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

Residual analysis methods for examining the fit of multi-dimensional point process models 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. One method involves rescaled residuals, obtained by transforming points along one coordinate to form a homogeneous Poisson process inside a random, irregular boundary. Another method involves thinning the point process according to the conditional intensity to form a homogeneous Poisson process on the original, untransformed space. 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 in question. The residual analysis highlights that the fit of the space-time ETAS model may be improved by allowing the parameters governing the triggering density to vary for earthquakes of different magnitudes. Such modifications may be 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 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 by Brillinger (1993) and Vere-Jones (1995). In particular, the

Epidemic-Type-Aftershock-Sequence (ETAS) model of Ogata (1988) has proven to be extremely useful in the description and modeling of earthquake occurrence times and locations.

The ETAS model examined in Ogata (1988) incorporates time and magnitude, but has no spatial component. In recent years extensions characterizing the space-time-magnitude behavior of earthquake occurrences have been offered (e.g. Ogata 1998, Zhuang et al. 2002), but little attention has been paid to the examination of the goodness-of-fit of such models. The focus of the current paper is to assess the fit of space-time ETAS models to earthquake occurrence data from Bear Valley, California. We present methods for assessing the fit of multi-dimensional point process models using residual analysis and show how these methods may be used to identify defects in space-time ETAS models and to suggest ways in which the models may be improved. In particular, in California, where most earthquake hypocenters lie on or very near pre-existing faults, incorporation of basic fault geometry into the spatial background rate in the ETAS model seems desirable, and the methods also suggest that some relaxation of the separability assumption in the space-time ETAS model may be appropriate.

Another aim of this paper is the development and application of residual analysis methods for multi-dimensional point process models. For many statistical models such as ordinary regression models, examination of residuals is straightforward and widely considered standard practice. For the case of multi-dimensional 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. We propose the inspection of both rescaled and thinned residuals. Following a review of some multidimensional point process models for earthquake occurrences in Section 2, seismological data from Bear Valley, California are described in Section 3. Thinned residuals are used in Section 4 to highlight deficiencies in standard spatial background rate estimates, and in Section 5 rescaled residuals are used to assess the separability of the magnitude distribution in the ETAS model.

Section 6 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 σ-finite random measures on a portion S of space-time-magnitude, taking values in the non-negative integers or infinity. One typically assumes the process to be completely simple, i.e. that with probability one all the points occur at distinct times. As any such process N may be uniquely characterized by its associated conditional rate process λ (Fishman and Snyder 1976; Daley and Vere-Jones 2003), 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, $\lambda(t, \mathbf{x}, m)$ 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.

Formally, a version of the conditional rate $\lambda(t, \mathbf{x}, m)$ 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 (Brémaud 1981; Schoenberg et al. 2002). See Jacod (1975) for conditions for existence and uniqueness of the conditional rate and its integrated form called the *compensator*. Note that although in the statistical literature (e.g. Daley and Vere-Jones 2003; 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.

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). The widely-used ETAS models of Ogata (1988, 1998) are based on the principle that earthquakes are clustered due to the occurrence of aftershocks, and furthermore aftershocks can in turn have aftershocks, etc. Given a collection of points $\{t_i, \mathbf{x}_i, m_i\}, i = 1, 2, \ldots$, the model may be written

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

The functions f and μ 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. Various forms for f, g, and μ are suggested in Ogata (1998) based on well-known seismological relations. 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). When the spatial coordinate $\mathbf{x}$ represents epicentral origin location (x, y) in the plane, a form for the triggering density g given in Ogata (1998) is

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

whose temporal component agrees with the modified Omori law (Utsu 1971). Functional forms for the background rate $\mu(\mathbf{x})$ are not typically given; instead μ is assumed constant or estimated by smoothing the main events in the catalog, e.g. using bi-cubic B-splines or kernel smoothing (Ogata 1998; Zhuang et al. 2002). In what follows we refer to the model governed by relations (1-3) and with background rate estimated by kernel smoothing simply as model (1).

Note that 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 μ 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 5 below. As an alternative we consider allowing the parameter vector θ to vary with magnitude, m. That is, we consider the model

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

where $\{M_j; j = 1, 2, \ldots, J\}$ is a partition of the observed magnitude range. Note that this revised model (4) is not inconsistent with the modified Omori law, which governs the decay of aftershocks over time for *all* magnitudes in the observed range.

3 Bear Valley Earthquake Data

The Bear Valley earthquake catalog obtained from the Council of the National Seismic System (CNSS) 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), between 1970 and 2000. The catalog is 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. Details about the data may be obtained from the CNSS at http://quake.geo.berkeley.edu.

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 Thinned spatial residuals and improved spatial background rate estimates

The fit of the spatial component of model (1) is conveniently investigated using thinned residuals, using a variation on the useful simulation technique of Lewis and Shedler (1979) and Ogata (1981). Given a space-time-magnitude point process N with conditional intensity λ, 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. When this thinning is done using the estimated conditional intensity $\hat{\lambda}$ in place of λ, the remaining points, which we call the thinned residual points, may be

examined for uniformity.

We consider the model (1) with background rate $\mu(x, y)$ estimated by kernel-smoothing of the larger events (those with magnitude at least 4.0). 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 μ 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 approximate a homogeneous Poisson process. A powerful test for the alternative hypothesis that the points are instead more (or less) spatially clustered than a homogeneous 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 K-functions (not shown) for thinned residuals from model (1) generally indicate excessive spatial clustering, i.e. spatial clustering in the data not adequately accounted for by 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 $\hat{\mu}(x, y)$ 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 (MLE) of the parameters in the model (1), is shown in Figure 4 along with the larger earthquakes used for the smoothing. The kernel smoothing using Euclidean distance is clearly too smooth in this case, so that points near the fault line, which occur at a high rate, are estimated to occur at a relatively low rate. Hence many of these points remain in the thinned residuals, causing excessive clustering in these residual processes. 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 fact, application of the method of Zhuang et al. (2002) still results in an overly smoothed background similar to that in Figure 4, 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 K-functions of the thinned residuals suggest modifying the estimation of $\mu(x, y)$ 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 $\mu(x, y)$ 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 $\mu(x, y)$ is shown in Figure 5. Hereafter we refer to the model governed by equations (1-3) and with μ estimated using the non-isotropic distance function as model (1').

