

On Mainshock Focal Mechanisms and the Spatial Distribution of Aftershocks

Ka Wong¹ and Frederic Paik Schoenberg^{1,2}

¹ UCLA Department of Statistics

8125 Math-Science Building,

Los Angeles, CA 90095–1554, USA.

² Corresponding author: frederic@stat.ucla.edu

Abstract

Despite previous observations that aftershocks lie preferentially along the fault planes of mainshocks, many branching models for earthquake occurrences do not directly incorporate information on the moment tensors of mainshocks in forecasting the spatial distribution of aftershocks, or do so in a very simplistic way using spatial kernels which have not been rigorously tested on earthquake data to date. We revisit the relationship between the mainshock moment tensor and aftershock location, analyzing strike-slip events in Southern California. Using data from the SCEDC catalog, we find ample evidence suggesting that the fault plane azimuth of a strike-slip earthquake has the highest concentration of aftershocks, a finding that is consistent with previous efforts. We propose a joint distribution to model the aftershock locations that is parameterized in terms of the distance and relative angle to the aftershocks. This model is compared to previously proposed models based on the normal distribution and the squared cosine function. Using residual analysis and weighted K-function as diagnostic measures, we find that both the normal and squared cosine models suffer from several serious problems, and that the joint distribution we propose has features similar to both alternative models but fits much better to Southern California earthquake data.

1 Introduction

Ogata (1998) proposed several spatial extensions to his earlier *epidemic type aftershock sequence* (ETAS) model, in a generalization of his widely-used temporal-magnitude model for earthquake occurrences (Ogata, 1988). Ogata (1998) made an effort to address the

need for an anisotropic aftershock spatial decay function based upon observations that aftershock sequences tend to follow an elliptical pattern rather than a circular one, and suggested the normal distribution as a model for the spatial distribution of aftershocks relative to the mainshocks that trigger them.

While the normal distribution may serve as an acceptable first-order approximation for the distribution of aftershock locations, there is scant evidence of its optimality. Furthermore, the model does not incorporate information on mainshock focal mechanisms which may have value in forecasting aftershock patterns, since it has been widely observed that aftershocks generally occur on or near the fault planes of their associated mainshocks. For instance, Willemann and Frohlich (1987) used the Anderson-Darling statistic to test the distribution of aftershock hypocenters on the focal sphere of mainshocks against a uniform distribution to show that for deep focus earthquakes, aftershocks that were greater than 20 km from the mainshock exhibit significant clustering in the plane of the Wadati-Benioff zone. Michael (1989) applied the same method to six shallow focus aftershock sequences in California and found significant clustering on the fault plane. Kagan (1992) adopted a more exploratory approach by using equal-area projection rather than the Anderson-Darling statistic. He too reached a similar conclusion regarding earthquake clustering along fault planes. Based on this result, Kagan and Jackson (1994) introduced an anisotropic function for their spatial smoothing kernel in their long-term earthquake forecast.

The aforementioned studies focused exclusively on the clustering of hypocenters in certain directions on the focal sphere. We argue that such analyses essentially ignore the

potentially important relationship between the *distance* from an earthquake’s hypocenter or epicenter to those of its aftershocks and the angular separation between an earthquake and its aftershocks. In this paper, we attempt to model both the angle and distance between earthquakes and their aftershocks in concert.

Our analysis focuses on strike-slip earthquakes with local magnitude $M_L > 3.0$ in Southern California between 1999 and 2006 as *mainshocks*. Due to the difficulty and subjectivity associated with the assessment of the precise branching structure in earthquake catalogs, we follow Zhuang et. al. (2002) in using a model-based method to identify aftershock sequences stochastically. Details of this method and the catalog are given in Section 2. In Section 3, we propose a semi-parametric model for the spatial distribution of aftershocks that is composed of a marginal distribution for distance and a conditional distribution for the relative angle from the mainshock. This model is compared to two alternatives that have been used to describe aftershock and seismic patterns respectively: the normal model and the spatial smoothing kernel of Kagan and Jackson (1994). In Section 4, we show that not only do aftershocks cluster around the fault plane of the mainshocks, but also their distances from the mainshock have a great impact on the distribution of their relative angles. We use point process residuals and the weighted K-function in Section 5 to assess the fit of the various models and show that the proposed semi-parametric model appears to offer superior fit to Southern California seismicity. A discussion and suggestions for future work are given in Section 6.

2 Data and Resources

The Southern California Earthquake Data Center (SCEDC) archives moment tensor estimates for earthquakes in Southern California (Clinton et al. 2006). We focus here on earthquakes in the catalog occurring between September 18, 1999, and Dec 31, 2006, with epicenters in Southern California and a moment magnitude of M3.0 or above. The published hypocenters in the SCEDC moment tensor catalog are used for the locations of both mainshocks and aftershocks. Based on comparison of the frequency-magnitude distribution of the catalog to the Gutenberg-Richter distribution, the catalog is believed to be complete for earthquakes above M3.0, though some events on the edge of the network can be missing due to the lack of a qualified moment tensor solution (Clinton et al. 2006). All qualified solutions have been labeled by previous researchers with a quality grade that reflects the precision of the inversion process, which in turn depends on the event's epicentral location relative to the monitoring stations, as well as its magnitude and depth (Clinton et. al. 2006). The three quality grades are A, B, and C, in decreasing order of their precision.

Because different types of earthquakes are likely to have very different patterns of aftershock activity, in this paper we consider only the aftershock activity surrounding strike-slip earthquakes of quality A or B, with the notion that the aftershock activity surrounding other types of events may be analyzed using similar methods in future work. As pointed out by Kagan and Jackson (1998), the simplicity of the geometry of strike-slip

faulting facilitates the description and interpretation of aftershock patterns: in a strike-slip event, the fault plane intersects the surface of the earth almost vertically and thus the fault plane azimuth is a fairly accurate representation of the fault plane itself.

Southern California is populated by a large number of right-lateral strike-slip faults. As a result, strike-slip events make up the majority of earthquakes in Southern California. In this paper, we categorize an earthquake mechanism as strike-slip if its neutral axis of the moment tensor (B-axis) is within 20° of the vertical (some previous authors have used a cutoff of 30° instead of 20°) (Frohlich 1992, 2001). A strike-slip will thus have a nearly vertical fault plane and a nearly horizontal rupture motion. Roughly 1/3 of earthquakes in the SCEDC catalog fall into this category.

In determining the mainshock-aftershock assignments, we adopt the model-based approach taken by Zhuang et. al. (2002). In this approach, one may assume an aftershock sequence model and assign the aftershock branching structure probabilistically according to the model. In particular, we take a space-time ETAS model (Ogata 1998) and assume the conditional intensity at point (t, x, y) is the sum of the background and triggering intensities,

$$\lambda(t, x, y | \mathcal{H}_t) = \mu(x, y) + \sum_{i: t_i < t} g(t, x, y | t_i, x_i, y_i),$$

where $\mu(\cdot)$ is a constant background rate and $g(\cdot | \cdot)$ is the response function. After fitting the catalog to the model with only strike-slips as trigger events, we may attribute event j to be an offspring of an earlier event i with probability $\rho_{i,j}$ defined as the ratio of the

response function of i to the conditional intensity at (t_j, x_j, y_j) , that is,

$$\rho_{i,j} = \frac{g(t_j, x_j, y_j | t_i, x_i, y_i)}{\lambda(t_j, x_j, y_j | \mathcal{H}_{t_j})}.$$

For any event i in the catalog, a stochastic realization of its aftershock sequence can be emulated by keeping all of its subsequent events with probability $\rho_{i,j}$ for $j = i + 1, i + 2, \dots, n$. Although this approach involves an extra step of fitting an ETAS model and leads to non-deterministic results, use of this method is an alternative to the often arbitrary and subjective decisions that can arise in determining the mainshock-aftershock assignments. Furthermore, the emulation process is repeated multiple times and is shown not to affect the main conclusions of this paper. We henceforth refer to a single realization of this Zhuang et al. (2002) assignment procedure. While our analysis is not concerned with discerning whether particular events are foreshocks, mainshocks, aftershocks, or swarms, for purposes of explanation in this paper, let us refer to the strike-slip earthquakes in the SCEDC catalog of quality A or B as *mainshocks*. According to these definitions, the catalog contains 87 distinct mainshocks, which collectively have a total of 1224 emulated aftershocks.

