key words: point process, separability, non-parametric tests, residual analysis, random time change, thinning, model evaluation.

[1] Department of Statistics, 8142 Math-Science Building, University of California, Los Angeles, 90095-1554. frederic@stat.ucla.edu

1 Introduction.

Multi-dimensional point process models, such as spatial-temporal point process models, have been increasingly used in a wide variety of applications to represent observations of discrete events such as earthquakes, wildfires, sightings of rare species, incidence of epidemics, etc. (see Ripley 1977, Diggle 1983, Schoenberg et al. 2002, and references therein). 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 location burning to that of another location burning does not change over time. Despite the importance of this assumption, the separability of such processes is rarely rigorously scrutinized. Ogata (1988) and more recently Schoenberg (2003) used parametric rescaling methods to observe departures from separability in the epidemic-type aftershock sequence (ETAS) model for earthquake occurrences, and separability tests have been constructed for time series and spatial autoregressive processes (see e.g. Shitan and Brockwell, 1995, and references therein), but if there are other examples of tests for separability in point processes, they are highly elusive.

The purpose of the present paper is to explore non-parametric methods for testing whether a multi-dimensional point process is separable. Following a brief review of point process terminology in Section 2, the problem is defined more precisely in terms of conditional intensities in Section 3. In Section 4, direct methods of testing separability based on comparing conditional intensity estimates are explored and in Section their performance is compared under various alternative hypotheses. Section 6 describes a different class of tests, based on assessment of the residuals of the point process in the transformed space after randomly rescaling using the method of Meyer (1971) and its generalizations. These tests are employed in Section 7 to detect departures from separability in wildfire data from Los Angeles County, followed by a summary and conclusions in Section 8.

2 Preliminaries.

A point process N is thought of as a random collection of points in some metric space $\mathcal{X}$, and is defined mathematically as a σ-finite random measure on $\mathcal{X}$, taking values in the non-negative integers or infinity. Thus $N(B)$ represents the number of points in a subset B of $\mathcal{X}$. We consider the case where $\mathcal{X}$ is a portion of space-time $\mathbf{R}^n$ and that the total number of observed points $N(\mathcal{X})$ is finite. We consider the first coordinate of the space to represent time and refer to the marks, i.e. the other $n-1$ coordinates, as *spatial*, though in applications they may not represent locations but instead other information about each point.

Formally, the point process N is considered adapted to some filtered probability space $(\Omega, \mathcal{F}_t, P)$, and the construct basic to point process modeling is the conditional intensity, λ, which is the unique (up to null sets) $\mathcal{F}$-predictable process such that $\int [dN(t, \mathbf{x}) - \lambda(t, \mathbf{x}) dt d\mathbf{x}]$ is a martingale. Assuming λ exists, it is typically construed as the limiting expected rate at which points accumulate at any location $(t, \mathbf{x}) = (t, x_1, ..., x_{n-1})$ of space-time, conditional on the history $\mathcal{F}_t^-$, which contains all previous information about the process prior to time t. As in most applications, we consider here the intensity conditional on one of the parameters, t. For point processes with simple ground process (i.e. for which with probability one, no two points occur at exactly the same time), the conditional intensity λ completely characterizes the finite-dimensional distributions of the process N; hence in modeling N it suffices to prescribe a model for λ. Important examples of point process models include the Poisson process, for which λ is deterministic, and Hawkes processes (Hawkes, 1971), which have the characteristic that a point at $(t, \mathbf{x})$ increases the conditional intensity thereafter; Hawkes processes are described in more detail below. For further introduction to point processes and conditional intensities, see Jacod (1975), Fishman and Snyder (1976), Cliff and Ord (1981), Diggle (1983), Daley and Vere-Jones (1988), Karr (1991), Andersen et al. (1993), Schoenberg et al. (2002a).

3 Separability

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

$$\lambda(t, x_1, ..., x_{n-1}) = \lambda_1(t)\lambda_2(x_1, ..., x_{n-1}), \tag{1}$$

where λ_2 is a fixed non-negative function and λ_1 is a non-negative $\mathcal{F}$-predictable process. If (1) holds we call the process *separable*. If furthermore the conditional intensity may be further reduced to the form

$$\lambda(t, x_1, ..., x_{n-1}) = \lambda_1(t) f_1(x_1), ..., f_{n-1}(x_{n-1}), \tag{2}$$

where λ_1 is again non-negative and $\mathcal{F}$-predictable and each f_i is a fixed non-negative function, then the process is *completely separable*. A hypothetical example of complete separability is given e.g. in Rathbun (1993): in such cases, the parameters governing each of the marginal processes may be estimated individually.

In most applications, not even the most ambitious modelers would claim that all dimensions, e.g. longitude and latitude, are separable, so complete separability is rarely assumed. However, separability is nearly always implicitly assumed in models for spatial-temporal and marked temporal point processes. An important example is the epidemic-type aftershock sequence (ETAS) model of Ogata (1988), commonly used to model the times and magnitudes of earthquake hypocenters. Note that in the ETAS model, although the conditional intensity of earthquakes at time t depends on the marks of previous earthquakes, the model is nevertheless separable since the magnitude distribution is not influenced by prior events; see Schoenberg (2003) for further elaboration.

One non-parametric way to investigate the validity of the null hypothesis (1) is the following: obtain a non-parametric (e.g. kernel, spline, wavelet) estimate $\bar{\lambda}_1$ of the temporal intensity λ_1 and another $\bar{\lambda}_2$ of the spatial intensity λ_2 (Vere-Jones, 1992; Brillinger, 1998). That is, in the case of the univariate and $n-1$-variate kernel estimates for example, let

$$\bar{\lambda}_1(t) = \int_{\mathcal{X}} k_1(t-u)dN(u, x_1, ..., x_{n-1}) \tag{3}$$

and

$$\bar{\lambda}_2(x_1, ..., x_{n-1}) = \int_{\mathcal{X}} k_{n-1}(x_1 - y_1, ..., x_{n-1} - y_{n-1})dN(t, y_1, ..., y_{n-1}), \tag{4}$$

where k_1 and k_{n-1} are one-dimensional and $n-1$-dimensional kernel densities, respectively.