The smoothing in Figure 5 is not nearly as smooth as that in Figure 4, and clearly corresponds much more closely with the spatial distribution of the actual events. After using the modified, anisotropic estimate of $\mu(x, y)$ and obtaining an intensity estimate $\hat{\lambda}$ by MLE, the value b, defined as the minimum of $\hat{\lambda}$ over the entire space S, is now miniscule, so each realization of thinned residuals contains only very few points (indeed, on average less than one). This suggests instead inspecting approximate residuals for model (1'), obtained 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 $k\hat{\lambda}(t_i, x_i, y_i, m_i)^{-1}/(\sum_{i=1}^{N(S)} \hat{\lambda}(t_i, x_i, y_i, m_i)^{-1})$ with the number k 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). Since the occurrance rate of a point at (t, x, y, m) is given by $\lambda(t, x, y, m)$ and since each such point is kept with probability inversely proportional to $\hat{\lambda}(t_i, x_i, y_i, m_i)$, the resulting process should again resemble a homogeneous Poisson process provided $\hat{\lambda}$ closely approximates the true conditional intensity, λ. Empirical 95% bounds for K-functions corresponding to 1000 realizations of such thinned residuals for the model (1') are shown in Figure 6, along with corresponding 95% bounds based on uniformly distributed points in S. The bounds in Figure 6 agree rather closely with those of the uniformly distributed points, and most of the clustering in the residual K-functions is no longer indicated. That is, on balance, the thinned residuals from model (1') do not appear to be substantially more clustered than points uniformly distributed in S.

5 Rescaled temporal-magnitude residuals and assessment of magnitude separability

The time-magnitude distribution of the earthquake process may be investigated using rescaled residuals based on the method due to Meyer (1971), who showed that for a collection $\{N_i\}$ of completely simple univariate point processes on the real half-line, provided $\int_0^\infty \lambda(t,i)dt = \infty$ for each i, 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 of unit rate. This type of horizontal 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. Results on vertical rescaling for have also been derived for spatial point processes (see Schoenberg 1999, and references therein), and Meyer's result also extends directly 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. The requirement that $\int_0^\infty \lambda(t,i)dt = \infty$ may be neglected in applications, since if the original process is observed within a compact space-time-magnitude range S, then the rescaled residual process is Poisson within a random, irregular boundary, as described in Schoenberg (1999). In the present case, because of the temporal volatility of the conditional intensity under the ETAS model for California earthquakes, the boundary corresponding to vertically-rescaled residuals (not shown) is highly irregular, making inspection of uniformity of the residuals quite difficult. The boundary for the horizontally-rescaled residuals, shown in Fig. 7, is not nearly as irregular, and the residual points in Fig. 7 appear to be relatively uniformly dispersed, though some clustering of points of magnitude 3.1 and 3.3 may be observed. Deviations from uniformity are difficult to discern directly by eye, so 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 Fig. 7. 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), we propose examining the second-order properties of the residuals along each horizontal line. For example, we 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 $(k_1(u,m) - k_2(u,m))/\sqrt{k_2(u,m)}$, where $k_1(u,m)$ is the number of pairs of residual points with magnitude m whose transformed times are within u of one another, and $k_2(u,m)$ is the the expected number of such pairs for a homogeneous Poisson process. The dotted curves in Fig. 8 show the estimated standardized second moment functions for model (1'), for the residuals of several different magnitudes, along with 95% bounds for the homogeneous Poisson process (dashed lines). 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.

Fig. 8 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. By contrast, the second moment function for the horizontally rescaled residuals obtained after fitting the non-separable model (4) by maximum likelihood are also displayed in Fig. 8 (solid curves), and do

not show the highly significant residual clustering characteristic of model (1'). In fitting (4), J is set to 2, M_1 to the magnitude range $[3.1, 3.25]$, M_2 to the set of all other magnitudes in the observed range, and the background rate μ is estimated in exactly the same way as model (1'). Note that the confidence bounds for the homogeneous Poisson process on the transformed space using model (1') were indistinguishable from those using the transformed space corresponding to model (4), so in Fig. 8, only the former is shown.

For comparison with Fig. 7, the horizontally-rescaled residuals for model (4) are shown in Fig. 9. One sees from Figs. 8 and 9 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.

6 Summary.

In the seismological application here it was demonstrated that anisotropic kernel smoothing enables better estimation of the spatially varying background rate $\mu(x, y)$ 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 virtually 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) 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, the possibility that the non-separability is an artifact of some feature of the data collection and magnitude estimation cannot be ruled out. Another 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.

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.

Acknowledgements. This material is based upon work supported by the National Science Foundation under Grant No. 9978318. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation. This research benefitted greatly from comments from an editor and anonymous reviewers, as well as from useful discussions with David Brillinger, David Vere-Jones, Yan Kagan, 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.

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

Brémaud, P. (1981), *Point Processes and Queues: Martingale Dynamics,* New York: Springer-Verlag.

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

Daley, D., and Vere-Jones, D. (2003), *An Introduction to the Theory of Point Processes,* 2nd ed. New York: Springer.

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.

Gutenberg, B. and Richter, C.F. (1944), "Frequency of Earthquakes in California," *Bulletin of the Seismological Society of America*, 34, 185–188.

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.

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.

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

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.

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.

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

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

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.