We consider an aftershock’s relative location to any mainshock with respect to the focal mechanism of the mainshock, as illustrated in Figure 1. The “beachball” is a representation of the focal mechanism of a right-lateral strike-slip. The location of an aftershock relative to a mainshock is measured by r and ϕ , where r is the epicentral distance between the two events, and ϕ is the angular separation between the aftershock and the mainshock’s fault

plane. (Here and in what follows, by *fault plane* we refer to the intersection of the Earth’s surface with the fault plane associated with a particular event.) The fault plane ambiguity is resolved by assuming all strike-slip faults in Southern California are right-lateral unless individual aftershock sequences clearly delineate a left-lateral fault upon examination. Gomberg (2003) observed strong directivity effects among a number of unilaterally-rupturing strike-slips with magnitude $M_s > 5.4$ in different tectonic environments. While events with directive ruptures may have asymmetrically triggered aftershocks, such events and their propagation directions are difficult to quantify in a large catalog of various-sized earthquakes. Therefore we will ignore possible directivity effects and treat fault plane as an axis without sense of direction. To tentatively differentiate between the compression zone and dilatation zone of the mainshock focal mechanism, one may define ϕ as the angle measured counter-clockwise from the aftershock’s epicenter to the nearest mainshock fault plane for a right-lateral strike-slip, and clockwise for a left-lateral event. Thus ϕ spans from 0 to π , and $\phi \in (0, \pi/2)$ and $\phi \in (\pi/2, \pi)$ represent the compression zone (containing the P-axis) and dilatation zone (containing the T-axis) respectively. While the Coulomb stress changes caused by the slip of the mainshock may be different in both zones (compression and dilatation), the numbers of aftershocks in both zones are found to be similar: 618 in dilatation zone and 606 in compression zone. Furthermore, a chi-square (χ^2) test indicates the observed aftershock patterns in the two zones are not statistically significant for this dataset. Indeed, when each zone is divided into a 3×3 grid and only cells with at least 5 points enter into the test, the χ^2 test as to whether the two distributions are signifi-

cantly different yields a χ^2 -statistic of 3.98 with 5 degrees of freedom, which corresponds to a p-value of 0.55. In light of such evidence, we do not distinguish the two zones in the remainder of this paper and restrict ϕ to $[0, \pi/2]$ by defining it as the absolute angular separation to the nearest fault plane.

3 Methods

3.1 Proposed TPWE model for the relative locations of aftershocks

We propose to model the distribution of the relative locations of aftershocks, in polar coordinates, as a product of two distributions: 1) a marginal tapered Pareto distribution for the distance r between mainshocks and their aftershocks, and 2) a wrapped exponential distribution for the relative angle ϕ between mainshocks and their aftershocks, given the distance r . The distribution may be written $f(r, \phi) = \frac{1}{r} \cdot f_r(r) \cdot f_{\phi|r}(\phi|r)$, where f_r and $f_{\phi|r}$ are each one-dimensional densities to be estimated, so that $\int f(r, \phi) r \, dr \, d\phi = 1$. We refer to this model as the Tapered Pareto Wrapped Exponential (TPWE) model in what follows.

The distribution of distance between mainshocks and aftershocks has been investigated in previous studies using various methods. Utsu (1969) noted that aftershock regions tend to be elliptical, and Ogata (1998) built upon this work, questioning whether the distance decay function is short range (i.e. normal) or long range (i.e. inverse power law) and proposing several moment-weighted models for alternatives. The inverse power law was

shown by Felzer et. al. (2006) to be a good description of aftershock distances between 0.2 km and 50 km. In a time-independent framework, Kagan and Jackson (1994) used a density proportional to $1/r$ to describe distances between mainshocks and aftershocks, in producing a long-term seismic hazard map.

More recently, several authors have begun using the tapered Pareto distribution to describe the distribution of phenomena such as seismic moments (Jackson and Kagan 1999; Vere-Jones et al. 2001) which obey some power-law type of behavior but which are not quite as heavy-tailed as the Pareto, and this tapered Pareto distribution has been used to describe the times and distances between successive earthquakes in Southern California (Schoenberg et al. 2008). The tapered Pareto has cumulative distribution function:

$$F_{tap}(x) = 1 - (a/x)^\beta \exp\left(\frac{a-x}{\theta}\right), a \leq x \leq \infty$$

Here θ is a threshold after which frequency begins to decay especially rapidly. Additional information concerning the density, characteristic function, moments, and other properties of the tapered Pareto can be found in Kagan and Schoenberg (2001).

While a variety of models for the density of r have been proposed, the form of $f_{\phi|r}(\phi|r)$ has been the subject of relatively scant scrutiny to date. As part of their spatial smoothing kernel, Kagan and Jackson (1998) used a directivity function, expressed as $D = 1 + \delta \cdot \cos^2(\phi)$, where δ measures the concentration of earthquake epicenters around the presumed fault plane. Here D does not serve exactly the same purpose as $f_{\phi|r}(\phi|r)$ because $f_{\phi|r}(\phi|r)$ concerns aftershocks, i.e. seismic activity at times *after* a mainshock, whereas D describes

the time-independent distribution of ϕ around any earthquake. Further, our investigations suggest that, for the SCEDC data, the distribution of ϕ appears to depend on r , so that the conditional distribution of ϕ given r may be more meaningful than the overall marginal distribution of ϕ . Due to lack of theoretical support for a particular functional form, we take a semi-parametric approach here. We propose a circular distribution called the wrapped exponential (*WE*), whose single parameter λ may be estimated locally within selected bins, as described below.

Wrapped distributions are useful for modeling angular variables such as ϕ , the relative angle between mainshocks and aftershocks (Mardia and Jupp 2000). Such a distribution is obtained by conceptually wrapping a distribution on the real line around the circumference of a unit circle. That is, if x is a real random variable with an arbitrary probability density function f , then the wrapped analogue of f has density

$$f_w(x_w) = \sum_{k=-\infty}^{\infty} f(x_w + 2k\pi), \quad x_w \in [0, 2\pi),$$

where

$$x_w = x(\bmod 2\pi)$$

This approach can be applied to any probability distribution to manufacture a large class of circular distributions, among which the wrapped Gaussian and wrapped Cauchy are examples. However, one shortcoming common to most such wrapped distributions is the lack of a closed form expression, which renders parameter estimation difficult. One exception

is the wrapped exponential (*WE*) in which the infinite series converges and has a remarkably simple solution (Jammalamadaka and Kozubowski, 2001). The *WE* has been used to model seismic events triggered by periodic processes (Jupp et al. 2004), and is obtained by applying the wrapping procedure to the exponential density, $f(x) = \lambda e^{-\lambda x}, x > 0$. Since ϕ in this analysis only spans $[0, \pi/2]$ by construction, it will cycle on a quarter circle instead of a full circle. On a quarter circle, the *WE* has density

$$f_{we}(\phi) = \frac{\lambda e^{-\lambda \phi}}{1 - e^{-\lambda \pi/2}}, \quad \phi \in [0, \pi/2].$$

Note that this is equivalent to the density of a truncated exponential random variable on the line, i.e. $f(X|X < \pi/2)$, due to the memoryless property of the exponential distribution.

