Title
Assessing Spatial Point Process Models Using Weighted K-functions: Analysis of California Earthquakes

Authors
Alejandro Veen
Frederic Paik Schoenberg

Publication Date
2011-10-25

Assessing Spatial Point Process Models Using Weighted K-functions: Analysis of California Earthquakes

Alejandro Veen[1] and Frederic Paik Schoenberg[2]

[1] UCLA Department of Statistics
8125 Math Sciences Building
Box 951554
Los Angeles, CA 90095-1554
U.S.A.
veen@stat.ucla.edu

[2] UCLA Department of Statistics
8125 Math Sciences Building
Box 951554
Los Angeles, CA 90095-1554
U.S.A.
frederic@stat.ucla.edu

We investigate the properties of a weighted analogue of Ripley's K-function which was first introduced by Baddeley, Møller, and Waagepetersen. This statistic, called the *weighted* or *inhomogeneous* K-function, is useful for assessing the fit of point process models. The advantage of this measure of goodness-of-fit is that it can be used in situations where the null hypothesis is not a stationary Poisson model. We note a correspondence between the weighted K-function and thinned residuals, and derive the asymptotic distribution of the weighted K-function for a spatial inhomogeneous Poisson process. We then present an application of the use of the weighted K-function to assess the goodness-of-fit of a class of point process models for the spatial distribution of earthquakes in Southern California.

1 Introduction

Ripley's K-function [Rip76], $K(h)$, is a widely used statistic to detect clustering or inhibition in point process data. It is commonly used as a test, where the null hypothesis is that the point process under consideration is a homogeneous Poisson process and the alternative is that the point process exhibits clustering or inhibitory behavior. Previous authors have described the asymptotic distribution of the K-function for simple point process models including the homogeneous Poisson case (see [Hei88], [Rip88] pp. 28–48, [Sil78]).

The K-function has also been used in conjunction with point process residual analysis techniques in order to assess more general classes of point process models. For instance, a point process may be rescaled (see [MD86], [Oga88], [Sch99]) or thinned [Sch03] to generate residuals which are approximately homogeneous Poisson, provided the model used to generate the residuals is correct. The K-function can then be applied to the residual process in order to investigate the homogeneity of the residuals, and the result can be interpreted as a test of the goodness-of-fit of the point process model in question. Hence, residual analysis of a point process model involves two steps, the transformation of the data into residuals and a subsequent test for whether the residuals appear to be well approximated by a homogeneous Poisson process.

Of course, other methods for assessing the homogeneity of a point process exist, including tests for monotonicity [Saw75], uniformity (see [DRS84], [Law88], [LL85],), and tests on the second and higher-order properties of the process (see [Bar64], [Dav77], [Hei91]). Likelihood statistics, such as Akaike's Information Criterion (AIC, [Aka74]) and the Bayesian Information Criterion (BIC, [Sch79]) are often used to assess more general classes of models; see e.g. [Oga98] for an application to earthquake occurrence models.

We focus here on Ripley's K-function, in particular on a modified version of the statistic which we call the *weighted K-function*, K_W, and which was first introduced as the *inhomogeneous K-function* in [BMW00]. It may be used to test a quite general class of null hypothesis models for the point process under consideration and it provides a direct test for goodness-of-fit, without having to assume homogeneity or to transform the points using residual analysis, the latter of which often introduces problems of highly irregular boundaries and large sampling variability when the conditional intensity in question is highly variable (see [Sch03]).

This paper is outlined as follows. In Section 2, the definitions of the ordinary and weighted K-functions are reviewed, a connection between K_W and thinned residuals is noted, and the asymptotic distribution of K_W is derived under certain conditions. The weighted K-function is then used in Section 3 to assess the goodness-of-fit for competing models for the spatial background rate of California earthquakes. Some concluding remarks are given in Section 4.

2 The Weighted K-function

In this Section, we derive its distributional properties of the weighted K-function, $K_W(h)$, under certain conditions. $K_W(h)$ is a weighted analogue of Ripley's K-function and it is similar to the mean of K-functions applied to a repeatedly thinned point pattern, denoted here as $K_M(h)$, an application of which can be found in [Sch03]. We begin with a review of Ripley's K-function.

