Title
Multi-dimensional Residual Analysis of Point Models for Earthquake Occurrences

Author
Schoenberg, Frederic P.

Publication Date
2003

Multi-dimensional residual analysis of point process models for earthquake occurrences.

Frederic Paik Schoenberg

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

phone: 310-794-5193

fax: 310-472-3984

email: frederic@stat.ucla.edu

Postal address: UCLA Dept. of Statistics

8142 Math-Science Building

Los Angeles, CA 90095–1554, USA.

Abstract

Methods of examining the fit of multi-dimensional point process models using residual analysis are proposed. One method involves rescaled residuals, obtained by transforming points along one coordinate to form a homogeneous Poisson process inside a random, irregular boundary. Both vertical and horizontal forms of this rescaling are discussed. We also present a different method of residual analysis, involving thinning the point process according to the conditional intensity to form a homogeneous Poisson process on the original, untransformed space. These methods for assessing goodness-of-fit are applied to point process models for the space-time-magnitude distribution of earthquake occurrences, using in particular the multi-dimensional version of Ogata's epidemic-type aftershock sequence (ETAS) model and a 30-year catalog of 580 earthquakes occurring in Bear Valley, California, as an example. The thinned residuals suggest that the fit of the model may be significantly improved by using an anisotropic spatial distance function in the estimation of the spatially varying background rate. Using rescaled residuals, it is shown that the temporal-magnitude distribution of aftershock activity is not separable, and that in particular, in contrast to the ETAS model, the triggering density of earthquakes appears to depend on the magnitude of the secondary events. The residual analysis highlights that the fit of the space-time ETAS model may be substantially improved by allowing the parameters governing the triggering density to vary for earthquakes of different magnitudes. Such modifications are important since the ETAS model is widely used in seismology for hazard analysis.

Key words: conditional intensity, spatial-temporal marked point process, ETAS model, seismology, separability.

1 Introduction.

Stochastic point process models for the space-time-magnitude distribution of earthquake occurrences have become essential components in the assessment of seismic hazard or risk, which are in turn critical for many purposes including civil engineering and insurance. Excellent reviews are provided in Vere-Jones (1970), Vere-Jones (1975), Brillinger (1982) and Brillinger (1993).

Unfortunately, the critical assessment of such models has been scant. An important exception is the seminal work of Ogata (1988), who used one-dimensional residual analysis to assess the temporal component of the Epidemic-Type Aftershock Sequence (ETAS) model, which has subsequently become very widely used in seismology. The ETAS model examined in Ogata (1988) incorporates time and magnitude, but has no spatial component. In recent years many models characterizing the space-time-magnitude behavior of earthquake occurrences have been offered, but little attention has been paid to the examination of the goodness-of-fit of such models. The aim of the current paper is to present some methods for assessing the fit of such multi-dimensional point process models using residual analysis and to show how these methods may be used to identify defects in models and to suggest ways in which models may be improved.

For many statistical models such as ordinary regression models, examination of residuals is straightforward and widely considered standard practice. For the case of point processes, consideration of residuals is a bit trickier, since mere subtraction of the mean from a point process does not result in a very useful diagnostic. Various forms of residual analysis are reviewed here, including vertical and horizontal rescaling and thinned residuals. Following a

review of some multidimensional point process models for earthquake occurrences in Section 2, these methods of residual analysis are discussed in Section 3. The methods are then applied in Section 4 to a sample earthquake catalog, resulting in some novel seismological observations and some suggested revisions for the models in Section 2. Section 5 contains some concluding remarks.

2 Space-time-magnitude point process models for earthquakes

Catalogs of earthquake occurrences are conveniently modeled as spatial-temporal marked point processes, i.e. as random measures on a portion S of space-time-magnitude, taking values in the non-negative integers $\mathbf{Z}^+$ (or infinity), and finite on any compact subset of S. One typically assumes the process to be simple, i.e. that with probability one all the points are distinct. As any such process N may be uniquely characterized by its associated conditional rate process λ (Fishman and Snyder 1976; Daley and Vere-Jones 1988), in modeling the process N one typically prescribes a model for λ.

For S a collection of times t, spatial locations $\mathbf{x}$, and magnitudes m, $\lambda(t, \mathbf{x}, z)$ may be thought of as the frequency with which events are expected to occur around the particular location and time and magnitude $(t, \mathbf{x}, m)$ in S, conditional on the entire prior history H_t of the point process up to time t. Though various types of conditional rate exist corresponding to conditioning on different forms of histories (see e.g. Merzbach and Nualart 1986), for most spatial-temporal marked point processes the prior temporal history H_t is the most

relevant for applications (Schoenberg et al. 2002), so we restrict our attention to this type of conditional rate in what follows.

Formally, a version of the conditional rate $\lambda(t, \mathbf{x}, z)$ associated with a spatial-temporal point process N may be defined as the limiting conditional expectation

$$\lim_{\Delta t, \Delta \mathbf{x}, \Delta m \downarrow 0} \frac{E[N\{(t, t+\Delta t) \times (\mathbf{x}, \mathbf{x}+\Delta \mathbf{x}) \times (m, m+\Delta m)\}|H_t]}{\Delta t \Delta \mathbf{x} \Delta m},$$

provided the limit exists. See Jacod (1975) for conditions for existence and uniqueness of the conditional rate and its integrated form called the *compensator*. Note that some authors instead define λ as

$$\lim_{\Delta t, \Delta \mathbf{x}, \Delta m \downarrow 0} \frac{P[N\{(t, t+\Delta t) \times (\mathbf{x}, \mathbf{x}+\Delta \mathbf{x}) \times (m, m+\Delta m)\} > 0|H_t]}{\Delta t \Delta \mathbf{x} \Delta m}.$$

For orderly point processes (processes where $\lim_{|A| \downarrow \emptyset} P\{N(A) > 1\}/|A| = 0$ for interval $A \subset S$), the two definitions are equivalent. Although in the statistical literature (e.g. Daley and Vere-Jones 1988; Karr 1991), λ is commonly referred to as the conditional *intensity* rather than conditional *rate*, to avoid confusion the term rate may be preferred in this context, since the term intensity is also used in seismology to describe the destructiveness of an earthquake.

In general, $\lambda(t, \mathbf{x}, m)$ depends not only on $t, \mathbf{x}, m$ but also on the times and locations of preceding events. In the simple case where $\lambda(t, \mathbf{x}, m)$ depends only on t, $\mathbf{x}$, and m, however, N is called a Poisson process, and if λ is constant over the entire space S then N is called a homogeneous Poisson process. For a thorough treatment of point processes, conditional intensities and their properties, see Daley and Vere-Jones, 1988.

There are numerous spatial-temporal-magnitude models for earthquake occurrences; for reviews see Kagan (1991), Vere-Jones (1992), Rathbun (1993), Kagan and Vere-Jones (1996),

and especially Ogata (1998). A widely used class of models for earthquake occurrences is the epidemic-type aftershock sequence (ETAS) model introduced by Ogata (1988), which as described in Ogata (1998) is a multi-dimensional extension of Hawkes' self-exciting point process (Hawkes 1971).

The space-time-magnitude ETAS model is based on the principle that earthquakes have aftershocks nearby in space-time and are therefore clustered, and furthermore these aftershocks can have aftershocks as well. Given a collection of points $\{t_i, \mathbf{x}_i, m_i\}, i = 1, 2, \ldots$, the model may be written as

$$\begin{aligned} \lambda(t, \mathbf{x}, m) &= f(m) \left[\nu(\mathbf{x}) + \int_0^t \int_{\mathbf{x}} \int_{m_0}^{m_1} g(t - t', ||\mathbf{x} - \mathbf{x}'||, m') dm' d\mathbf{x}' dt' \right] \\ &= f(m) \left[\nu(x, y) + \sum_{i: t_i < t} g(t - t_i, ||\mathbf{x} - \mathbf{x}_i||, m_i) \right]. \end{aligned} \tag{1}$$

The functions f and ν govern the magnitude-frequency and background seismicity rate, respectively, and the function g, called the *triggering density*, describes how the rate of earthquakes increases after an earthquake, and how this increase in seismicity decays over time and space, and as a function of the magnitude of the triggering event. There are various forms for f, g, and ν suggested in Ogata (1998) based on well-known seismological relations, including the Gutenberg-Richter law describing the frequency of earthquakes of different magnitudes, and the modified Omori law governing the temporal behavior of aftershock activity. For instance, one typically takes for the magnitude-frequency term

$$f(m) \propto exp\{-\beta(m - m_0)\}, \tag{2}$$

where m_0 is the minimal magnitude threshold for the earthquake catalog, in agreement with the Gutenberg-Richter relation (Gutenberg and Richter 1944). For the case where the spatial

coordinate $\mathbf{x}$ represents epicentral origin location (x, y) in the plane, forms for the triggering density g given in Ogata (1998) include

$$g(t, x, y, m) = \frac{K_0 exp\{\alpha(m - m_0)\}}{(t + c)^p (x^2 + y^2 + d)^q}, \tag{3}$$

$$g(t, x, y, m) = \frac{K_0}{(t + c)^p} exp\left\{-\frac{x^2 + y^2}{2dexp\{\alpha(m - m_0)\}}\right\},$$

and

$$g(t, x, y, m) = \frac{K_0}{(t + c)^p} \left(\frac{x^2 + y^2}{exp\{\alpha(m - m_0)\}} + d\right)^{-q}.$$

The temporal component in these three forms of g agree with the the modified Omori law (Utsu 1971). Functional forms for the background rate $\nu(\mathbf{x})$ are not typically given; instead ν is assumed constant or estimated by smoothing the main events in the catalog, e.g. using kernel smoothing or bi-cubic B-splines (Ogata 1998; Zhuang et al. 2002).

Note that the model (1) is separable with respect to magnitude and the spatial-temporal coordinates, in the sense that the frequency of earthquakes of magnitude m at any time and location depends only on the overall density of earthquakes of magnitude m given by $f(m)$ and on the overall rate of events dictated by ν and g, but not on the interaction of the two. That is, while the triggering density g depends on the magnitude of the *triggering* event, it does not depend on the magnitude of the event m whose rate is being calculated in the formula for $\lambda(t, \mathbf{x}, m)$. This separability property is an important feature of the model (1) that is explored in Section 4 below.

3 Residual analysis

Given a space-time-magnitude point process model for earthquakes, one typically selects the parameter vector θ for the model by maximizing the log-likelihood function

$$L(\theta) = \int_S log[\lambda(\theta)]dN - \int_S \lambda(\theta)d\mu,$$

and selection between models is often done simply by minimizing some information criterion such as the Akaike Information Criterion or Bayesian Information Criterion (see e.g. Ogata and Tanemura 1984; Ogata 1998; Lu et al. 1999; Schoenberg and Bolt 2000).