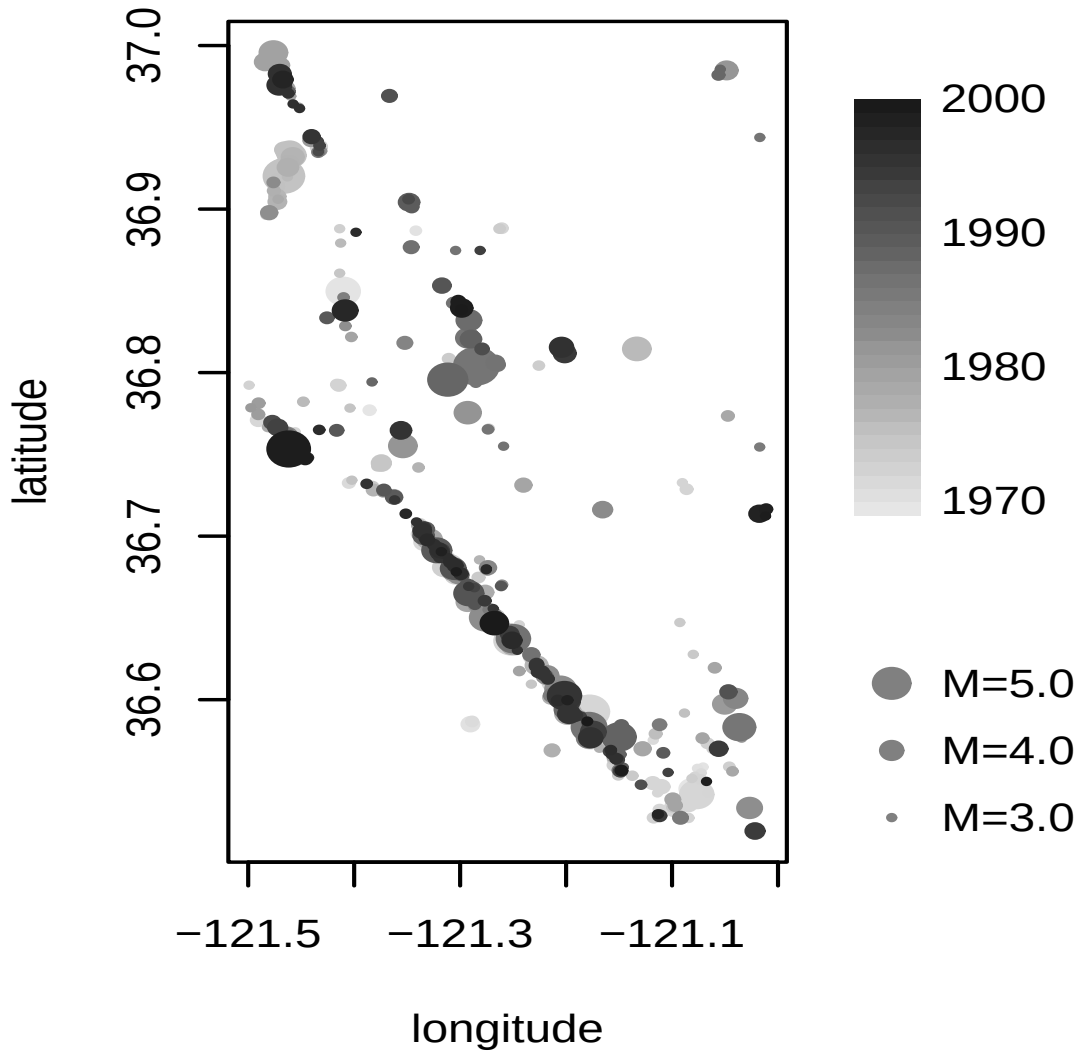

Figure 1: Locations, times and magnitudes of moderate-sized ($M \geq 3.5$) earthquakes in Bear Valley, CA, between 1970 and 2000.