Next, find a non-parametric spatial-temporal estimate $\hat{\lambda}$ of $\lambda(t, x_1, ..., x_{n-1})$, e.g. the kernel estimate

$$\hat{\lambda}(t, x_1, ..., x_{n-1}) = \int_{\mathcal{X}} k_n(t-u, x_1 - y_1, ..., x_{n-1} - y_{n-1})dN(u, y_1, ..., y_{n-1}), \tag{5}$$

where k_n is an n-dimensional kernel density.

One may then compare the resulting spatial-temporal intensity estimates $\hat{\lambda}(t, x_1, ..., x_{n-1})$ and $\tilde{\lambda}(t, x_1, ..., x_{n-1}) = \bar{\lambda}_1(t)\bar{\lambda}_2(x_1, ..., x_{n-1})/N(\mathcal{X})$. Thus $\tilde{\lambda}$ is a *separable* non-parametric intensity estimate satisfying (1), while $\hat{\lambda}$ may be non-separable.

Much has been written about optimally selecting kernel densities and bandwidths and correcting for edge effects. We refer the reader to Vere-Jones (1992). Of concern in the present article is not the construction of suitable non-parametric intensity estimates, but rather how to test for separability after such estimates have been obtained. We note, however, that for approximating the conditional intensity of a Hawkes process or other clustered or inhibitory process, it may be desirable to use a one-sided kernel density rather than a symmetric one. The conditional intensity of a Hawkes process is typically left-continuous but with a discontinuity at each point of the process, and the same is true of kernel intensity estimates when a left-continuous kernel density with support on $(0, \infty)$ is used.

4 Direct tests of separability.

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

$$S_1 = \sup\{|\hat{\lambda}(t, \mathbf{x}) - \tilde{\lambda}(t, \mathbf{x})|/\sqrt{\tilde{\lambda}(t, \mathbf{x})}; (t, \mathbf{x}) \in \mathcal{X}\} \tag{6}$$

or

$$S_2 = \inf\{|\hat{\lambda}(t, \mathbf{x}) - \tilde{\lambda}(t, \mathbf{x})|/\sqrt{\tilde{\lambda}(t, \mathbf{x})}; (t, \mathbf{x}) \in \mathcal{X}\}. \tag{7}$$

Other options include the Cramer-von Mises-type statistic

$$S_3 = \int_0^T \int_{\mathbf{x}} [\hat{\lambda}(t, \mathbf{x}) - \tilde{\lambda}(t, \mathbf{x})]^2 d\mathbf{x} dt, \tag{8}$$

or the log-likelihood ratio statistic

$$S_4 = \int_{\mathcal{X}} [\log\{\hat{\lambda}(t, \mathbf{x})\} - \log\{\tilde{\lambda}(t, \mathbf{x})\}] dN - \int_0^T \int_{\mathbf{x}} [\hat{\lambda}(t, \mathbf{x}) - \tilde{\lambda}(t, \mathbf{x})] d\mathbf{x} dt. \tag{9}$$

Abnormally large values of any of these test statistics indicates 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 $\mathcal{X}$ where N has no points.

5 Performance of direct tests.

The performance of the tests of Section 4 may be investigated under various alternatives to (1). One such alternative is that the interaction between t and $\mathbf{x}$ is additive rather than multiplicative, i.e.

$$\lambda(t, \mathbf{x}) = \lambda_1(t) + \lambda_2(\mathbf{x}), \tag{10}$$

where λ_1 is a predictable non-negative $\mathcal{F}$-adapted process and λ_2 is a fixed non-negative function.

To determine which test statistic seems most sensitive to this type of alternative, many realizations of point processes in $\mathbf{R}^2_+$ were simulated according to (10). For each of the simulations, kernel estimates $\hat{\lambda}$ and $\tilde{\lambda}$ were generated, and the bootstrap distribution of the statistics listed above were obtained. A typical example is shown in Figs. 1-2. In Fig. 1a, a realization of a Poisson process on the unit square with non-separable, additive intensity (10) and

$$\lambda_1(t) = \exp(a_1 + b_1 t); \quad \lambda_2(x) = \exp(a_2 + b_2 x), \tag{11}$$

with $(a_1, a_2, b_1, b_2) = (3, 3, 3, 1)$, is shown. The model is clearly non-separable, since near time $t = 0$, it is much more likely that a point has a large value of x, but near time $t = 1$, the distribution of x becomes nearly uniform. Figs. 1b and 1c shows kernel intensity estimates $\hat{\lambda}$ and $\tilde{\lambda}$ for the simulation in Fig. 1. One sees in Fig. 1c the symmetry of the estimate $\tilde{\lambda}$; it is precisely this symmetry that is mandated by the assumption of separability.

Fig. 2 shows the the 6 test statistics, S_1 through S_6, applied to the simulated points in Fig. 1a, along with the bootstrap distributions of each of the statistics. For instance, in Fig. 2a, the dark vertical line indicates that for the realization in Fig. 1a, the statistic S_1 takes a value of approximately 16.2, and the histogram shows the distribution of the statistic S_1 applied to 10,000 point processes each simulated independently with conditional intensity $\tilde{\lambda}$ shown in Fig. 1c. Thus the histogram shows the distribution of S_1 under the separability assumption in (1), and the approximate one-sided p-value indicates the proportion of these simulated processes for which the statistic S_1 is greater than 16.2. Figs. 2b-f show the statistics S_2 through S_6 applied to the simulated points in Fig. 1, along with their corresponding bootstrap distributions and p-values. One sees that in this case the statistic S_3 is substantially more sensitive than the others to the departure from separability in the additive model (10).

We analyzed many different non-separable models and in every case the results were similar to that above: each time, the Cramer-von-Mises type statistics S_3 was most powerful at detecting non-separability (examples may be seen at www.stat.ucla.edu/~frederic/sep). Even for a very highly non-separable model such as

$$\lambda(t, x) = 200 \left(\mathbf{1}_{\{t \in [4,6]\}} + \mathbf{1}_{\{x \leq .3\}} \right),$$

where the change in the distribution of the x-values over time is readily apparent, the test statistics S_1 and S_2 did not signal any problems, with bootstrap p-values of 0.6210 and 0.3967, respectively. Similarly, statistics S_4, S_5 and S_6 were not as sensitive to the departure from separability as the Cramer-von-Mises statistic S_3, for which the bootstrap p-value was 0.0003. The results for other non-separable Poisson processes were similar, suggesting that the statistic S_3, which integrates the squared difference between the separable and non-separable intensity estimates over all times and locations, is a more powerful test statistic than the others under Poisson alternatives.