Although such likelihood criteria are useful for comparing the relative fit of competing models, they are not useful for assessing the absolute goodness-of-fit of models or for highlighting ways in which models may be improved. For such purposes, residual analysis may be of service.

Point process residual analysis involves rescaling or thinning the original point process to obtain a new point process that is homogeneous Poisson. A basic method for forming residual point processes by rescaling is due to Meyer (1971), who showed that for a collection $\{N_i\}$ of simple univariate (i.e. purely temporal) point processes on the real half-line such that no two processes share a point at the same time, if one rescales the points by moving each point (t, i) to the point $(\int_0^t \lambda(t', i)dt', i)$, then one obtains a sequence of independent homogeneous Poisson processes, each having unit rate. Alternative proofs and variations of this result have been given by Papangelou (1972), Brémaud (1972), Aalen and Hoem (1978), Brown and Nair (1986), and Kallenberg (1990). Papangelou (1972) gave the following interpretation in the univariate case: "Roughly, moving in $[0, \infty)$ so as to meet expected future points at a rate of one per time unit (given at each instant complete knowledge of the past), we meet them

at the times of a Poisson process." Meyer's method is referred to as *horizontal* rescaling, since for each point, only its temporal (or horizontal) coordinate is changed. This type of rescaling was shown to be extremely useful by Ogata (1988), who performed residual analysis of an ETAS model for the temporal-magnitude (i.e. non-spatial) behavior of earthquakes. In what follows, we focus on the multi-dimensional case. Meyer's result extends quite readily to spatial-temporal marked point processes; i.e. by transforming each point $(t_i, \mathbf{x}_i, m_i)$ to $(\int_0^t \lambda(t', i)dt', \mathbf{x}_i, m_i)$, one again obtains an independent sequence of homogeneous Poisson processes of unit rate, as described in Vere-Jones and Schoenberg (2002).

Alternatively, one may choose to rescale the points vertically, i.e. to focus on one non-temporal dimension of the point process and rescale each point along that dimension according to the conditional intensity λ. For simple point processes on $\mathbf{R}^d$, Schoenberg (1999) generalized results of Merzbach and Nualart (1986) and Nair (1990) to show that such transformation, moving e.g. $(t_i, \mathbf{x}_i, m_i)$ to $(t_i, \mathbf{x}_i, \int_0^{m_i} \lambda(t_i, \mathbf{x}_i, m)dm)$ again results in a homogeneous Poisson process of rate one. When the original process is observed within a window such as $[t_0, t_1] \times [x_0, x_1] \times [y_0, y_1] \times [m_0, m_1]$, for example, the transformed process is a homogeneous unit-rate Poisson process within a boundary of irregular shape (see Schoenberg 1999). As with horizontally rescaled residuals, one may assess the model by examining the residuals to see if they appear homogeneous throughout the space. There are a wide variety of tests for this purpose; see e.g. Diggle (1979), Ripley (1979), Lawson (1988), Heinrich (1991), Andersen et al. (1993) and references therein for examples.

One potential drawback to vertical and horizontal rescaling is that inhomogeneity in the residual process with respect to the transformed coordinates may be difficult to interpret,

particularly when one spatial coordinate is transformed. Thus, instead of rescaling spatial coordinates, a different method for obtaining a homogeneous Poisson residual process is by thinning, using a variation on the useful simulation technique of Lewis and Shedler (1979) and Ogata (1981). One may obtain a homogeneous Poisson process with rate b by keeping each point (t_i, x_i, y_i, m_i) in the original point process independently with probability $b/\lambda(t_i, x_i, y_i, m_i)$, where b is the minimum of $\lambda(t, x, y, m)$ over the entire observation region, S. Like rescaled residuals, the resulting thinned residuals may be examined for uniformity versus various alternatives such as trend, clustering, etc. In addition, because of the stochastic component in the generation of the thinned residuals, one may repeatedly generate many realizations of thinned residuals, each distributed according to a homogeneous Poisson process with rate b. (Note that these individual thinned Poisson processes are not independent of one another, however.)

If the minimal rate b is very small, each realization of thinned residuals may contain too few points to allow for testing for homogeneity. In such cases, one may modify the thinning procedure above and instead obtain an approximation of the thinned residuals by generating some number k of residual points by randomly selecting k of the original $N(S)$ points such that the point (t_i, x_i, y_i, m_i) is chosen with probability $k\lambda(t_i, x_i, y_i, m_i)^{-1}/(\sum_{i=1}^{N(S)} \lambda(t_i, x_i, y_i, m_i)^{-1})$. Provided k is small relative to $N(S)$, the distribution of the resulting residuals will closely approximate that of ordinary thinned residuals.

4 Application to Bear Valley Data

Our sample earthquake catalog consists of 580 earthquakes of magnitude 3.0 and higher (up to magnitude 5.5) occuring along a 35 kilometer portion of the San Andreas Fault around Bear Valley, California (latitude 36.5 to 37.0, longitude -121.5 to -121.0). The data are described in Schoenberg and Bolt (2000), where it is noted that this dataset is typical of a catalog used as a basis for seismic hazard calculations. The catalog was obtained from the Council of the National Seismic System at quake.geo.berkeley.edu and details about the data may be obtained there.

The space-time-magnitude distribution of the Bear Valley earthquakes is shown in Figure 1. The locations are epicentral origin estimates, the magnitude of each event is represented in the Figure by the size of the circle, and the time by its darkness. One sees in Figure 1 that most of the events occur approximately along a narrow strip, which corresponds to a portion of the San Andreas Fault. The time-magnitude distribution of the points is highlighted in Figure 2, where one sees the increased frequency with which the smaller earthquakes occur. No obvious trend in the magnitude distribution over time is easily discernable in Figure 2; this issue of separability of the marginal distributions is investigated further below.

4.1 Thinned spatial residuals

The spatial fit of the model is conveniently investigated using thinned residuals as described in Section 3 above. We consider the model (1) with magnitude distribution (2), trigger density (3), and background rate $\nu(x, y)$ estimated by kernel-smoothing of the larger events (those with magnitude at least 4.0); in what follows we simply refer to this as model (1). After

fitting this model by maximum likelihood to obtain an estimate $\hat{\theta}$ of the parameter vector $\theta = (\beta, K_0, \alpha, c, p, d, q, \rho, a)$, where ρ is the bandwidth in the kernel smoothing for ν and a is the constant of proportionality in (2), the thinned residuals were obtained. That is, the original dataset was thinned by keeping each point (t_i, x_i, y_i, m_i) with probability $b/\hat{\lambda}(t_i, x_i, y_i, m_i)$, where $\hat{\lambda}(t_i, x_i, y_i, m_i) = \lambda(t_i, x_i, y_i, m_i; \hat{\theta})$ and $b = \min_S \hat{\lambda}(t_i, x_i, y_i, m_i)$. Because b is rather small due to small values of λ at certain locations, each realization of thinned residuals consists of only a few points. Figure 3 shows a typical example of thinned residuals based on the model (1).

Examination of one realization for homogeneity is a very low-power method for assessing the model, but one may readily generate many realizations of thinned residuals and examine them collectively. Under the null hypothesis that the model is correctly specified, each realization should be distributed according to a homogeneous Poisson process. A powerful test for the alternative hypothesis that the points are instead over- or under-clustered spatially relative to a Poisson process is the estimated K-function, which indicates the proportion of pairs of points per unit area available that are within distance d of one another. That is, for any spatial distance d, $K(d)$ is estimated as

$$AN^{-2} \sum_{\substack{i<j \\ ||\mathbf{x}_i - \mathbf{x}_j|| < d}} s(\mathbf{x}_i, \mathbf{x}_j),$$

where A is the area of the observation window, N is the number of observed points, and $s(\mathbf{x}_i, \mathbf{x}_j)^{-1}$ is the proportion of area of the ball centered at $\mathbf{x}_i$ passing through $\mathbf{x}_j$ that falls within the observation window (see Ripley 1981). The solid lines in Figure 4 delimit the middle 95%-range of the estimated K-functions for 1000 realizations of thinned residuals based on model (1). For comparison, the shaded area in Figure 4 shows the corresponding

middle 95%-range of estimated K-functions for 1000 realizations of points uniformly spread over the spatial region. That is, for each realization j consisting of n_j thinned residual points, a realization of n_j uniformly distributed points is generated, and the shaded region in Figure 4 shows the range of the estimated K-functions for these uniformly distributed points.

From Figure 4 one can see that the two sets of realizations seem in general agreement as far as the range of K-function values are considered. However, a noticeable discrepancy is that for small values of d, the upper 95% bound for the thinned residuals is higher than that for the uniformly distributed points; this indicates excessive spatial clustering in the thinned residuals. In other words, Figure 4 indicates spatial clustering in the original dataset that is not adequately accounted for in the model (1). One may suspect that the source of this clustering is the tendency for the points in Figure 1 to fall near the fault line, as previously mentioned. In fitting the model (1), the estimated background rate $\hat{\nu}(x, y)$ is determined by kernel-smoothing the larger earthquakes in the dataset, hence locations at greater Euclidean distances from larger earthquakes have lower estimated background rate. The estimated background rate, corresponding to the maximum likelihood estimation of the parameters in the model (1), is shown in Figure 5 along with the larger earthquakes used for the smoothing. It should be noted that kernel smoothing using variable bandwidths, as proposed in Zhuang et al. (2002), does not solve the problem of residual clustering in Figure 4; in fact, application of the method of Zhuang et al. (2002) still results in an overly smoothed background similar to that in Figure 5, since the kernel smoothes are still isotropic and have even larger bandwidths around the outlying points.

The obvious pattern in Figure 1 and the clustering in the residuals as observed in Figure

4 suggest modifying the estimation of $\nu(x, y)$ in model (1) by using a non-isotropic distance function that takes into account the approximate colinearity of most of the earthquake locations. For instance, one may simply fit a line ℓ (e.g. by regression) to the locations of the larger earthquakes and then replace Euclidean distance in the smoothing procedure for $\nu(x, y)$ with a new distance function such that distances between two points orthogonal to ℓ are weighted γ times the distance in the direction of ℓ. Like the other model parameters, the parameter γ may be estimated by maximum likelihood. The resulting estimate of the background rate $\nu(x, y)$ is shown in Figure 6.

The smoothing in Figure 6 is not nearly as smooth as that in Figure 5, and clearly corresponds much more closely with the spatial distribution of the actual events. Using the modified estimate of the background rate and the maximum likelihood estimates of the other parameters in the model (1), one may reconstruct numerous realizations of thinned residuals and their corresponding K-functions. After using the modified, anisotropic estimate of $\nu(x, y)$, the value b, defined as the minimum value of $\hat{\lambda}$ over the entire space, is now miniscule, so each realization of thinned residuals contains only very few points (indeed, on average less than one). Thus for the purposes of comparison with Figure 4, approximate residuals for the modified model, hereafter referred to as model (1'), are obtained according to the procedure described in the previous Section, with the number of points per residual process a Poisson random variable with mean equal to the mean number of residual points when thinning according to the original model (1). The 95% bounds for these K-functions corresponding to 1000 realizations of the thinned residuals for the modified model (1') are shown in Figure 7, along with corresponding 95% bounds based on uniformly distributed

points, just as in Figure 4. The bounds in Figure 7 agree somewhat more closely with those of the uniformly distributed points, and a good portion of the clustering indicated in Figure 4 is no longer indicated.