2.1 Ripley's K-function and Variants

Consider a Poisson process of intensity λ on a connected subset $\mathcal{A}$ of the plane $\mathbf{R}^2$ with finite area A, and let the N points of the process be labelled $\{p_1, p_2, \ldots, p_N\}$. Ripley's K-function $K(h)$ is typically defined as the average number of further points within h of any given point divided by the overall rate λ, and is most simply estimated via

$$\hat{K}(h) = \frac{1}{\hat{\lambda} N} \sum_r \sum_{s \neq s} \mathbf{1}(|p_r - p_s| \leq h), \tag{1}$$

where $\hat{\lambda} = N/A$ is an estimate of the overall intensity, $\mathbf{1}(\cdot)$ is the indicator function and h is some inter-point distance of interest. As pointed out in [SS00], one can also estimate $K(h)$ using an estimator for the squared intensity $\tilde{\lambda^2} = N(N-1)/A^2$:

$$\tilde{K}(h) = \frac{1}{\tilde{\lambda^2} A} \sum_r \sum_{s \neq s} \mathbf{1}(|p_r - p_s| \leq h). \tag{2}$$

In applications, estimates of K are typically calculated for several different choices of h. For a homogeneous Poisson process, the expectation of $\hat{K}(h)$ is πh^2 (similarly for $\tilde{K}(h)$). Values which are higher than this expectation indicate clustering, while lower values indicate inhibition. However, it should be noted that a point pattern can be clustered at certain scales and inhibitory at others. Note also that two very different point processes may have identical K-functions, as $K(h)$ only takes the first two moments into account. An example of such a situation can be found in [BS84].

Under the null hypothesis that the point process is homogeneous Poisson with rate λ, $\hat{K}(h)$ is asymptotically normal:

$$\hat{K}(h) \approx N\left(\pi h^2, \frac{2\pi h^2}{\lambda^2 A}\right), \tag{3}$$

as the area of observation A tends to infinity (see p. 642 of [Cre93] or pp. 28–48 of [Rip88]). As is pointed out in [SS00], it is crucial to use an estimate of λ or λ^2 rather than their true values, even if they are known. Situations where the true intensity is known can arise in simulation studies, where one may feel tempted to plug in the true value for the intensity in (1) or (2). Somewhat surprisingly, however, using the true value for λ or λ^2 will actually inflate the variance of $\hat{K}(h)$ by a factor of $1 + 2\pi h^2 \lambda$ (see [Hei88]).

Several variations on $\hat{K}(h)$ have been proposed. Many deal with corrections for boundary effects, as found in [Rip76], [JD81], and [Ohs83]. Variance-stabilizing transformations of estimated K-functions which are more easily interpretable have been proposed (see [Bes77]), such as $\hat{L}(h)$ and $\hat{L}(h) - h$ where

$$\hat{L}(h) = \sqrt{\frac{\hat{K}(h)}{\pi}}\ . \tag{4}$$

2.2 Definition and Distribution of the Weighted K-function

Suppose that a given planar point process in a connected subset $\mathcal{A}$ of $\mathbf{R}^2$ with finite area A may be specified by its conditional intensity with respect to some filtration on $\mathcal{A}$, for $(x, y) \in \mathcal{A}$ (see [DVJ03]). The point process need not be Poisson; in the simple case where the point process is Poisson, however, the conditional intensity and ordinary intensity coincide. Suppose that the conditional intensity of the point process is given by $\lambda(x, y)$.

The *weighted K-function*, used to assess the model $\lambda_0(x, y)$, may be defined as

$$K_W(h) = \frac{1}{\lambda_*^2 A} \sum_r w_r \sum_{s \neq r} w_s \mathbf{1}(|p_r - p_s| \leq h) \tag{5}$$

where $\lambda_* := \inf\{\lambda_0(x, y); (x, y) \in \mathcal{A}\}$ is the infimum of the conditional intensity over the observed region for the model to be assessed and $w_r = \lambda_*/\lambda_0(p_r)$, where $\lambda_0(p_r)$ is the modelled conditional intensity at point p_r.

One can think of the weighted K-function as a combination of Ripley's K-function and the thinning method used for residual analysis in [Sch03]. In [Sch03], $K(h)$ is repeatedly applied to thinned data where the probability of retaining a point is inversely proportional to the conditional intensity at that point. The computation of the weighted K-function $K_W(h)$ uses these same retaining probabilities as weights for the points in order to offset the inhomogeneity of the process. By incorporating all pairs of the observed points, rather than only the ones that happen to be retained after an iteration of random thinning, the statistic $K_W(h)$ eliminates the sampling variability in any finite collection of random thinnings. Indeed, simulations appear to indicate that $K_W(h)$ has approximately the same distribution as $K_M(h)$, the mean of K-functions on a repeatedly thinned point pattern, as the number of random thinnings approaches infinity.