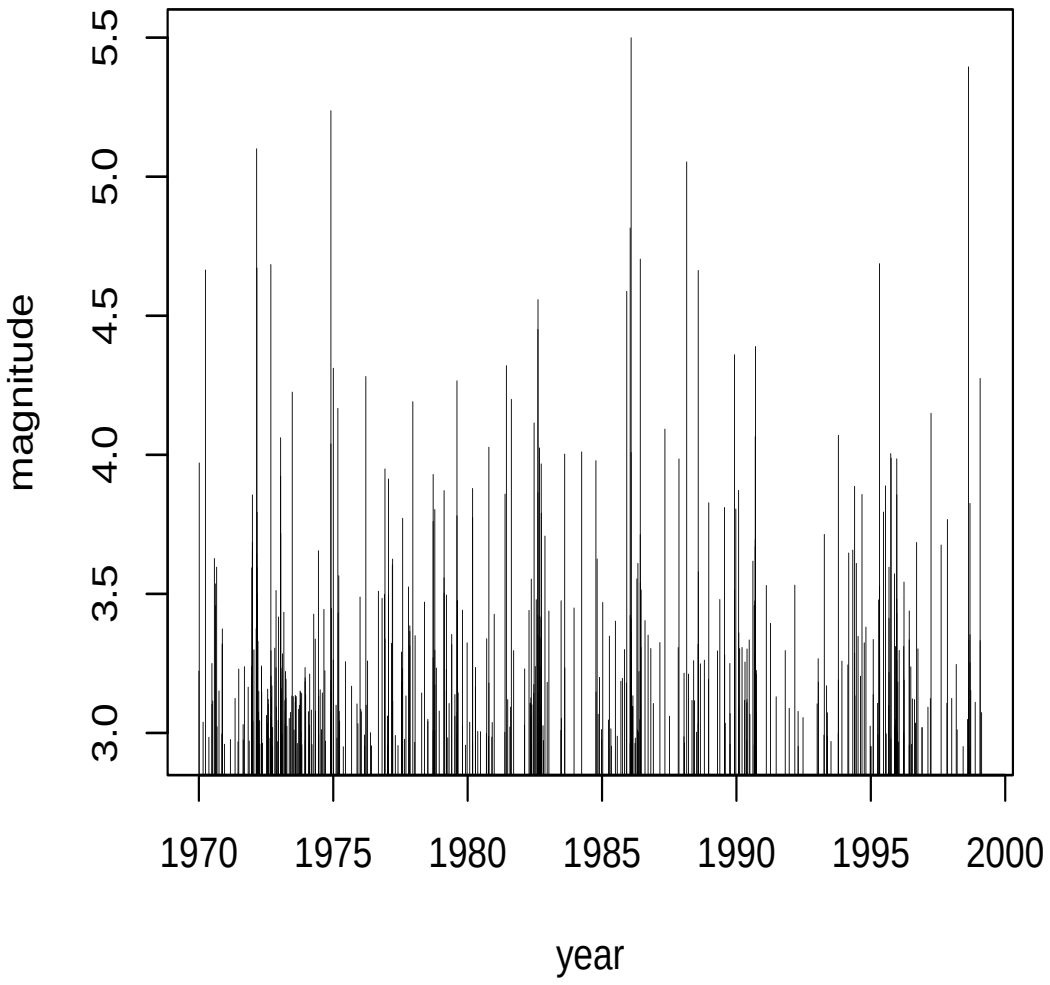

Figure 2: Times and magnitudes of Bear Valley, CA earthquakes.

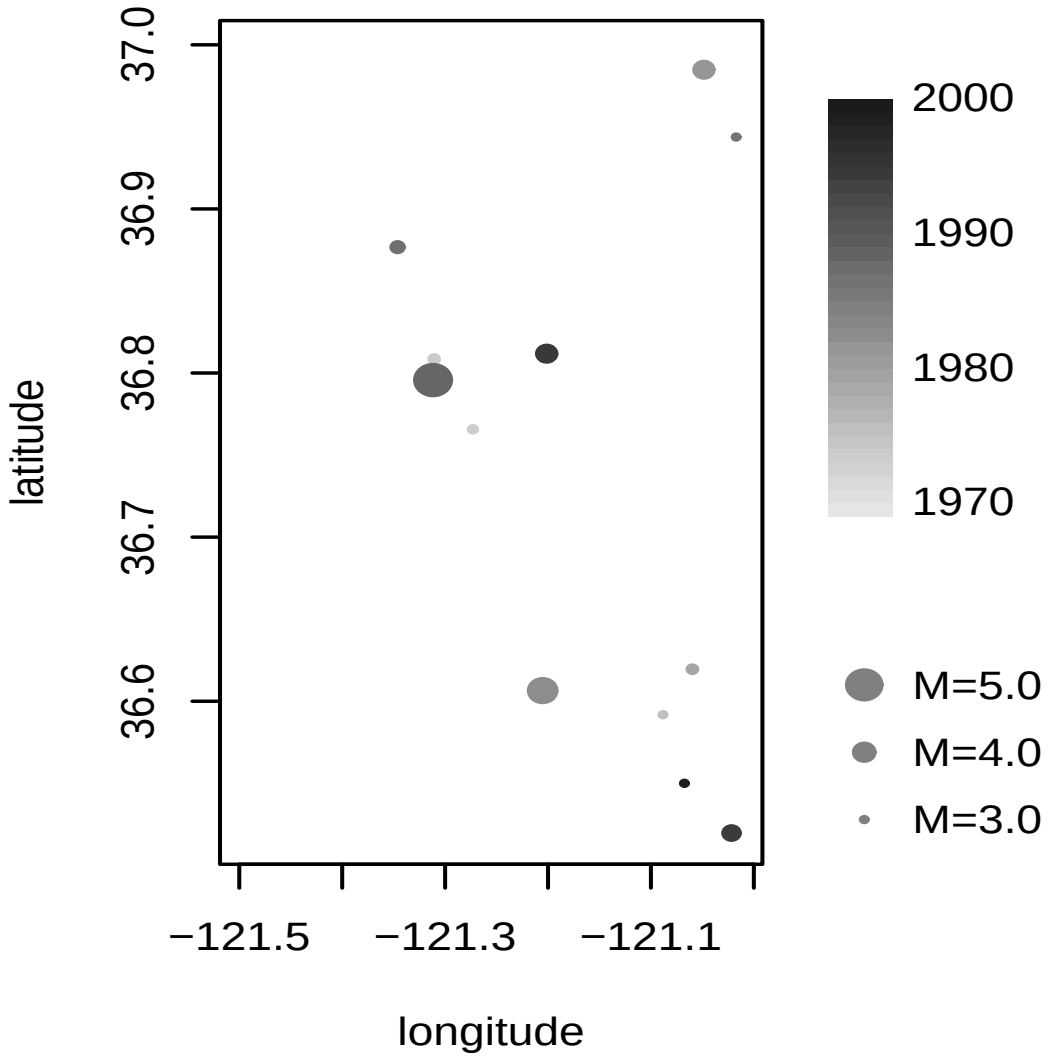

Figure 3: Sample thinned residual process, for model (1).