Although the test statistic S_3 is quite powerful at detecting broad, gradual variations in the conditional mean of x, it is not exceedingly powerful at detecting non-separability in the form of clustering or inhibition. For instance, an alternative of interest is that the point process N is a clustered process, e.g. a Hawkes process (Hawkes, 1971). Such a process in $\mathbf{R}^2_+$ has conditional intensity of the form

$$\lambda(t, \mathbf{x}) = f(t, \mathbf{x}) + \int_{\mathcal{X}} g\left(t - u, \mathbf{d}(\mathbf{x}, \mathbf{y})\right) dN(u, \mathbf{y}), \tag{12}$$

where $\mathbf{d}$ is a spatial distance function, and f and g are deterministic functions from $\mathcal{X}$ and $\mathbf{R}^2_+$, respectively, to $\mathbf{R}_+$. The function g is typically required to be zero for $t < u$, so that points can influence the rate of accumulation of points at future times but not in the past. In addition, the requirement that f and g be non-negative is usually made to ensure that λ is non-negative everywhere, and the integral in the right hand side of (12) is often assumed to be less than unity to ensure the stability of the process (Bremaud et al., 2002). It is important to note that not all clustering is in violation of the separability condition (1). For instance, many commonly used Hawkes models, such as the ETAS model of Ogata (1988), exhibit temporal clustering, but the model is still separable provided that despite variations in the rate of points per unit time, the distribution of the *marks* remains unchanged. When temporal and *spatial* clustering interact, however, then the existence of a point at (t, x) may change the mark distribution so that in the near future the likelihood of a point having a mark near x increases, and this violates condition (1).

We applied the statistics listed above to many different Hawkes processes; some results are shown at www.stat.ucla.edu/∼frederic/sep . In every case S_3 was most powerful among the six statistics listed above. For a Hawkes process on the unit square with conditional intensity

$$\lambda(t, x) = a + b \sum_{t' < t} \exp\{c(t' - t) - d|x - x'|\}, \tag{13}$$

and with the parameters $(a, b, c, d) = (50, 1000, 50, 50)$ selected so that the clustering is quite extreme, the distribution of the x-values changes dramatically over time, in violation of (1). Fig. 3a shows a simulation of such a process. In this case, for an appropriate test statistic it should be extremely unlikely under the null hypothesis to obtain a value larger than that observed. However, the statistic S_3 was larger in 28 of the $10,000$ simulated separable processes with conditional intensity $\tilde{\lambda}$ than for the data in Fig. 3a.

S_3 is similarly weak in detecting non-separable inhibition. For example, an inhibitory analog of (13) is a process with conditional intensity

$$\lambda(t, x) = \left[a - b \sum_{t' < t} \exp\{c(t' - t) - d|x - x'|\} \right]^+ . \tag{14}$$

Here a point at (t, x) decreases the conditional intensity nearby, and the positivity restriction is imposed merely to ensure that the conditional intensity is non-negative. A realization of the model (14) on the unit square with parameters $(a, b, c, d) = (400, 4000, 50, 50)$ is given in Fig. 3b; again, the parameters were selected so that the inhibitory nature of the process could be readily observed. Note that again the process is highly non-separable since the distribution of x-values changes over time, with x-values near previous x-values that have occurred recently becoming very unlikely. However, in this case the non-separability is extremely localized, and statistics such as S_3 are ineffective at detecting this type of non-separability: in fact 6,300 of the 10,000 simulations of separable processes with conditional intensity $\tilde{\lambda}$ had S_3 statistics larger than that applied to the realization in Fig. 3b, and the other statistics listed above fared no better. An alternative test that is more sensitive to non-separable inhibition and clustering is described in Section 6 below.

6 Residual Test of Separability.

An alternative way to test a point process for separability is to inspect the residuals of the process after rescaling the process to obtain a process which, 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 multivariate point process with simple ground process, if one transforms each point $(t, \mathbf{x})$ by moving it to $(A(t), \mathbf{x})$, what results is a sequence of independent Poisson processes of unit rate. Meyer's theorem has been generalized by Merzbach and Nualart (1986), Nair (1990), and Schoenberg (1999) to the case of vertically rescaled spatial point processes in $\mathbf{R}_+^n$. In this vertical rescaling, one focuses on one non-temporal coordinate x_{n-1}, and each point $(t, x_1, x_2, ..., x_{n-2}, x_{n-1})$ is shifted to $\left(t, x_1, x_2, ..., x_{n-2}, \int_0^{x_{n-1}} \lambda(t, x_1, x_2, ..., x_{n-2}, y)dy\right)$. Assuming the original process has simple ground process, the resulting process is Poisson with unit rate (Schoenberg, 1999).

This method of rescaling may be used in conjunction with non-parametric (e.g. kernel) intensity estimation to construct a non-parametric test for separability, as follows. Given a non-parametric separable intensity estimate such as $\tilde{\lambda}$ described in Section 3, one may vertically rescale the process, moving each point $(t, x_1, x_2, ..., x_{n-2}, x_{n-1})$ to

$$\left(t, x_1, x_2, ..., x_{n-2}, \int_0^{x_{n-1}} \tilde{\lambda}(t, x_1, x_2, ..., x_{n-2}, y)dy\right).$$

If the process is indeed separable, then $\tilde{\lambda}$ should closely approximate λ and thus the rescaled 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, if one is interested in detecting clustering, the estimated L-function may be useful (Ripley, 1977; Ripley, 1979; Diggle 1983; Baddeley et al. 2000). The estimated 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. 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. For instance, in $\mathbf{R}^2$, using Euclidean distance, $\hat{L}(u)$ is defined as

$$\hat{L}(u) = \sqrt{\hat{K}(u)/\pi} - u. \tag{15}$$

Fig. 3c shows the residuals when this method of vertical rescaling is applied to the simulated Hawkes process of Fig. 3a. The obvious departures from uniformity in Fig. 3c indicate departure from separability of the original point process. This is further clarified in

Fig. 3e which shows the estimated L-function for the residual points, after further rescaling the axes in Fig. 3c so that the scales of t and x are commensurate. The plot of $\hat{L}$ in Fig. 3e highlights the intense clustering in the rescaled points. The dashed lines in Fig. 3e show 95%-confidence bounds from 10,000 independent simulations of homogeneous processes, each with the same number of points as the original process of Fig. 3a; for each simulation the points are distributed uniformly within the irregular boundary of Fig. 3c and the resulting estimated L-function is obtained. At its peak the estimated L-function shown in Fig. 3e has an estimated p-value of 0, i.e. not one of the 10,000 simulated homogeneous processes had a value of $\hat{L}(u)$ exceeding that of the residual process in Fig. 3c, which suggests that the $\hat{L}$-test applied to the vertically-rescaled residuals represents a more powerful test for non-separable clustering compared with S_3 or the other tests described in Section 4.

