Submitted to the Annals of Applied Statistics
arXiv: arXiv:0000.0000

A NOTE ON NONPARAMETRIC ESTIMATES OF SPACE-TIME HAWKES POINT PROCESS MODELS FOR EARTHQUAKE OCCURRENCES

By Eric Warren Fox *, Frederic Paik Schoenberg * and Joshua Seth Gordon *

*University of California, Los Angeles**

Space-time Hawkes point process models for the conditional rate of earthquake occurrences traditionally make many parametric assumptions about the form of the triggering function for the rate of aftershocks following an earthquake. Marsan and Lengliné (2008) developed a completely nonparametric method that provides an estimate of a stationary background rate for mainshocks, and a histogram estimate of the triggering function. At each step of the procedure the model estimates rely on computing the probability each earthquake is a mainshock or aftershock of a previous event. The focus of this paper is the improvement and assessment of Marsan and Lengliné's method in the following ways: (a) the proposal of novel ways to incorporate a nonstationary background rate; (b) adding error bars to the histogram estimates which capture the sampling variability and bias in the estimation of the underlying seismic process. A simulation study is designed to evaluate and validate the ability of our methods to recover the triggering function and spatially varying background rate. An application to earthquake data from the Tohoku District in Japan is discussed at the end, and the results are compared to a well established parametric model of seismicity for this region.

1. Introduction. Hawkes point process models of earthquake seismicity usually rely heavily on parametric assumptions about the triggering function for the spatial-temporal rate of aftershock activity following an earthquake. Some important examples are the parametric forms of the Epidemic Type Aftershock Sequences (ETAS) model of Ogata (1998). Marsan and Lengliné (2008) proposed a more flexible nonparametric approach for estimating Hawkes process models of seismicity which makes no a-priori assumptions about the shape of the triggering function, and provides a data-driven estimate instead. Their method, named Model Independent Stochastic Declustering (MISD), is an iterative algorithm that alternates between first estimating the probability each earthquake in the catalog is either a mainshock or aftershock and second, updating a stationary background rate for mainshock activity and a probability weighted histogram estimate for the triggering function.

Nonparametric methods for estimating point process models have shown a wide rage

Keywords and phrases: Point processes, nonparametric estimation, Hawkes process, MISD, ETAS model, earthquake forecasting.

of applications, especially in situations where the form of the intensity function is unknown and difficult to determine. Using wavelets Brillinger (1998) described a technique for estimating the conditional intensity and second order intensity with applications to neurophysiology and seismology. Adelfio and Chiodi (2013, 2015) considered a semi-parametric estimation procedure that simultaneously estimates a nonparametric background rate and parametric triggering function for a space-time Hawkes process model of seismicity. Marsan and Lengliné (2008) applied the fully nonparametric MISD method to a Southern California earthquake catalog to estimate the spatial-temporal rates aftershock activity following an earthquake of given magnitude. They also demonstrated the application of their routine for stochastically declustering earthquake catalogs to isolate mainshocks and remove aftershock clusters. Nichols and Schoenberg (2014) used MISD as a diagnostic tool to evaluate the dependency between the magnitude of an earthquake and the magnitudes of its aftershocks. By repeatedly applying the MISD algorithm to stochastically assign earthquakes as either mainshocks or aftershocks they created confidence intervals for the average magnitude of aftershocks following an earthquake of given magnitude. In an application to criminology, Mohler et al. (2011) developed a Monte-Carlo based nonparametric method similar to MISD to estimate a space-time point process model for the occurrence rate of burglaries in a Los Angeles district. They demonstrated that this approach leads to improved hotspot maps for flagging times and locations where burglaries are likely to occur. An interesting result of this study is that crimes spur other crimes nearby in space and time, much as earthquakes trigger local aftershock sequences.

The focus of this paper is the improvement and assessment of the nonparametric method of Marsan and Lengliné (2008) for estimating space-time Hawkes point process models of earthquake occurrences. Along these lines, our primary goals are,

1. The proposal of novel ways to incorporate a nonstationary background rate into the MISD algorithm.
2. Adding error bars to the histogram estimates of the triggering function which quantify the sampling variability and bias in the estimation of the underlying seismic process.

The original MISD algorithm assumes that the background rate for mainshocks is a stationary Poisson process in time and space. While an estimate of the mean mainshock rate over an observation region is useful, the expansion of MISD to incorporate a nonstationary background component is an important next step and improvement by allowing for localized estimates mainshock activity related to variations in the underlying tectonic field and the locations major faults. Moreover, an estimate of a spatially varying background rate can be used to identify regions with a persistent and heightened incidence of large seismic events, independent of aftershock clustering features which diminish over time.

Nichols and Schoenberg (2014) proposed a way to adjust MISD to incorporate nonstationarities in the background process by initially kernel smoothing over all events in the catalog and weighing each event by its corresponding kernel estimate. However, a main shortcoming of this approach is that nonstationarities in the background rate are only

defined on the observed data and not at each pixel of the observation window. Moreover, the authors of this work were primarily interested in applying the method to evaluate the dependence between the magnitudes of earthquakes and their aftershocks, and the explicit assessment or validation of the proposed estimation technique was not addressed.

In this paper, we propose two novel ways to incorporate a spatially varying background rate into the MISD method. First, we discuss a histogram estimator approach, which is a natural extension of the stationary rate estimator of Marsan and Lengliné (2008). Second, we apply the variable kernel estimator, used by Zhuang, Ogata and Vere-Jones (2002) for semi-parametric estimation, into the context of MISD. We validate and assess new methodology by simulating earthquake catalogs from a space-time model (ETAS) and evaluating the ability of each method to recover the true form of the nonstationary background rate and triggering function governing the simulations.

Simulation is also a powerful tool for understanding the statistical properties of the histogram estimators of the triggering function. By repeatedly simulating and re-estimating an earthquake model error bars can be computed which capture the sampling distributions of the estimates. An easily calculated analytic approximation for the error bars found through simulation is discussed at the end of the paper.

This paper is organized as follows: In Section 2, we provide an overview of space-time point process models of seismicity. In Section 3, we describe our modified version of the MISD algorithm, and propose a couple new ways to incorporate a nonstationary background rate. In Section 4, we validate and assess our methods with simulation studies, and discuss boundary issues. In Section 5, we apply our method to an earthquake dataset from the Tohoku District in Japan. In the Discussion Section we summarize and speculate about our results and suggest future directions and applications for this research.

2. Space-time Point Process Models. Consider a marked space-time point process $N(t,x,y)$ representing the times, locations, and magnitudes, $\{(t_i,x_i,y_i,m_i) : i = 1\cdots,N\}$, of earthquake occurrences. Space-time point process models of seismicity are usually specified with a conditional intensity function which is defined as the infinitesimal expected rate at which events occur around (t,x,y) given the history of the process $H_t = \{(t_i,x_i,y_i,m_i) : t_i < t\}$. That is,

$$\begin{aligned}&\lambda(t,x,y|H_t)\\ &= \lim_{\Delta t,\Delta x,\Delta y\downarrow 0} \frac{E[N\{(t,t+\Delta t)\times(x,x+\Delta x)\times(y,y+\Delta y)\}|H_t]}{\Delta t\Delta x\Delta y}.\end{aligned} \tag{1}$$

Conditional intensities are a natural way to model point processes as all finite-dimensional distributions of a simple point process are uniquely determined by its condition intensity (Daley and Vere-Jones, 2003).

In seismology, one typically models the conditional intensity in (1) as a Hawkes-type

self-exciting point process taking the following form:

$$\begin{aligned} \lambda(t,x,y,m|H_t) &= J(m)\lambda(t,x,y|H_t) \\ \lambda(t,x,y|H_t) &= \mu(x,y) + \sum_{\{i:t_i<t\}} \nu(t-t_i, x-x_i, y-y_i; m_i). \end{aligned} \tag{2}$$

For example, models of this type, referred to as Epidemic Type Aftershock Sequences (ETAS) models, were introduced by Ogata (1988) for the description of earthquake catalogs. Such models categorize earthquake occurrences into two types: mainshocks and aftershocks. The rate of mainshocks occurring over a spatial region is modeled by the background intensity $\mu(x,y)$, which is assumed a nonstationary Poisson process in space and stationary in time. The rate of aftershock activity following an earthquake occurring at (t_i, x_i, y_i) with magnitude m_i is modeled by the triggering function ν, which is often assumed Gaussian or power-law in parametric models. The summation term gives the contribution of all previously occurring events in the catalog to the overall rate of seismicity at time t and location (x,y). The distribution of earthquake magnitudes $J(m)$ is typically assumed independent of all other model components, and follows an exponential distribution according to the well known magnitude frequency law of Gutenberg and Richter (1944). Note that model (2) specifies a space-time branching process since any earthquake occurrence (including an aftershock) is capable of triggering its own aftershock sequence.