4.2 Rescaled temporal-magnitude residuals

The time-magnitude distribution of the earthquake process may be investigated using rescaled residuals. Figure 8 presents the residuals for the model (1'), integrated over all spatial locations, using vertical rescaling. Because of the epidemic-type nature of the model, the conditional rate λ is extremely high immediately following large events and decays quickly thereafter. Thus it is extremely difficult to discern from Figure 8 whether the points are uniformly scattered within this very irregularly-shaped boundary, even if the points are plotted on logarithmic scale as in Figure 9. In addition, estimates of clustering statistics such as the K-function based on these vertically rescaled residuals are largely dominated by boundary effects.

A convenient alternative is to rescale the points horizontally, i.e. to fix the magnitude of each point and rescale its temporal coordinate. The resulting horizontally rescaled residuals are plotted in Figure 10. Once again the boundary is somewhat irregular, though not nearly as much so as in Figure 8.

The residuals in Figure 10 appear to be relatively uniformly dispersed, though deviations are difficult to discern. In order to investigate further, one may inspect the residual points corresponding to each magnitude value, i.e. the residuals along each horizontal line in Figure 10. If the model is correctly specified, then the points on each line should be distributed as

independent Poisson processes. A natural alternative hypothesis is that points on certain lines are clustered; this corresponds to the notion that the aftershock triggering process is not separable with respect to time and magnitude as specified in equation (1). That is, for certain times, the proportion of earthquakes in certain magnitude ranges may be higher than this proportion at other times. This would result in clustering in the horizontally rescaled residuals within these magnitude ranges.

To test for nonseparability (clustering), one can examine the second-order properties of the residuals along each horizontal line. For example, one may estimate the second moment, or one-dimensional analog of the K-function of Ripley (1981). For the residual process, one may estimate the standardized second moment for each magnitude m and each transformed time-lag u as the quantity $(k_1 - k_2)/\sqrt{(}k_2)$, where k_1 is the number of pairs of residual points with magnitude m whose transformed times are within u of one another, and k_2 is the the expected number of such pairs for a homogeneous Poisson process. Inspection of the estimated standardized second moment of the residuals for each magnitude reveals that the model (1') fits extremely well for most magnitudes, especially magnitudes greater than 3.5. However, there is excessive clustering at small distances for magnitudes between 3.1 and 3.25, inclusive; this clustering is not evident, however, for magnitudes smaller than 3.1. Figure 11 displays the estimated standardized second moment functions, along with 95% bounds for the Poisson process, for the residuals of several different magnitudes.

Figure 11 suggests that the residuals in certain magnitude ranges are overly clustered; hence the assumption of temporal-magnitude separability in model (1') appears to be invalidated. To remedy this, we consider a variant of model (1') such that the parameter vector

θ is permitted to vary with magnitude, m. That is, we fit the model

$$\lambda(t,x,y,m) = \tag{4}$$
$$\sum_{j=1}^{J} \mathbf{I}_{\{m\in M_j\}} a_j exp\{-\beta_j(m-m_0)\} \left[\nu(x,y) + \sum_{i:t_i<t} \frac{K_j exp\{\alpha_j(m_i-m_0)\}}{(t-t_i+c_j)^{p_j}\{(x-x_i)^2+(y-y_i)^2+d_j\}^{q_j}}\right]$$

where $\{M_j; j = 1, 2, \ldots, J\}$ is a partition of the observed magnitude range, and $\nu(x,y)$ is obtained by anisotropic kernel smoothing of the larger events as in (1'). Note that this revised model (4) is not inconsistent with the modified Omori law in (3), which governs the decay of aftershocks over time for *all* magnitudes in the observed range.

The horizontally rescaled residuals obtained after fitting the revised model (4) by maximum likelihood are displayed in Figure 12, and the estimated standardized second moments of the residuals for various magnitudes are displayed in Figure 13. In fitting (4), J is set to 2, M_1 to the magnitude range $[3.1, 3.25]$, and M_2 to the set of all other magnitudes in the observed range $[3, 5.5]$. One sees that after fitting the model (4), the clustering in the residuals for magnitudes 3.1 and 3.2 is no longer present, and in general the residuals appear to be scattered quite uniformly across the transformed boundary. The modified form (4) apparently provides an adequate description of the clustering in the original dataset.

5 Summary.

The residual analysis methods discussed here may be used to supplement likelihood criteria such as the AIC in assessing the goodness-of-fit for multi-dimensional point process models. Unlike likelihood criteria, examination of thinned and rescaled residuals enables hypothesis testing against various alternatives such as clustering and trend, as well as graphical depiction

of when and where models appear to fit well or poorly.

Thinned residuals appear to be especially useful for examining the spatial components of spatial-temporal marked point process models, since such residuals do not require transformation of the spatial coordinates. For epidemic-type processes such as the ETAS model for earthquake occurrences, horizontally rescaled residuals may be preferred over vertically rescaled residuals, since in the latter case the residuals are observed in a very highly irregular boundary due to the high volatility of the conditional rate function over time. By comparison, the conditional rate tends not to vary as wildly with magnitude; hence the horizontally rescaled residuals are observed within a much smoother and more regular boundary, facilitating their direct examination by eye as well as the estimation of statistical properties such as the second moment or higher-order properties of the residuals.

In the seismological application here it was demonstrated that anisotropic kernel smoothing enables better estimation of the spatially varying background rate $\nu(x, y)$ in the space-time-magnitude ETAS model when applied to earthquakes in Bear Valley, California. This is no great surprise because of the apparent fault structure in this region, which is accounted for in the anisotropic smoothing. More surprising, however, is the result that the assumption of separability of the magnitude distribution, assumed not only in the ETAS model but in all or at least nearly all existing models for earthquake occurrences, appears to be invalidated in this case. It is shown that for earthquakes in certain magnitude ranges (here, magnitude 3.1 to 3.25 or so) the previous triggering events appear to affect the rate of earthquakes in this magnitude range differently than for other events. Hence, in future use of models such as ETAS for purposes of seismic hazard estimation and other uses, it may be advisable to

allow the model parameters to vary across different magnitude scales.

Our conclusions are not seriously affected by the problem of boundary effects, since in each case our estimates of second moment properties are compared with simulated Poisson processes observed on identical boundaries. However, one potential issue deserving further consideration is the problem of artifactual regularity of the residuals when the model parameters are estimated rather than known. Indeed, if the model parameters are known, then the residual processes discussed here are distributed exactly as homogeneous Poisson processes, but if the parameters are estimated, then the residuals are typically slightly more regular than Poisson; see e.g. Schoenberg (2002) for details. Though some authors (e.g. Khamaladze 1988, Heinrich 1991, Andersen et al. 1993) discuss the role of estimation on the distributions of certain test statistics for point processes, the extent to which the regularity of residuals, as well as dependence of successive iterations of thinned residuals, affects statistics such as the K-function of the residuals is an important topic for future consideration.

Acknowledgements. This material is based upon work supported by the National Science Foundation under Grant No. 9978318. This research benefitted greatly from useful discussions with David Brillinger, David Vere-Jones, and Pierre Brémaud. Thanks to Robert Uhrhammer, the Northern California Earthquake Data Center (NCEDC), the Northern California Seismic Network, U.S. Geological Survey, Menlo Park, and the Berkeley Seismological Laboratory, University of California, Berkeley, for generosity in organizing and sharing earthquake catalogs.

References.

Aalen, O. and Hoem, J. (1978), "Random Time Changes for Multivariate Counting Processes," *Scandinavian Actuarial Journal*, 81–101.

Andersen, P., Borgan, O., Gill, R., and Keiding, N. (1993), *Statistical Models Based on Counting Processes,* New York: Springer.

Brémaud, P.M. (1972), "A Martingale Approach to Point Processes," Memorandum ERL-M345, Electronics Research Laboratory, University of California, Berkeley, Berkeley, CA.

Brillinger, D.R. (1982), "Seismic Risk Assessment: Some Statistical Aspects," *Earthquake Prediction Research*, 1, 183–195.

Brillinger, D.R. (1993), "Earthquake Risk and Insurance," *Environmetrics*, 4, 1–21.

Brown, T. and Nair, M. (1988), "A Simple Proof of the Multivariate Random Time Change Theorem for Point Processes," *Journal of Applied Probability*, 25, 210–214.

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

Diggle, P. (1979), "On Parameter Estimation and Goodness-of-fit Testing for Spatial Point Patterns," *Biometrics*, 35, 87-101.

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

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

Heinrich, L. (1991). "Goodness-of-fit Tests for the Second Moment Function of a Stationary Multidimensional Poisson Process," *Statistics*, 22, 245–278.

Jacod, J. (1975), "Multivariate Point Processes: Predictable Projections, Radon-Nikodym Derivatives, Representation of Martingales," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 31, 235-253.

Kagan, Y.Y. (1991), "Likelihood Analysis of Earthquake Catalogs," *Journal of Geophysical Research*, 106, 135–148.

Kagan, Y. and Vere-Jones, D. (1996), "Problems in the Modelling and Statistical Analysis of Earthquakes," in: *Lecture Notes in Statistics (Athens Conference on Applied Probability and Time Series Analysis)*, 114, eds. C.C. Heyde, Yu.V. Prorohov, R. Pyke, and S.T. Rachev, New York: Springer, pp. 398–425.

Kallenberg, O. (1990), "Random Time Change and an Integral Representation for Marked Stopping Times," *Probability Theory*, 86, 167–202.

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

Khamaladze, E. (1988), "An Innovation Approach to Goodness-of-fit Tests in R^m," *Annals of Statistics*, 16, 1503–1516.

Lawson, A. (1988), "On Tests for Spatial Trend in a Non-homogeneous Poisson Process," *Journal of Applied Statistics*, 15, 225–234.

Lewis, P., and Shedler, G. (1979), "Simulation of Nonhomogeneous Poisson Processes by Thinning," *Naval Research Logistics Quarterly*, 26, 403-413.

Lu, C.S., Harte, D., and Bebbington, M. (1999), "A Linked Stress Release Model for Historical Japanese Earthquakes: Coupling among Major Seismic Regions," *Earth Planets Space*, 51, 907-916.

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

Meyer, P. (1971), "Demonstration Simplifiée d'un Théorème de Knight," In *Séminaire de Probabilités* **V**, Université Strasbourg, Lecture Notes in Mathematics, 191, 191–195.