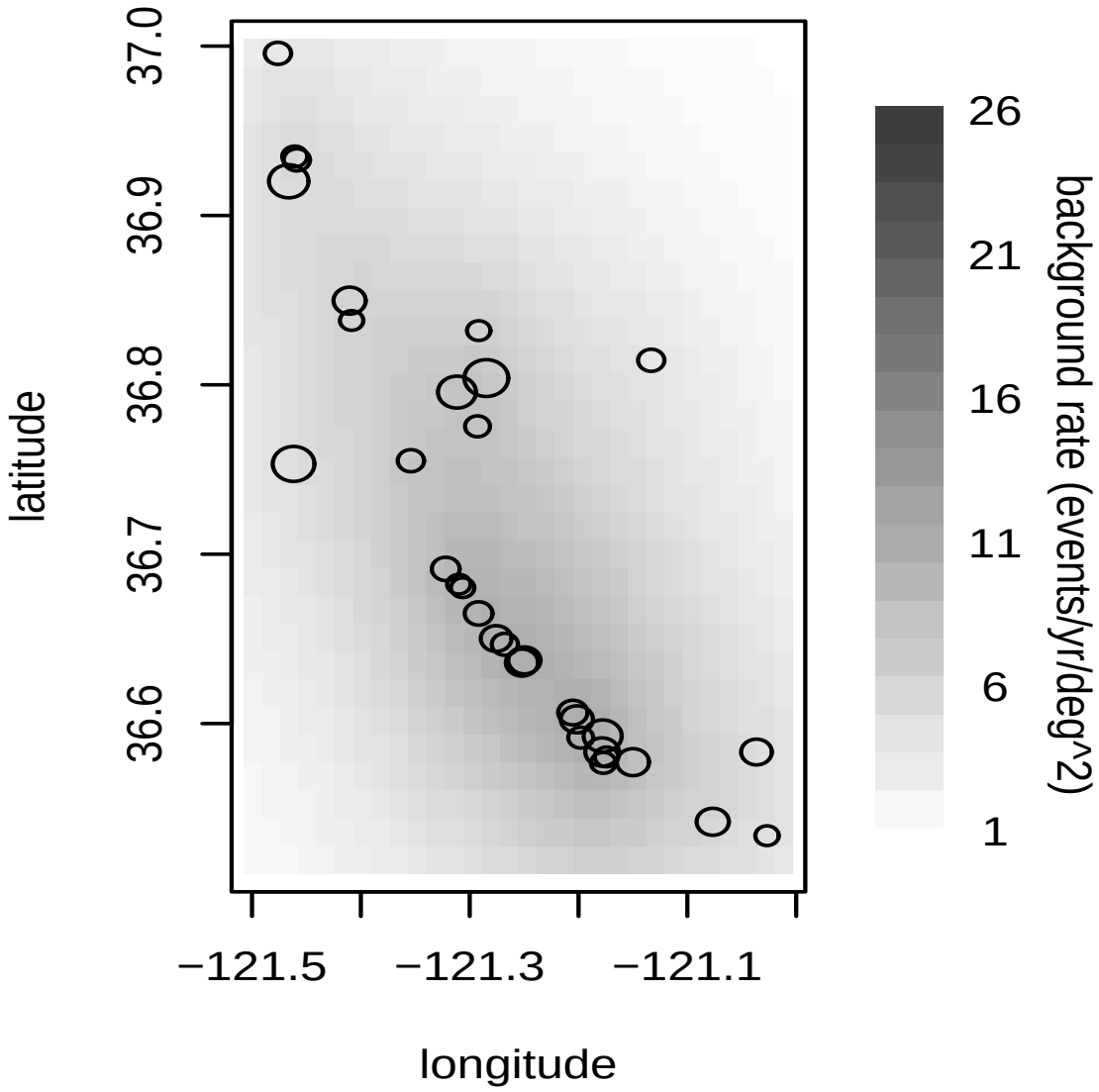

Figure 4: Spatially-varying background rate $\mu(x, y)$, estimated by isotropic kernel smoothing of the $M \geq 4$ events. Circles reperesent $M \geq 4$ events. The magnitude scale corresponding to the size of the circles is same as for Figure 1.