Ogata (1998) considered many parameterizations of the response function of (2) which take the following standard form:

$$\nu(t-t_i, x-x_i, y-y_i; m_i) = \kappa(m_i) g(t-t_i) f(x-x_i, y-y_i; m_i). \tag{3}$$

Here $\kappa(m_i)$ is the magnitude productivity function, which gives the expected number of aftershocks following an earthquake of magnitude m_i. The temporal component g is a probability density function governing the rate of aftershocks following an earthquake at time t_i. The spatial component f is a probability density function for the spatial distribution of aftershocks occurring around an earthquake with epicenter (x_i, y_i). The dependence of the spatial response function on the magnitude m_i is built into some models.

One example of a parametrization of the triggering function for ETAS is given by:

$$\kappa(m) = Ae^{\alpha(m-m_c)}, \tag{4}$$

$$g(t) = (p-1)c^{(p-1)}(t+c)^{-p}, \tag{5}$$

$$f(x,y) = \frac{(q-1)d^{q-1}}{\pi}(x^2+y^2+d)^{-q}, \tag{6}$$

where m_c is the magnitude cut-off for the catalog, $t>0$, and (A, α, p, c, q, d) are parameters to be estimated. Here g corresponds to the modified Omori formula (see Utsu, Ogata and Matsu'ura (1995) for details), and f is isotropic (rotation invariant) with a long range power-law decay rate.

The parameters of model (2) can be estimated by maximizing the log-likelihood function (Ogata, 1998) with respect to the parameters of the model:

$$log(L) = \sum_{i=1}^{N} \log(\lambda(t_i, x_i, y_i|H_t)) - \int_0^T \int \int_S \lambda(t, x, y|H_t) dx dy dt, \tag{7}$$

where $S \times [0, T]$ is the space-time observation window. In practice (7) can be maximized using the EM algorithm (Veen and Schoenberg, 2008) or numerical routines (Schoenberg, 2013) that optimize an approximation to (7). The nonstationary background component, $\mu(x, y)$, can be estimated with nonparametric techniques such as bi-cubic B-splines (Ogata, 1998) or kernel smoothing (Zhuang, Ogata and Vere-Jones, 2002; Musmeci and Vere-Jones, 1992). The techniques for estimating $\mu(x, y)$ are often implemented in conjunction with a declustering algorithm used to isolate mainshocks.

Marsan and Lengliné (2008) proposed the MISD algorithm to nonparametrically estimate the triggering function ν and stationary background rate $\mu(x, y) = \mu$ for the space-time Hawkes process model (2). Marsan and Lengliné (2010) showed that their method is an EM-type algorithm under the assumption that the background rate is stationary and the triggering function is piecewise constant. For the E-step, the branching structure of the process is estimated by computing the probabilities, for each pair (i, j) of earthquakes, of earthquake i having directly triggered earthquake j, as well as the probability of being a mainshock for each observed earthquake. For the M-step, the estimated branching structure is used to update an estimate of the stationary background rate and triggering function with probability weighted histogram estimators. The two-step procedure is repeated until the algorithm converges. A similar method is discussed in Mohler et al. (2011) using a Monte-Carlo based approach that alternates between sampling a realization of the estimated branching structure and updating estimates of the background rate and triggering function using kernel density estimation on the sampled data.

3. Nonparametric Methods. This section discusses the nonparametric method of Marsan and Lengliné (2008) to estimate the space-time Hawkes process model (2) using histogram estimators. We make the following modifications to the original algorithm:

1. We incorporate a nonstationary background rate;
2. We assume the separability of the triggering function into components for magnitude, time, and distance;
3. We perform histogram density estimation on the temporal and spatial triggering components $g(t)$ and $f(r)$, where $r = \sqrt{x^2 + y^2}$.

The above modifications make the method consistent with estimating the standard form of the triggering function in (3). As in Marsan and Lengliné (2008), we assume the spatial triggering component is isotropic, that is $f(x, y) = f(x^2 + y^2)$; this means the rate of aftershock activity following an earthquake only depends on the distance r from the

earthquake's epicenter and not direction (circular aftershock regions). Also, to be consistent with model (2), the background component $\mu(x, y)$ is assumed nonstationary in space and stationary in time.

3.1. *Histogram Estimators.* Let P be a $N \times N$ lower triangular probability matrix with entries,

$$p_{ij} = \begin{cases} \text{probability earthquake } i \text{ is an aftershock of } j, & i > j \\ \text{probability earthquake } i \text{ is a mainshock}, & i = j \\ 0, & i < j \end{cases} \tag{8}$$

$$P = \begin{bmatrix} p_{11} & 0 & 0 & \cdots & 0 \\ p_{21} & p_{22} & 0 & \cdots & 0 \\ p_{31} & p_{32} & p_{33} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ p_{N1} & p_{N2} & p_{N3} & \cdots & p_{NN} \end{bmatrix} \qquad P^{(0)} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1/2 & 1/2 & 0 & \cdots & 0 \\ 1/3 & 1/3 & 1/3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ 1/N & 1/N & 1/N & \cdots & 1/N \end{bmatrix}$$

The only constraint for matrix P is $\sum_{j=1}^{N} p_{ij} = 1$. The rows must sum to 1 since each earthquake in the branching process is either a mainshock or an aftershock of a previously occurring earthquake. $P^{(0)}$ is one possible initialization. For this matrix, $\sum_{i=1}^{N} p_{ii}$ can be interpreted as the estimated number of mainshocks, while $\sum_{i=1}^{N} \sum_{j=1}^{i-1} p_{ij}$ (sum of the non-diagonal elements) is the estimated number of aftershocks.

Below is the MISD algorithm of Marsan and Lengliné (2008) with the modifications specified in the beginning of this section. For the spatial component, we specify a histogram density estimator of $h(r) = 2\pi r f(r)$ since $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy = \int_{0}^{\infty} 2\pi r f(r) dr = 1$; here $h(r)$ represents the underlying probability density function for the distance r between an earthquake and its aftershock.

Algorithm 1.

1. Initialize $P^{(0)}$, set iteration index $v = 0$.
2. Estimate nonstationary background rate $\mu(x, y)$:

$$\mu_{k,l}^{(v)} = \frac{1}{T \Delta x \Delta y} \sum_{D_{k,l}} p_{ii}^{(v)}, \quad k = 0, \cdots, n_x^{bins} - 1, \; l = 0, \cdots, n_y^{bins} - 1.$$

3. Estimate triggering components $\kappa(m)$, $g(t)$, and $h(r)$:

$$\kappa_k^{(v)} = \frac{\sum_{A_k} p_{ij}^{(v)}}{N_k^{mag}}, \quad k = 0, \cdots, n_m^{bins} - 1;$$

$$g_k^{(v)} = \frac{\sum_{B_k} p_{ij}^{(v)}}{\Delta t_k \sum_{i=1}^{N} \sum_{j=1}^{i-1} p_{ij}^{(v)}}, \quad k = 0, \cdots, n_t^{bins} - 1;$$

$$h_k^{(v)} = \frac{\sum_{C_k} p_{ij}^{(v)}}{\Delta r_k \sum_{i=1}^{N} \sum_{j=1}^{i-1} p_{ij}^{(v)}}, \quad k = 0, \cdots, n_r^{bins} - 1.$$

4. Update probabilities $P^{(v+1)}$, letting r_{ij} be the epicentral distance between earthquakes i and j and $f^{(v)}(r_{ij}) = h^{(v)}(r_{ij})/(2\pi r_{ij})$:

$$p_{ij}^{(v+1)} = \frac{\kappa^{(v)}(m_j) g^{(v)}(t_i - t_j) f^{(v)}(r_{ij})}{\mu^{(v)}(x_i, y_i) + \sum_{j=1}^{i-1} \kappa^{(v)}(m_j) g^{(v)}(t_i - t_j) f^{(v)}(r_{ij})} \text{ for } i > j,$$

$$p_{ii}^{(v+1)} = \frac{\mu^{(v)}(x_i, y_i)}{\mu^{(v)}(x_i, y_i) + \sum_{j=1}^{i-1} \kappa^{(v)}(m_j) g^{(v)}(t_i - t_j) f^{(v)}(r_{ij})}.$$

5. If $\max_{i,j} |p_{ij}^{(v+1)} - p_{ij}^{(v)}| < \epsilon$, where $i \leq j$, then the algorithm has converged (in practice we take $\epsilon = 10^{-3}$). Otherwise, set $v \leftarrow v + 1$ and repeat steps 2–5 until convergence.