We conjecture that, provided the conditional intensity λ is sufficiently smooth, $K_W(h)$ will be asymptotically normal as the area of observation A approaches infinity. Indeed, for the Poisson case where λ is locally constant on distinct subregions whose areas $A_i^{(n)}$ are large relative to the interpoint distance h_n, we have the following result.

Theorem 1. *Let $N^{(n)}$ be a sequence of inhomogeneous Poisson processes with intensities $\lambda^{(n)}$ and weighted K-functions $K_W^{(n)}$, defined on connected subsets $\mathcal{A}^{(n)} \subset \mathbf{R}^2$ of finite areas $A^{(n)}$. Suppose that for each n, the observed region $\mathcal{A}^{(n)}$ can be broken up into disjoint subregions $\mathcal{A}_1^{(n)}, \mathcal{A}_2^{(n)}, \ldots, \mathcal{A}_{I_n}^{(n)}$ each having area $A_i^{(n)} = A^{(n)}/I_n$, and that the intensity $\lambda_i^{(n)}$ is constant within $\mathcal{A}_i^{(n)}$. Suppose also that for some scalar $\lambda_{\min}$, $0 < \lambda_{\min} \leq \lambda_i^{(n)} < \infty$ for all i, n. In addition, suppose that, as $n \to \infty$, $I_n \to \infty$ and $h_n^2/A_i^{(n)} \to 0$. Further, assume that the boundaries of $\mathcal{A}_i^{(n)}$ are sufficiently regular that the number of pairs of points (p_r, p_s) with $|p_r - p_s| \leq h_n$ such that p_r and p_s are in distinct* *subregions is small, satisfying* $R^{(n)} := \frac{1}{A^{(n)}} \sum_{p_r, p_s} \frac{\mathbf{1}(|p_r - p_s| \leq h_n)\mathbf{1}(i \neq j)}{\lambda_i^{(n)} \lambda_j^{(n)}} \to 0$ *in probability as* $n \to \infty$*, where the sum is over all* $p_r \in A_i^{(n)}, p_s \in A_j^{(n)}$*. Then* $K_W^{(n)}(h_n)$ *is asymptotically normal as* $n \to \infty$*:*

$$\frac{K_W^{(n)}(h_n) - \pi h_n^2}{\sqrt{\frac{2\pi h_n^2}{A^{(n)} H\left((\lambda^{(n)})^2\right)}}} \approx N(0,1),$$

where $H\left((\lambda^{(n)})^2\right)$ *represents the harmonic mean of the squared intensity within the observed region* $\mathcal{A}^{(n)}$*.*

Proof. We first show that $K_W^{(n)}(h_n)$ can be represented as the arithmetic mean of K-functions computed individually on each of the squares $i = 1, 2, \ldots, I_n$, plus the remainder term $R^{(n)}$ defined above:

$$K_W^{(n)}(h_n) = \frac{1}{\lambda_*^2 A^{(n)}} \sum_r w_r \sum_{s \neq r} w_s \mathbf{1}(|p_r - p_s| \leq h_n) \tag{6}$$

$$= \frac{1}{\lambda_*^2 A^{(n)}} \sum_{i=1}^{I_n} \frac{\lambda_*^2}{(\hat{\lambda}_i^{(n)})^2} \sum_{r_i} \sum_{s_i \neq r_i} \mathbf{1}(|p_{r_i} - p_{s_i}| \leq h_n) + R^{(n)} \tag{7}$$

$$= \frac{1}{I_n} \sum_{i=1}^{I_n} \frac{1}{(\hat{\lambda}_i^{(n)})^2 A_i^{(n)}} \sum_{r_i} \sum_{s_i \neq r_i} \mathbf{1}(|p_{r_i} - p_{s_i}| \leq h_n) + R^{(n)}$$

$$= \frac{1}{I_n} \sum_{i=1}^{I_n} \hat{K}_i^{(n)}(h_n) + R^{(n)} \tag{8}$$

Since the intensity $\lambda_i^{(n)}$ is constant on each square $\mathcal{A}_i^{(n)}$, the weights w_r, w_s assigned to a pair of points in $\mathcal{A}_i^{(n)}$ within distance h_n are each $\lambda_*^2/(\hat{\lambda}_i^{(n)})^2$, which is used in going from (6) to (7). Thus, since $R^{(n)}$ converges to zero in probability by assumption, the distribution of the weighted K-function is equivalent to that of the mean of the I_n ordinary K-functions in (8).