Nair, M. (1990), "Random Space Change for Multiparameter Point Processes," *Annals of Probability*, 18, 1222–1231.

Ogata, Y. (1981), "On Lewis' Simulation Method for Point Processes," *IEEE Transactions on Information Theory*, IT-27, 23-31.

Ogata, Y. (1988), "Statistical Models for Earthquake Occurrences and Residual Analysis for Point Processes," *Journal of the American Statistical Association*, 83, 9-27.

Ogata, Y. (1998), "Space-time Point Process Models for Earthquake Occurrences," *Annals of the Institute of Statistical Mathematics*, 50, 379–402.

Ogata, Y. and M. Tanemura (1984), "Likelihood Analysis for Spatial Point Patterns," *Proceedings of the International Biometric Conference*, 12, 276-286.

Papangelou, F. (1972), "Integrability of Expected Increments of Point Processes and a Related Random Change of Scale," *Transactions of the American Mathematics Society*, 165, 483–506.

Rathbun, S.L. (1993), "Modeling Marked Spatio-temporal Point Patterns," *Bulletin of the International Statistics Institute*, 55, 379–396.

Ripley, B. (1979), "Tests of 'Randomness' for Spatial Point Patterns," *Journal of the Royal Statistical Society, Series B*, 41, 368-374.

Ripley, B. (1981), *Spatial Statistics.* New York: Wiley.

Schoenberg, F. (1999), "Transforming Spatial Point Processes into Poisson Processes," *Stochastic Processes and their Applications*, 81, 155–164.

Schoenberg, F.P. (2002), "On Rescaled Poisson Processes and the Brownian Bridge," *Annals of the Institute of Statistical Mathematics*, 54, 445–457.

Schoenberg, F., and Bolt, B. (2000), "Short-term Exciting, Long-term Correcting Models for Earthquake Catalogs," *Bulletin of the Seismological Society of America*, 90 849–858.

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

Vere-Jones, D. (1970), "Stochastic Models for Earthquake Occurrence," *Journal of the Royal Statistical Society, Series B*, 32, 1–62.

Vere-Jones, D. (1975), "Statistical Aspects of the Estimation of Earthquake Risk," *Australian Journal of Statistics*, 17, 134–147.

Vere-Jones, D. (1995), "Forecasting Earthquakes and Earthquake Risk," *International Journal of Forecasting*, 11, 503–538.

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.

Vere-Jones, D. and Schoenberg, F.P. (2002), "Rescaling Marked Point Processes," *Festschrift for Daryl Daley*, in preparation.

Zhuang, J., Ogata, Y., and Vere-Jones, D. (2002), "Stochastic Declustering of Space-Time Earthquake Occurrences," *Journal of the American Statistical Association*, 97, 369–380.

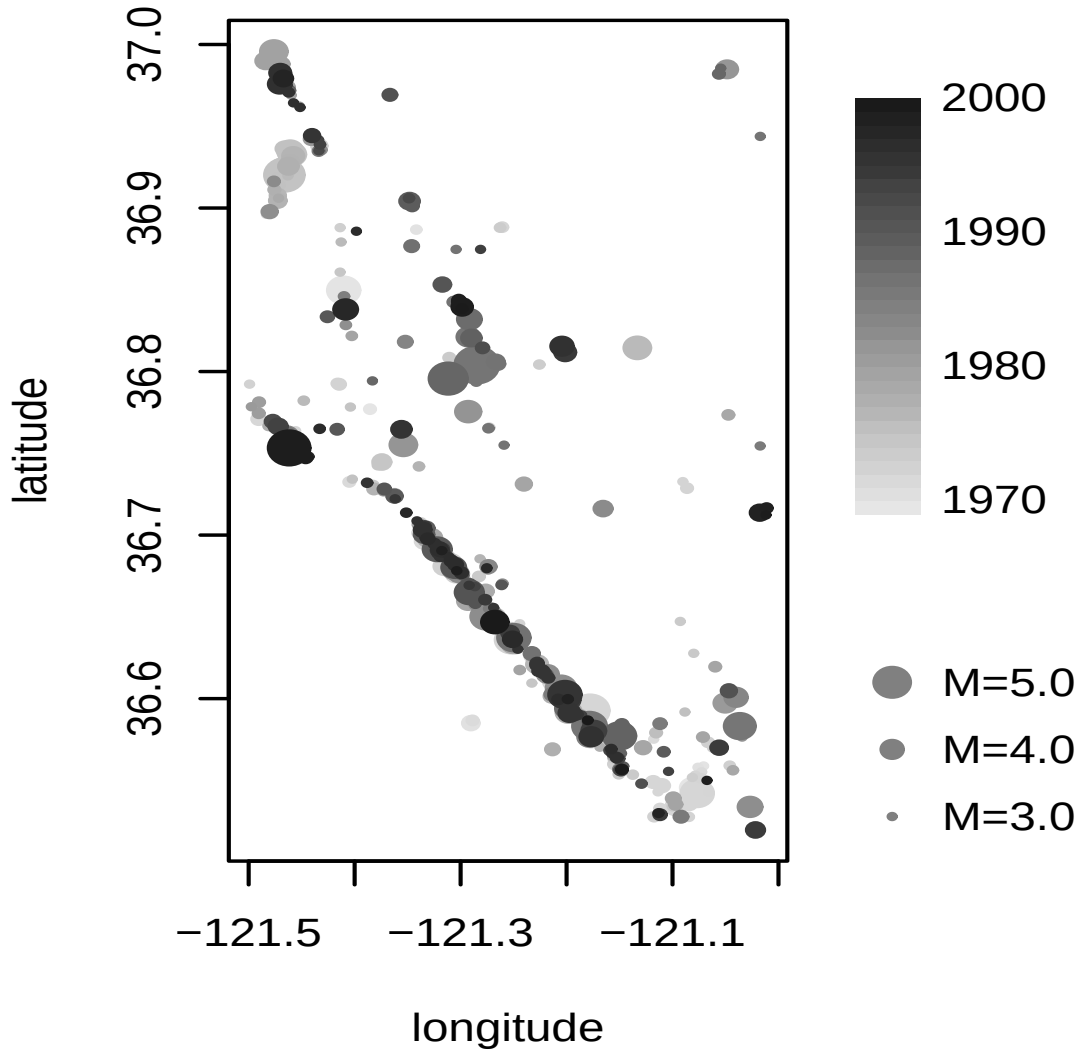

Figure 1: Locations, times and magnitudes of moderate-sized ($M \geq 3.5$) earthquakes in Bear Valley, CA, between 1970 and 2000.

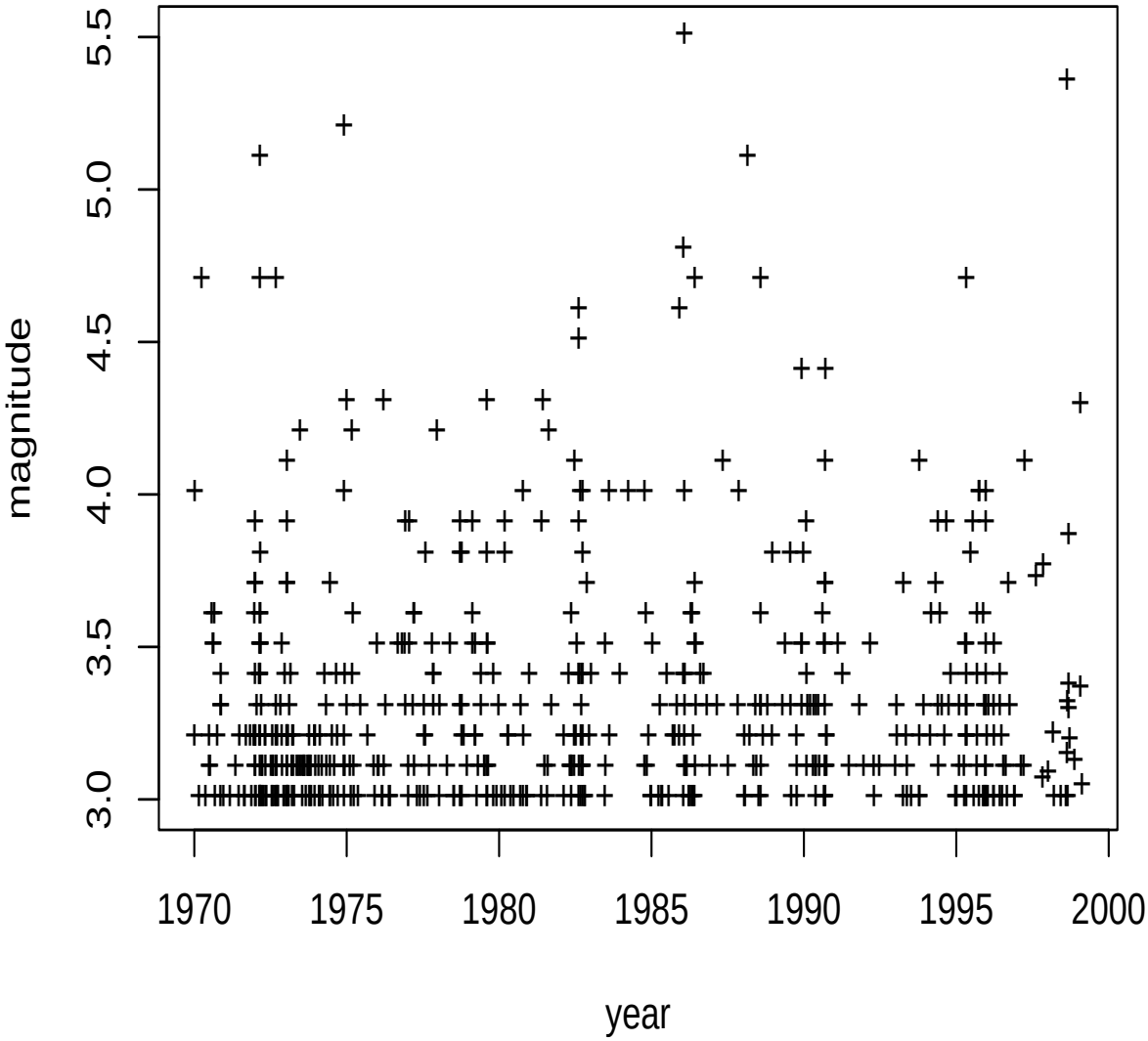

Figure 2: Times and magnitudes of Bear Valley, CA earthquakes.