For step 2 of Algorithm 1, the notation is defined as follows:

- n_x^{bins} and n_y^{bins} are the number of bins along the x and y axis for the 2-dimensional histogram estimator of $\mu(x, y)$ ($n_x^{bins} \cdot n_y^{bins}$ bins total).
- $D_{k,l} = \{i : k\Delta x < x_i \leq (k+1)\Delta x, l\Delta y < y_i \leq (l+1)\Delta y\}$ where Δx and Δy are the fixed bin widths along the x and y axes.

For step 3 of Algorithm 1, the notation is defined as follows:

- n_m^{bins}, n_t^{bins}, and n_r^{bins} are the number of bins for the the histogram estimators of the magnitude κ, temporal g, and spatial h components of the triggering function.
- $A_k = \{(i, j) : \delta m_k < m_j \leq \delta m_{k+1}, i > j\}$ is the set of indices of all pairs of earthquakes whose mainshock magnitudes fall within the k^{th} bin $(\delta m_k, \delta m_{k+1}]$ of the histogram estimator for $\kappa(m)$, where $\Delta m_k = \delta m_{k+1} - \delta m_k$ is the bin width.
- $N_k^{mag} = \sum_{j=1}^{N} I(\delta m_k < m_j \leq \delta m_{k+1})$ is the number of earthquakes whose magnitudes fall within the interval $(\delta m_k, \delta m_{k+1}]$.
- $B_k = \{(i, j) | \delta t_k < t_i - t_j \leq \delta t_{k+1}, i > j\}$, is the set of indices of all pairs of earthquakes whose time differences fall within the k^{th} bin $(\delta t_k, \delta t_{k+1}]$ of the histogram estimator for $g(t)$, where $\Delta t_k = \delta t_{k+1} - \delta t_k$ is the bin width.
- $C_k = \{(i, j) | \delta r_k < r_{ij} \leq \delta r_{k+1}, i > j\}$ is the set of indices of all pairs of earthquakes whose epicentral distances r_{ij} fall within the k^{th} bin $(\delta r_k, \delta r_{k+1}]$ of the histogram estimator for $h(r)$, where $\Delta r_k = \delta r_{k+1} - \delta r_k$ is the bin width.

In step 2 of Algorithm 1 the nonstationary background rate is estimated with a histogram estimator which is a generalization of the stationary estimator in the original MISD algorithm. In our modified method, the spatial observation window S is partitioned into equally sized cells of width Δx and height Δy. The estimated rate within each cell is given by the sum of the background probabilities, p_{ii}, corresponding to earthquakes occurring within that cell, and then dividing the sum by $\Delta x \cdot \Delta y \cdot T$ to give the rate of mainshocks per unit area per unit time. Note, the histogram estimator in step 2 reduces to the stationary case in Marsan and Lengliné (2008) when $n_x^{bins} = n_y^{bins} = 1$ and $\Delta x \cdot \Delta y = S$ (i.e. only one cell equal to the spatial observation window is specified). Also note that the estimator of g is itself a density since $\sum_{k=0}^{n_t^{bins}-1} \Delta t_k \hat{g}_k = 1$, and similarly for the histogram estimator of h.

The assumption of separability allows for robust computation of model components by substantially reducing the number of bins needed to estimate the model (only a one-dimensional support is needed for the histogram estimator of each triggering component). Furthermore, since we perform histogram density estimation on g and f the output of Algorithm 1 has meaningful interpretation as in Ogata (1998). For instance, the histogram estimate of the magnitude productivity $\hat{\kappa}_k(m)$ has the natural interpretation as the estimated mean number of aftershocks directly triggered by an earthquake with magnitude m falling in the k^{th} magnitude bin $(\delta m_k, \delta m_{k+1}]$.

3.2. *Variable Kernel Estimation.* A shortcoming of the histogram method for estimating the background rate in Algorithm 1 is the implicit assumption of constancy within each bin. If a large mainshock occurs, then the contribution of that event to the background seismicity is limited to the bin in which the event is contained. If a bin does not contain any earthquake events, then the estimated rate of mainshocks in that bin is zero. Hence, the method does not allow for the estimate to vary smoothly over the spatial observation region and is highly dependent on the choice of the partition. This motivates considering a kernel smoothing approach, where the background rate estimate only depends on the choice of the smoothing parameter (bandwidth) and varies continuously over the pixels in the spatial window.

As an alternative to the histogram approach (Algorithm 1, step 2) for estimating the nonstationary background rate, we adopt the variable bandwidth kernel estimator used by Zhuang, Ogata and Vere-Jones (2002):

$$\mu(x,y) = \gamma \tau(x,y), \tag{9}$$

$$\tau(x,y) = \frac{1}{T} \sum_{i=1}^{N} p_{ii} k_{d_i}(x - x_i, y - y_i). \tag{10}$$

Here the index i runs through all the events in the catalog, γ is a scaling factor, and k is the Gaussian kernel function,

$$k_{d_i}(x,y) = \frac{1}{2\pi d_i^2} \exp\left(-\frac{x^2+y^2}{2d_i^2}\right).$$

The kernel is weighted by p_{ii}, the probability that event i is a mainshock, and has a varying bandwidth d_i specified for each event in the catalog. The bandwidth d_i is computed by finding the radius of the smallest disk centered at (x_i, y_i) that contains at least n_p other events, and is greater than some small value ϵ representing the location error. Zhuang, Ogata and Vere-Jones (2002) suggest taking n_p between 15–100 and $\epsilon = 0.02$ degrees. A variable bandwidth estimate is preferred since a large fixed bandwidth over-smooths areas with clustered events, and a small fixed bandwidth under-smooths areas with sparsely located events.

In Zhuang, Ogata and Vere-Jones (2002) the estimate (9) is part of a semi-parametric model for ETAS, with parameters estimated via maximum likelihood. Since our approach is completely nonparametric, the scaling factor γ for the estimate of the background rate needs to be carefully defined. This leads to the following algorithm for estimating the space-time Hawkes process model (2) with a variable kernel estimator for the background seismicity:

Algorithm 2.

1. Initialize $P^{(0)}$ and compute d_i for each event $i = 1, \cdots, N$.
2. Estimate nonstationary background rate $\mu(x, y)$:

$$\mu^{(v)}(x, y) = \frac{\sum_{i=1}^{N} p_{ii}^{(v)}}{Z^{(v)}} \tau^{(v)}(x, y).$$

3. Follow Steps 3–5 in Algorithm 1.

The normalizing factor $Z^{(v)}$ at iteration v is chosen so that

$$\frac{1}{Z^{(v)}} \int_0^T \int \int_S \tau^{(v)}(x, y) dx dy dt = 1,$$

and consequently,

$$\int_0^T \int \int_S \mu^{(v)}(x, y) dx dy dt = \sum_{i=1}^{N} p_{ii}^{(v)},$$

where $\sum_i p_{ii}^{(v)}$ is the estimated number of mainshocks occurring in the space-time observation window. In practice, $Z^{(v)}$ can be found by first computing $\tau^{(v)}(x, y)$ as defined in (10) at each pixel, and then evaluating the integral of $\tau^{(v)}(x, y)$ over $S \times [0, T]$ with a Reimann sum over those pixels.

4. Simulation Results.

4.1. *Histogram Estimator Method.* In this section we assess the performance of the nonparametric method described in Algorithm 1 to recover an earthquake model from synthetic catalogs. For this study, earthquake occurrences are simulated from the ETAS

model with parametric triggering function given by (4, 5, 6). The parameter values are the maximum likelihood estimates $(A, \alpha, p, c, d, q) = (0.322, 1.407, 1.121, 0.0353, 0.0159, 1.531)$ from Table 2, row 8 of Ogata (1998) (parameters estimated from earthquake data over a $36 \sim 42°$N latitude and $141 \sim 145°$E longitude region off the east coast of Tohoku District, Japan with time span 1926–1995). Earthquake magnitudes are generated independently of other model components according to an exponential density $J(m) = \beta e^{-\beta(m-m_c)}$ with $\beta = ln(10)$ (equivalent to a Gutenberg-Richter b-value equal to 1). The observation window for the simulation is $S \times T = [0, 4] \times [0, 6] \times [0, 25000]$, and the magnitude cut-off is $m_c = 0$. The nonstationary background rate is specified by partitioning the spatial observation window S into 4 equally sized cells with the varying rates shown in Figure 2(a). An example of a simulated realization is shown in Figure 1. For a description of the simulation procedure for ETAS please see Algorithm C from Zhuang, Ogata and Vere-Jones (2004).