The *WE* has a single shape parameter λ . When $\lambda = 0$, the *WE* corresponds to a uniform distribution on its support. As λ increases, the skewness of the distribution increases, and as λ approaches ∞ , the distribution degenerates to a point mass at $\phi = 0$. With only one parameter, the *WE* provides a very good fit to the conditional distribution of ϕ given r . Evidence suggests, however, that the conditional distribution of ϕ changes depending on the value of r . One way to model its dependency on r is to let the parameter λ vary as a function of r . We propose to estimate the value of λ locally by maximum likelihood (ML) successively on different bins, each containing n pairs of points, sorted according to the distance r between mainshock and aftershock; in order to estimate λ more accurately within each bin, we use all possible mainshock-aftershock pairs, weighting each pair by its probability $\rho_{i,j}$ of being an actual mainshock-aftershock pairing. We experimented with

several choices of n and selected $n = 200$ in order to achieve a satisfactory bias-variance tradeoff.

It is a subject of debate whether the distribution of aftershocks is dependent on the mainshock magnitude (Kagan 2002, Huc et. al. 2003, Davidsen et. al. 2005). We attempt to account for possible scaling effects by using an additional scaled distance in this analysis. We estimate the surface rupture length (L) of a fault from empirical relationships (Wells et. al. 1994) and express the scaled distance in terms of fault lengths, r/L . The marginal distribution of r/L and the conditional distribution of ϕ with respect to r/L are studied in the same manner as in the case for r .

3.2 Alternative models

Two alternative models are considered in this paper for comparison. The first model is the normal model and the second is the spatial smoothing kernel of Kagan and Jackson (1994). The normal model has been proposed to describe the relative locations of aftershocks and thus is certainly applicable to the problem in hand. Kagan and Jackson's model, by contrast, was suggested in a slightly different context, as explained below, but may nevertheless constitute a relevant model for comparison.

The normal model is often used as a conveniently simple spatial distribution for aftershock sequences (Rhoades and Evison 1993, Kagan 2002, Ogata 1998). A slight modification was made by Ogata (1998), who introduced an anisotropic function based on the normal model as a spatial extension of his earlier ETAS model (Ogata 1988), and pro-

posed a separate normal model to be fit to each aftershock sequence in the de-clustered catalogue, using an anisotropic metric that replaces the Euclidean metric. Compelling evidence to support the normal model is the commonly seen elliptical shape of aftershock zones (Rhoades and Evison 1993).

Kagan and Jackson (1994) estimated the long-term rate densities for earthquakes as a weighted sum of smoothing kernels, each centered at the epicenter of a previous i th earthquake, using information on the earthquakes’ focal mechanisms. Adopting current notation, the density at point (x, y) is estimated in Kagan and Jackson (1994) as

$$f(x, y) = \sum_i f_i(r_i, \phi_i, M_i) + s$$

where $s = 0.02$ is a small constant to allow for “surprises” far from past earthquakes. Of interest here is their proposed smoothing kernel

$$f_i(r_i, \phi_i, M_i) = A \cdot (M_i - M_{cut}) \cdot [1 + \delta \cdot \cos^2(\phi_j)] \cdot \frac{1}{r_i},$$

where A is a normalization constant, M_i is the magnitude of earthquake, and δ is a parameter controlling the degree of azimuthal concentration in a direction relative to the earthquakes focal mechanism (Kagan and Jackson 1994). We refer to this model as the *KJ* model in what follows. Kagan and Jackson used their own expert knowledge to choose δ . In a region where the concentration of aftershocks along the fault plane is high, δ should be assigned a high value, and δ should be small if the earthquakes are dispersed relatively isotropically. The choice of smoothing kernel was selected as a result of the analysis in Kagan (1992) of focal mechanisms and the distribution of hypocenters on the focal sphere.

Although the model was motivated by analysis of subduction zone earthquakes, Kagan et al. (2007) recently applied this model to seismicity in southern California using $\delta = 100$. In the current analysis, the value of δ is selected via maximum likelihood.

3.3 Diagnostics

We focus on a subset of relative mainshock-aftershock locations in a 20×20 km² window within which compare the goodness-of-fit for the different models. One method of assessing the goodness-of-fit of a spatial density for point process is by examining the residuals over various quadrats, as suggested in Baddeley et al. (2005). That is, one partitions the space into cells and calculates the residuals in each cell, which may be standardized in various ways. For instance, the residual R_i corresponding to cell i may be defined as

$$R_i = \frac{N_i - E_i}{\sqrt{E_i}},$$

where N_i is the observed number of points in the cell, and E_i is the expected number of points defined. E_i can be found by integrating the estimated intensity \hat{f} over the cell i . This is effectively Pearson residuals in the Poisson log-linear regression context. By construction, the residuals are standardized to have mean 0 and standard deviation approximately 1. Outliers and systematic patterns in the residuals may indicate lack of fit.

The K -function described by Ripley (1981) is commonly used to detect excessive clustering or inhibition in a point process. The function $K(h)$ is defined as the average number of additional points within h of any given point, divided by the overall rate. The null

hypothesis is that the underlying point process is homogeneous Poisson. In cases where the hypothesis is not uniform, each point may be weighted according to the rate of the point-process in question, yielding the *weighted* or inhomogeneous K -function (Baddeley et al. 2000). The weighted K -function has been used by Veen and Schoenberg (2005) to assess the spatial distributions in point process models for earthquakes in Southern California. To test the null hypothesis that the spatial intensity of points (mainshock-aftershock pairs) in region D is $f_0(x, y)$, the weighted K -function may be defined as

$$K_W(h) = \frac{1}{f_*^2 N} \sum_i w_i \sum_{j \neq i} w_j \mathbf{1}(|p_i - p_j| \leq h),$$

where N is the total number of observed pairs of points, $f_* := \inf\{f_0(x, y); (x, y) \in D\}$ is the infimum of the density over the observed region, $\mathbf{1}(\cdot)$ is the indicator function, and $w_r = f_*/f_0(p_r)$, where $f_0(p_r)$ is the modeled density of pairs of points at vector distance p_r apart. Veen and Schoenberg (2005) verified that for the Poisson case where f_0 is locally constant on distinct subregions whose areas are large relative to the interpoint distance h_n , the *weighted* K -function is approximately normal with mean πh^2 and variance $2\pi h^2 A/[E(N)]^2$, where A is the size of the area being studied and N is the number of points observed in A . One common issue in applying the K -function or weighted K -function is the problem of boundary correction. One method of edge correction is to make mirror images of the points (where each point is a mainshock-aftershock pair) along the boundaries over which these points are observed, as suggested e.g. in Ripley (1981). Since all of our observed points are restricted to the first quadrant on the plane, we reflect each point

along both the x- and y-axes.

Statistical inference is drawn from simulation-based confidence bounds. 4,000 samples are taken and a weighted K-function is estimated for each sample. Based on these weighted K-functions, we form 95% pointwise confidence bounds to make statistical inferences.

4 Results

4.1 Fit of proposed models

Figure 2 shows a subset of mainshock-aftershock pairs as described in Section 2. Not surprisingly, it is immediately noticeable that the concentration of points is much higher near the x-axis (i.e. the fault plane) than elsewhere. One may observe that aftershocks seem to be distributed nearly uniformly in all directions when they are close to the respective mainshock, whereas when they are further away, they tend to lie more predominantly along the azimuth of the fault plane (i.e. along the x-axis). One may also observe the discreteness in the observed distances r between mainshocks and their aftershocks, for pairs that are very close together; this is a result of the resolution of measurements recorded in the catalog. Note that rounding errors in the locations, compounded by estimation errors in epicenter locations, may translate into large errors in the estimation of ϕ , especially for small values of r , where a tiny change in location will translate into a large change in ϕ . Willemann and Frohlich (1987) and Michael (1989) discarded aftershocks within 5 km of each mainshock, in order to avoid dealing with these noisy observations.