Under the conditions of the theorem, $\hat{K}_i^{(n)}$ is asymptotically normal from [Rip88], and since the point process on $\mathcal{A}_i^{(n)}$ is homogeneous Poisson with rate $\lambda_i^{(n)} \geq \lambda_{\min} > 0$, the variance of $\hat{K}_i^{(n)}$ is bounded above by the variance of a homogeneous Poisson process on $\mathcal{A}_i^{(n)}$ with rate $\lambda_{\min}$. This implies that the collection of random variables $\left\{ \frac{\hat{K}_i^{(n)}(h_n) - \pi h_n^2}{I_n \sqrt{Var\left(\hat{K}_i^{(n)}(h_n)\right)}} \right\}$ satisfies the Lindeberg condition (see e.g. p. 98 of [Dur91]), and therefore the mean $\frac{1}{I_n} \sum_{i=1}^{I_n} \hat{K}_i^{(n)}(h_n)$ is asymptotically normal. The variance of $K_W^{(n)}(h) = Var\left(\frac{1}{I_n} \sum_{i=1}^{I_n} \hat{K}_i^{(n)}(h)\right) + o(n)$, which can be computed as

$$
\begin{aligned}
Var\left(\frac{1}{I_n}\sum_{i=1}^{I_n}\hat{K}_i^{(n)}(h)\right) &= \frac{1}{I_n^2}\sum_{i=1}^{I_n} Var\left(\hat{K}_i^{(n)}(h)\right) \\
&= \frac{1}{I_n^2}\sum_{i=1}^{I_n}\frac{2\pi h^2}{(\lambda_i^{(n)})^2 A_i^{(n)}} \\
&= \frac{2\pi h^2}{A^{(n)} H\left((\lambda^{(n)})^2\right)}, \qquad (9)
\end{aligned}
$$

where (9) follows from the fact that $A_i^{(n)} = A^{(n)}/I_n$.

□

Note that a variance-stabilized version of the weighted K-function can be defined in analogy with (4), namely:

$$L_W(h) = \sqrt{\frac{K_W(h)}{\pi}}. \qquad (10)$$

3 Application

The test statistic $K_W(h)$ in (5) is applicable to a very general class of planar point process models. We investigate their application to models for the spatial background rate for the occurrences of Southern California earthquakes.

3.1 Data Set

Data on Southern California earthquakes are compiled by the Southern California Earthquake Center (SCEC). The data include the occurrence times, magnitudes, locations, and often waveforms and moment tensor solutions, based on recordings at an array of hundreds of seismographic stations located throughout Southern California, including over 50 stations in Los Angeles County alone. The catalog is maintained by the Southern California Seismic Network (SCSN), a cooperative project of the California Institute of Technology and the United States Geological Survey. The data are available to the public; information is provided at http://www.data.scec.org.

We focus here on the spatial locations of a subset of the SCEC data occurring between 01/01/1984 and 06/17/2004 in a rectangular area around Los Angeles, California, between longitudes $-122°$ and $-114°$ and latitudes $32°$ and $37°$ (approximately $733\,km \times 556\,km$). The data set consists of earthquakes with magnitude not smaller than 3.0, of which 6,796 occurred within the given 21.5-year period. The epicentral locations of these earthquakes are shown in Fig. 1.

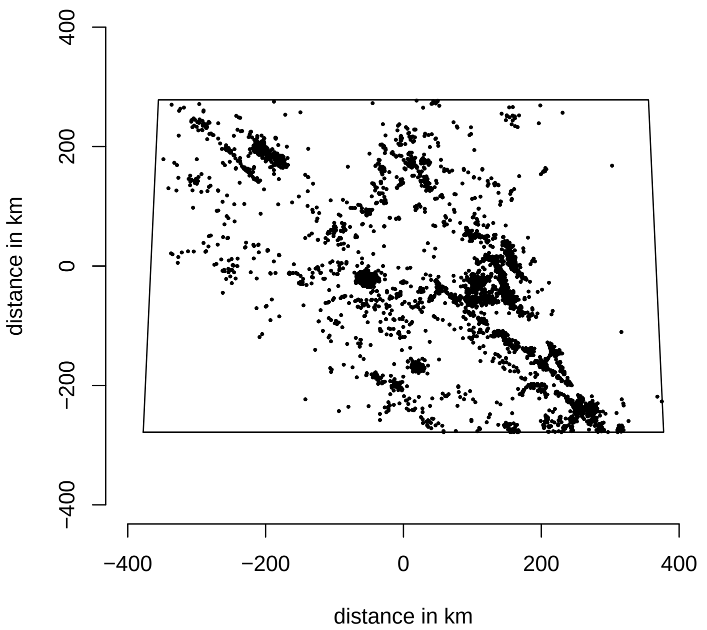

Fig. 1. Earthquakes in Southern California 1984-2004: The data set consists of 6796 earthquakes with magnitude 3.0 or larger