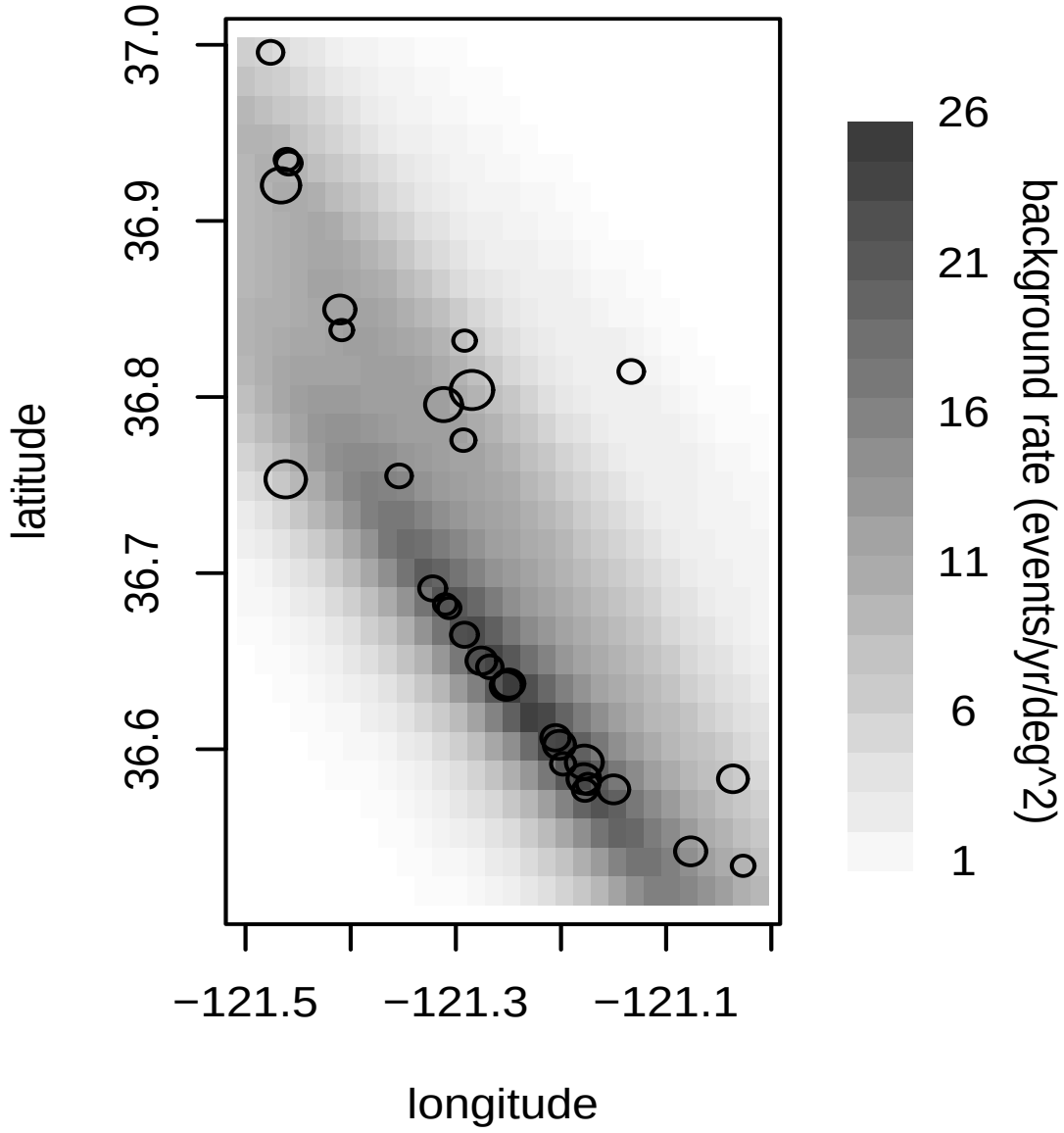

Figure 5: Anisotropic kernel estimate of the spatially-varying background rate $\mu(x, y)$, based on smoothing the $M \geq 4$ events.

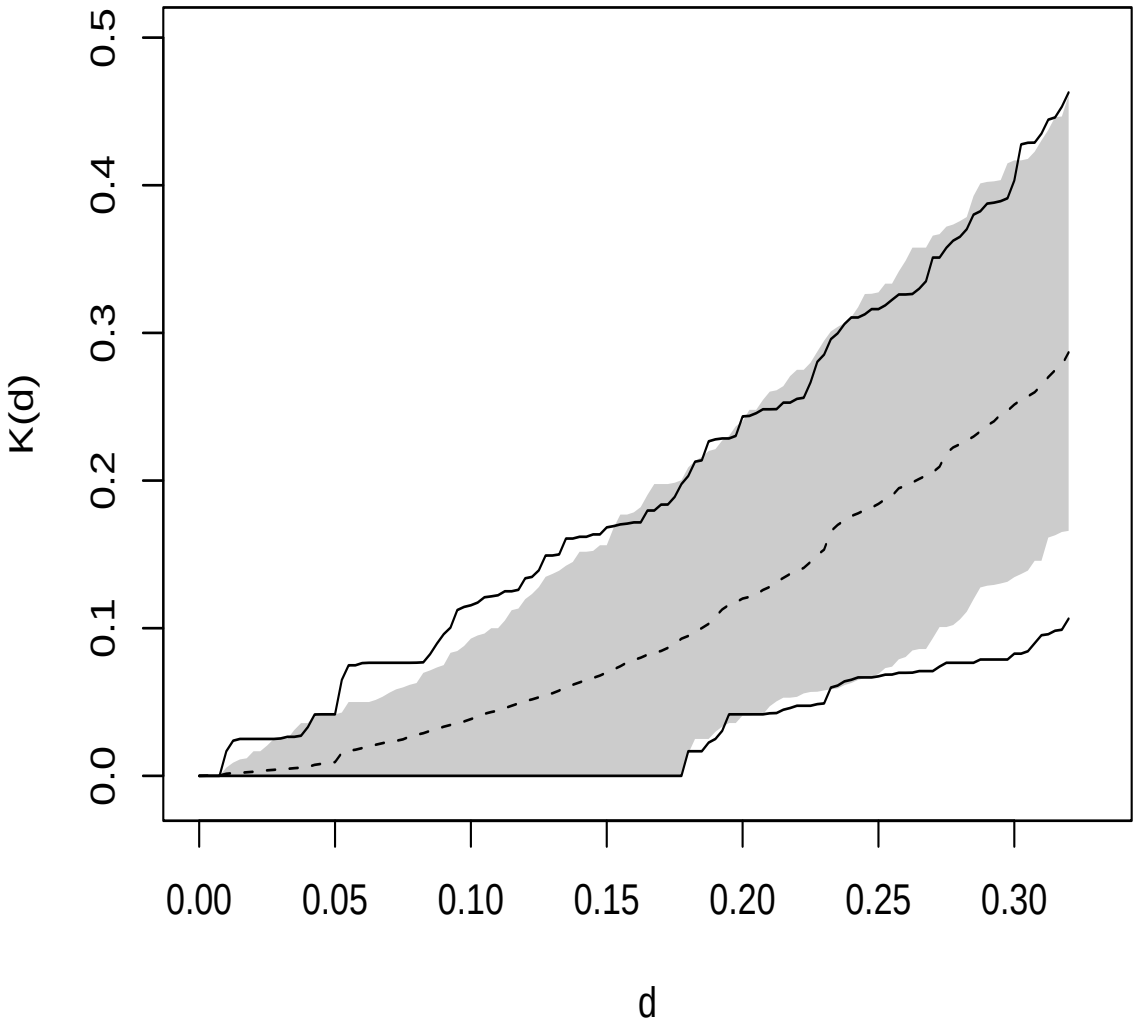

Figure 6: Middle 95%-ranges of estimated K-functions for 1000 thinned residuals based on model (1') [solid curve] and for 1000 homogeneous Poisson processes on the same space [shaded area]. The dashed curve indicates the mean estimated K-function of the thinned residuals of model (1').