Figure 3 shows the survival functions of r , a fitted Pareto (i.e. inverse power law), and a fitted tapered Pareto distribution, on log-log scale. The diagram indicates the Pareto offers a good fit to the data for $r < 50$ km, coinciding with the observation made by Felzer et. al. (2006). On the other hand, the tapered Pareto distribution fits the data equally well in the same range, however, its drastically better fit to the tail distribution renders it a preferred alternative.

Figure 4 displays the conditional histogram of ϕ for three different ranges of r : a) $10 \text{ km} \leq r < 15 \text{ km}$, b) $5 \text{ km} \leq r < 10 \text{ km}$, and c) $0 \text{ km} \leq r < 5 \text{ km}$. For r less than 5 km, the distribution of ϕ is seen almost uniform. As noted earlier, for small r values, ϕ is very sensitive to errors in the locations. Hence a possible explanation for its uniform distribution in this range of r is that it is dominated by noise. In higher ranges of r values, the skewness of the distribution increases as ϕ tend to concentrate at low values. Overlaid on the histograms are fitted WE densities. We can see that the WE is able to capture the shape of the conditional histograms in different regions and seems to fit each of the densities rather well.

The local behavior of ϕ is sensitive to the value of r , as shown in Figure 5, which displays the local weighted maximum likelihood estimates of λ plotted against the mean distance in the corresponding bins. One sees that when r is small, the estimates of λ are unstable and seem again dominated by noise. As r increases, λ climbs steadily until roughly $r = 18$ km before it declines slowly. For the purposes of interpolation, extrapolation, or forecasting, one may seek a parametrization of the estimates of λ . The F-distribution provides a good

approximation to the shape of the estimates of λ . Let $f_{v_1, v_2}(x)$ be the density function of the F-distribution with v_1 and v_2 degrees of freedom, the function $2.7 \times f_{10, 600}(x/22)$ is found to be the best fit by least squares among functions of comparable form. The fitted function is superimposed on the estimates of λ in Figure 5.

The marginal distribution of scaled distance r/L and the corresponding condition distribution of ϕ are shown in Figures 6 and 7 respectively. The fit of the tapered Pareto to r/L is slightly worse than is the case for r but still appears reasonable. In comparison, the Pareto distribution systemically deviates from the empirical survival function over the entire range of data, rendering it a misfit. The conditional distribution of ϕ with respect to r/L shows strong resemblance to its counterpart for unscaled distance. The estimates of λ are low in the upper and lower ranges of r/L and are largest at roughly 40 fault lengths. The shape of the estimates seems again well capture by an F-distribution. Although it is difficult to infer from Figures 6 and 7 whether the distribution of aftershocks is dependent on mainshock magnitude, both figures seem to suggest that TPWE is applicable to both *scaled* and *unscaled* distances, at least for the range of data considered. For this reason, we focus exclusively on r in the remainder of this paper, with the notion that the same can be done for r/L .

Figure 8 displays the density surfaces on logarithmic scale for the fitted KJ, normal, and TPWE models, to unscaled relative mainshock-aftershock locations. A comparison between these surfaces reveals characteristic differences and similarities between the models. The KJ model has a rather sharp peak that reaches about the same height as the normal model.

The KJ density decays very slowly outward and, as a result, retains a substantial density through the entire region. Its shape mimics a 2-petalled rose that expands along the x-axis and tightens along the y-axis. The normal model in contrast is comparatively smooth near the origin and has relatively flat tails, obtaining densities that are very close to zero outside the visible contours. The contour lines themselves, for the normal model, all have an elliptical shape that seems to resemble aftershock zones. The TPWE model can perhaps be viewed as a hybrid of the KJ and normal models. On the one hand, it possess a sharp peak, like the KJ model (though the TPWE's peak density is much higher than that of the KJ model). On the other hand, its tail is quite thin and the visible contour lines cover roughly the same area as the normal density. In addition, the bow-tie shape of the contours of the TPWE model seems to have features that encompass characteristics of both the KJ and normal models. While the TPWE model is not meant to be a compromise of the other two models by construction, it nevertheless shares some similarity with both of them.

4.2 Diagnostics

The absolute values of the quadrat residuals of the KJ, normal, and TPWE models are shown in Figure 9, on a logarithmic scale in order to facilitate visualization. One sees immediately that the normal model has several outlying residuals of very large size. This is due to the occurrence of mainshock-aftershock pairs at relative distances where the normal model assigns a density very close to zero. By contrast, the KJ model assigns a substantial density to these outliers. However, it does so at the cost of having a substantial density

throughout the entire region. As a result, the model is over-predicting in most of the upper half of the top-left panel of Fig. 8, at relative distances where very few mainshock-aftershock pairs were observed. The proposed TPWE joint distribution seems to achieve a balance between the other two. On the one hand, the outlying residuals are several orders of magnitude smaller than those in the normal model; on the other hand, the TPWE model's residuals are much smaller than the KJ model in relative locations where observations are rare. Near the origin, both the normal and KJ models tend to under-predict the density. For instance, in the KJ model, there is a vertical cluster of large residuals in the region above the origin, indicating a systematic lack of fit. A similar problem is seen in the residuals of the normal model. The peaks in these two densities are too low, relative to the observed mainshock-aftershock pairs. The TPWE model, by contrast, has much smaller residuals near the origin, indicating superiority of fit.

The weighted K -functions for the three models are shown in Figure 10. Also plotted are the theoretical mean and simulation-based pointwise 95% confidence bounds. There is serious departure from the confidence bounds in both the KJ and normal models, indicating statistically significant lack-of-fit. The weighted K -function for the KJ model is below the lower threshold of the 95% confidence bounds for all values of h , as a result of its wide-area over-estimation of aftershock density. In the normal model, the weighted K -function is plotted on a logarithmic scale because the estimates of the K_W are orders of magnitude above the upper bounds of the 95% confidence intervals. Such a dramatic departure is a result of serious underestimation of the density at the origin as well as a few outlier

locations where aftershocks are observed. The TPWE does not seem to have systematic over-estimation or under-estimation of the density of relative distances between mainshocks and their aftershocks, nor is there any serious indication in Figure 10 of clustering or inhibition of the mainshock-aftershock pairs relative to this joint distribution.

5 Conclusion

The tapered Pareto / wrapped exponential (TPWE) model appears to adequately describe the locations of aftershocks relative to their corresponding mainshocks, and vastly outperforms competing models such as the KJ and normal models. It must be emphasized, however, that this analysis was performed using only southern California strike-slip earthquakes of quality A or B as candidates for *mainshocks*. In other seismic regions, especially in areas where the faulting is more heterogeneous, the TPWE model might not fit well, and an important direction for future work is the investigation of the fit of such models in other seismically active zones. It should also be noted that earthquakes are treated as point sources in this analysis. For large earthquakes whose epicenters are a poor representation of their locations and where the validity of point source approximation is doubtful, TPWE may not be applicable.

One motivation for the current investigation is to improve existing models for forecasting seismicity. Models which are currently used for such purposes, such as the ETAS models of Ogata (1998), tend to use rather simplistic forms for the spatial kernel, and

the goodness-of-fit of these spatial kernels have been too sparsely investigated to date. Integration of the proposed TPWE model into ETAS models and other branching point process models for earthquakes is an important subject for future research. In addition, as depth measurements for earthquakes become increasingly accurate and other features of earthquake faults become discernable, including possible directivity of aftershocks, such effects should also be implemented in branching models for earthquake forecasting.