We simulate and re-estimate the ETAS model 200 times to assess the variability in the estimates over multiple realizations of earthquake catalogs from the specified model. The results for the estimation of the nonstationary background rate are shown in Figure 2(b). Each cell in Figure 2(b) shows the 0.025 and 0.975 quantiles of the estimates, and the mean of the estimates which correspond to the cell's grey scale level. Figure 2(b) reveals that the nonparametric method (Algorithm 1) is able to recover the sharp differences between the rates in each cell with reasonably small errors. While the means of the estimates are close to the true rates (Figure 2(a)), a bias is apparent, as the 95% coverage intervals fall consistently above the true values. In the next section we show that this over-estimation is due to boundary effects induced by excluding aftershocks that occur outside the space-time observation window.

The results for the estimation of the magnitude, temporal, and distance components of the triggering function from the 200 realizations of ETAS are shown in Figure 3. The histogram density estimates of $g(t)$ and $h(r)$ are plotted on log-log scales with equally spaced logarithmic bins since the true densities are power-law. The bins for the histogram estimator of the magnitude productivity $\kappa(m)$ are also equally spaced.

The method is able to recover the form of each component of the triggering function since the true value governing the simulations is contained within the 95% coverage error bars for most bins. The error bars for the estimates of $g(t)$ and $h(r)$ reveal that the estimation is most accurate in the middle range. The high variability in the estimates for bins corresponding to small time differences t and distances r is not surprising since the partition is logarithmically scaled, and therefore these bin widths are very small. The error bars at the right-tail ends of the distributions of $g(t)$ and $h(r)$ do not cross and underestimate the true densities. In the next section we show that this estimation bias is due to boundary issues. The error bars for the estimates of the magnitude productivity function $\kappa(m)$ increase with magnitude, although this is expected since earthquake magnitudes are exponentially distributed and therefore only a few large magnitude events occur in each simulation.

4.2. *Boundary Issues.* When simulating earthquake catalogs from the ETAS model the mainshocks are restricted to occur within the space-time observation window $S \times [0, T]$. However, the times and locations of aftershocks, simulated from the triggering function components g and f, may occur outside of this boundary. In the last section, we neglected boundary effects, and only used simulated data occurring within the space-time observation window to estimate the model using Algorithm 1.

To evaluate the boundary effects on the estimation we include simulated aftershocks which occur within a distance ϵ_r of the spatial boundary and a time ϵ_t of the temporal boundary, i.e. all aftershocks occurring within $[-\epsilon_r, 6 + \epsilon_r] \times [-\epsilon_r, 4 + \epsilon_r] \times [0, 25000 + \epsilon_t]$. We then run Algorithm 1 on the expanded simulation data, and slightly modify step 4 so that $\mu(x_i, y_i) = 0$ if event (t_i, x_i, y_i, m_i) falls outside of $S \times [0, T]$.

To measure the change in performance of Algorithm 1 on estimating the nonstationary background rate as we increase ϵ_r and ϵ_t we use the root-mean-square deviation (RMSD):

$$\sqrt{\frac{1}{n_x^{bins} n_y^{bins}} \sum_{i,j} (\hat{\mu}_{ij} - \mu_{ij})^2}. \tag{11}$$

Here $\hat{\mu}_{ij}$ and μ_{ij} are the estimate and true value for the background rate in the (i, j) cell respectively. We simulate the ETAS model 10 times using the same parameters and background rate as in Section 4.1, with mainshocks again restricted to $S \times [0, T] = [0, 4] \times [0, 6] \times [0, 25000]$, but aftershocks allowed to occur outside that region. For each simulation, the RMSD is computed for increasing values of ϵ_r and ϵ_t. Figure 4 shows the mean RMSD from the 10 realizations at selected values of ϵ_r and ϵ_t; the vertical lines represent a standard deviation in RMSD above and below the mean. The incorporation of aftershocks falling outside the space-time observation window significantly improves the performance of the estimation of the background rate. The RMSD appears to level off when $\epsilon_r = 10^{0.5} = 3.16$ and $\epsilon_t = 10000$.

Figure 5 shows the results from simulating and re-estimating ETAS with Algorithm 1 200 times with a boundary correction of $\epsilon_r = 1000$ and $\epsilon_t = 10^6$. Again, we simulate events with the same parameters and space-time window as Section 4.1. The only difference is that in the estimation we use aftershocks occurring within a distance $\epsilon_r = 1000$ and time $\epsilon_t = 10^6$ of the boundary of the observation window. Since the temporal and distance components of the triggering function used to generate the data are power-law it is possible for aftershocks to occur at very far distances and times from the observation window.

The background rate estimate in Figure 5 is a substantial improvement over the estimate in Figure 2, which neglected boundary effects. The true rates are contained in the 95% coverage intervals for each cell in Figure 5. Moreover, the consistent over-prediction of the rates evident in Figure 2 is no longer present, and the results suggest the bias in the cell means is negligible once the boundary effects are accounted for. The error bars for the triggering function components in Figure 5 also show substantial improvement when compared to Figure 3. The histogram estimates for $g(t)$ and $h(r)$ contain the true density

values for large time differences t and distances r. Moreover, accounting for boundary effects expands the reach of the estimation (histogram estimates at bins beyond $r = 10$ and $t = 25000$) and reduces the error at the tail ends. Lastly, the error bars for the estimates of the magnitude productivity $\kappa(m)$ in Figure 5 appear more centered around the true value than in Figure 3.

4.3. *Variable Kernel Estimation Method.* In this section we use simulation to assess the ability of Algorithm 2 to recover the components of the space-time Hawkes process model (2) with a smoothly varying background rate. Here we simulate from a parametric ETAS model with the same triggering function and parameter values as Section 4.1. However, instead of the background rate in Figure 2(a) with stationary rates in each cell on a 2×2 grid, we simulate from the smoother background rate shown in Figure 7(a). This nonstationary background rate was generated by performing fixed bandwidth kernel density estimation over the locations of 883 earthquakes of magnitude 5.0 or greater, longitude $141 \sim 145°$E, latitude $36 \sim 42°$N, and time between 16 Jan 2007 to 28 Dec 2014.[1] To simulate from the kernel smoothed background rate in Figure 7(a) we use the thinning procedure of Lewis and Shedler (1979) and set the expected number of background events equal to 2000.

Figure 7(b) shows the probability weighted variable kernel estimate (Algorithm 2, step 2) of the nonstationary background rate from a single simulated realization of the ETAS model. The epicentral location and space-time plots of the simulated earthquake data used for this estimate are shown in Figure 6. The kernel estimate of the background rate depends on the smoothing parameter n_p (Section 3.2). Here we choose $n_p = 50$, since this value gives the lowest RMSD (11) for $n_p \in \{10, 15, \cdots, 95, 100\}$. The kernel estimates are evaluated on a 100×100 pixel grid (making $n_x^{bins} = n_y^{bins} = 100$ when evaluating (11)).

As discussed in Section 3.2, the nonparametric estimation of ETAS is sensitive to boundary effects. As a boundary correction for the estimation with Algorithm 2, we allow for aftershocks occurring within $\epsilon_r = 3$ degrees and $\epsilon_t = 3000$ days of the space-time boundary $S \times [0, T] = [0, 4] \times [0, 6] \times [0, 25000]$. Note that the selected values, $\epsilon_r = 3$ and $\epsilon_t = 3000$, correspond to where the RMSD in Figure 4 begins to level off. The panels in Figure 6 show the boundary (dashed rectangles) and simulated aftershocks occurring in the specified region outside the boundary.

This estimate in Figure 7(b) resembles the overall form of the true background intensity Figure 7(a) and recovers many of the mainshock hotspots. However, near location (2.06, 2.33), a hotspot appears to have been erroneously estimated, i.e. a false positive has been identified. This is due to the large magnitude event ($m > 4$) that occurred in the simulation at this location, as denoted by the asterisk in Figure 6(a). The mean of 200 estimates of the background rate from 200 simulated realizations of ETAS is shown in Figure 7(c), and appears to closely resemble the true background rate. Hence, while there may be

[1]Data gathered from http://www.quake.geo.berkeley.edu/anss/catalog-search.html with spatial observation window the same as Ogata (1998).

discrepancies for estimates from a single realization due to sampling variation, the variable kernel estimator appears to be unbiased since the mean of the estimates from repeated simulation is close to the true background intensity. Moreover, the pointwise 0.025 and 0.975 quantiles for 200 estimates of the number of background events, given by $(1943.7, 2219.4)$, contains the true value of 2000 background events specified for the simulation.