3.2 Analysis

Spatial background rates are commonly estimated by seismologists by smoothing the larger events only. For instance [Oga98] suggests anisotropic kernel smoothing of larger events in order to estimate the spatial background intensity for all earthquakes. In this application, we investigate various spatial background seismicity rate estimates involving kernel smoothings of only the 2030 earthquakes of magnitude 3.5 and higher, by using $K_W(h)$ to assess their fit to the earthquake data set. The local seismicity at location (x, y) may be estimated using a bivariate kernel smoothing $\mu(x, y)$ of the events of magnitude at least 3.5. Figure 2 shows such a kernel smoothing, using an anisotropic bivariate normal kernel with a bandwidth of 8 km and a correlation of -0.611. That is,

$$\mu(x, y) = \sum_{r=1}^{N} f(x - x_r, y - y_r), \tag{11}$$

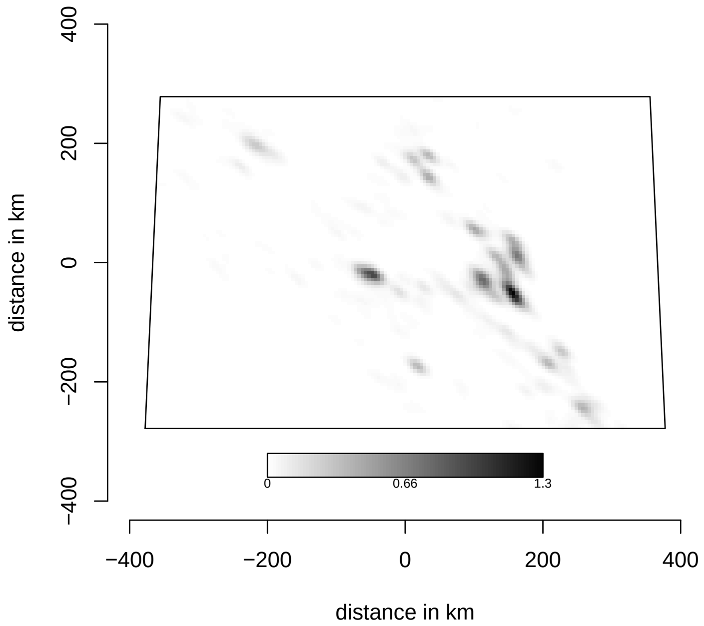

Fig. 2. Kernel smoothing of seismicity in Southern California 1984-2004: An anisotropic bivariate normal kernel with a bandwidth of 8 km ($\rho = -0.611$, $\sigma_x = \sigma_y = 8$ km) is applied to 2030 earthquakes with magnitude not smaller than 3.5

where the sum is over all points (x_r, y_r) with magnitude $m_r \geq 3.5$, and f is the bivariate normal density centered at the origin with standard deviation $\sigma_x = \sigma_y = 8$ km and correlation $\rho = -0.611$. This correlation is estimated using the empirical correlation of the values of x_r and y_r, and the bandwidth is selected by inspection. The agreement of Figs. 1 and 2 does not seem grossly unreasonable.

Since such a kernel smoothing uses only the observed seismicity over the last 20 years (a relatively small time period by geological standards), one may wish to allow for the possibility of seismicity in regions where no earthquakes of magnitude 3.5 or higher have recently been observed. One way to do this is by estimating the spatial background intensity via a weighted average of the kernel-smoothed seismicity of magnitude at least 3.5 and a positive constant representing an estimate of the spatial background intensity under the assumption that the process is homogeneous Poisson. Hence we consider the estimate of the form

$$\hat{\lambda}_a(x, y) = a\mu(x, y) + (1 - a)\nu, \tag{12}$$

where $\nu = N/A$ is the estimated conditional intensity for a homogeneous Poisson model and a is some constant with $0 \leq a \leq 1$.