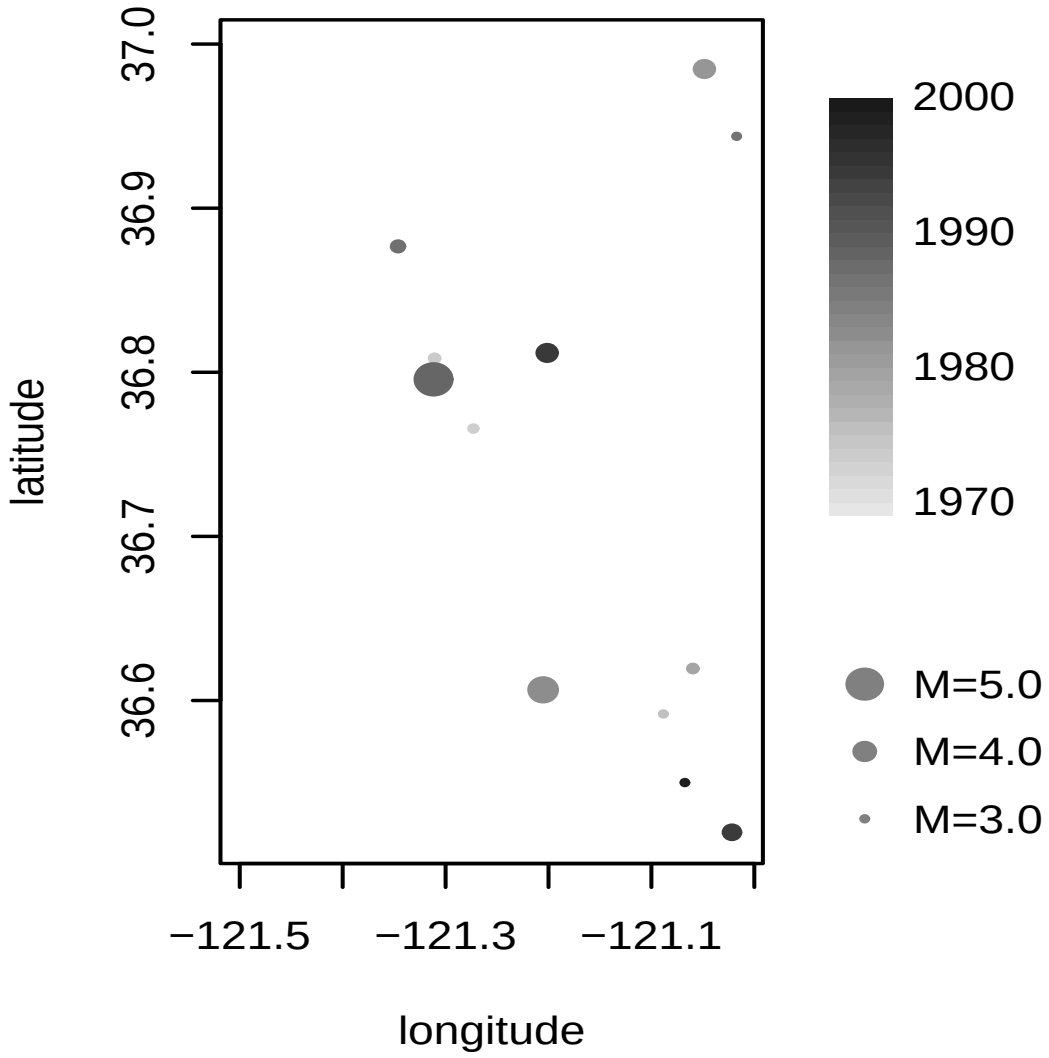

Figure 3: Sample thinned residual process, for model (1).

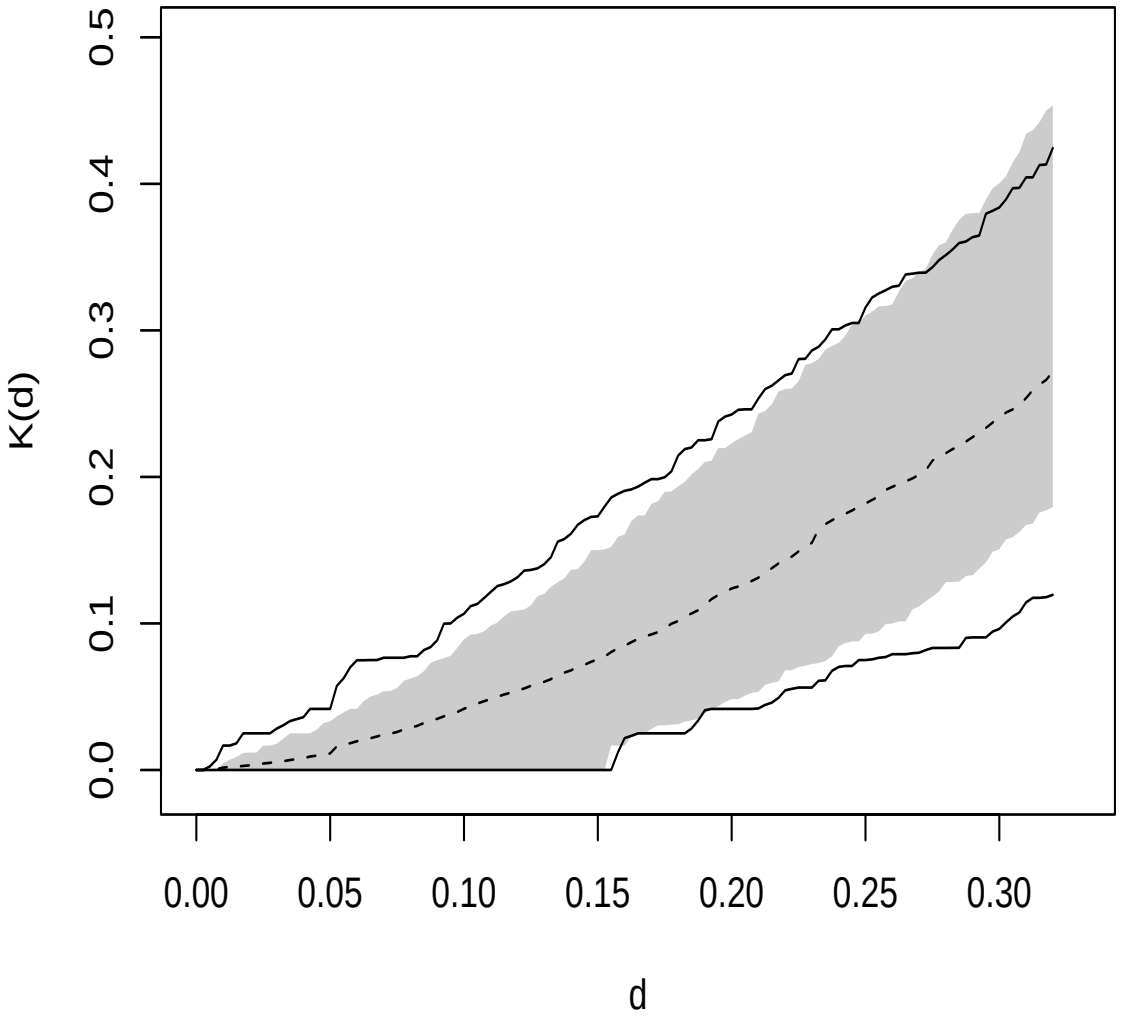

Figure 4: Ranges of estimated K-functions for thinned residuals and for Poisson processes. Solid lines demark middle 95%-range of estimated K-functions for 1000 realizations of thinned residuals based on model (1). Shaded area demarks middle 95%-range of estimated K-functions for 1000 homogeneous Poisson processes on the same space.

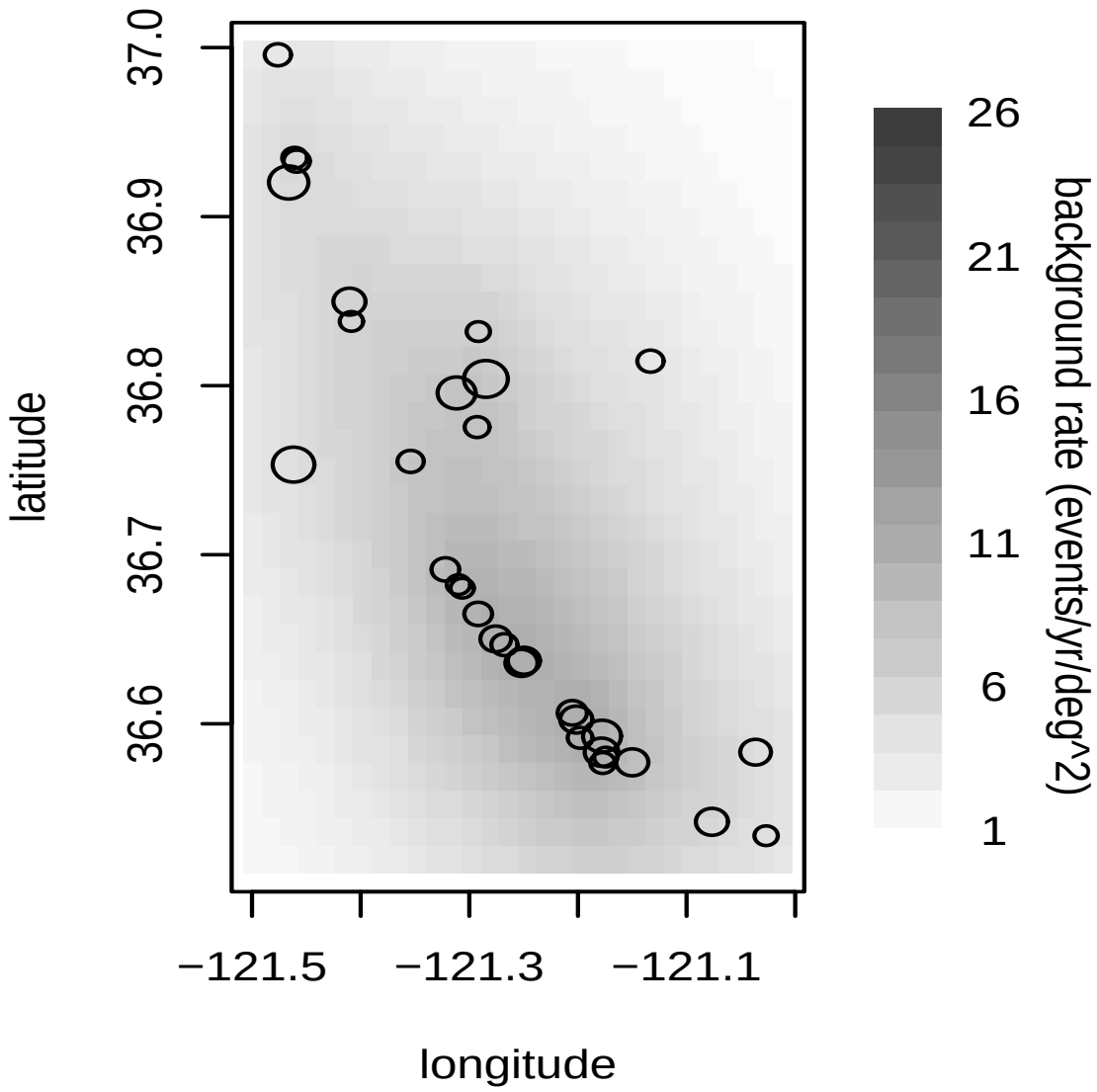

Figure 5: Spatially-varying background rate $\nu(x, y)$, estimated by isotropic kernel smoothing of the $M \geq 4$ events. Circles reperesent $M \geq 4$ events. The magnitude scale corresponding to the size of the circles is same as for Figure 1.