The $\hat{L}$-function applied to vertically-rescaled residuals is also very powerful at detecting inhibitory behavior in violation of the separability hypothesis (1). Fig. 3d displays the residuals resulting from vertically rescaling the simulated inhibitory process (14) of Fig. 3b. From Fig. 3d one may discern that the residual points are less clustered than one would expect of a homogeneous Poisson process, and this is confirmed in Fig. 3f which shows $\hat{L}$ applied to these residual points, along with 95% confidence bounds obtained from 10,000 simulated homogeneous processes within the boundary of Fig. 3d. In addition to detecting clustering, the estimated L-function applied to the residuals obtained using the non-parametric separable intensity estimate appears to be very sensitive to inhibitory behavior in violation of assumption (1).

7 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 et al., 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 (Keeley and Keeley, 1988; Yool et al., 1985; Naveh, 1994). Data on wildfires have been systematically recorded in Los Angeles County since the late 19th century; the records on fires burning greater than 10 ha are believed to be nearly complete dating back to about 1950. Fig. 4a shows data on 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). Information on numerous covariates for this data has been recorded and analyzed (see for example Schoenberg et al., 2002b), but for the purpose of this example we focus exclusively on the times and sizes of the wildfires shown in Fig. 4a. Point process models used to describe such datasets, as well as other forestry data, are typically separable (Stoyan and Pettinen, 2000), though this assumption is generally not checked.

Nonseparable and separable kernel intensity estimates of the LACFD data are shown in Figs. 4b and 4c. Though some clustering appears to be present, no other obvious departures are immediately observable at a glance from Figs. 4a and 4b: the distribution of wildfire sizes does not appear to change drastically over time. This is confirmed by Fig. 4d which shows the statistic S_3 applied to the wildfire data, along with the bootstrap distribution of this statistic based on resampling 1,000 separable point processes with intensities as in Fig. 4c. Since S_3 is quite powerful at detecting non-separability 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 dataset are not excessive.

On the other hand, there does seem to be significant clustering, in violation of the separability hypothesis, as demonstrated in Fig. 5. Fig. 5a shows the vertically rescaled residuals, using the separable conditional intensity estimate $\tilde{\lambda}$ of Fig. 4c. The estimated L-function of the rescaled points is shown in Fig. 5b, along with 95%-confidence bounds based on resampling points uniformly distributed within the vertically rescaled boundary. From Fig. 5b one sees that the clustering in the residuals is highly significant, especially for rescaled distances of 0.3 to 0.5; these correspond to differences between points in the original dataset in the range of 1.215 to 2.025 years.

Further inspection of Figs. 4a and 5a can illuminate the main sources of clustering in the residuals. In vertical rescaling, the y-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 and those occurring in years where 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 Fig. 4a occurring in the year 1970 and with areas of 4.3 and 4.6 log km^2. In Fig. 4a 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 placed very close together: their y-coordinates in Fig. 5a are both approximately 15.0. Similarly, the four points between 1970 and 1971 with sizes in the range 2.5 to 3.0 log km^2 in Fig. 4a are rescaled in Fig. 5a so that they are highly clustered, with rescaled y-coordinates between 12.9 and 13.2. Another example is the cluster of four fires occurring in 1997, of sizes 0.8 to 0.9 log km^2. Because 1997 was a year with relatively few fires, $\tilde{\lambda}$ is rather low in this year, so these fires are clustered together in the residual plot of Fig. 5a, with all four residual points having a y-coordinate of very nearly 5.0.

8 Discussion.

While the lack of a very significant gradual change in the wildfire area distribution over time is not surprising, the significant, non-separable 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, since the resampled processes used for the confidence bounds in Fig. 5b each consist of points uniformly distributed over the identical boundary as that of the residual points. Nor can the apparent non-separability reasonably be attributed to errors or rounding in the dataset: 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 dataset also cannot be due to the insuitability of a model, since the methods used in its detection were non-parametric.

One possible explanation is that the clustering in the wildfire dataset may be partly due to climatic and temporal variations. For instance, 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); fires of certain sizes may be especially likely under particular weather conditions. Clustering may similarly result from variations in human response or ignition patterns, which may cause more fires of a given size to occur in some years 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; Guo and Rundel, 1997), clusters of fires of similar size that happen to occur at one time may tend to repeat, resulting in multiple clusters throughout the dataset. 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, and hence it should not be surprising to observe some significant non-separability in a dataset; rather it would be quite surprising if the separability condition were met. This last conclusion has serious implications for multi-dimensional 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 estimated L-function applied to vertically-rescaled residuals appear to be quite powerful tests for separability in multi-dimensional 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 non-separability in the form of clustering and inhibition quite well. The application of separability tests to other wildfire datasets, in other areas, and to other point process datasets in general, should be performed in the future.

References.

Andersen, P.K., Borgan, O., Gill, R.D., & Keiding, N. (1993). *Statistical Models Based on Counting Processes.* Springer, NY.

Baddeley, A.J., Moller, J., & Waagepetersen, R. (2000). Non- and semi-parametric estimation of interaction in inhomogeneous point patterns. *Statistica Neerlandica* **54**(3), 329–350.

Bremaud, P., Nappo, G., & Torrisi, G.L. (2002). Rate of convergence to equilibrium of marked Hawkes processes. *J. Appl. Probab.* **39**, 123-136.

Brillinger, D.R. (1998). Some wavelet analyses of point process data. in *31st Asilomar Conference on Signals, Systems and Computers*, pp. 1087–1091. IEEE Computer Society, Los Alamitos, California.

Cliff, A. & Ord, J. (1981). *Spatial Processes: Models and Applications.* Pion: London.

Cressie, N.A. (1993). *Statistics for spatial data, revised edition.* Wiley, New York.