The histogram estimates and corresponding 95% coverage error bars in Figure 8 appear to successfully describe the true triggering components. This demonstrates the ability of Algorithm 2 to recover the nonstationary background rate with a variable kernel estimator and triggering function with histogram estimators. There are slight discrepancies between the histogram estimates of the triggering function and the true values due to boundary effects. Most noticeably, the 95% coverage error bars for the estimates of $g(t)$ and $h(r)$ do not contain the true density values at the right-tail ends of the distributions. Boundary correction values larger than $\epsilon_r = 3$ and $\epsilon_t = 3000$ may result in more accurate estimates, as in the asymptotic case shown in Figure 5. However, the selected values seem sufficient for estimating the background intensity.

5. Application to Japan Dataset. We apply the MISD algorithm (Algorithm 2) to earthquake data from the ANSS catalog `http://www.quake.geo.berkeley.edu/anss/catalog-search.html`. The dataset contains 6075 earthquakes of magnitude 4.0 or greater occurring over a 10 year period between 5 Jan 2005 – 31 Dec 2014. The spatial widow is a $141 \sim 145°$E longitude and $36 \sim 42°$N latitude region off the east coast of the Tohoku District in northern Japan. This is the same spatial region analyzed in Ogata (1998), although the time window in this study is different. An epicentral and space-time plot of the data is show in Figure 9, with the asterisk corresponding to the 2011 magnitude 9.0 Tohoku earthquake.

The variable kernel estimate of the background rate (Algorithm 2, step 2) is shown in Figure 10. Here we chose the smoothing parameter $n_p = 50$, corresponding to the best choice for the simulation study in Section 4.3. Figure 10 is an important plot for assessing seismic risk since it shows the estimate of the underlying spatial Poisson processes $\mu(x, y)$ for maishock activity which persists over time in the region. In total, the algorithm estimated there to have been 809 mainshocks, or 13.3% of the total seismicity; this suggests that most of the events in the dataset are aftershocks, temporally and spatially linked to previously occurring earthquakes.

The histogram estimates of the components of the triggering function are shown in Figure 11. The grey error bars approximate ± 2 standard errors, capturing the sampling variation for the histogram estimates in each bin (see Appendix A for the derivation of the analytic standard errors). The estimates of $g(t)$ and $h(r)$ both exhibit power-law type distributions, and the error bars appear similar to the ones obtained in the simulation study (Section 4.3). Note, the estimates at the right-tail ends of these distributions ($t > 1000$ days and $r > 1$ degree) are perhaps unreliable and underestimate the truth due to boundary effects, as demonstrated in the simulation study (Section 4.2).

The estimate of the magnitude productivity function $\kappa(m)$ appears to follow an exponential form. The error in the estimation of the productivity increases with magnitude, as also demonstrated in the simulation study (Section 4.3). In the dataset there are only 3 events of magnitude 7.4 or greater, and hence large sampling variation for the estimates of the mean productivity for large magnitude events. The estimate in the last bin was estimated with only one event, namely the magnitude 9.0 Tohoku earthquake. It appears that the magnitude productivity for this event is underestimated; perhaps this is due to boundary effects since many of the aftershocks may have occurred outside the observation window.

Superimposed on Figure 11 are the parametric estimates of the ETAS model (4, 5, 6) for this same region from Table 2, row 11 of Ogata (1998). Amazingly, the parametric and nonparametric estimates agree closely. This suggests that seismicity in this region is well captured by an ETAS model with power-law $g(t)$ and $f(r)$, and exponential $\kappa(m)$. Since our dataset was gathered over a different time window than Ogata (1998), the results also suggest that properties of aftershock sequences in this region are rather invariant over time.

Note that in Figure 11 the nonparametric estimate of the triggering density $g(t)$ is slightly higher than what Ogata previously estimated for small time intervals t . This could perhaps be attributable to increased accuracy of seismometers in this region detecting aftershocks occurring shortly after large earthquakes more accurately than previously.

6. Discussion. The results of this article demonstrate that the MISD algorithm performs remarkably well at nonparametrically estimating space-time Hawkes process models (2) for earthquake occurrences. By repeatedly simulating and re-estimating a known parametric earthquake model (ETAS), we verified and evaluated novel ways to incorporate a nonstationary background rate into the method. Moreover, the error bars added to the histogram estimates of the triggering function captured the true values and showed reasonable sampling variation in the estimates over most bins. While the simulation results showed bias in the estimation of the background rate and right-tail ends of the triggering function components, this problem became noticeably less severe once boundary effects were taken into account.

A striking result in the application to earthquake data from the Tohoku region in Japan is that the nonparametric estimate matched closely with a previously estimated parametric form of the ETAS model. This further justifies the ETAS model as an adequate model of seismicity for the selected region in Japan.

The parametric forms for point process models in seismology are the result of many decades of refinement. However, for any given seismic region, a multitude of different parameterizations of ETAS may be considered. The nonparametric methods discussed in this paper can serve as a diagnostic to assess which parametrization is a good fit to the data. In other applications of self-exciting point processes, such as crime or finance, there is a less established literature on parametric models. In such applications, nonparametric estimation can be a powerful exploratory tool in determining a suitable parameterization of the

triggering function. The error bars on the histograms estimates can be used for statistical inference, and to identify places where the nonparametric estimate is more or less reliable as either a diagnostic or exploratory tool.

APPENDIX A: ANALYTIC ERROR BARS

Here we provide a derivation of an analytic approximation for computing standard errors for the histogram estimators of the triggering function components (error bars in Figure 11, Section 5). We proceed by first deriving an approximation of the standard error for the histogram estimator of $g(t)$, and then note that the standard errors for the histogram estimators of $\kappa(m)$ and $h(r)$ can be approximated similarly. Please use Section 2.1 as a reference for much of the notation in this appendix, and note that p_{ij} refers to the triggering probability (8) after Algorithm 2 has converged.

Let $t \in (\delta t_k, \delta t_{k+1}]$ and $\hat{g}(t) = g_k$ be the histogram density estimator of $g(t)$. Now suppose S_k is a random variable representing the number of triggered events in bin k, i.e. the number of aftershocks occurring between $(\delta t_k, \delta t_{k+1}]$ days after the earthquakes that directly trigger them. Then S_k follows a binomial distribution with number of trials n_t equal to the true number of triggered events (aftershocks) for the process, and success probability θ_k^g equal to the true probability an aftershock occurs between $(\delta t_k, \delta t_{k+1}]$ days after the earthquake that directly triggers it. Since we do not know the true values for the binomial parameters we estimate them with $\hat{n}_t = \sum_{i=1}^{N} \sum_{j=1}^{i-1} p_{ij}$ and $\hat{\theta}_k^g = \sum_{B_k} p_{ij}/\hat{n}_t$. Hence, an approximation of the variance of the histogram density estimator $g_k = S_k/(\Delta t_k n_t)$ is given by:

$$\widehat{Var}(g_k) = \frac{(\hat{\theta}_k^g)(1-\hat{\theta}_k^g)}{\hat{n}_t \Delta t_k^2}$$

Similarly, we can approximate the variances for the other histogram estimators:

$$\widehat{Var}(\kappa_k) = \frac{\hat{n}_t(\hat{\theta}_k^\kappa)(1-\hat{\theta}_k^\kappa)}{(N_k^{mag})^2}$$

$$\widehat{Var}(h_k) = \frac{(\hat{\theta}_k^h)(1-\hat{\theta}_k^h)}{\hat{n}_t \Delta r_k^2}$$

where $\hat{\theta}_k^\kappa = \sum_{A_k} p_{ij}/\hat{n}_t$ and $\hat{\theta}_k^h = \sum_{C_k} p_{ij}/\hat{n}_t$.