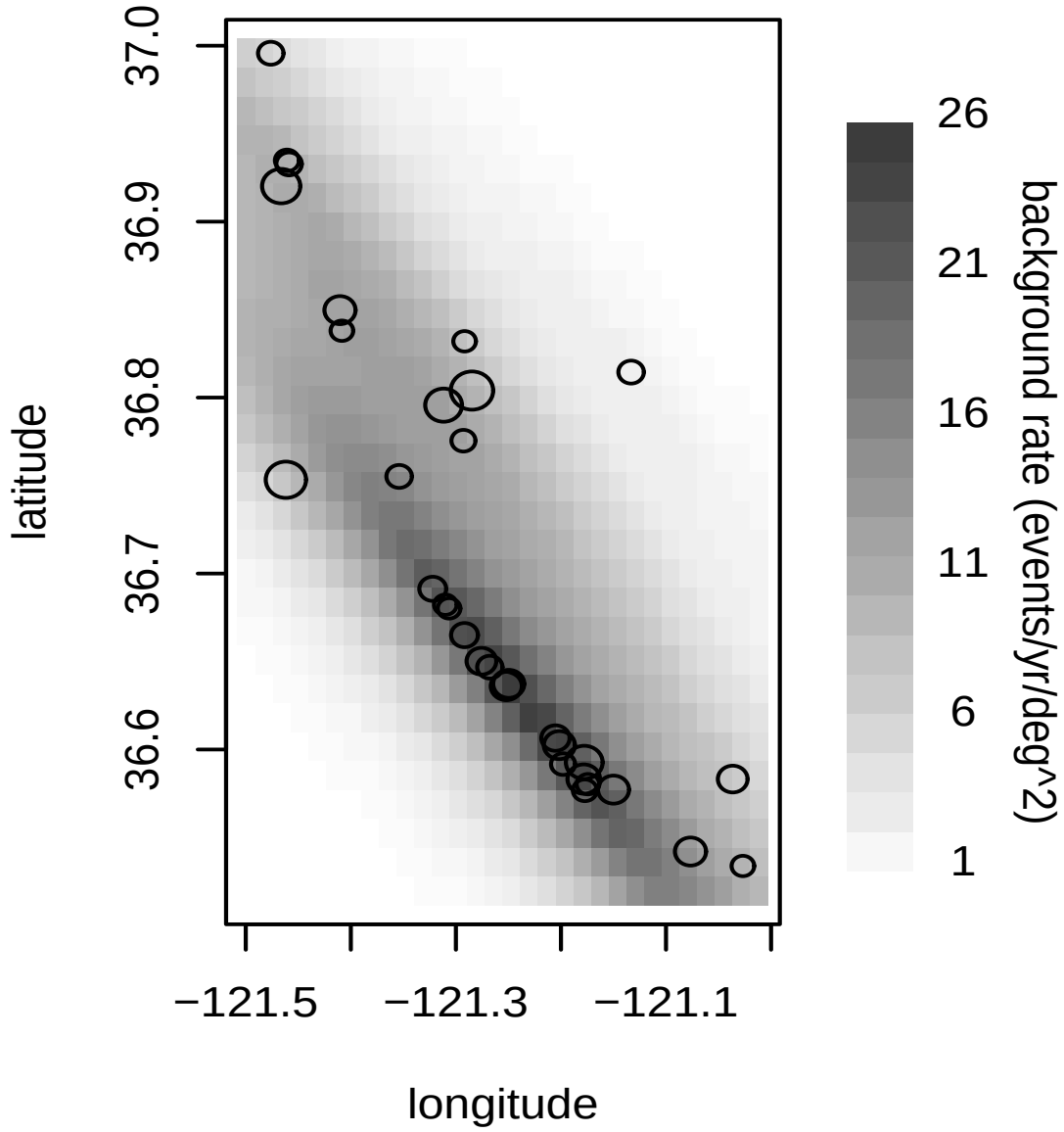

Figure 6: Anisotropic kernel estimate of the spatially-varying background rate $\nu(x, y)$, based on smoothing the $M \geq 4$ events.

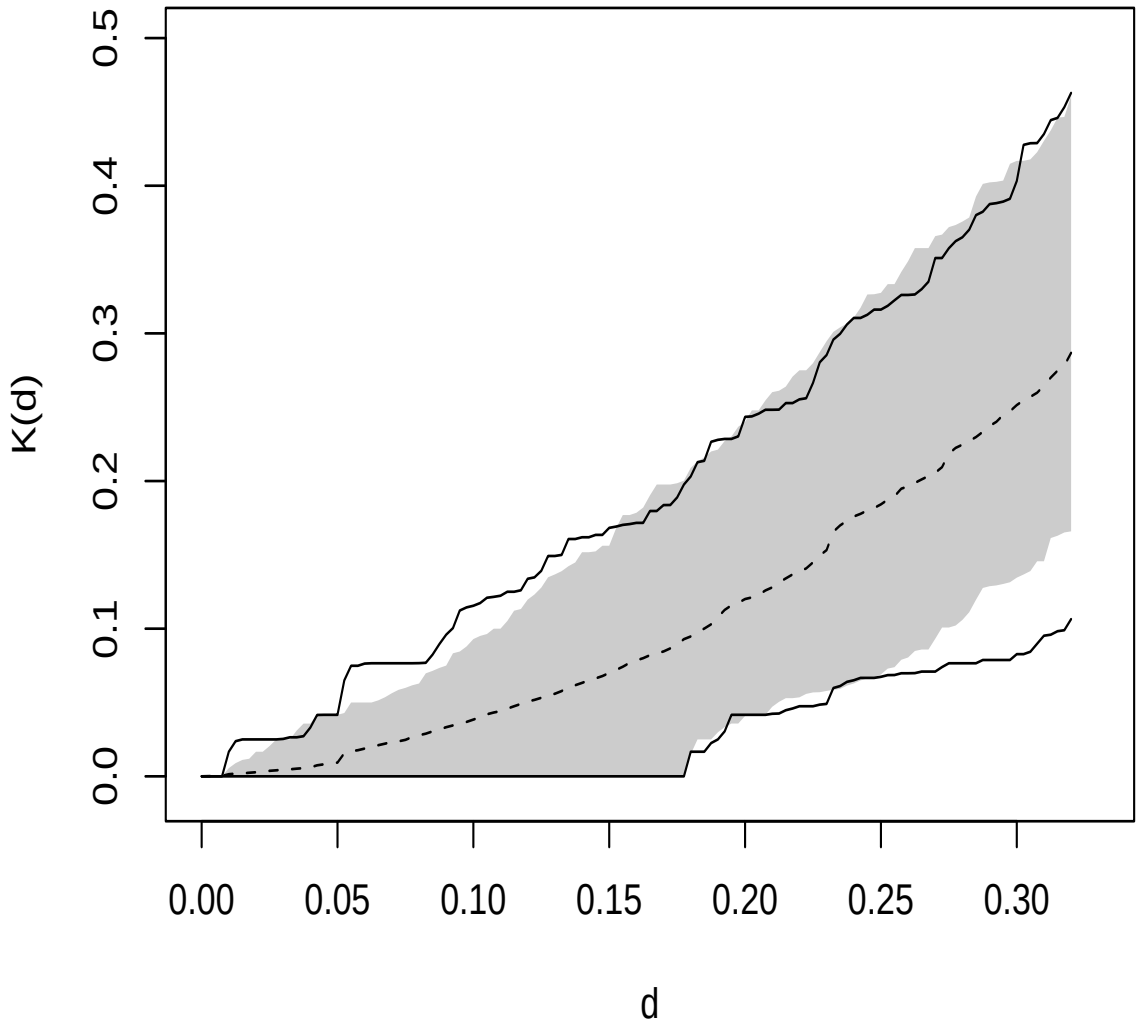

Figure 7: Middle 95%-ranges of estimated K-functions for 1000 thinned residuals based on model (1') [solid lines] and for 1000 homogeneous Poisson processes on the same space [shaded area].