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

Diggle, P. (1983). *Statistical Analysis of Spatial Point Patterns.* Academic Press: London.

Fishman, P. M. & Snyder, D. L. (1976). The Statistical Analysis of Space-time Point Processes. *IEEE Transactions on Information Theory*, **IT-22**(3), 257-274.

Guo, Q. & Rundel, P. (1997). Self-thinning during early post-fire chaparral succession: mechanisms and implications. *Ecology* **79**, 579–586.

Hanes, T. (1971). Succession after fire in the chaparral of Southern California. *Ecological Monographs*, **41**,27-52.

Hawkes, A.G. (1971). Point Spectra of some Mutually Exciting Point Processes. *Journal of the Royal Statistical Society, Series B* **33**, 438–443.

Jacod, J. (1975). Multivariate point processes: predictable projections, Radon-Nikodym derivatives, representation of martingales. *Z. Wahrsch. Verw. Geb.* **31**, 235-253.

Johnson, E. & Gutsell, S. (1994). Fire frequency models, methods and interpretations. *Advances in Ecological Research* **25**, 239-287.

Karr, A. (1991). *Point Processes and Their Statistical Inference, second ed.* New York: Dekker.

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.* Wiley: Chicester, England.

Keeley, J.E. & Fotheringham, C.J. (2001). History and management of crown-fire ecosystems: a summary and response. *Conservation Biology* **15**, 1561–1567.

Keeley, J.E. & Keeley, S. (1988). Chaparral. In: Barbour, M.G., and Billings, D.W. (eds.). *North American Terrestrial Vegetation.* Cambridge Univ. Press, Cambridge, 165-207.

Lewis, P. & Shedler, G. (1979). Simulation of nonhomogeneous Poisson processes by thinning. *Naval Res. Log. Quart.* **26**, 403-413.

Merzbach, E. & Nualart, D. (1986). A Characterization of the Spatial Poisson Process and Changing Time. *Annals of Probability* **14**, 1380–1390.

Meyer, P. (1971). Démonstration simplifée 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**(4), 1252–1262.

Nair, M. (1990). Random Space Change for Multiparameter Point Processes. *Ann. Probab.*, **18**, 1222–1231.

Naveh, Z. (1994). The role of fire and its management in the conservation of Mediterranean ecosystems and landscapes. In: *The Role of Fire in Mediterranean-Type Ecosystems.* ed. Moreno, J. and Oechel, W. Springer-Verlag: New York.

Ogata, Y. (1981). On Lewis' simulation method for point processes. *IEEE Trans. Inf. Theory* **27**, 23-31.

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

Pyne, S., Andrews, P., & Laren, R. (1996). *Introduction to wildland fire.* Wiley: New York.

Rathbun, S.L. (1993). Modeling Marked Spatio-temporal Point Patterns. *Bull. Int. Statist. Inst.*, **55**(2), 379–396.

Ripley, B. (1977). Modelling spatial patterns. *JRSS-B,* **39**, 172-192.

Ripley, B. (1979). Tests of 'randomness' for spatial point patterns. *JRSS-B*, **41**, 368-374.

Schoenberg, F.P. (1999). Transforming Spatial Point Processes into Poisson Processes. *Stoch. Proc. Appl.*, **81**, 155–164.

Schoenberg, F.P. (2003). Multi-dimensional residual analysis of point process models for earthquake occurrences. *J. Amer. Statist. Assoc.*, in review.

Schoenberg, F.P., Brillinger, D.R., & Guttorp, P.M. (2002a). Point processes, spatial-temporal. In *Encyclopedia of Environmetrics*, eds. A. El-Shaarawi and W. Piegorsch, New York: Wiley, vol. 3, pp. 1573–1577.

Schoenberg, F.P., Peng, R., Huang, Z., & Rundel, P. (2002b). Detection of nonlinearities in the dependence of burn area on fuel age and climatic variables. *Int. J. Wildland Fire* **12**(1), 1–10.

Shitan, M. & Brockwell, P.J. (1995). An asymptotic test for separability of a spatial autoregressive model. *Commun. Statist. Theory Meth.* **24**(8), 2027–2040.

Solow, A. (1992). Model-checking in non-stationary Poisson processes. *Appl. Stoch. Models and Data Anal.* **8**, 129-132.

Stoyan, D. & Penttinen, A. (2000). Recent applications of point process methods in forestry statistics. *Stat. Sci.* **15**(1), 61–78.

Vere-Jones, D. (1992). Statistical Methods for the Description and Display of Earthquake Catalogs. in *Statistics in Environmental and Earth Sciences*, eds. A. Walden and P.

Guttorp, London: Edward Arnold, pp. 220 - 236.

Whelan, R. (1995). *The Ecology of Fire.* Cambridge University Press: Cambridge.

Yool, S., Eckhardt, D., & Cosentino, M. (1985). Describing the brushfire hazard in Southern California. *Anal. Assoc. Am. Geograph.* **75**, 431-442.