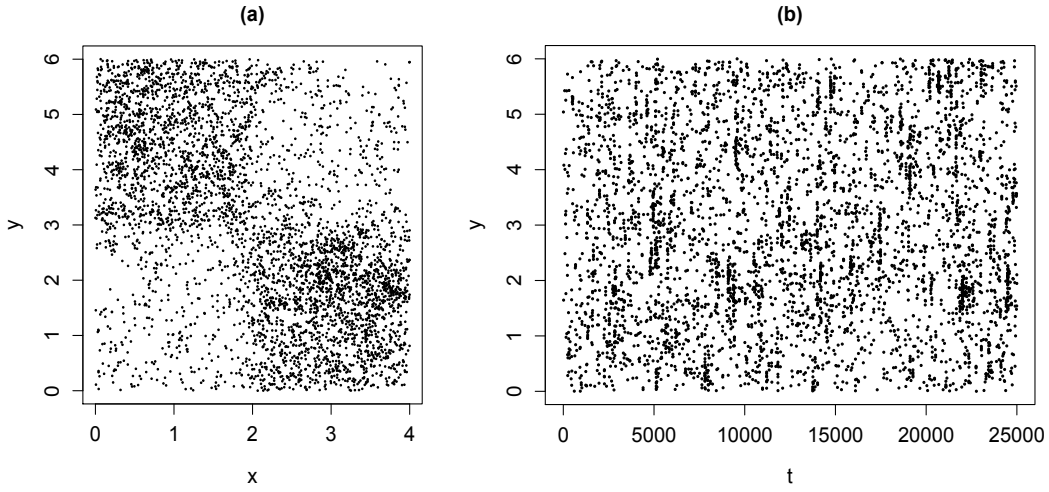

FIG 1. *Simulated realization of ETAS model (4–6) with background rate varying in each quadrant; (a) epicentral locations, and (b) space-time plot of simulated earthquakes.*

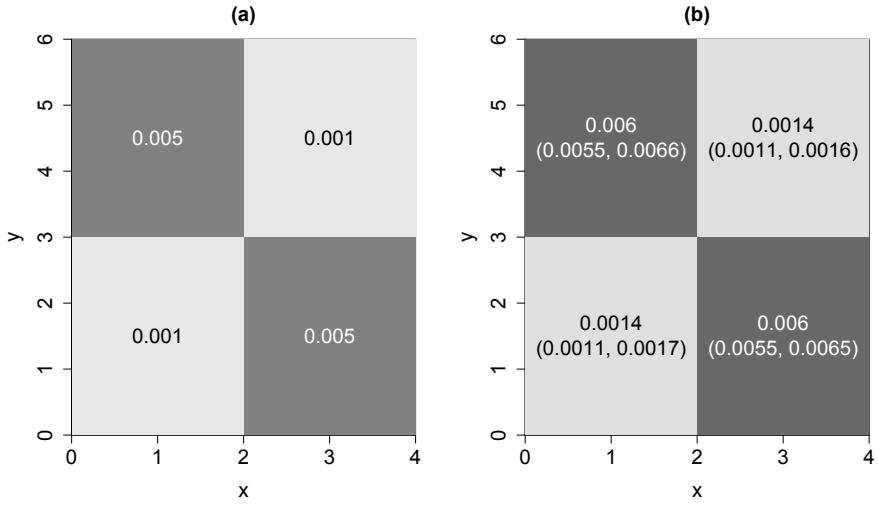

FIG 2. *(a) True background rate for simulation study in Section 4.1. (b) Results for estimating the background rate with Algorithm 1 from 200 simulations of ETAS. The means of the estimates printed in each cell correspond to the grey scale levels; the intervals are the 0.025 and 0.975 quantiles for the estimates in each cell.*

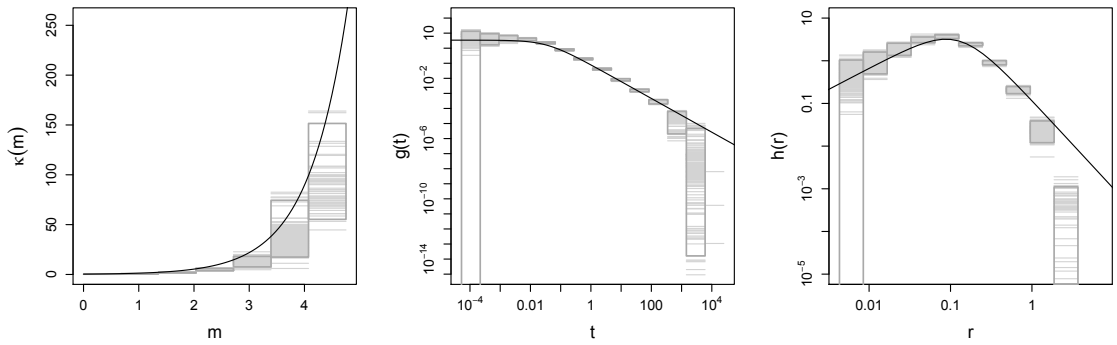

FIG 3. *Magnitude, temporal, and distance components for triggering function from the simulation study in Section 4.1. The black curves are the true triggering components used for the simulation. The light grey horizontal lines in each bin are the histogram estimates from the 200 simulations of ETAS; the dark grey boxes are the 95% coverage intervals (error bars) for the estimates in each bin (i.e. pointwise 0.025 and 0.975 quantiles).*

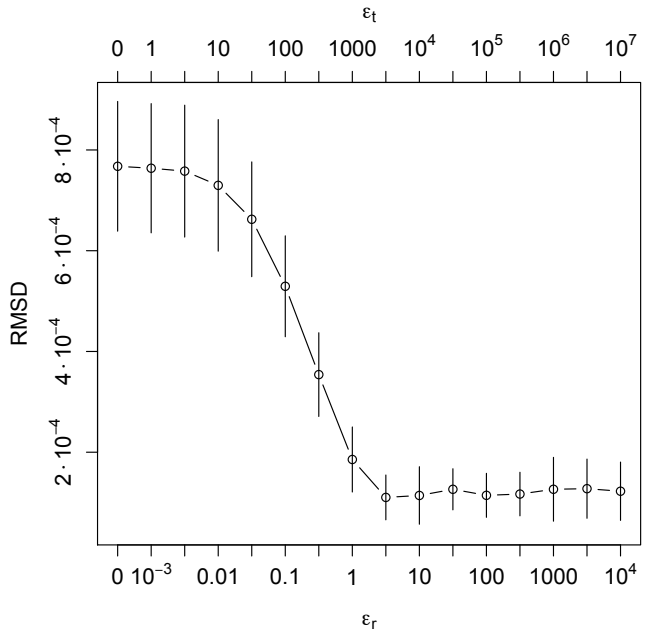

FIG 4. *RMSD of the background rate, equation (11), for increasing values of ϵ_r and ϵ_t. RMSDs are averaged from 10 realizations of ETAS; the vertical bars cover one standard deviation above and below the mean.*