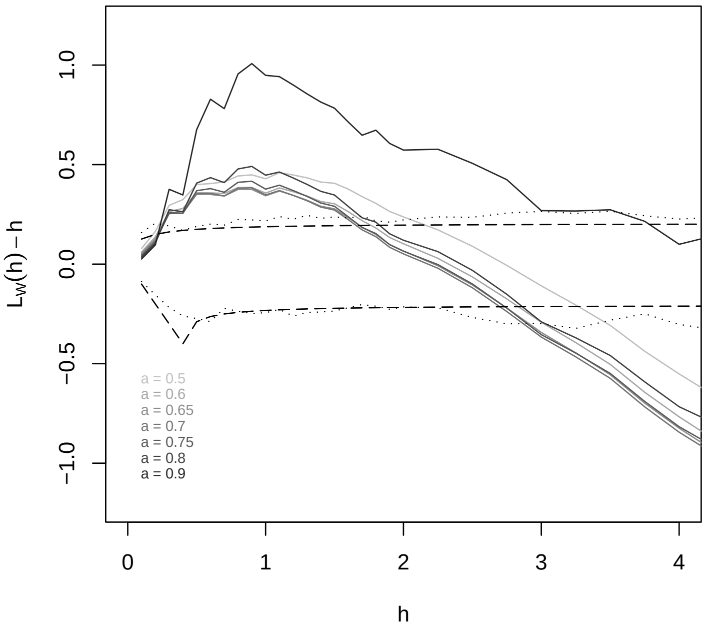

Fig. 3. Weighted L-function for competing models: The difference between the weighted L-function and its expectation h is shown for different values of a in the background intensity model $\hat{\lambda}_a$ as described in (12). The dashed and dotted lines are 95% bounds for $L_W(h) - h$ using model $\hat{\lambda}_{a=0.7}$ based on the theoretical result of Theorem 1 (dashed) and simulations (dotted).

Instead of plotting the weighted K-function for visual inspection, we will show the difference between $L_W(h)$ as given by (10) and its expectation h, because the latter highlights the departures of the estimate from its hypothetical expectation. Figure 3 shows $L_W(h) - h$ applied to several spatial intensity estimates, each of the form (12), using different values for the parameter a. For the competing estimates $\hat{\lambda}_a$, a takes on the values 0.5, 0.6, 0.65, 0.7, 0.75, 0.8, and 0.9, where a darker line color indicates a higher value of a. The lower values of a give more weight to the homogeneous background rate than higher values of a.

High values of a, such as $a = 0.9$ or greater, fit very poorly to the data, especially for small values of h, as shown in Figure 3. For such values of a, the intensity estimate gives most of the weight to the kernel smoothing, so that pairs of small earthquakes in areas where there were no earthquakes of magnitude greater than or equal to 3.5 have a very small probability and are hence given enormous weight in the computation of K_W. Similarly, for values of $a = 0.6$ or less, the intensity estimate gives too much weight to the homogeneous Poisson component and too little to the kernel smoothing of the large events, so that the resulting model underpredicts the intense clustering in the data occurring around the larger events.

For larger values of h, $L_W(h)$ tends to be smaller than its expectation. This is due to the fact that all the models $\hat{\lambda}_a$ inspected in this work include a background intensity component which is too high in those areas of Fig. 1 where no earthquakes occur. Under any of the proposed models, one would expect more earthquakes very far from the regions of high seismicity than actually occurred, and the absence of pairs of such earthquakes leads to values of $L_W(h)$ which are significantly smaller than expected.

In order to pick the best model $\hat{\lambda}_a$, attention should be focused on the smaller values of h, especially since the assumption in Theorem 1 that λ be locally constant is clearly invalidated if many pairs of points which are within distance h have very different estimated intensities. For small values of h, Theorem 1 may not be grossly inappropriate since the models for λ are continuous in this example. As shown in Fig. 3, $L_W(h) - h$ seems to decrease towards its expectation for most small values of h, indicating a rather satisfactory fit for values of a approaching $a = 0.7$ from either direction. This value of a appears to offer better fit than other values of a (and certainly is far better than the conventional $a = 1.0$). However, even for $a = 0.7$, for h in the range of 0.3km to 1.3km, the values of $L_W(h) - h$ exceed the 95% bounds for $L_W(h) - h$, which are shown as dashed and dotted lines in Fig. 3.

The dashed lines in Fig. 3 are derived using the result in Theorem 1 for model $\hat{\lambda}_{a=0.7}$. The dotted lines are based on empirical 95% bounds for $L_W(h) - h$ based on 150 simulations of model $\hat{\lambda}_{a=0.7}$. The simulated bounds line up quite well with the theoretical bounds, which indicates that the conditions of the theorem are sufficiently satisfied in our application. In particular, the observed area seems to be sufficiently large, the intensity sufficiently smooth (at least for the values of h used in this work), and boundary effects do not seem to affect the estimation of $K_W(h)$ in any substantial way.