Figure Captions

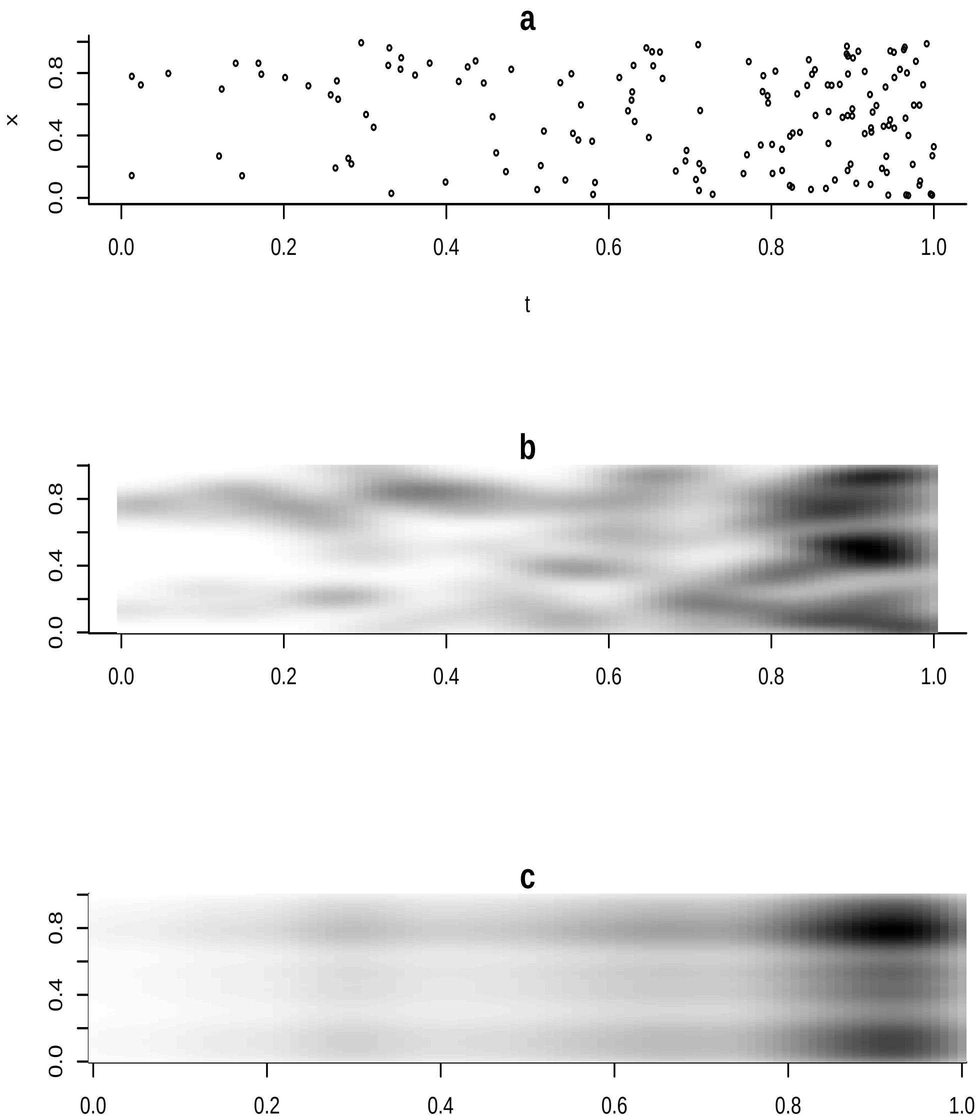
Figure 1: a) Simulated Poisson process with additive, exponential intensity (10-11); b) Non-separable kernel intensity estimate $\hat{\lambda}$; c) Separable intensity estimate $\tilde{\lambda}$.

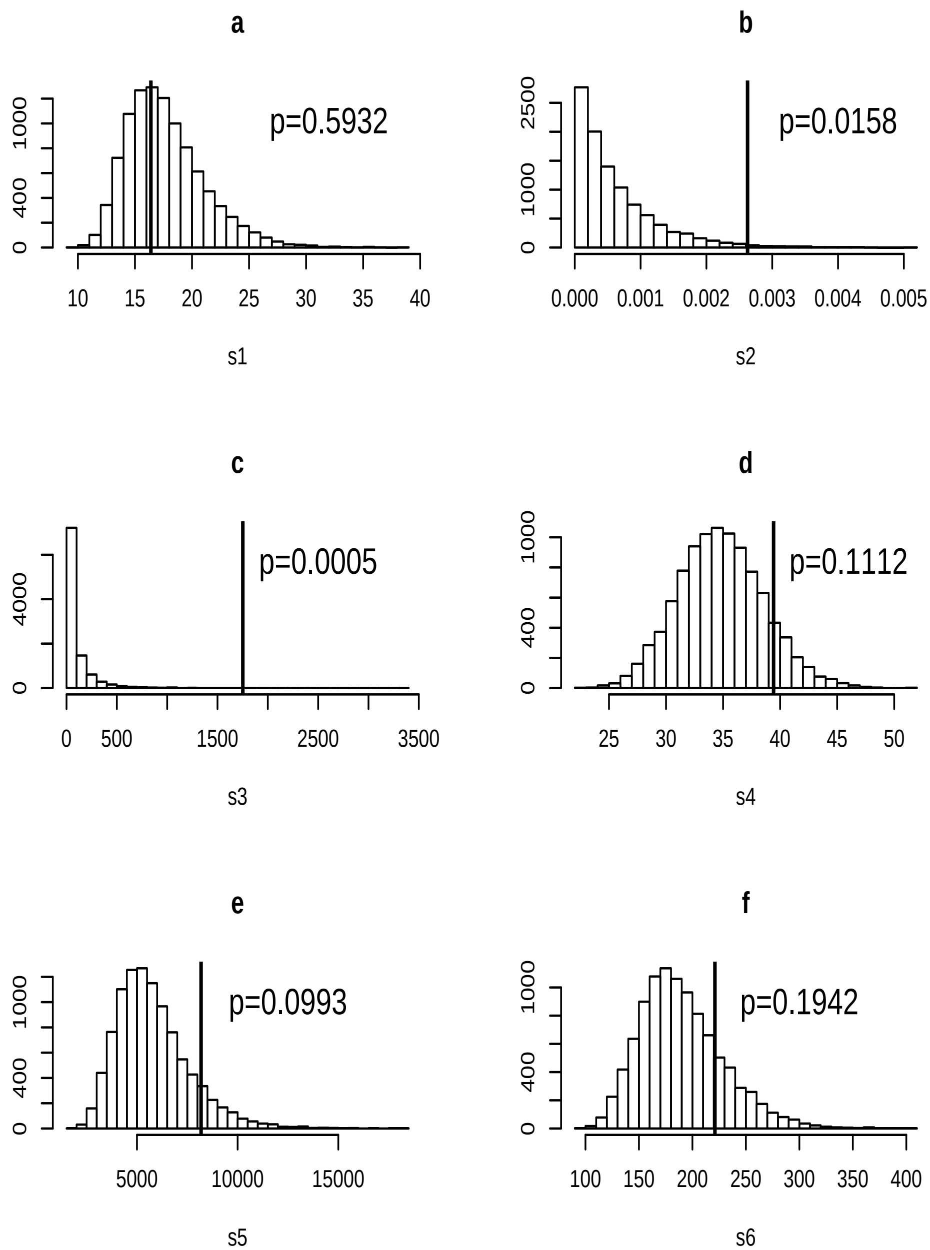
Figure 2: Test statistics applied to additive exponential model, along with bootstrap distributions and one-sided p-values.