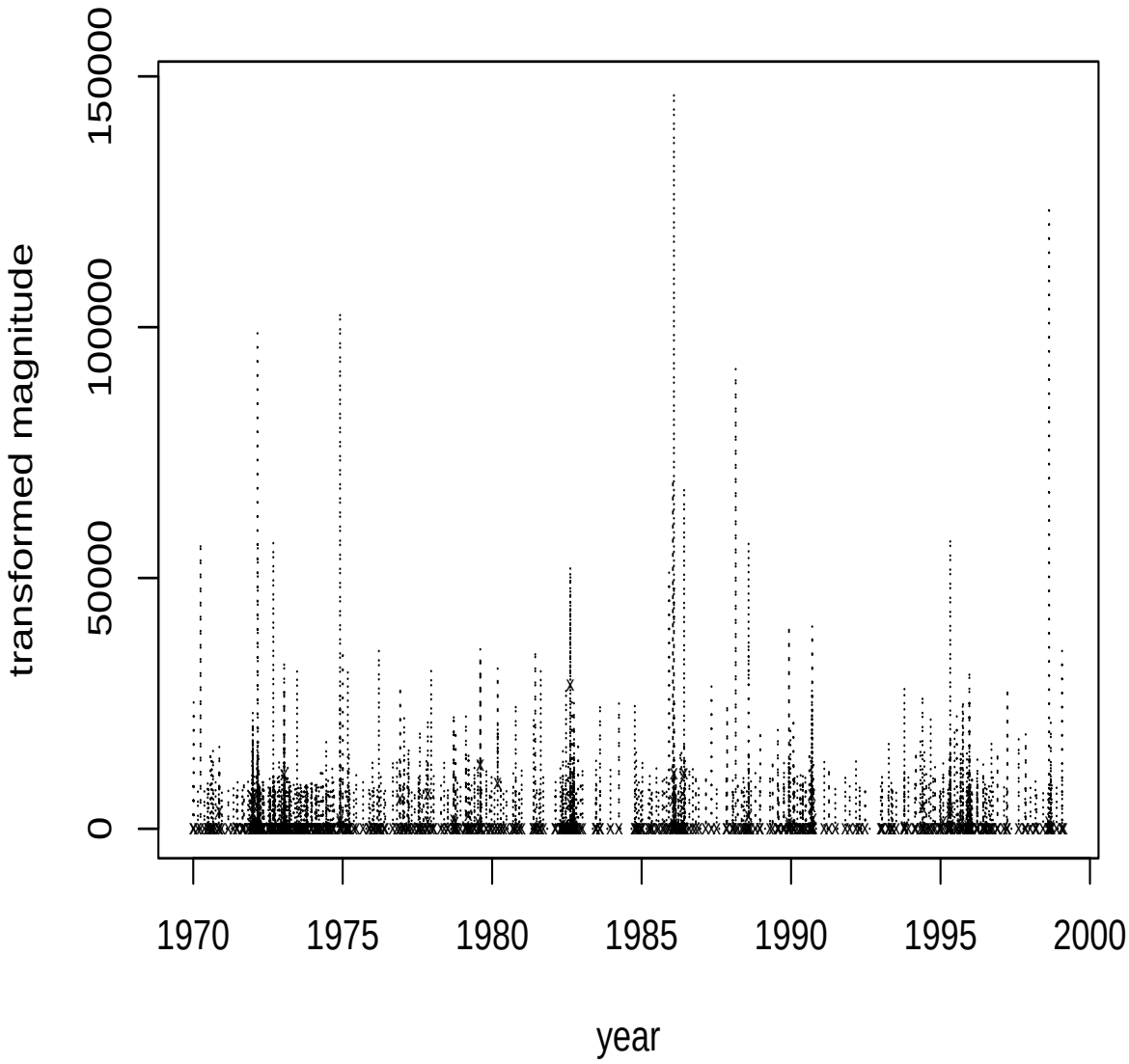

Figure 8: Vertically rescaled residuals based on model (1'), in time-magnitude plane.

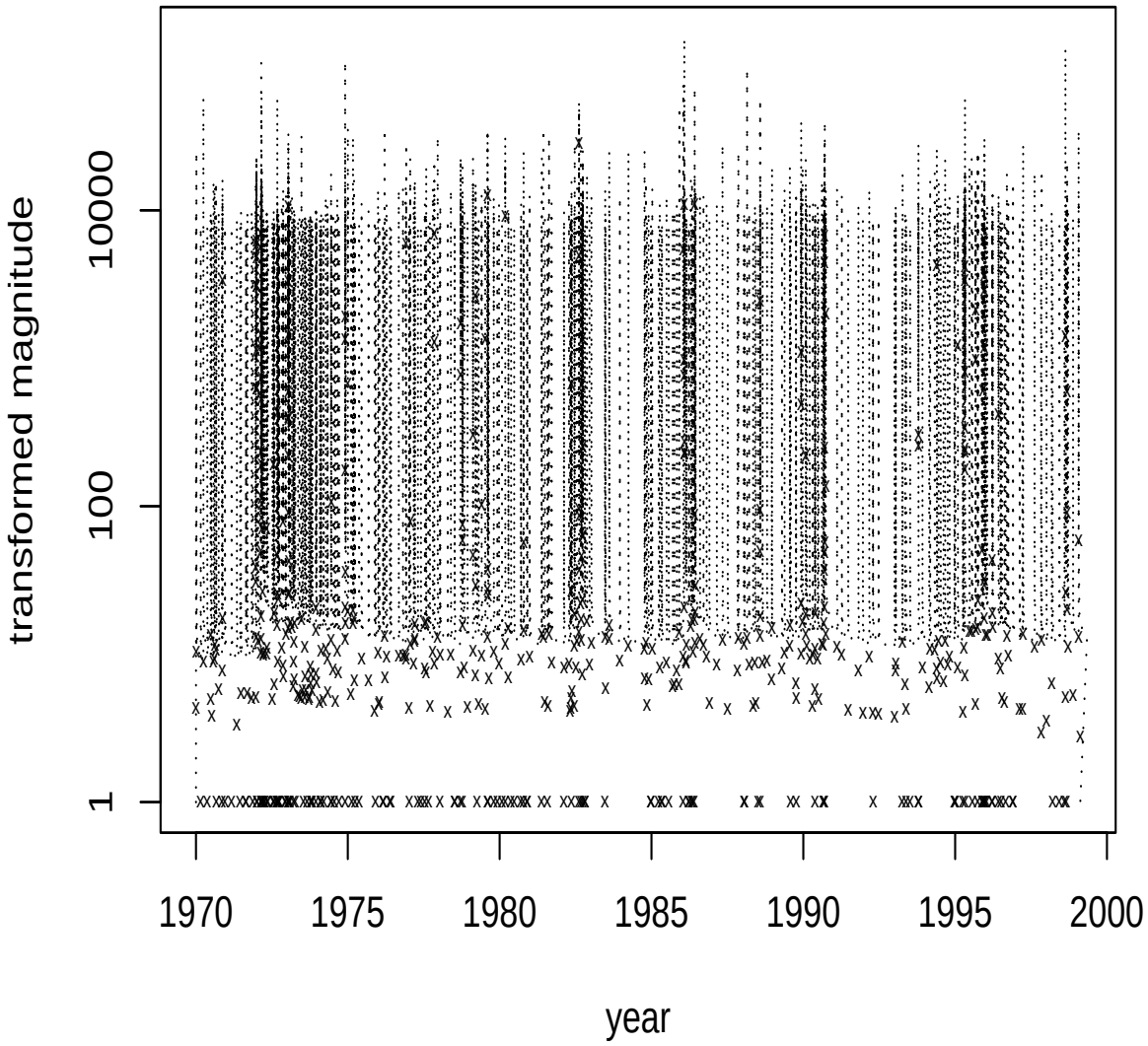

Figure 9: Vertically rescaled residuals based on model (1'), with the rescaled magnitudes displayed on a logarithmic scale.

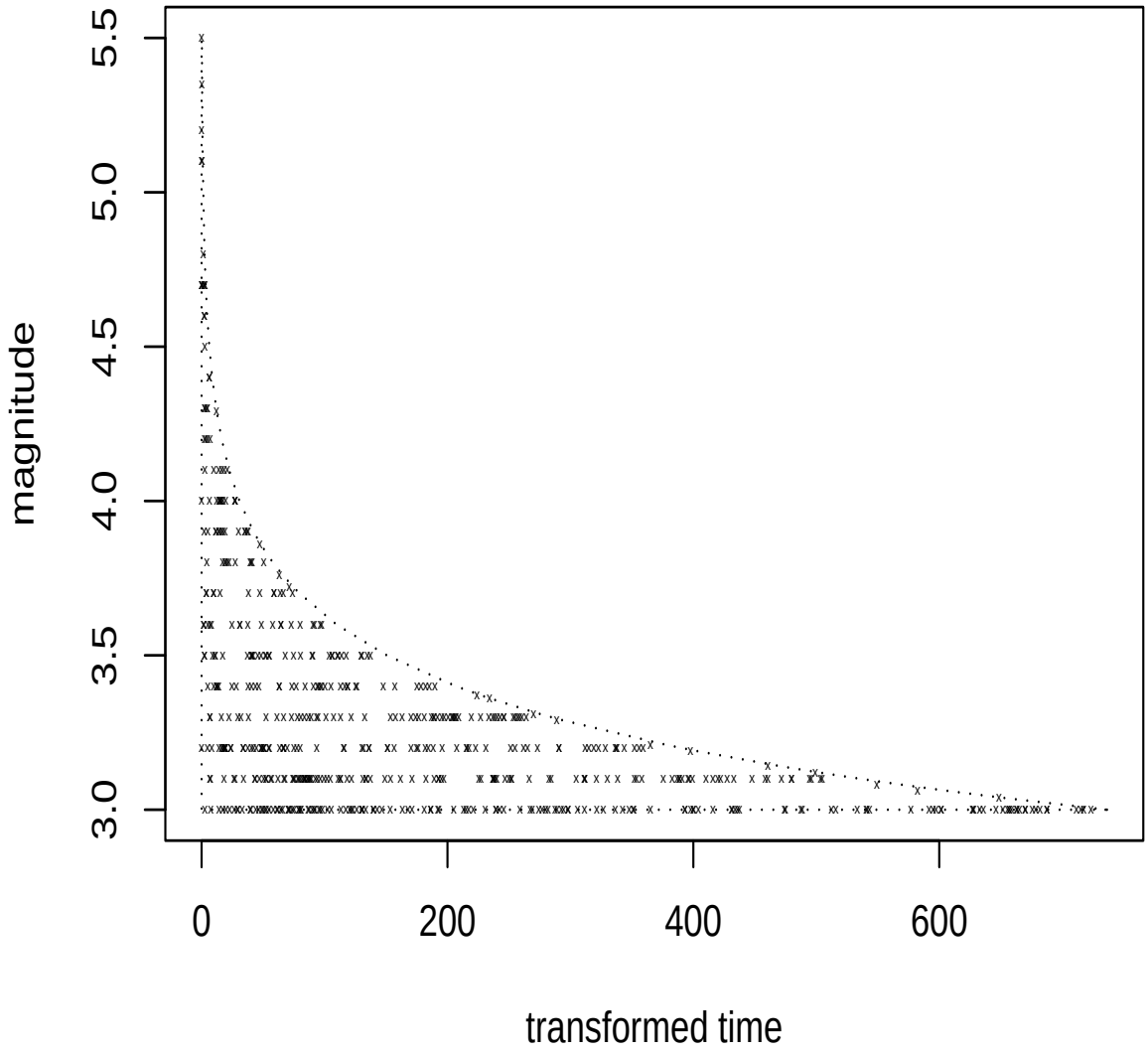

Figure 10: Horizontally rescaled residuals based on model (1'), in the time-magnitude plane.