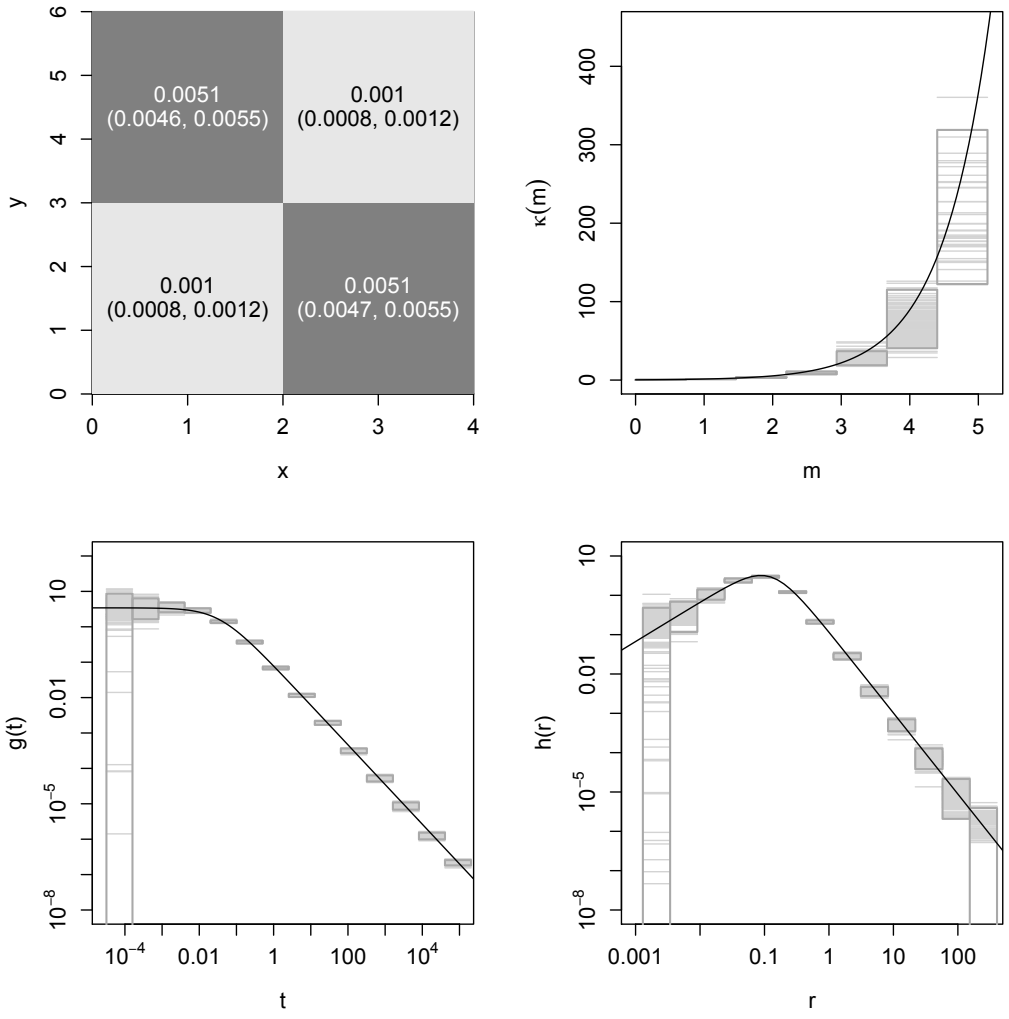

FIG 5. *Estimates of the background rate and triggering function components from 200 ETAS simulations, with boundary correction for aftershock activity* $\epsilon_r = 1000$ *and* $\epsilon_t = 10^6$.

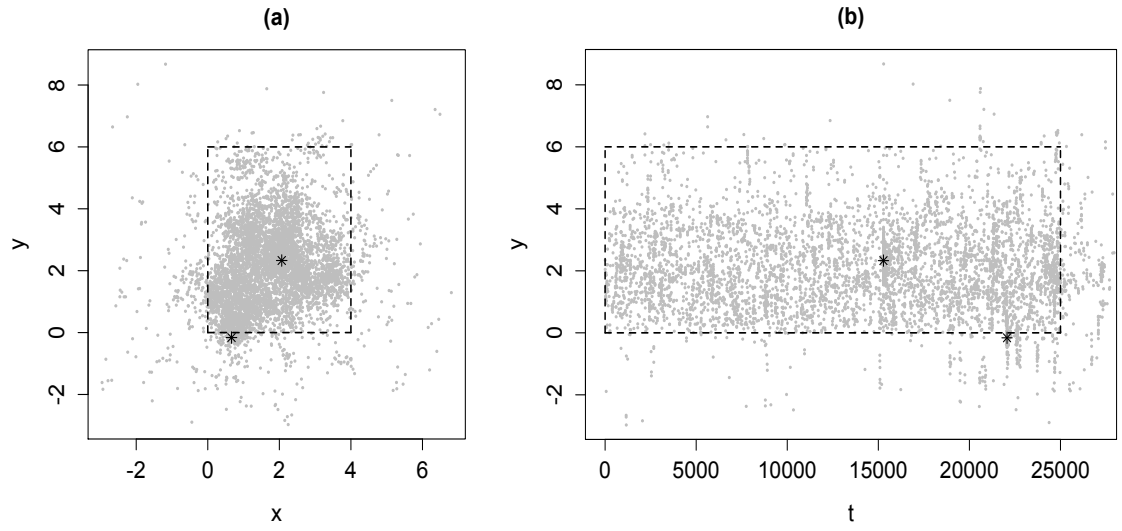

FIG 6. *Simulated realization of ETAS model (4–6) with smooth nonstationary background rate; (a) epicentral locations, and (b) space-time plot of simulated earthquakes. The dotted rectangles in each plot are the spatial and temporal boundaries for the observation window* $S \times [0, T] = [0, 4] \times [0, 6] \times [0, 25000]$. *Aftershocks occurring within a distance* $\epsilon_r = 3$ *and time* $\epsilon_t = 3000$ *of the boundary are plotted outside the rectangle. The asterisks denote events with magnitudes* $m > 4$.

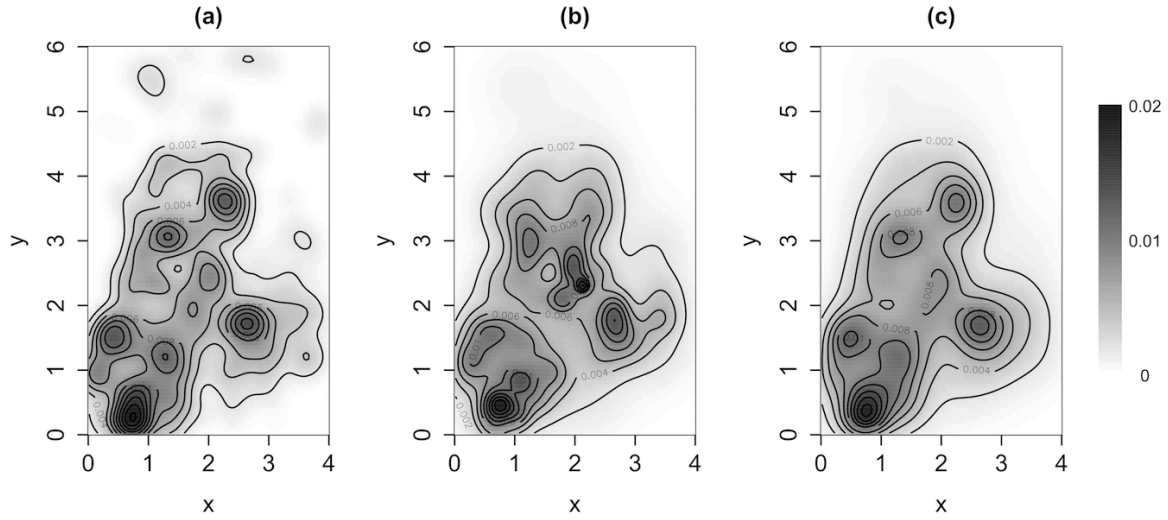

FIG 7. *(a) True background rate for simulation study in Section 4.2. (b) Estimate of background rate from one simulated realization of ETAS and, (c) mean estimate from 200 realizations.*

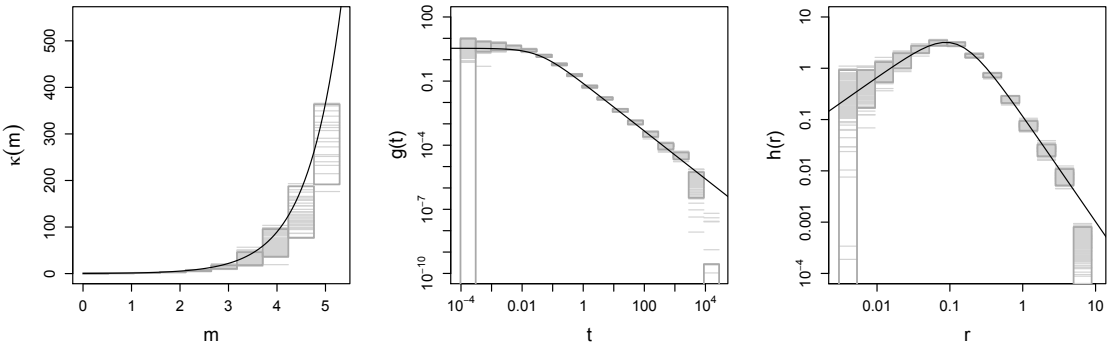

FIG 8. *Magnitude, temporal, and distance components for triggering function from the simulation study in Section 4.3. The black curves are the true triggering components used for the simulation. The light grey horizontal lines in each bin are the histogram estimates from the 200 simulations; the dark grey boxes are the 95% coverage error bars for the estimates.*