In summary, the data set contains many more small earthquakes in areas far removed from any of the larger events than predicted by a kernel smoothing of the larger events only, and clearly contains much more clustering than would be predicted by a homogeneous Poisson model. However, there is significant short-range clustering of the smaller earthquakes that occur in these locations not covered by the larger events, which explains the positive departure of $L_W(h)$ for small ranges of h. At the same time, the total number of earthquakes occurring in these remote areas is small; that is, the preponderance of these smaller earthquakes are occurring much closer to the large events than one would expect from a homogeneous Poisson process, which explains why $L_W(h)$ is smaller than expected for larger values of h. Although a mixture of a kernel smoothing of the larger events and a homogeneous Poisson background appears to fit much better than either of these individually, no such mixture can thoroughly account for the observed patterns mentioned above.

4 Concluding Remarks

The application of the weighted K-function to spatial background rate estimates for Southern California seismicity shows the power of K_W in testing for goodness-of-fit. The weighted K-function is easily able to detect the major departures from the data for simple kernel or Poisson estimates of the spatial distribution of earthquakes. In addition, even for the optimally-chosen mixture model for the background events, the weighted K-function is able to detect deficiencies and to indicate potential areas for improvement.

K_W has some advantages to alternative goodness-of-fit procedures like thinning or re-scaling, especially in situations where the intensity on the observed region has high variability. For the mixture estimate with a=0.7, for instance, estimates of $\hat{\lambda}_a$ ranged from 0.0049978 to 0.96792. With intensity estimates varying over such a wide range, the application of thinning procedures can by quite problematic. Since the estimated lowest intensity is very low, only very few points will be kept after a random iteration of thinning, which introduces a high degree of sampling variability. Re-scaling procedures, on the other hand, would lead to highly irregular boundaries, which would make it rather difficult to compute any test statistics on the re-scaled process.

In contrast to standard kernel smoothing of the larger events in the catalog, the method of spatial background rate estimation which mixes the kernel estimate with a homogeneous constant rate appears to offer somewhat superior fit to the SCEC dataset. This suggests that spatial background rate estimates in commonly used models for seismic hazard, such as the epidemic-type aftershock sequence (ETAS) model of [Oga98], might possibly be improved in this way as well. Seismologically, the results are consistent with the notion that Southern California earthquakes, though certainly far more likely to occur on known faults, can potentially occur on unknown faults as well, and these faults may be quite uniformly dispersed. The results suggest that a spatial background rate estimate incorporating both of these possibilities could provide improved fit to existing models for seismic hazard. Such a modification may be especially relevant given the occurrences in California of blind (i.e. previously unknown) faults such as the one which ruptured during the Northridge earthquake in 1994, causing at least 33 deaths and 138 injuries as well as extensive public and private property damage [PAKB$^+$98].

Further study is needed in order to confirm the seismological results suggested herein, for several reasons. First, it remains to be seen whether the features observed here may be reproduced elsewhere or are particular to Southern California. Second, in the estimation of the intensities of the form (12), the bandwidth and choice of kernel were not optimally selected, but chosen rather arbitrarily. Another issue worth mentioning is that the earthquakes of magnitude greater than 3.5 were used both in the fitting and in the testing. This is in keeping with common practice in seismology, though in statistical terms this is certainly non-standard. Also note that the clustering of small earthquakes in areas where the model assigns low intensity, as suggested by the high values of $L_W(h) - h$ for small h in Figure 3, may or not be causal clustering. That is, these high values of $L_W(h) - h$ may be attributable to clustering of these small earthquakes not accounted for by any mixture model of type (12), or may instead be attributable to inhomogeneity of the process not accounted for by the model. However, the weighted K-function cannot discriminate between these alternatives. It is similarly unclear how robust the estimator $K_W(h)$ is to various departures from our assumptions, and in particular whether the weighted K-function is more or less robust than alternative measures of goodness-of-fit, such as thinned and re-scaled residuals. This is an important subject for future research. In addition, the problem of boundary effects in the estimation of the weighted K-function has not been addressed in this paper. Instead, we have attempted to give a simplified presentation in introducing $K_W(h)$ and its application. It should be noted, however, that exactly the same standard boundary-correction techniques which are used for the ordinary K-function (see Sect. 2.1) can be used for the weighted K-function as well. Fortunately, in our application the fraction of points within distance h of the boundary was so small for all values of h considered as to make such considerations rather negligible.