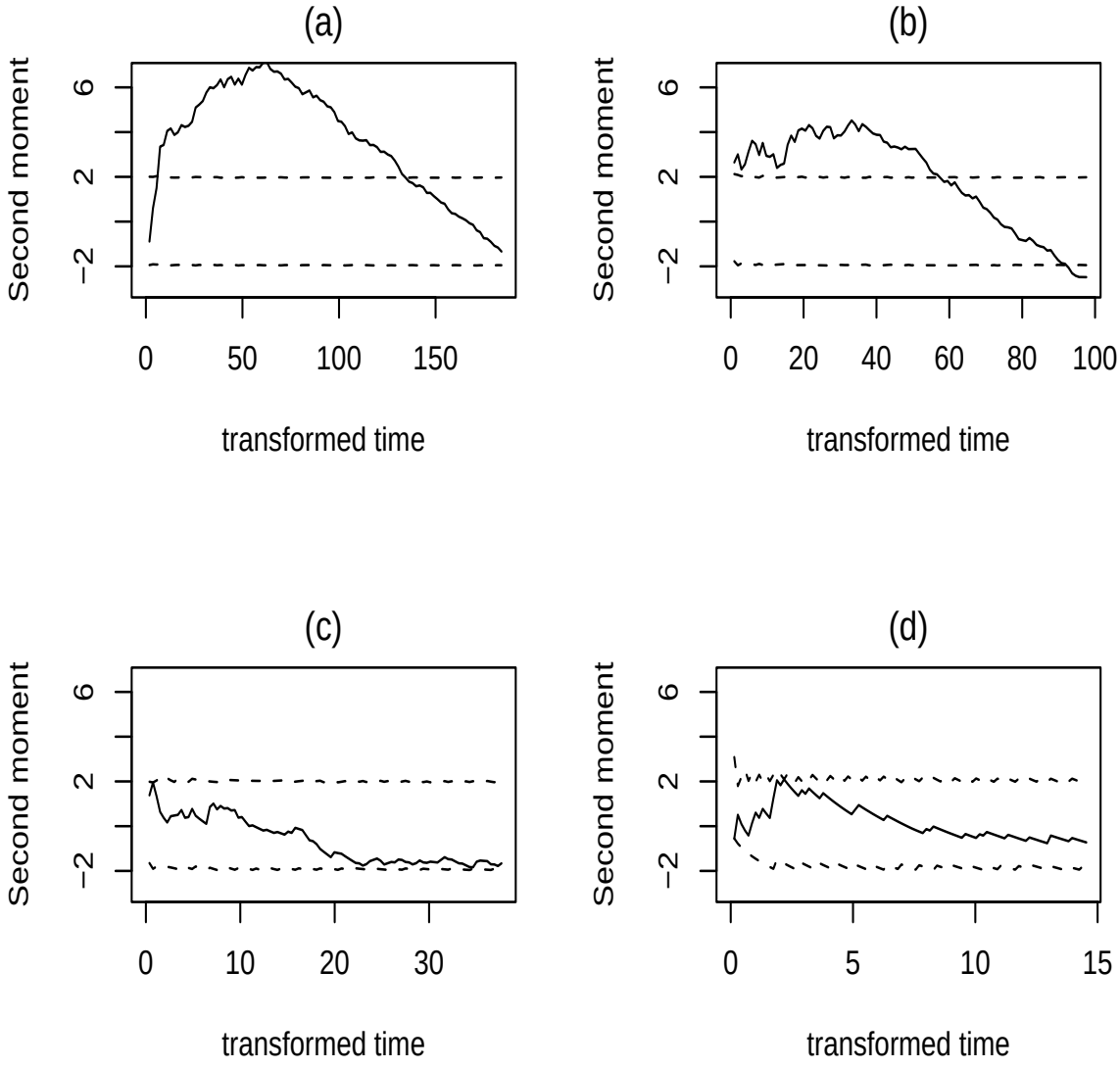

Figure 11: Estimated standardized second moments of horizontally rescaled residuals based on model (1'), for (a) magnitude 3.1; (b) magnitude 3.2; (c) magnitude 3.5; (d) magnitude 3.8.

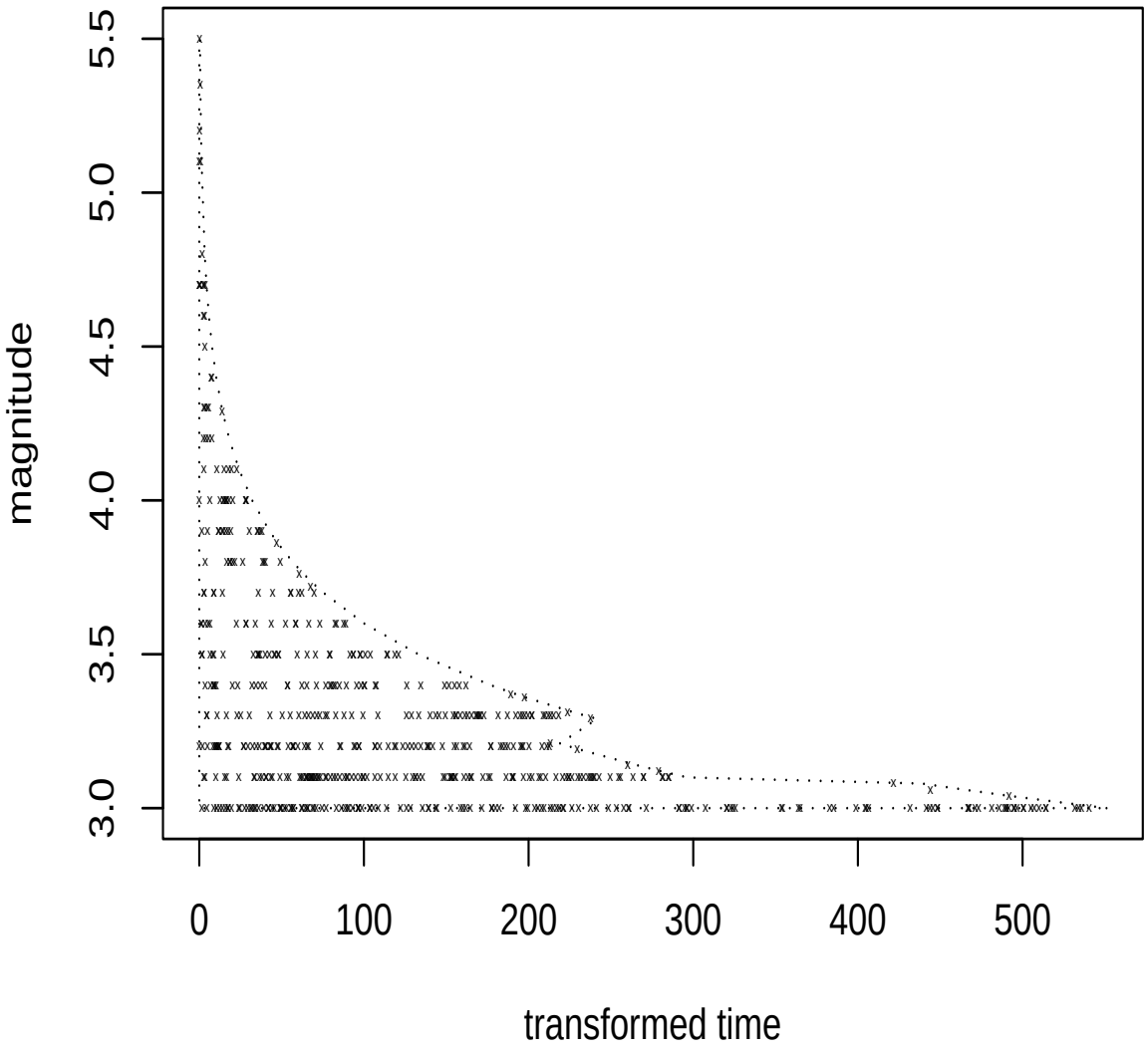

Figure 12: Horizontally rescaled residuals based on model (4).

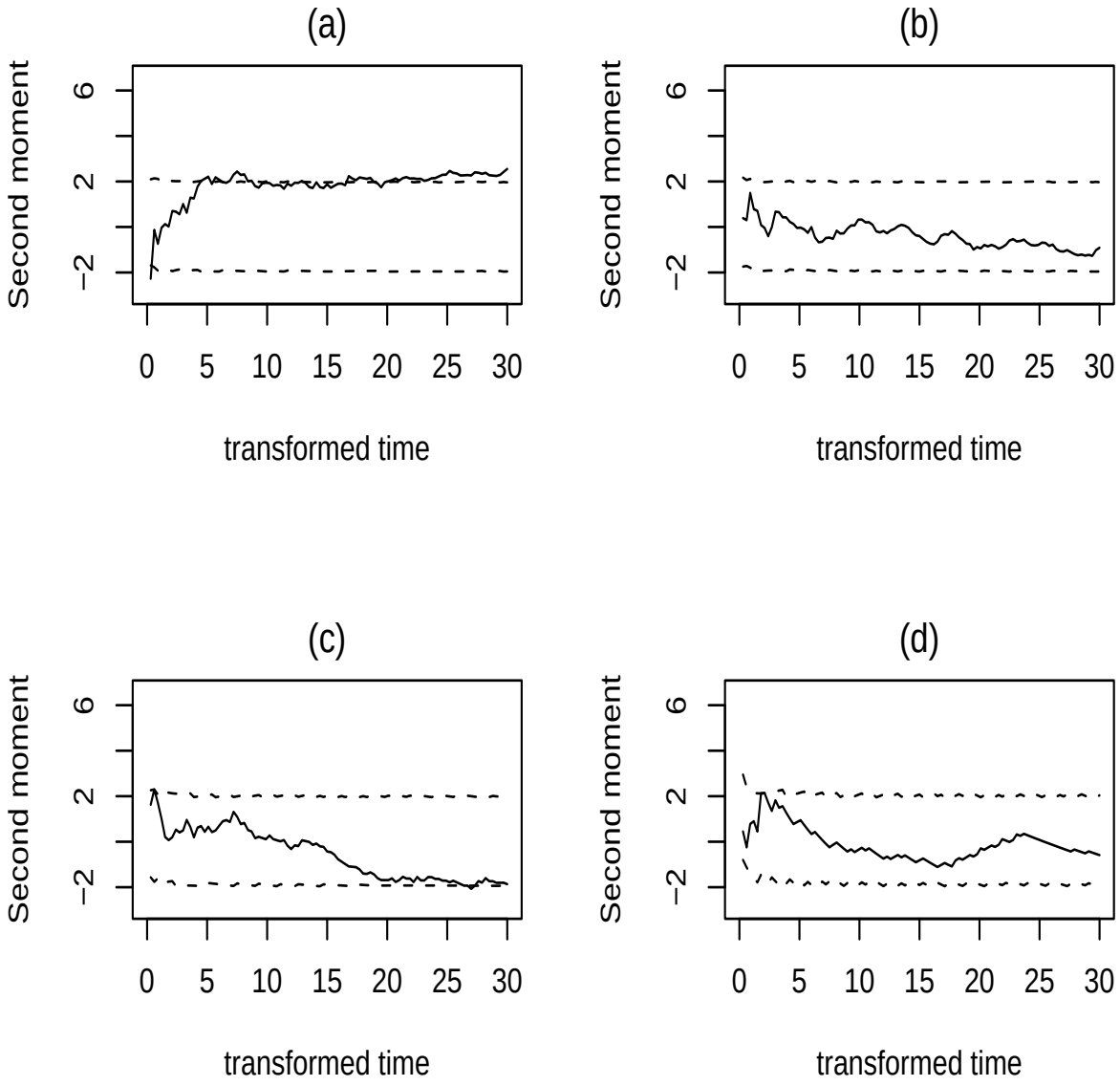

Figure 13: Estimated standardized second moments of horizontally rescaled residuals based on model (4), for (a) magnitude 3.1; (b) magnitude 3.2; (c) magnitude 3.5; (d) magnitude 3.8.