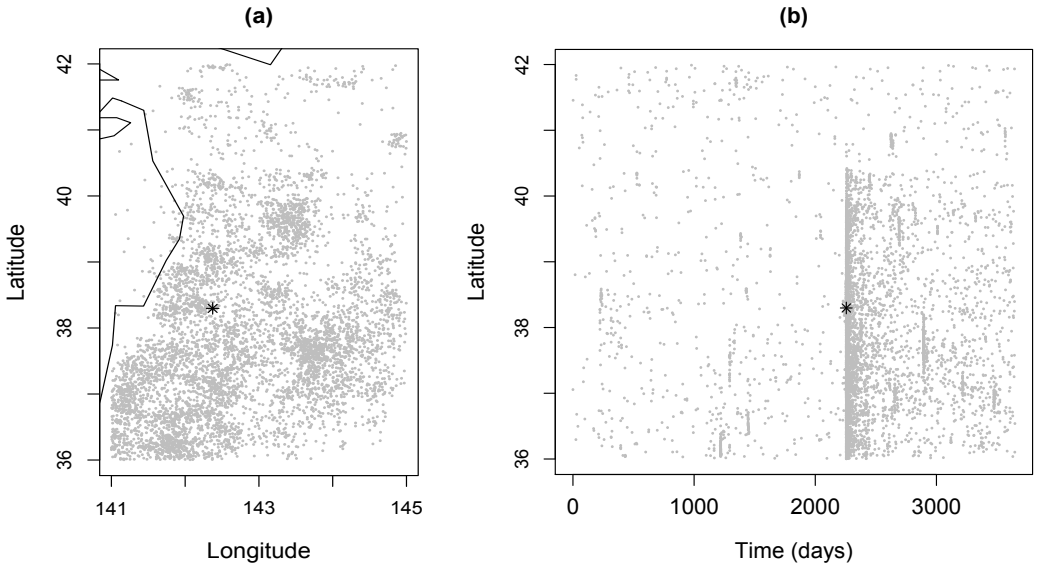

FIG 9. *Epicentral locations (a) and space-time plot (b) of earthquakes, magnitude 4.0 or greater, occurring off the east coast of the Tohoku District, Japan. The asterisk corresponds to the 2011 Tohoku earthquake of magnitude 9.0.*

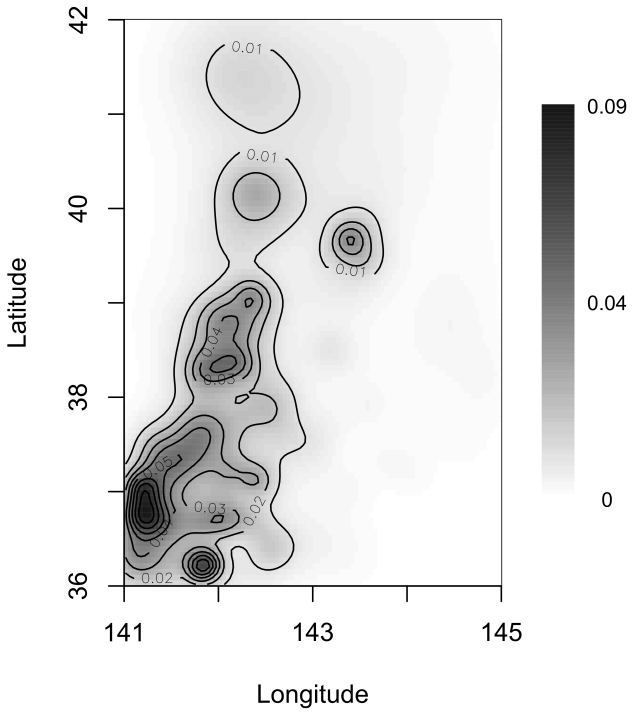

FIG 10. *Estimate of background rate (Algorithm 2, step 2) for Japan earthquake dataset (Section 5). Rate values are in events/day/degree*^{2}.

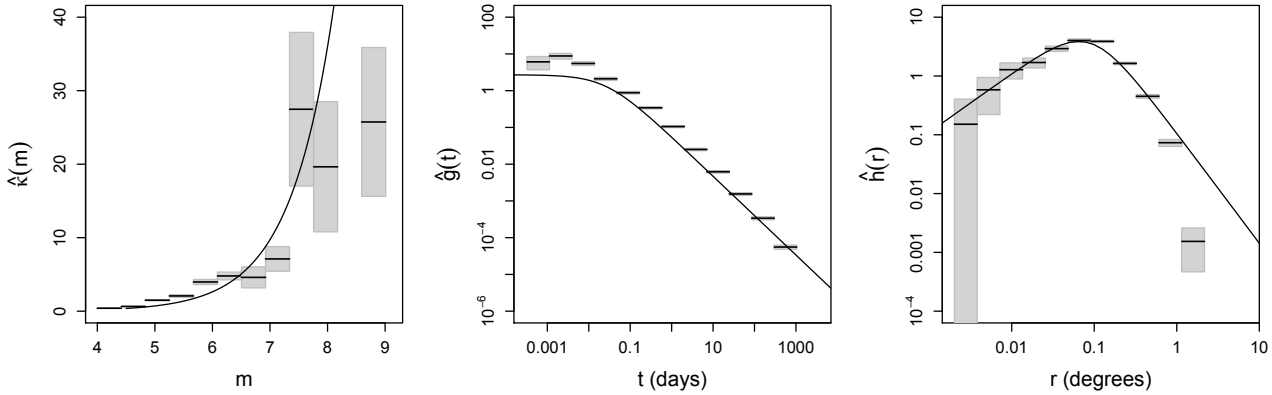

FIG 11. *Magnitude, temporal, and distance components for triggering function estimated from the Japan earthquake dataset (Section 5). The black horizontal lines are the estimates in each bin. The grey boxes are the error bars covering* ± 2 *standard errors. The black curves are the parametric estimates from Ogata (1998) in the same region.*

REFERENCES

Adelfio, G. and Chiodi, M. (2013). Mixed estimation technique in semi-parametric space-time point processes for earthquake description. In *Proceedings of the 28th International Workshop on Statistical Modelling 8-13 July, 2013, Palermo* **1** 65–70.

Adelfio, G. and Chiodi, M. (2015). Alternated estimation in semi-parametric space-time branching-type point processes with application to seismic catalogs. *Stochastic Environmental Research and Risk Assessment* **29** 443–450.

Brillinger, D. R. (1998). Some wavelet analyses of point process data. In *ASILOMAR CONFERENCE ON SIGNALS SYSTEMS AND COMPUTERS* **2** 1087–1091. COMPUTER SOCIETY PRESS.

Daley, D. J. and Vere-Jones, D. (2003). *An Introduction to the Theory of Point Processes, Volume 1: Elementary Theory and Methods*, Second ed. Springer, New York.

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

Lewis, P. A. and Shedler, G. S. (1979). Simulation of nonhomogeneous Poisson processes by thinning. *Naval Research Logistics Quarterly* **26** 403–413.

Marsan, D. and Lengliné, O. (2008). Extending earthquakes' reach through cascading. *Science* **319** 1076–1079.

Marsan, D. and Lengliné, O. (2010). A New Estimation of the Decay of Aftershock Density WIth Distance to the Mainshock. *J. Geophys. Res.* **115**.

Mohler, G. O., Short, M. B., Brantingham, P. J., Schoenberg, F. P. and Tita, G. E. (2011). Self-Exciting Point Process Modeling of Crime. *Journal of the American Statistical Association* **106** 100-108.

Musmeci, F. and Vere-Jones, D. (1992). A space-time clustering model for historical earthquakes. *Annals of the Institute of Statistical Mathematics* **44** 1–11.

Nichols, K. and Schoenberg, F. P. (2014). Assessing the dependency between the magnitudes of earthquakes and the magnitudes of their aftershocks. *Environmetrics* **25** 143–151.

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.

Schoenberg, F. P. (2013). Facilitated estimation of ETAS. *Bulletin of the Seismological Society of America* **103** 601–605.

Utsu, T., Ogata, Y. and Matsu'ura, R. S. (1995). The centenary of the Omori formula for a decay law of aftershock activity. *Journal of Physics of the Earth* **43** 1–33.

Veen, A. and Schoenberg, F. P. (2008). Estimation of Space–Time Branching Process Models in Seismology Using an EM–Type Algorithm. *Journal of the American Statistical Association* **103** 614–624.

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

Zhuang, J., Ogata, Y. and Vere-Jones, D. (2004). Analyzing earthquake clustering features by using stochastic reconstruction. *Journal of Geophysical Research: Solid Earth (1978–2012)* **109**.

E. W. Fox
F. P. Schoenberg
J. S. Gordon
Department of Statistics
University of California
8125 Math Sciences Building
Los Angeles, California 90095-1554
E-mail: eric.fox@stat.ucla.edu
frederic@stat.ucla.edu
joshuagordon@g.ucla.edu