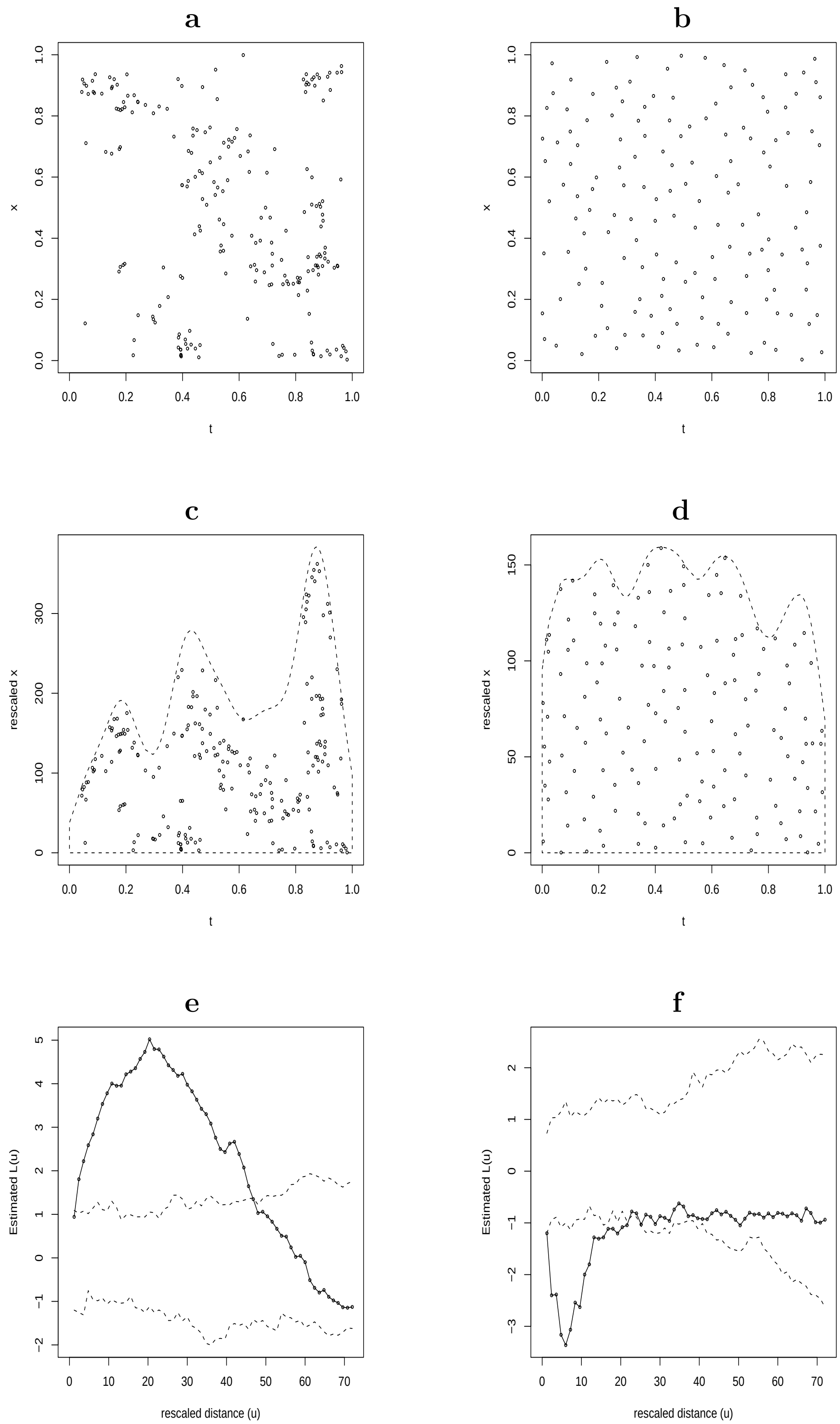
Figure 3: a) Simulated Hawkes process with conditional intensity (13); b) Simulated inhibitory process with conditional intensity (14); c) Vertical rescaling of Hawkes process using the separable intensity estimate $\tilde{\lambda}$; d) Vertical rescaling of inhibitory process using the separable intensity estimate $\tilde{\lambda}$; e) L-function applied to vertically rescaled residuals of the Hawkes process; f) L-function applied to vertically rescaled residuals of the inhibitory process.