Acknowledgement. We thank Yan Kagan, David Jackson, and the anonymous referees for their very useful suggestions. This material is based upon work supported by the National Science Foundation and the Southern California Earthquake Center under Grant No. 129569.

References

- [1] Baddeley, A., Mller, J., and Waagepetersen, R. (2000). Non and semi-parametric estimation of interaction in inhomogeneous point patterns. *Statistica Neerlandica*, 54(3), 329-350.
- [2] Baddeley, A., Turner, R., Moeller, J., and Hazelton, M. (2005). Residual analysis for spatial point processes. *Journal of the Royal Statistical Society B*, 67(5), 617–666.
- [3] Clinton, J. F., Hauksson, E., and Solanki, K. (2006). An evaluation of the SCSN moment tensor solutions: robustness of the M_w magnitude scale, style of faulting, and automation of the method, *BSSA*, 96(5), 1689-1705.
- [4] Davidsen, J. and Paczuski, M. (2005). Analysis of the spatial distribution between successive earthquakes, *Phys. Rev. Lett.*, 94, Art. Num. 048501.
- [5] Felzer, K. and Brodsky, E. (2006). Decay of aftershock density with distance indicates triggering by dynamic stress, *Nature*, 441, 735-738.
- [6] Frohlich, C. (1992). Triangle diagrams: ternary graphs to display similarity and diversity of earthquake focal mechanisms, *Physics of the Earth and Planetary Interiors*, 75, 193-198.
- [7] Frohlich, C. (2001). Display and quantitative assessment of distributions of earthquake focal mechanisms, *Geophys. J. Int.*, 144, 300-308.

- [8] Gomberg, J., Bodin, P., and Reasenberg, P. (2003). Observing earthquakes triggered in the near field by dynamic deformations, *BSSA*, 93(1), 118-138.
- [9] Huc, M. and Main, I. G. (2003). Anomalous stress diffusion in earthquake triggering: Correlation length, time-dependence, and directionality, *J. Geophys. Res.* 108(B7), 2324.
- [10] Jackson, D.D. and Kagan, Y.Y. (1999). Testable earthquake forecasts for 1999. *Seismological Research Letters* 70(4), 393-403.
- [11] Jammalamadaka, S.R. and Kozubowski, T.J. (2001). A wrapped exponential circular model, *Proc. of AP Academy of Sciences*, 5(1), 43-56.
- [12] Jupp, T., Pyle, D., Mason, B., and Dade, B. (2004). A statistical model for the timing of earthquakes and volcanic eruptions influenced by periodic processes, *Journal of Geophysical Research*, 109(B2), B02206.1–B02206.16.
- [13] Kagan, Y.Y. (1992). Correlations of earthquake focal mechanisms, *Geophys. J. Int.*, 110, 305-320.
- [14] Kagan, Y.Y. (2002). Aftershock zone scaling, *BSSA*, 92(2), 641-655.
- [15] Kagan, Y.Y. and Jackson D.D. (1994). Long-term probabilistic forecasting of earthquakes, *Journal of Geophysical Research*, 99(B7), 13,685-13,700.

- [16] Kagan, Y.Y. and Jackson D.D. (1998). Spatial aftershock distribution: Effect of normal stress, *Journal of Geophysical Research*, 103(B10), 24453–24267.
- [17] Kagan, Y.Y. Jackson D.D., and Rong Y. (2007). A Testable five-year forecast of moderate and large earthquakes in Southern California based on smoothed seismicity, *Seismological Research Letters*, 78(1), 94-98.
- [18] Kagan, Y. and Schoenberg, F. (2001). Estimation of the upper cutoff parameter for the tapered Pareto distribution. *J. Appl. Prob.* 38A, Supplement: Festschrift for David Vere-Jones, D. Daley, editor, 158–175.
- [19] Mardia, K.V. and Jupp, P.E. (2000). *Directional Statistics*, Wiley, New York
- [20] Michael, A. (1989). Spatial patterns of aftershocks of shallow focus earthquakes in California and implications for deep focus earthquakes, *Journal of Geophysical Research*, 94(B5), 5615-5626.
- [21] Ogata, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes, *JASA*, 83(401), 9-27.
- [22] Ogata, Y. (1998). Space-time point-process models for earthquake occurrences, *Annals of the Institute of Statistical Mathematics*, 50(2), 379-402.
- [23] Rhoades, D.A. and Evison, F.F. (1993). Long-range earthquake forecasting based on a single predictor, *Geophysical Journal of the Royal Astronomical Society*, 59, 43–56.

- [24] Ripley, B. (1981). *Spatial Statistics*, Wiley, New York.
- [25] Schoenberg, F.P., Barr, C., and Seo, J. (2008). The distribution of Voronoi cells generated by Southern California earthquake epicenters, *Environmetrics*, 19, 1-14.
- [26] Utsu, T. (1969). Aftershocks and earthquake statistics (I): some parameters which characterize an aftershock sequence and their interaction. *Journal of the Faculty of Science, Hokkaido University, Ser. VII (geophysics)*, 3, 129–195.
- [27] Veen, A. and Schoenberg, F.P. (2005). Assessing spatial point process models for California earthquakes using weighted K-functions: analysis of California earthquakes. in *Case Studies in Spatial Point Process Models*, editors: Baddeley, A., Gregori, P., Mateu, J., Stoica, R., and Stoyan, D. (eds.), Springer, NY, 293–306.
- [28] Vere-Jones, D., Robinson, R., and Yang, W.Z. (2001). Remarks on the accelerated moment release model, *Geophysical Journal International*, 144(3), 517-531.
- [29] Wells, D. and Coppersmith, K. (1994). New empirical relationships among magnitude, rupture length, rupture width, rupture area, and surface displacement, *BSSA*, 84(4), 974-1002.
- [30] Willemann, R.J. and Frohlich, C. (1987). Spatial patterns of aftershocks of deep focus earthquakes, *Journal of Geophysical Research*, 92(B13), 13927-13943.
- [31] Zhuang, J., Ogata, Y., and Vere-Jones, D. (2002). Stochastic declustering of space-time earthquake occurrences, *JASA*, 97(458), 369-380.

Authors' affiliations and addresses.

Ka Wong

Ph.D. student, UCLA Department of Statistics

8125 Math-Science Building

University of California, Los Angeles, CA 90095–1554, USA.

Frederic Paik Schoenberg

Professor, UCLA Department of Statistics

8125 Math-Science Building

Los Angeles, CA 90095–1554, USA.

Figure captions.

Figure 1: Schematic showing an aftershock's relative location with respect to a right-lateral strike-slip mainshock's focal mechanism.

Figure 2: Scatterplot of mainshock-aftershock relative locations, with respect to the mainshock's fault plane. Only a subset is displayed in a $20 \times 20 \text{ km}^2$ window.

Figure 3: Survival function $(1 - F\{r\})$ for mainshock-aftershock distances, r .

Figure 4: Histograms of relative angle (ϕ) between mainshocks and aftershocks, arranged according to distance r between mainshocks and aftershocks. Top panel: $10 \text{ km} \leq r < 15 \text{ km}$; middle panel: $5 \text{ km} \leq r < 10 \text{ km}$; bottom panel: $0 \text{ km} \leq r < 5 \text{ km}$.

Figure 5: Estimates of the parameter λ in the wrapped exponential (WE) distribution, as a function of distance r between mainshock and aftershock. The gray dots are estimates obtained by bins of 200 mainshock-aftershock pairs each; the black dotted curve shows the parameterization using the density function of an F-distribution.