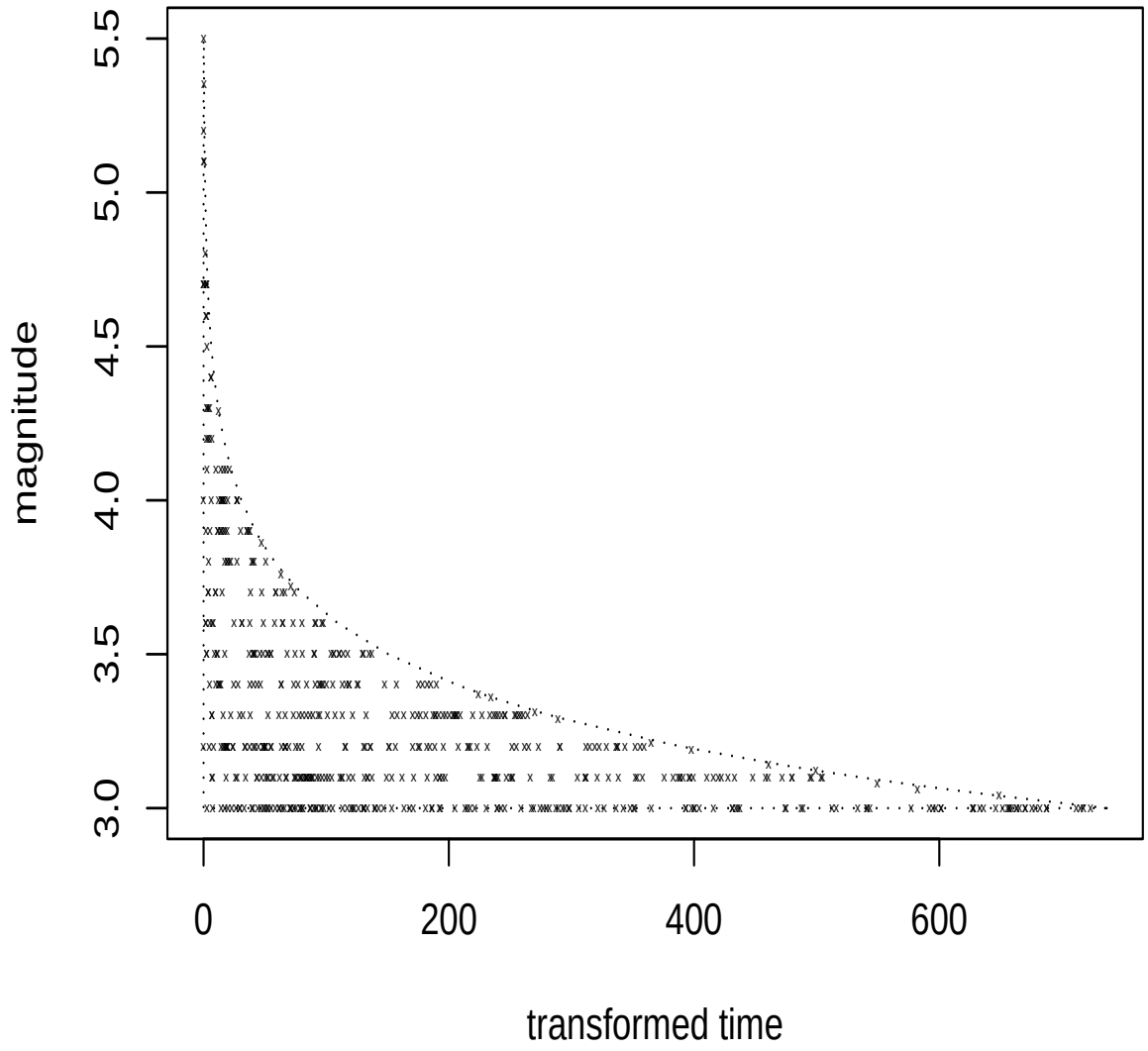

Figure 7: Horizontally rescaled residuals based on model (1'), in the time-magnitude plane.