5 Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant No. 0306526. We thank Dave Jackson, Yan Kagan, and anonymous referees for helpful comments, and the Southern California Earthquake Center for its generosity in sharing their data.

References

[Aka74] H. Akaike. A new look at statistical model identification. *IEEE Transactions on Automatic Control*, AU–19:716–722, 1974.

[Bar64] M. Bartlett. The spectral analysis of two-dimensional point processes. *Biometrika*, 51:299–311, 1964.

[Bes77] J.E. Besag. Comment on “modelling spatial patterns” by b.d. ripley. *Journal of the Royal Statistical Society, Series B*, 39:193–195, 1977.

[BMW00] A.J. Baddeley, J. Møller, and R. Waagepetersen. Non- and semi-parametric estimation of interaction in inhomogeneous point patterns. *Statistica Neerlandica*, 54(3):329–350, 2000.

[BS84] A.J. Baddeley and B.W. Silverman. A cautionary example on the use of second-order methods for analyzing point patterns. *Biometrics*, 40:1089–1093, 1984.

[Cre93] N.A.C. Cressie. *Statistics for spatial data, revised edition.* Wiley, New York, 1993.

[Dav77] R. Davies. Testing the hypothesis that a point process is poisson. *Adv. Appl. Probab.*, 9:724–746, 1977.

[DRS84] J. Dijkstra, T. Rietjens, and F. Steutel. A simple test for uniformity. *Statistica Neerlandica*, 38:33–44, 1984.

[Dur91] R. Durret. *Probability: Theory and Examples.* Wadsworth, Belmont, CA, 1991.

[DVJ03] D. Daley and D. Vere-Jones. *An Introduction to the Theory of Point Processes, 2nd edition.* Springer, New York, 2003.

[Hei88] L. Heinrich. Asymptotic gaussianity of some estimators for reduced factorial moment measures and product densities of stationary poisson cluster processes. *Statistics*, 19:87–106, 1988.

[Hei91] L. Heinrich. Goodness-of-fit tests for the second moment function of a stationary multidimensional poisson process. *Statistics*, 22:245–278, 1991.

[JD81] Ohser J. and Stoyan D. On the second-order and orientation analysis of planar stationary point processes. *Biom. J.*, 23:523–533, 1981.

[Law88] A. Lawson. On tests for spatial trend in a non-homogeneous poisson process. *J. Appl. Stat.*, 15:225–234, 1988.

[LL85] B. Lisek and M. Lisek. A new method for testing whether a point process is poisson. *Statistics*, 16:445–450, 1985.

[MD86] E. Merzbach and Nualart. D. A characterization of the spatial poisson process and changing time. *Ann. Probab.*, 14:1380–1390, 1986.

[Oga88] Y. Ogata. Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical Association*, 83:9–27, 1988.

[Oga98] Y. Ogata. Space-time point-process models for earthquake occurrences. *Ann. Inst. Statist. Math.*, 50:379–402, 1998.

[Ohs83] J. Ohser. On estimators for the reduced second-moment measure of point processes. *Math. Operationsf. Statist., Ser. Statistics*, 14:63–71, 1983.

[PAKB+98] C. Peek-Asa, J.F. Kraus, L.B. Bourque, D. Vimalachandra, J. Yu, and J. Abrams. Fatal and hospitalized injuries resulting from the 1994 northridge earthquake. *Int. J. Epidemiology*, 27 (3):459–465, 1998.

[Rip76] B.D Ripley. The second-order analysis of stationary point processes. *J. Appl. Probab.*, 13:255–266, 1976.

[Rip88] B.D Ripley. *Statistical Inference for Spatial Processes.* Cambridge University Press, Cambridge, 1988.

[Saw75] J.G. Saw. Tests on intensity of a poisson process. *Communications in Statistics*, 4 (8):777–782, 1975.

[Sch79] G. Schwartz. Estimating the dimension of a model. *Annals of Statistics*, 6:461–464, 1979.

[Sch99] F. Schoenberg. Transforming spatial point processes into poisson processes. *Stochastic Processes and their Applications*, 81:155–164, 1999.

[Sch03] F.P. Schoenberg. Multi-dimensional residual analysis of point process models for earthquake occurrences. *Journal of the American Statistical Association*, 98:789–795, 2003.

[Sil78] B.W. Silverman. Distances on circles, toruses and spheres. *J. Appl. Probab.*, 15:136–143, 1978.

[SS00] D. Stoyan and H. Stoyan. Improving ratio estimators of second order point process characteristics. *Scandinavian Journal of Statistics*, 27:641–656, 2000.