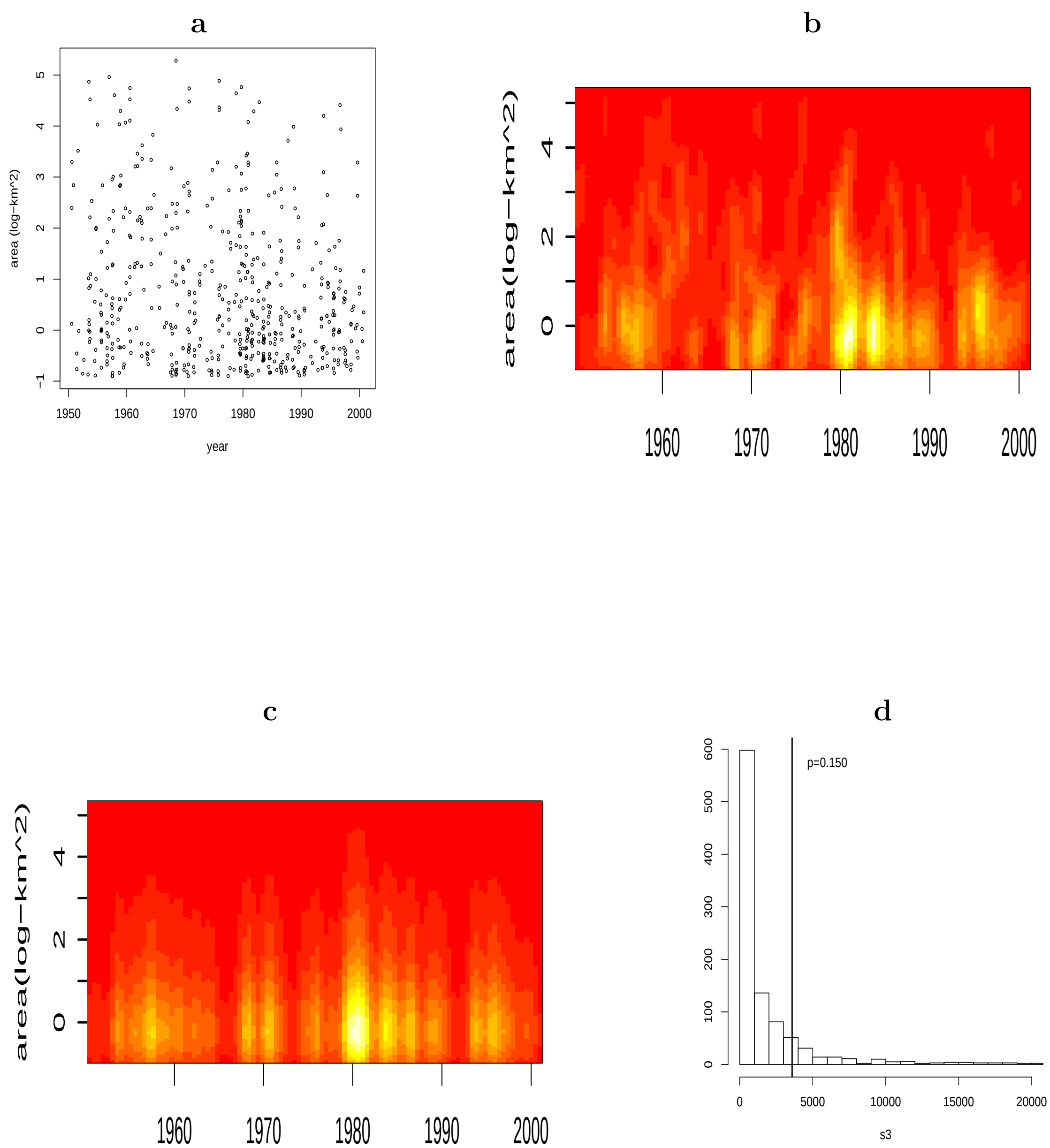
Figure 4: a) Times and burn areas greater than 10 ha from January 1950 to January 2001 in Los Angeles County, from LACFD wildfire data; b) Non-separable kernel intensity estimate $\hat{\lambda}$; c) Separable kernel intensity estimate $\tilde{\lambda}$; d) S_3 applied to LACFD wildfire data, along with bootstrap distribution from simulations of separable intensity estimate $\tilde{\lambda}$.

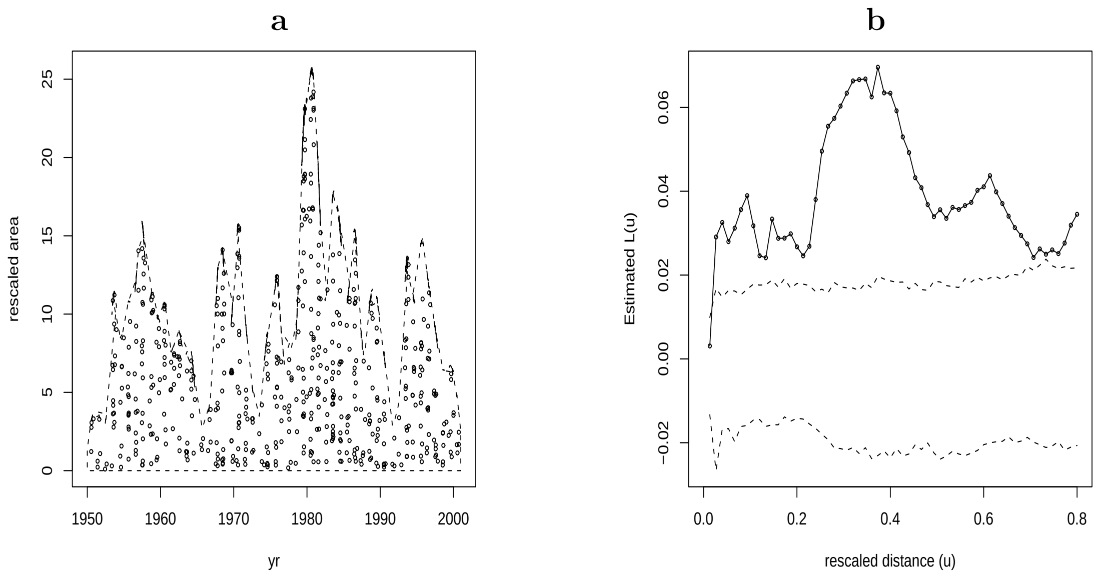
Figure 5: a) Vertical rescaling of LACFD wildfire data using the separable intensity estimate $\tilde{\lambda}$; b) L-function applied to vertically rescaled residuals.


a
x
0.0
0.4
0.8
0.0
0.2
0.4
0.6
0.8
1.0
t
b
0.0
0.4
0.8
0.0
0.2
0.4
0.6
0.8
1.0
c
0.0
0.4
0.8
0.0
0.2
0.4
0.6
0.8
1.0


a
p=0.5932
0
400
1000
10
15
20
25
30
35
40
s1
b
p=0.0158
0
1000
2500
0.000
0.001
0.002
0.003
0.004
0.005
s2
c
p=0.0005
0
4000
0
500
1500
2500
3500
s3
d
p=0.1112
0
400
1000
25
30
35
40
45
50
s4
e
p=0.0993
0
400
1000
5000
10000
15000
s5
f
p=0.1942
0
400
1000
100
150
200
250
300
350
400
s6


a
1.0
0.8
0.6
0.4
0.2
0.0
x
0.0 0.2 0.4 0.6 0.8 1.0
t
b
1.0
0.8
0.6
0.4
0.2
0.0
x
0.0 0.2 0.4 0.6 0.8 1.0
t
c
300
200
100
0
rescaled x
0.0 0.2 0.4 0.6 0.8 1.0
t
d
150
100
50
0
rescaled x
0.0 0.2 0.4 0.6 0.8 1.0
t
e
5
4
3
2
1
0
−1
−2
Estimated L(u)
0 10 20 30 40 50 60 70
rescaled distance (u)
f
2
1
0
−1
−2
−3
Estimated L(u)
0 10 20 30 40 50 60 70
rescaled distance (u)


a
b
c
d
area (log−km^2)
year
1950
1960
1970
1980
1990
2000
area(log−km^2)
p=0.150
s3
0
5000
10000
15000
20000
100
200
300
400
500
600


a
b
rescaled area
yr
Estimated L(u)
rescaled distance (u)