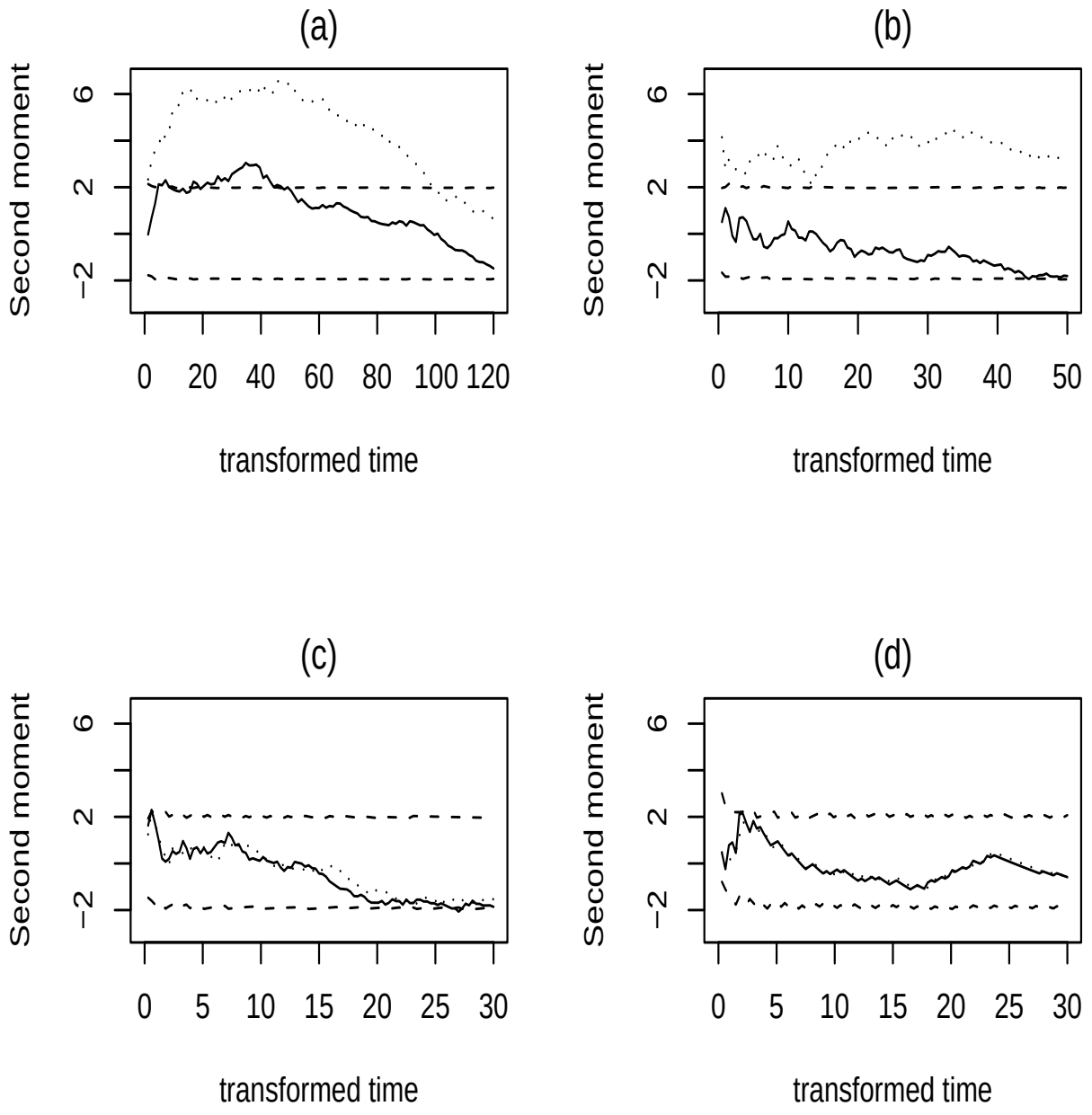

Figure 8: Estimated standardized second moments of horizontally rescaled residuals, for (a) magnitude 3.1; (b) magnitude 3.2; (c) magnitude 3.5; (d) magnitude 3.8. In each case the dotted curve corresponds to model (1'), the solid curve corresponds to model (4), and the dashed lines indicate 95% confidence bounds for the homogeneous Poisson process on the transformed space corresponding to model (1').

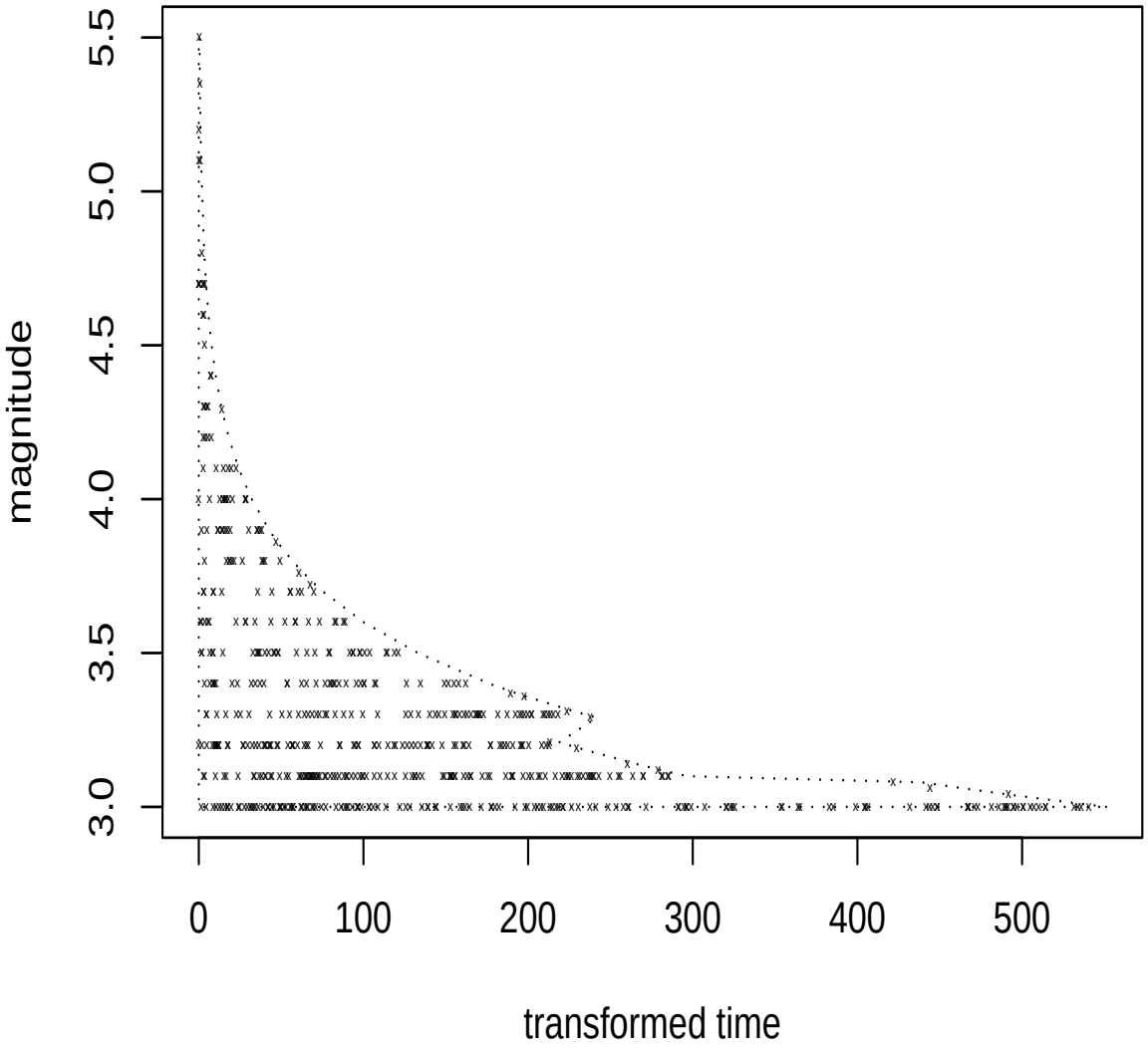

Figure 9: Horizontally rescaled residuals based on model (4).