Figure 6: Survival function $(1 - F\{r/L\})$ for normalized mainshock-aftershock distances, r/L .

Figure 7: Estimates of the parameter λ in the wrapped exponential (WE) distribution, as a function of scaled distance r/L . The gray dots are estimates obtained by bins of 200 mainshock-aftershock pairs each; the black dotted curve shows the parameterization using the density function of an F-distribution.

Figure 8: Density plots on logarithmic scale corresponding to three models for mainshock-aftershock relative locations. Left: KJ model; middle: normal model; right: TPWE model.

Figure 9: Quadrat residuals from each of the three models for mainshock-aftershock relative locations. Left: KJ model; middle: normal model; right: TPWE model.

Figure 10: Weighted K -functions corresponding to three models for mainshock-aftershock relative locations. Top: KJ model; middle: normal model; bottom: TPWE model.

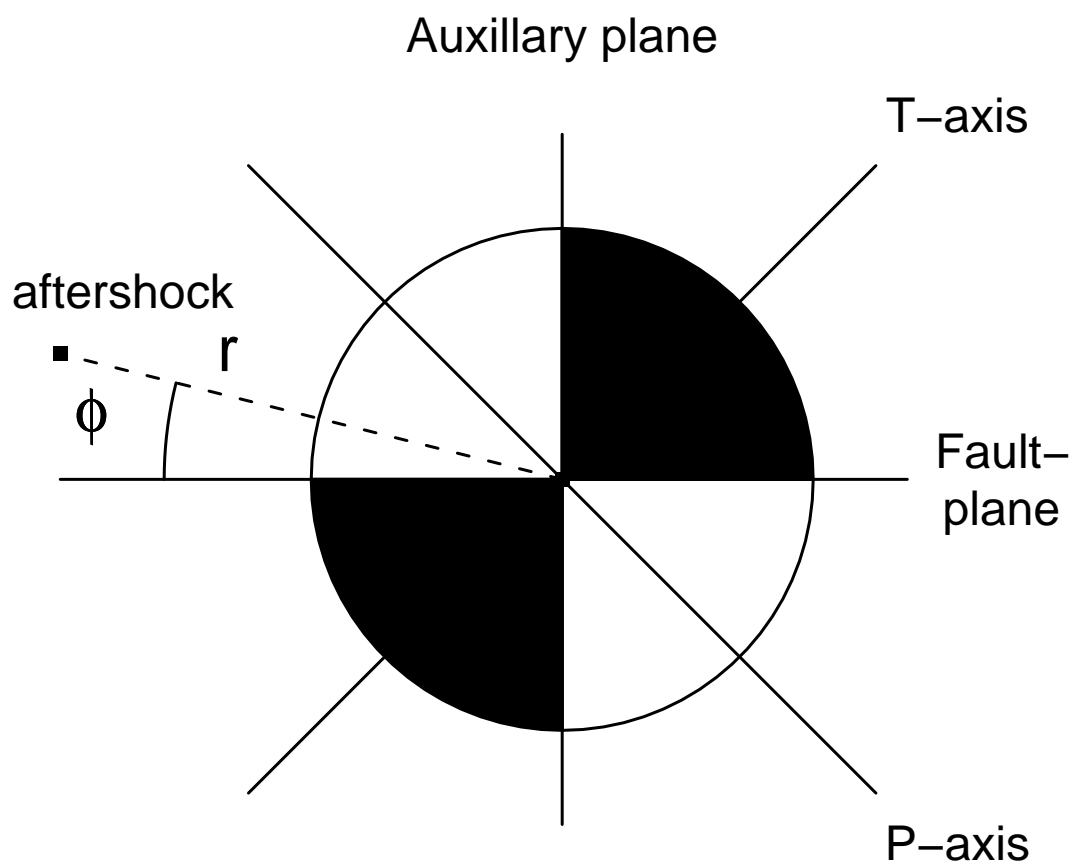


Figure 1:

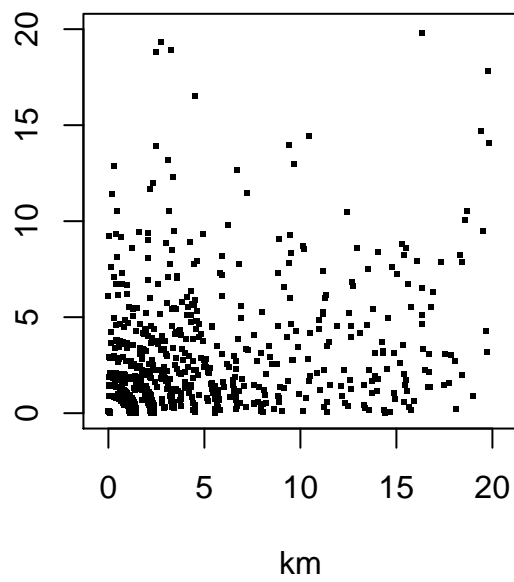


Figure 2:

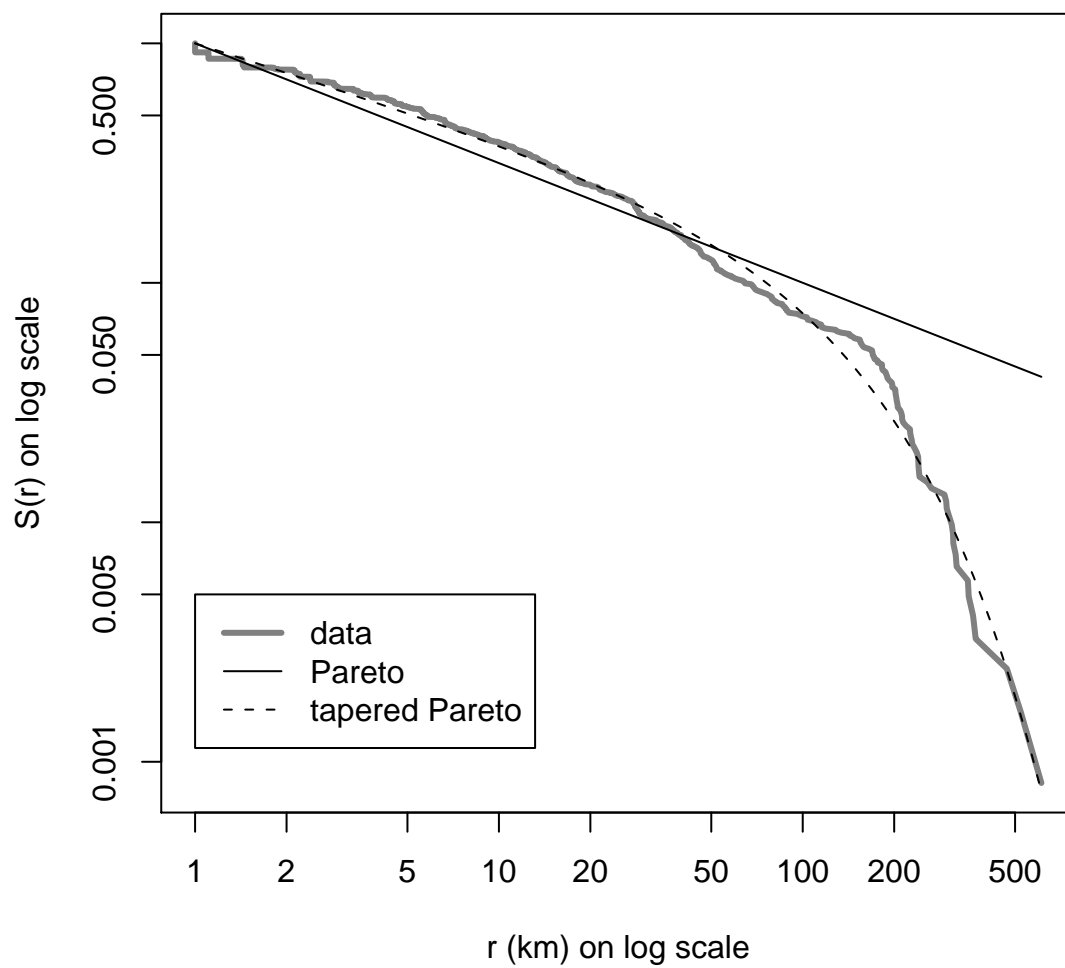


Figure 3:

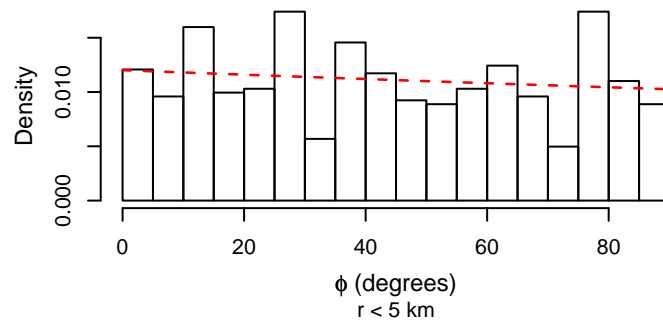
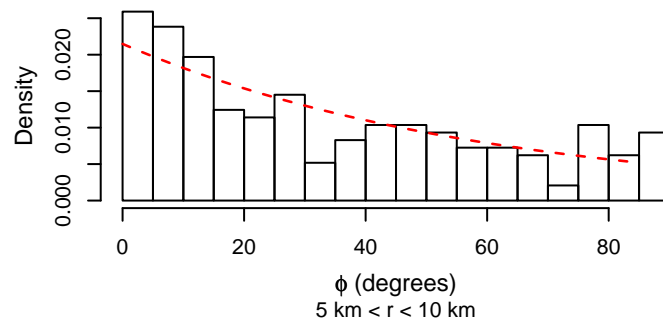
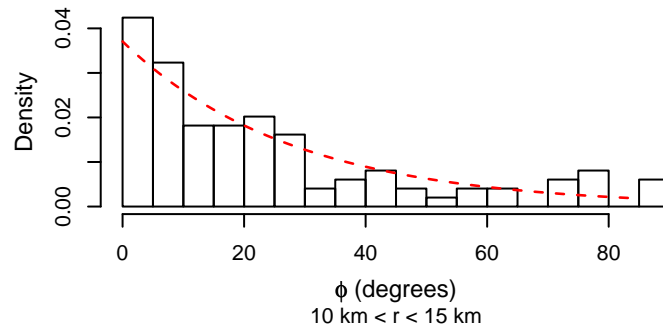


Figure 4:

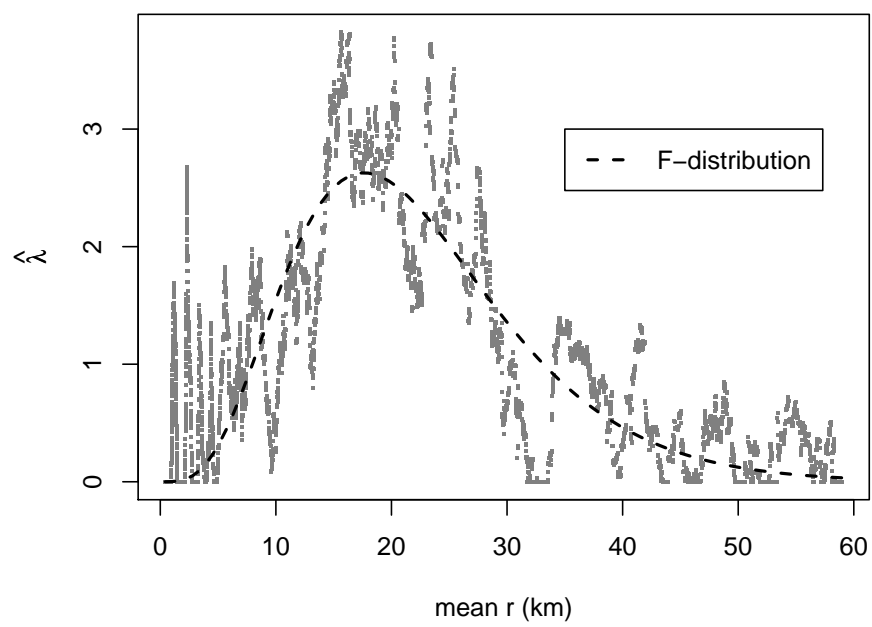


Figure 5:

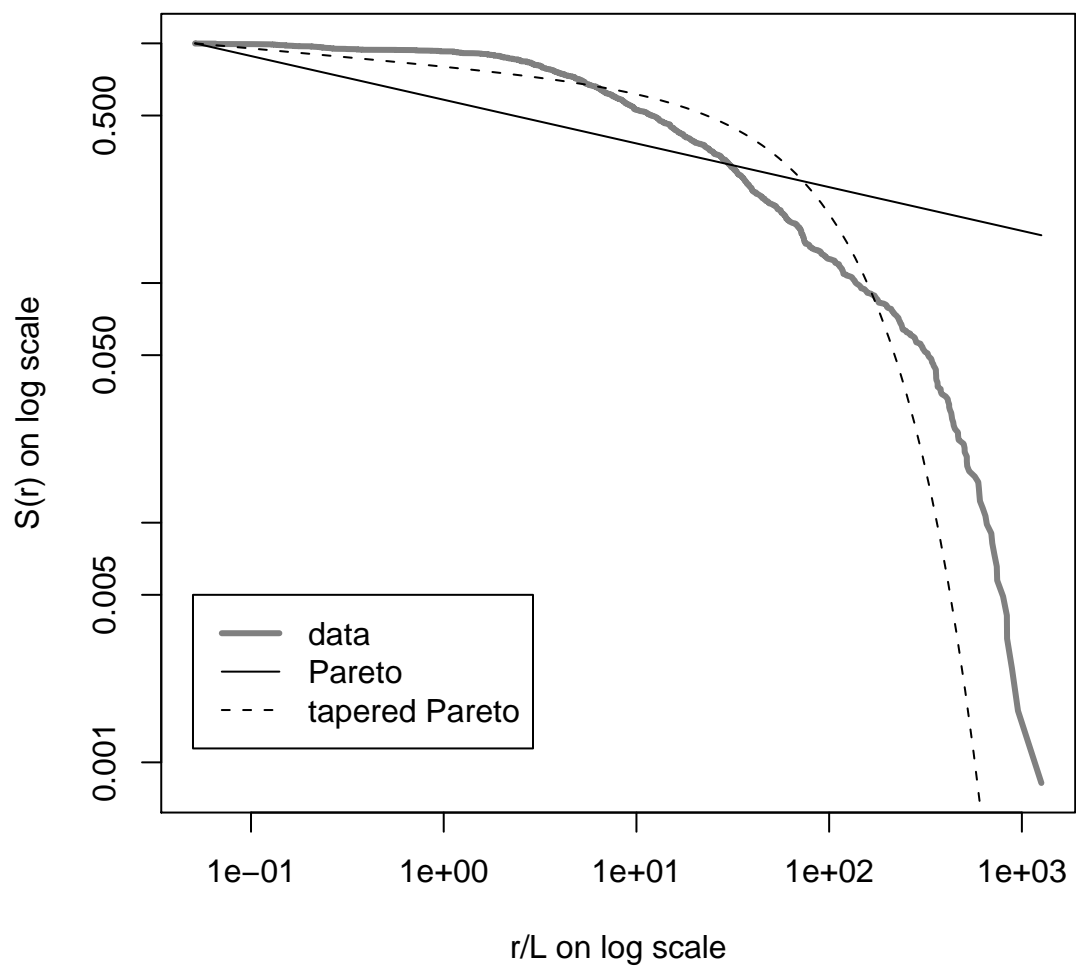


Figure 6:

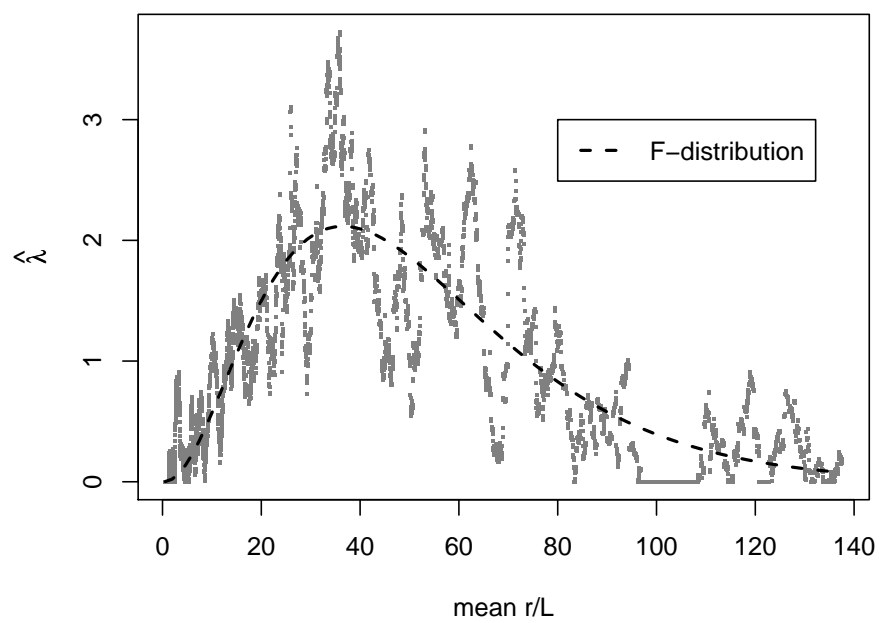


Figure 7:

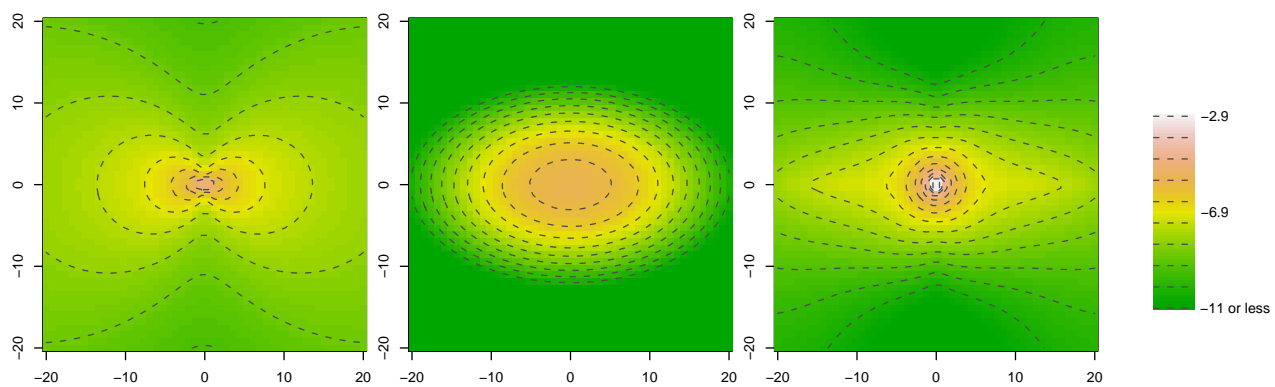


Figure 8:

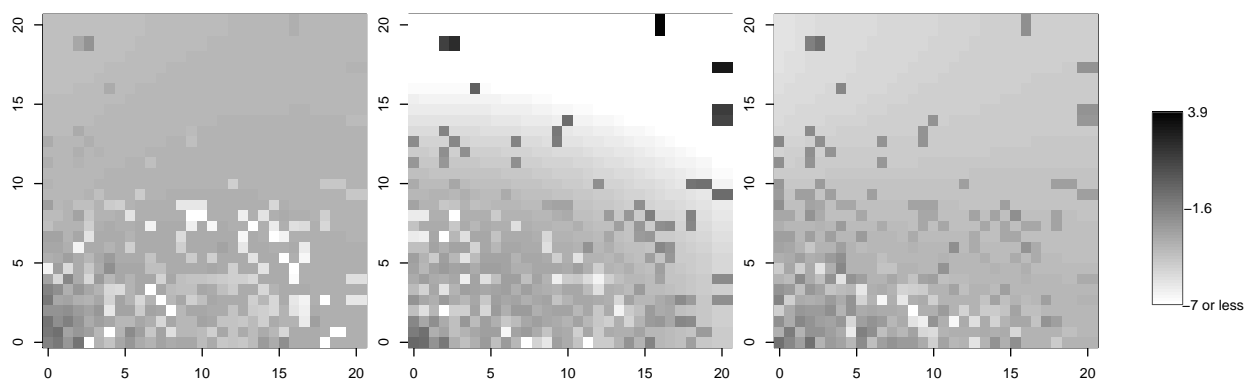


Figure 9:

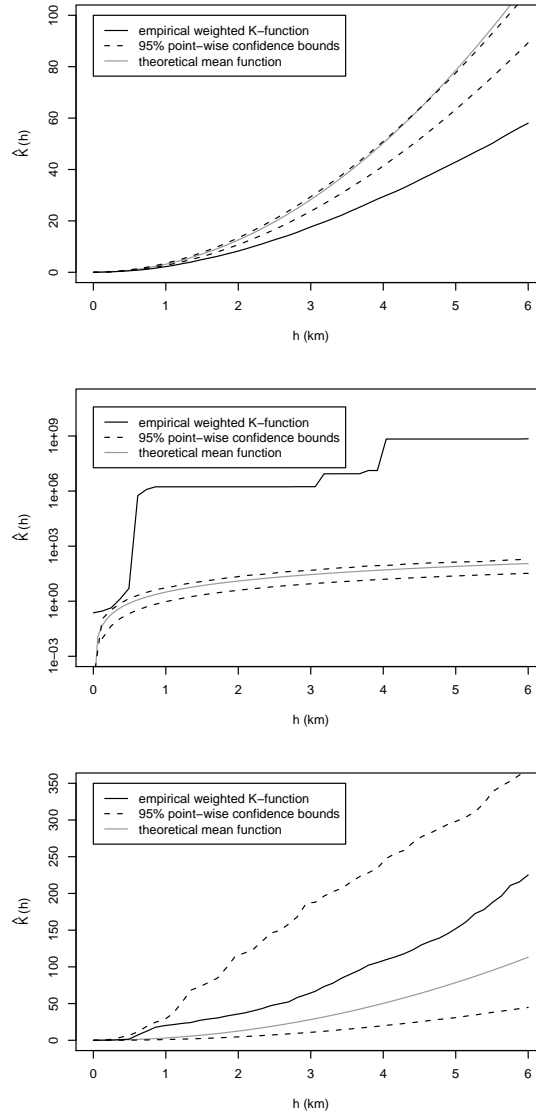


Figure 10: