

SPATIALLY INHOMOGENEOUS BACKGROUND RATE ESTIMATORS AND UNCERTAINTY QUANTIFICATION FOR NONPARAMETRIC HAWKES POINT PROCESS MODELS OF 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. As an alternative, Marsan and Lengliné [*Science* **319** (2008) 1076–1079] developed a completely nonparametric method that provides an estimate of a homogeneous 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 spatially inhomogeneous background rate; (b) adding error bars to the histogram estimates which quantify the sampling variability 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 [Hawkes (1971)] 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 the following: first, estimating the probability each earthquake in the catalog is either a mainshock or aftershock; and second,

Received May 2015; revised June 2016.

Key words and phrases. Point processes, nonparametric estimation, Hawkes process, MISD, ETAS model, earthquake forecasting.

updating a homogeneous 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 range of applications, especially in situations where the form of the intensity function is unknown and difficult to determine. Brillinger (1998) described a technique using wavelets for estimating the conditional intensity and second order intensity with applications to neurophysiology and seismology. Adelfio and Chiodi (2013, 2015) considered a semiparametric 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 of aftershock activity following an earthquake of given magnitude. They also demonstrated the application of their routine to stochastically decluster earthquake catalogs to isolate mainshocks and remove aftershock clusters. 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. 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.

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 as follows:

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

Note that in this paper we are interested in modeling spatial variations and not temporal variations in the background rate.

The original MISD algorithm assumes that the background rate for mainshocks is a constant 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 an inhomogeneous background component is an important next step and

improvement by allowing for localized estimates of mainshock activity related to variations in the underlying tectonic field and the locations of 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 inhomogeneity in the background process by initially kernel smoothing over all events in the catalog and weighting each event by its corresponding kernel estimate. However, a main shortcoming of this approach is that kernel smoothing over all events in a Hawkes process results in high variance and inaccuracy, as triggered events influence the estimated background rate locally. 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 constant rate estimator of Marsan and Lengliné (2008). Second, we apply the variable bandwidth kernel estimator, used by Zhuang, Ogata and Vere-Jones (2002) for semi-parametric estimation, into the context of MISD. We validate and assess this new methodology by simulating earthquake catalogs from a space–time ETAS model and evaluating the ability of each method to recover the true form of the inhomogeneous background rate and triggering function governing the simulation.

For the second goal of this paper, uncertainty quantification for the histogram estimators of the triggering function, we propose a bootstrap procedure. By repeatedly simulating and re-estimating a fitted nonparametric model, we can construct error bars which quantify the sampling variability of the histogram estimators, including the variability caused by the uncertainty of the mainshock-aftershock assignment of earthquakes.

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 two new ways to incorporate an inhomogeneous background rate. In Section 3, we also describe a bootstrap procedure for computing error bars for histogram estimates of the triggering function. 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 Section 6, we summarize and speculate about our results and suggest future directions 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, $\{(t_i, x_i, y_i, m_i) : i = 1, \dots, N\}$, 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,

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

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 conditional 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:

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

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 an inhomogeneous Poisson process in space and constant 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:

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

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 parameterization of the triggering function for ETAS is given by

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

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

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

where m_c is the magnitude cutoff 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:

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

where $S \times [0, T]$ is the space–time observation region. 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 inhomogeneous background component, $\mu(x, y)$, is often estimated with nonparametric techniques. For instance, [Musmeci and Vere-Jones \(1992\)](#) and [Zhuang, Ogata and Vere-Jones \(2002\)](#) employed a kernel smoothing estimator. Alternatively, [Ogata \(1998\)](#) used bi-cubic B-splines on mainshock events identified by the magnitude-based clustering algorithm, an iterative procedure the places space–time windows around the largest events in the catalog and then removes all other events (aftershocks) inside those windows. In this paper we opt to focus on the weighted kernel method of [Zhuang, Ogata and Vere-Jones \(2002\)](#) since it provides a way to incorporate the estimated probability that each event is a mainshock.

[Marsan and Lengliné \(2008\)](#) proposed the MISD algorithm to nonparametrically estimate the triggering function ν and homogeneous 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 constant and the triggering rates are 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 homogeneous 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 an inhomogeneous background rate;
2. We assume the separability of the triggering function into components for magnitude, time and distance;
3. We perform histogram estimation on the magnitude productivity function $\kappa(m)$, temporal triggering density $g(t)$ and spatial triggering density $f(r)$ (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 inhomogeneous in space and constant in time.

3.1. Histogram estimators. Let P be an $N \times N$ lower triangular probability matrix with entries

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

$$P = \begin{bmatrix} p_{11} & 0 & 0 & \cdots & 0 \\ p_{21} & p_{22} & 0 & \cdots & 0 \\ p_{31} & p_{32} & p_{33} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ p_{N1} & p_{N2} & p_{N3} & \cdots & p_{NN} \end{bmatrix},$$

$$P^{(0)} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1/2 & 1/2 & 0 & \cdots & 0 \\ 1/3 & 1/3 & 1/3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1/N & 1/N & 1/N & \cdots & 1/N \end{bmatrix}.$$

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; others will be considered in Section 4.4. 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 nondiagonal elements) is the estimated number of aftershocks.

Below is the MISD algorithm of Marsan and Lengliné (2008) with the modifications specified at 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 inhomogeneous background rate $\mu(x, y)$:

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

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

$$\begin{aligned} \kappa_k^{(v)} &= \frac{\sum A_k p_{ij}^{(v)}}{N_k^{\text{mag}}}, \quad k = 1, \dots, n_m^{\text{bins}}; \\ g_k^{(v)} &= \frac{\sum B_k p_{ij}^{(v)}}{\Delta t_k \sum_{i=1}^N \sum_{j=1}^{i-1} p_{ij}^{(v)}}, \quad k = 1, \dots, n_t^{\text{bins}}; \\ h_k^{(v)} &= \frac{\sum C_k p_{ij}^{(v)}}{\Delta r_k \sum_{i=1}^N \sum_{j=1}^{i-1} p_{ij}^{(v)}}, \quad k = 1, \dots, n_r^{\text{bins}}. \end{aligned}$$

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

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

5. If $\max_{i,j} |p_{ij}^{(v+1)} - p_{ij}^{(v)}| < \varepsilon$, where $i \leq j$, then the algorithm has converged (in practice, we take $\varepsilon = 10^{-3}$). 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 $\mu(x, y)$ ($n_x^{\text{bins}} \cdot n_y^{\text{bins}}$ bins total).
- $D_{kl} = \{i : (k-1)\Delta x < x_i \leq k\Delta x, (l-1)\Delta y < y_i \leq l\Delta y\}$, 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 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.
- $N_k^{\text{mag}} = \sum_{j=1}^N I(\delta m_k < m_j \leq \delta m_{k+1})$ 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.

The iterations of Algorithm 1 can be interpreted as an EM-type algorithm. In Appendix A of the supplement [Fox, Schoenberg and Gordon (2016)] we specify the expected complete data log-likelihood for the space–time Hawkes process (2) and show that the estimators in steps 2 and 3 maximize an approximation to this likelihood if we assume that the background and triggering rates are piecewise constant.

In step 2 of Algorithm 1 the inhomogeneous background rate is estimated with a histogram estimator which is a generalization of the homogeneous estimator in the original MISD algorithm. In our modified method, the spatial observation region 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 that the histogram estimator in step 2 reduces to the homogeneous case in Marsan and Lengliné (2008) when $n_x^{\text{bins}} = n_y^{\text{bins}} = 1$ and $\Delta x \cdot \Delta y = S$ (i.e., only one cell equal to the spatial observation region is specified). Also note that the estimator of g is itself a density since $\sum_{k=1}^{n_t^{\text{bins}}} \Delta t_k \hat{g}_k = 1$, 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 $\hat{\kappa}_k$ 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 observation region.

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

$$(9) \quad \mu(x, y) = \gamma \tau(x, y),$$

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

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 $\varepsilon = 0.02$ 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 semiparametric 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 $i = 1, \dots, N$.

2. Estimate inhomogeneous 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 region. 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 Riemann sum over those pixels.

3.3. Uncertainty quantification for the triggering function. In this section we propose a procedure that uses the bootstrap [Davison and Hinkley (1997)] to compute error bars for the histogram estimates of the triggering function. The procedure works by simulating many realizations (bootstrap datasets) from a nonparametric Hawkes process model estimated with Algorithms 1 or 2. Then the nonparametric model is re-estimated from the bootstrap datasets to get many replicates of the histogram estimates of the triggering function. The error bars are computed as the percentiles of the bootstrap replicates of the histogram estimates in each bin. The procedure is written out formally below.

ALGORITHM 3. 1. Simulate from the estimated nonparametric model $\hat{\lambda}(t, x, y|H_t)$ over observation region $S \times [0, T]$ to get a bootstrap earthquake dataset $b^* = \{(t_i^*, x_i^*, y_i^*, m_i^*) : i = 1, \dots, N^*\}$.

2. Re-estimate the nonparametric model from bootstrap dataset b^* to get a bootstrap replicate $\hat{\lambda}^*(t, x, y|H_t)$ of the nonparametric model.

3. Repeat previous steps M times (in practice, we take $M = 200$).
4. Compute bootstrap error bars for the histogram estimates of the triggering function using the $\alpha/2$ and $1 - \alpha/2$ percentiles of the bootstrap replicates of the triggering components:

$$\begin{aligned} (\hat{\kappa}_{k;\alpha/2}^*, \hat{\kappa}_{k;1-\alpha/2}^*), & \quad k = 1, \dots, n_m^{\text{bins}}, \\ (\hat{g}_{k;\alpha/2}^*, \hat{g}_{k;1-\alpha/2}^*), & \quad k = 1, \dots, n_t^{\text{bins}}, \\ (\hat{h}_{k;\alpha/2}^*, \hat{h}_{k;1-\alpha/2}^*), & \quad k = 1, \dots, n_r^{\text{bins}}, \end{aligned}$$

where k refers to the bin index for the histogram estimate of each triggering component.

The simulation algorithm used to generate a bootstrap dataset from a fitted nonparametric Hawkes process model is provided in Appendix B of the supplement [Fox, Schoenberg and Gordon (2016)].

The bootstrap intervals from Algorithm 3 quantify the sampling variability of the histogram estimators over multiple realizations of the self-exciting point process. This includes the variability due to estimating the mainshock-aftershock labels probabilistically with the EM algorithm since the probabilities are re-estimated from the bootstrap datasets. Note that we use what are commonly referred to as percentile intervals in step 4 of Algorithm 3, and that alternative types of bootstrap intervals exist [e.g., see Davison and Hinkley (1997)].

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 equations (4)–(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 $36^\circ\sim 42^\circ\text{N}$ latitude and $141^\circ\sim 145^\circ\text{E}$ 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 region for the simulation is $S \times [0, T] = [0, 4] \times [0, 6] \times [0, 25,000]$, and the magnitude cutoff is $m_c = 0$. The inhomogeneous background rate is specified by partitioning the spatial observation region S into 4 equally sized cells with the varying rates shown in Figure 1. An example of a simulated realization is shown in Figure 2. For a description of the simulation procedure for ETAS, please see Algorithm C from Zhuang, Ogata and Vere-Jones (2004).

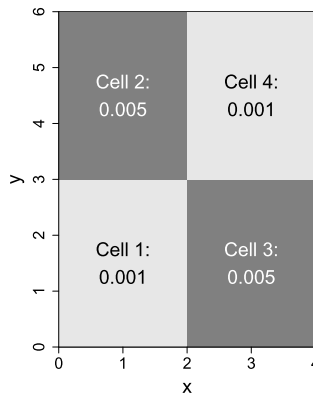


FIG. 1. Inhomogeneous background rate used for the simulation study in Section 4.1; true rates are displayed in each cell (numbered 1–4 for convenience).

We evaluate the performance of the nonparametric method by examining the bias, standard deviation and root-mean-square error (RMSE) of the estimates over 200 simulated realizations of the specified ETAS model. For the inhomogeneous background rate these quantities are computed as

$$(11) \quad \text{Avg}(\hat{\mu}_{kl}) = \frac{1}{200} \sum_{i=1}^{200} \hat{\mu}_{kl}^i,$$

$$(12) \quad \text{Bias}(\hat{\mu}_{kl}) = \text{Avg}(\hat{\mu}_{kl}) - \mu_{kl},$$

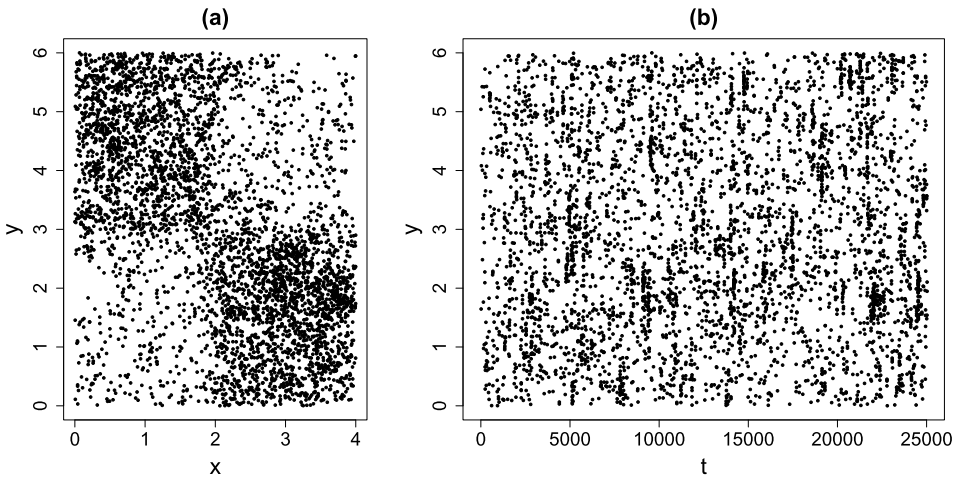


FIG. 2. Simulated realization of ETAS model (4)–(6) with background rate varying in each quadrant; (a) epicentral locations, and (b) space–time plot of simulated earthquakes.

TABLE 1

Average, bias, standard deviation and RMSE for the estimates of the inhomogeneous background rate from the simulation study in Section 4.1. Cell numbers correspond to Figure 1

| Cell | Truth | Avg | Bias | SD | RMSE |
|------|-------|---------|---------|---------|---------|
| 1 | 0.001 | 0.00137 | 0.00037 | 0.00014 | 0.00039 |
| 2 | 0.005 | 0.00597 | 0.00097 | 0.00030 | 0.00101 |
| 3 | 0.005 | 0.00597 | 0.00097 | 0.00031 | 0.00101 |
| 4 | 0.001 | 0.00138 | 0.00038 | 0.00013 | 0.00040 |

$$(13) \quad \text{SD}(\hat{\mu}_{kl}) = \sqrt{\frac{1}{200} \sum_{i=1}^{200} (\hat{\mu}_{kl}^i - \text{Avg}(\hat{\mu}_{kl}))^2},$$

$$(14) \quad \text{RMSE}(\hat{\mu}_{kl}) = \sqrt{\frac{1}{200} \sum_{i=1}^{200} (\hat{\mu}_{kl}^i - \mu_{kl})^2},$$

where $\hat{\mu}_{kl}^i$ and μ_{kl} are the estimate from the i th simulation and true value of the background rate in the (k, l) cell, respectively. The results for the estimation of the inhomogeneous background rate, shown in Table 1, reveal that the nonparametric method (Algorithm 1) is able to recover the sharp differences between the rates in each cell with reasonably small RMSE. However, the rates are overestimated in each cell, and the bias contributes more to the RMSE than the standard deviation of the estimates. In the next section we show that this overestimation bias is due to boundary effects in the simulation induced by excluding aftershocks that occur outside the space–time observation region.

Figure 3 shows the histogram estimates of the magnitude, temporal and spatial components of the triggering function from 200 realizations of the specified ETAS model. The estimates of $g(t)$ and $h(r)$ have equally spaced bins on a logarithmic scale and are plotted on a log–log scale since the true densities are power law. The estimates of $\kappa(m)$ have equally spaced magnitude bins and are plotted on a log scale. Overall, the histogram estimates in Figure 3 appear to successfully recover the shape of each component of the triggering function governing the simulation (i.e., the black curves in this figure).

We quantify the performance of the histogram estimator of each triggering function component by examining its bias, standard deviation and RMSE computed as

$$(15) \quad \text{Avg}(\hat{g}_k) = \frac{1}{200} \sum_{i=1}^{200} \hat{g}_k^i,$$

$$(16) \quad \text{Bias}(\hat{g}_k) = \text{Avg}(\hat{g}_k) - \bar{g}_k,$$

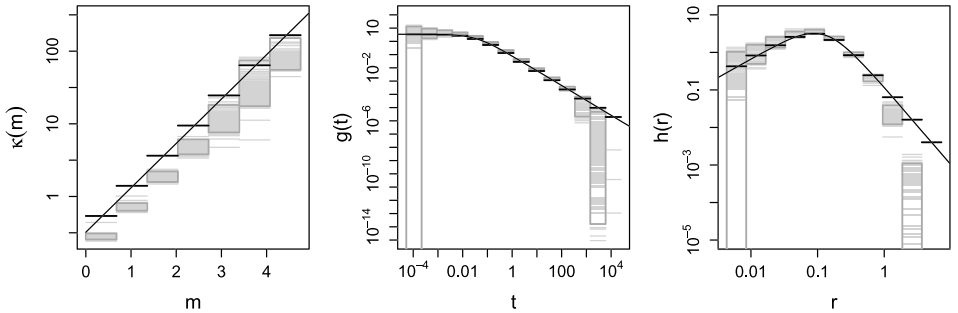


FIG. 3. Magnitude, temporal and distance components for the triggering function from the simulation study in Section 4.1. The black curves are the true triggering components governing the simulation, and the black horizontal lines are the binned average values of these functions. The gray horizontal lines are the histogram estimates from 200 simulations of ETAS; the dark gray boxes correspond to the pointwise 0.025 and 0.975 percentiles of the estimates in each bin.

(17)
$$SD(\hat{g}_k) = \sqrt{\frac{1}{200} \sum_{i=1}^{200} (\hat{g}_k^i - \text{Avg}(\hat{g}_k))^2},$$

(18)
$$RMSE(\hat{g}_k) = \sqrt{\frac{1}{200} \sum_{i=1}^{200} (\hat{g}_k^i - \bar{g}_k)^2},$$

where \hat{g}_k^i is the estimate of $g(t)$ over bin k for the i th simulation, and \bar{g}_k is the average value of the $g(t)$ over bin k [i.e., $\bar{g}_k = \frac{1}{\Delta t_k} \int_{\delta t_k}^{\delta t_{k+1}} g(t) dt$]; equations (15)–(18) are defined similarly for \hat{h}_k and $\hat{\kappa}_k$. The values for these performance measures are given in Tables 2–4.

Tables 3 and 4 reveal that the density estimates of $g(t)$ and $h(r)$ are most reasonable in the midrange of each histogram’s support. The large standard deviations

TABLE 2
Average, bias, standard deviation and RMSE for the histogram estimates of $\kappa(m)$ from the simulation study in Section 4.1. Bias and RMSE are evaluated using the binned average values of $\kappa(m)$, which are provided in the column labeled Truth

| Bin | Truth | Avg | Bias | SD | RMSE |
|---------------|-------|-------|--------|--------|-------|
| (0, 0.679] | 0.539 | 0.28 | −0.26 | 0.0183 | 0.26 |
| (0.679, 1.36] | 1.4 | 0.724 | −0.677 | 0.0516 | 0.679 |
| (1.36, 2.04] | 3.64 | 1.88 | −1.76 | 0.16 | 1.77 |
| (2.04, 2.72] | 9.47 | 4.9 | −4.57 | 0.603 | 4.61 |
| (2.72, 3.39] | 24.6 | 12.7 | −11.9 | 2.61 | 12.2 |
| (3.39, 4.07] | 64 | 34.5 | −29.4 | 12.9 | 32.1 |
| (4.07, 4.75] | 166 | 88.4 | −77.8 | 23.9 | 81.4 |

TABLE 3

Average, bias, standard deviation and RMSE for the histogram estimates of $g(t)$ from the simulation study in Section 4.1. Bias and RMSE are evaluated using the binned average values of $g(t)$, which are provided in the column labeled Truth

| Bin | Truth | Avg | Bias | SD | RMSE |
|---------------------|----------|----------|-----------|----------|----------|
| (0.000052, 0.00022] | 3.41 | 5.53 | 2.12 | 3.71 | 4.27 |
| (0.00022, 0.00091] | 3.37 | 4.93 | 1.57 | 1.96 | 2.51 |
| (0.00091, 0.0038] | 3.19 | 4.86 | 1.67 | 1.02 | 1.95 |
| (0.0038, 0.016] | 2.62 | 3.96 | 1.34 | 0.417 | 1.4 |
| (0.016, 0.066] | 1.51 | 2.26 | 0.748 | 0.157 | 0.764 |
| (0.066, 0.27] | 0.534 | 0.795 | 0.261 | 0.05 | 0.265 |
| (0.27, 1.1] | 0.132 | 0.196 | 0.0643 | 0.0123 | 0.0655 |
| (1.1, 4.8] | 0.0282 | 0.0412 | 0.013 | 0.00271 | 0.0132 |
| (4.8, 20] | 0.00577 | 0.00811 | 0.00234 | 0.000689 | 0.00244 |
| (20, 83] | 0.00117 | 0.00157 | 0.000403 | 0.000153 | 0.000431 |
| (83, 345] | 0.000236 | 0.000292 | 5.55E-05 | 4.65E-05 | 7.24E-05 |
| (345, 1439] | 4.77E-05 | 3.18E-05 | -1.59E-05 | 1.56E-05 | 2.23E-05 |
| (1439, 5998] | 9.63E-06 | 3.67E-07 | -9.26E-06 | 1.37E-06 | 9.36E-06 |
| (5998, 25,000] | 1.94E-06 | 3.21E-11 | -1.94E-06 | 4.5E-10 | 1.94E-06 |

in the first two rows of Tables 3 and 4 are due to the very small logarithmically spaced bin sizes. Tables 3 and 4 also show that the density estimates become negatively biased over bins on the right-tail ends of $g(t)$ and $h(r)$. This can also be seen clearly in Figure 3. The role of boundary effects on the estimates of the tail ends of the triggering densities will be investigated in the next section.

TABLE 4

Average, bias, standard deviation and RMSE for the histogram estimates of $h(r)$ from the simulation study in Section 4.1. Bias and RMSE are evaluated using the binned average values of $h(r)$, which are provided in the column labeled Truth

| Bin | Truth | Avg | Bias | SD | RMSE |
|------------------|---------|----------|----------|----------|---------|
| (0.0043, 0.0085] | 0.426 | 0.511 | 0.0847 | 0.283 | 0.296 |
| (0.0085, 0.017] | 0.826 | 0.988 | 0.162 | 0.284 | 0.327 |
| (0.017, 0.033] | 1.55 | 1.89 | 0.337 | 0.288 | 0.443 |
| (0.033, 0.064] | 2.58 | 3.08 | 0.5 | 0.256 | 0.562 |
| (0.064, 0.13] | 3.11 | 3.68 | 0.567 | 0.194 | 0.599 |
| (0.13, 0.25] | 2.14 | 2.45 | 0.309 | 0.1 | 0.325 |
| (0.25, 0.49] | 0.846 | 0.909 | 0.0625 | 0.0495 | 0.0797 |
| (0.49, 0.95] | 0.246 | 0.209 | -0.0369 | 0.0215 | 0.0427 |
| (0.95, 1.9] | 0.064 | 0.0241 | -0.0398 | 0.00702 | 0.0405 |
| (1.9, 3.7] | 0.0161 | 0.000118 | -0.016 | 0.000303 | 0.016 |
| (3.7, 7.2] | 0.00402 | 1.24E-10 | -0.00402 | 1.76E-09 | 0.00402 |

Tables 3 and 4 also show a positive bias in the histogram density estimates for $t \in (0.000052, 345)$ and $r \in (0.0043, 0.49)$. The reason for this is that the histogram density estimates are constrained to integrate to one over their finite supports, while the true densities governing the simulation have infinite supports and integrate to one over the positive real line. For instance, the true temporal density $g(t)$ has 80% of its mass below 25,000, while the density estimate $\hat{g}(t)$ has 100% of its mass below 25,000.

Table 2 and Figure 3 show that the nonparametric method recovers the underlying exponential form of the magnitude productivity $\kappa(m)$ with reasonable sampling variation. However, the productivity function is underestimated in each bin; one explanation is that many aftershocks in the simulation fall outside the observation region and are not included in the estimation. Table 2 also shows that the standard deviation of the estimates increases with magnitude, although this is expected as there are only a few large magnitude events in each simulation with which to estimate the productivity.

To illustrate the performance of the bootstrap procedure for uncertainty quantification (Section 3.3), Figure S1 and Table S1 in Appendix C of the supplement [Fox, Schoenberg and Gordon (2016)] show a realization of the bootstrap intervals for one nonparametric estimate of the specified ETAS model. Note that the bootstrap intervals only quantify the sampling variability of the estimates, and are therefore affected by the biases in the estimates caused by boundary effects.

4.2. Boundary issues. When simulating earthquake catalogs from the ETAS model the mainshocks are restricted to occur within the space–time observation region $S \times [0, T]$. However, the times and locations of aftershocks, simulated from the triggering function components g and h , 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 region 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, that is, all aftershocks occurring within $[-\varepsilon_r, 4 + \varepsilon_r] \times [-\varepsilon_r, 6 + \varepsilon_r] \times [0, 25,000 + \varepsilon_t]$. 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 inhomogeneous background rate as we increase ε_r and ε_t , we use the RMSE evaluated over all cells (k, l) :

$$(19) \quad \sqrt{\frac{1}{n_x^{\text{bins}} n_y^{\text{bins}}} \sum_{k,l} (\hat{\mu}_{kl} - \mu_{kl})^2}.$$

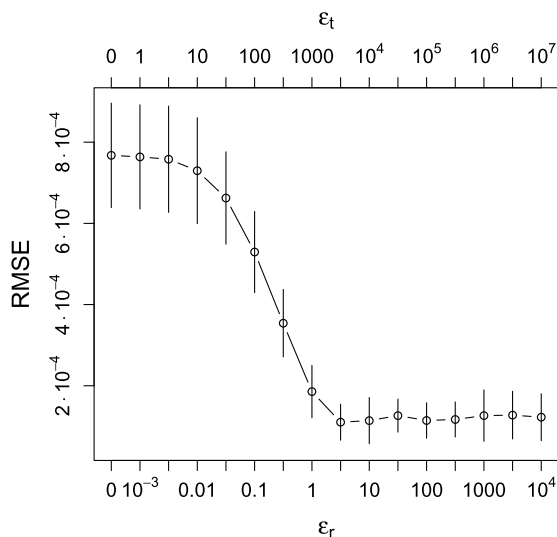


FIG. 4. RMSE of the background rate, equation (19), for increasing values of ε_r and ε_t . RMSEs are averaged from 10 realizations of ETAS; the vertical bars cover one sample standard deviation in the RMSE values, above and below the mean.

We simulate the ETAS model 10 times using the same parameters and background rate as in Section 4.1, with mainshocks again restricted to $S \times [0, T] = [0, 4] \times [0, 6] \times [0, 25,000]$, but aftershocks allowed to occur outside that region. For each simulation, the RMSE (19) is computed for increasing values of ε_r and ε_t . Figure 4 shows the mean RMSE from the 10 realizations at selected values of ε_r and ε_t ; the vertical lines represent a standard deviation in RMSE above and below the mean. The incorporation of aftershocks falling outside the space–time observation region significantly improves the performance of the estimation of the background rate. The RMSE appears to level off when $\varepsilon_r = 10^{0.5} = 3.16$ and $\varepsilon_t = 10^{3.5} = 3162.28$.

Tables 5–8 show the results from simulating and re-estimating ETAS with Algorithm 1 200 times with a boundary correction of $\varepsilon_r = 1000$ and $\varepsilon_t = 10^6$. Again, we simulate events with the same parameters and space–time region as Section 4.1. The only difference is that in the estimation we use aftershocks occurring within a distance $\varepsilon_r = 1000$ and time $\varepsilon_t = 10^6$ of the boundary of the observation region.

The measures in Table 5 show substantial improvement in the estimation of the inhomogeneous background rate when compared to Table 1, which neglected boundary effects. The consistent over-prediction of the rates is no longer present, and the bias in each cell is negligible once the boundary effects are accounted for.

The histogram estimates of the triggering function in Figure 5 also show much improvement when compared to Figure 3. After applying the boundary correction, the large negative biases in the right-tail ends of densities $g(t)$ and $h(r)$ are no longer present, and the estimates of the magnitude productivity $\kappa(m)$ appear more centered around the true value.

TABLE 5

Average, bias, standard deviation and RMSE for the estimates of the inhomogeneous background from the simulation study with boundary correction discussed in Section 4.2. Cell numbers correspond to Figure 1

| Cell | Truth | Avg | Bias | SD | RMSE |
|------|-------|----------|-----------|----------|----------|
| 1 | 0.001 | 0.000989 | −0.000011 | 0.000096 | 0.000096 |
| 2 | 0.005 | 0.005034 | 0.000034 | 0.000209 | 0.000212 |
| 3 | 0.005 | 0.005038 | 0.000038 | 0.000209 | 0.000213 |
| 4 | 0.001 | 0.001001 | 0.000001 | 0.000094 | 0.000094 |

Tables 6–8 also quantify a substantial reduction in the bias and RMSE of the triggering estimates when compared to Tables 2–4.

Note that a positive bias is present in the histogram estimates of $g(t)$ (Table 7) even after applying the boundary correction $\varepsilon_r = 1000$ and $\varepsilon_t = 10^6$. However, this is perhaps not surprising since the power-law density $g(t)$ governing the simulation has 12.5% of its mass above 10^6 . Further simulation studies have also demonstrated that applying a larger boundary correction ($\varepsilon_t > 10^6$ and $\varepsilon_r > 1000$) further reduces the bias in the estimation of the triggering densities and magnitude productivity.

Illustratively, Figure S2 and Table S2 in Appendix C of the supplement [Fox, Schoenberg and Gordon (2016)] show the bootstrap intervals (Section 3.3) for one nonparametric estimate the specified ETAS model with boundary correction. Since the boundary correction reduces the biases in the estimates, the corresponding bootstrap intervals are also substantially less biased (i.e., when compared to Figure S1 and Table S1).

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

TABLE 6

Average, bias, standard deviation and RMSE for the histogram estimates of $\kappa(m)$ from the simulation study with boundary correction discussed in Section 4.2. Bias and RMSE are evaluated using the binned average values of $\kappa(m)$, which are provided in the column labeled Truth

| Bin | Truth | Avg | Bias | SD | RMSE |
|---------------|-------|-------|--------|--------|-------|
| (0, 0.679] | 0.539 | 0.423 | −0.116 | 0.0123 | 0.117 |
| (0.679, 1.36] | 1.4 | 1.08 | −0.322 | 0.0414 | 0.325 |
| (1.36, 2.04] | 3.64 | 2.77 | −0.87 | 0.155 | 0.884 |
| (2.04, 2.72] | 9.47 | 7.1 | −2.37 | 0.546 | 2.43 |
| (2.72, 3.39] | 24.6 | 18.4 | −6.25 | 2.3 | 6.66 |
| (3.39, 4.07] | 64 | 45.2 | −18.8 | 10.3 | 21.4 |
| (4.07, 4.75] | 166 | 128 | −37.8 | 33.9 | 50.8 |

TABLE 7

Average, bias, standard deviation and RMSE for the histogram estimates of $g(t)$ from the simulation study with boundary correction discussed in Section 4.2. Bias and RMSE are evaluated using the binned average values of $g(t)$, which are provided in the column labeled Truth

| Bin | Truth | Avg | Bias | SD | RMSE |
|---------------------|----------|----------|----------|----------|----------|
| (0.000052, 0.00022] | 3.41 | 3.77 | 0.358 | 2.39 | 2.42 |
| (0.00022, 0.00091] | 3.37 | 3.88 | 0.516 | 1.11 | 1.22 |
| (0.00091, 0.0038] | 3.19 | 3.63 | 0.443 | 0.59 | 0.738 |
| (0.0038, 0.016] | 2.62 | 3.05 | 0.426 | 0.249 | 0.494 |
| (0.016, 0.066] | 1.51 | 1.76 | 0.245 | 0.0953 | 0.263 |
| (0.066, 0.27] | 0.534 | 0.616 | 0.0819 | 0.0262 | 0.086 |
| (0.27, 1.1] | 0.132 | 0.152 | 0.0197 | 0.00717 | 0.021 |
| (1.1, 4.8] | 0.0282 | 0.0322 | 0.00406 | 0.00158 | 0.00436 |
| (4.8, 20] | 0.00577 | 0.00662 | 0.000853 | 0.000372 | 0.00093 |
| (20, 83] | 0.00117 | 0.00132 | 0.000155 | 9.42E-05 | 0.000182 |
| (83, 345] | 0.000236 | 0.000273 | 3.65E-05 | 2.39E-05 | 4.36E-05 |
| (345, 1439] | 4.77E-05 | 5.6E-05 | 8.3E-06 | 5.78E-06 | 1.01E-05 |
| (1439, 5998] | 9.63E-06 | 1.11E-05 | 1.49E-06 | 1.45E-06 | 2.08E-06 |
| (5998, 25,000] | 1.94E-06 | 2.22E-06 | 2.76E-07 | 2.61E-07 | 3.8E-07 |

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 1 with constant rates in each cell on a 2×2 grid, we simulate from the smoother background rate shown in Figure 6(a). This inhomogeneous background rate was

TABLE 8

Average, bias, standard deviation and RMSE for the histogram estimates of $h(r)$ from the simulation study with boundary correction discussed in Section 4.2. Bias and RMSE are evaluated using the binned average values of $h(r)$, which are provided in the column labeled Truth

| Bin | Truth | Avg | Bias | SD | RMSE |
|------------------|---------|---------|-----------|---------|----------|
| (0.0043, 0.0085] | 0.426 | 0.402 | -0.0237 | 0.161 | 0.163 |
| (0.0085, 0.017] | 0.826 | 0.839 | 0.0126 | 0.188 | 0.188 |
| (0.017, 0.033] | 1.55 | 1.54 | -0.00707 | 0.171 | 0.171 |
| (0.033, 0.064] | 2.58 | 2.57 | -0.0119 | 0.16 | 0.16 |
| (0.064, 0.13] | 3.11 | 3.12 | 0.00374 | 0.127 | 0.127 |
| (0.13, 0.25] | 2.14 | 2.17 | 0.0251 | 0.0776 | 0.0815 |
| (0.25, 0.49] | 0.846 | 0.851 | 0.00498 | 0.0351 | 0.0355 |
| (0.49, 0.95] | 0.246 | 0.245 | -0.00143 | 0.0149 | 0.0149 |
| (0.95, 1.9] | 0.064 | 0.0628 | -0.00121 | 0.00582 | 0.00595 |
| (1.9, 3.7] | 0.0161 | 0.0155 | -0.00061 | 0.00207 | 0.00216 |
| (3.7, 7.2] | 0.00402 | 0.00387 | -0.000151 | 0.00063 | 0.000648 |

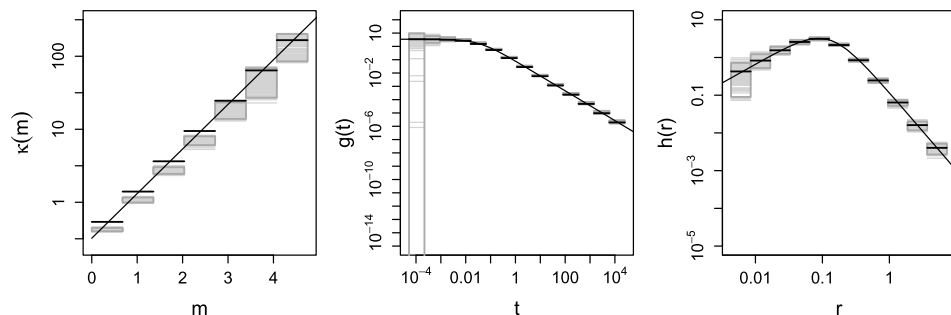


FIG. 5. Estimates of the triggering function components from 200 ETAS simulations, with boundary correction for aftershock activity $\varepsilon_r = 1000$ and $\varepsilon_t = 10^6$.

generated by performing fixed bandwidth kernel density estimation over the locations of 883 earthquakes of magnitude 5.0 or greater, longitude $141^\circ\sim 145^\circ\text{E}$, latitude $36^\circ\sim 42^\circ\text{N}$, and time between 16 Jan 2007 to 28 Dec 2014 [data gathered from <http://www.quake.geo.berkeley.edu/anss/catalog-search.html> with the same spatial observation region as Ogata (1998)]. To simulate from the kernel smoothed background rate, in Figure 6(a) we use the thinning procedure of Lewis and Shedler (1979) and set the expected number of background events equal to 2000.

Figure 6(b) shows the probability weighted variable kernel estimate (Algorithm 2, step 2) of the inhomogeneous 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 7. 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 RMSE (19)

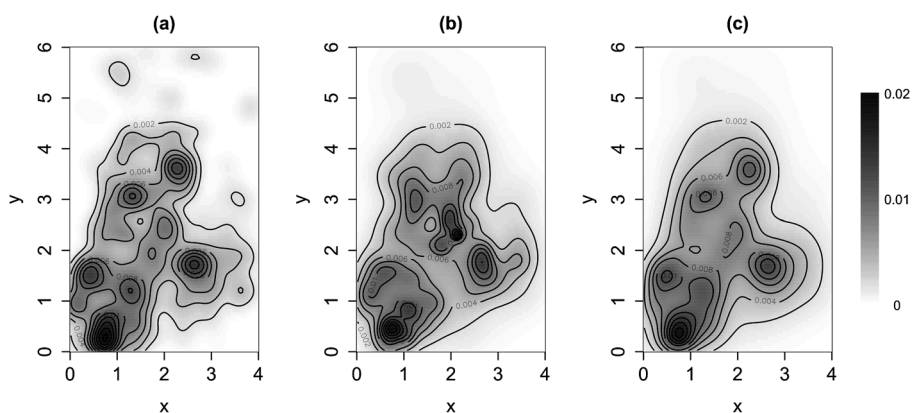


FIG. 6. (a) True background rate for simulation study in Section 4.3. (b) Estimate of background rate from one simulated realization of ETAS, and (c) mean estimate from 200 realizations.

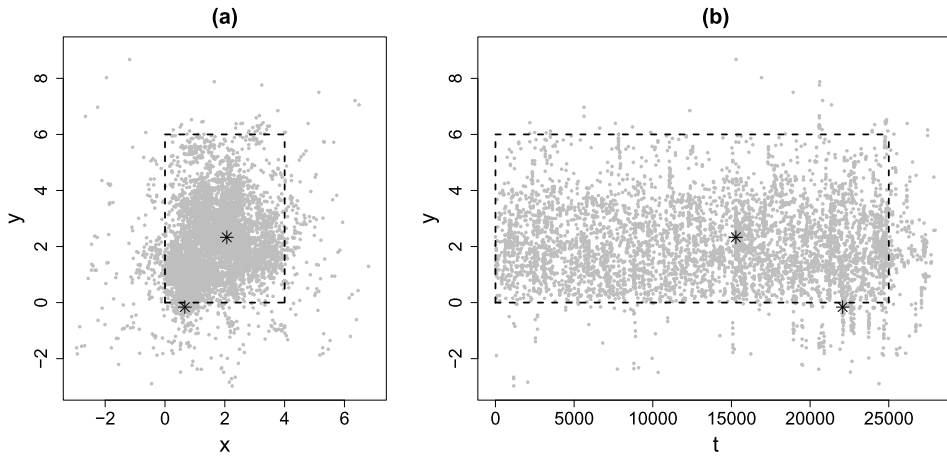


FIG. 7. Simulated realization of ETAS model (4)–(6) with smooth inhomogeneous background rate; (a) epicentral locations, and (b) space–time plot of simulated earthquakes. The dashed rectangles in each plot are the spatial and temporal boundaries for the observation region $S \times [0, T] = [0, 4] \times [0, 6] \times [0, 25,000]$. Aftershocks occurring within a distance $\varepsilon_r = 3$ and time $\varepsilon_t = 3000$ of the boundary are plotted outside the rectangle. The asterisks denote events with magnitudes $m > 4$.

for $n_p \in \{10, 15, \dots, 95, 100\}$. The kernel estimates are evaluated on a 100×100 pixel grid [making $n_x^{\text{bins}} = n_y^{\text{bins}} = 100$ when evaluating (19)].

As discussed in Section 4.2, 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 $\varepsilon_r = 3$ degrees and $\varepsilon_t = 3000$ days of the space–time boundary $S \times [0, T] = [0, 4] \times [0, 6] \times [0, 25,000]$. Note that the selected values, $\varepsilon_r = 3$ and $\varepsilon_t = 3000$, correspond to where the RMSE in Figure 4 begins to level off. The panels in Figure 7 show the boundary (dashed rectangles) and simulated aftershocks occurring in the specified region outside the boundary.

The estimate in Figure 6(b) resembles the overall form of the true background intensity [Figure 6(a)] and recovers many of the mainshock hotspots. However, near location (2.06, 2.33), a hotspot appears to have been erroneously estimated, that is, 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 7(a). The mean of 200 estimates of the background rate from 200 simulated realizations of ETAS is shown in Figure 6(c) and appears to closely resemble the true background rate. Hence, while there may be discrepancies for estimates from a single realization due to sampling variation, the mean of the variable kernel estimates from repeated simulation appears to correctly recover all major hotspots in the true background intensity.

The average of the estimated total number of mainshock events over the 200 simulations is 2074.85 with a bias of 74.85, standard deviation of 76.22, and root-mean-square error of 106.83. Note that while a positive bias is present, it is less

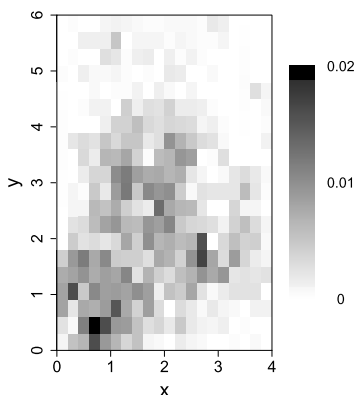


FIG. 8. Estimate of the smoothly varying background rate using the histogram estimation method (Algorithm 1, step 2) on one simulated realization of the ETAS model specified in Section 4.3.

than 4% of the average estimate, and the standard deviation of the estimates is slightly larger than the bias (which is not the case when neglecting boundary effects as in Table 1).

Figure 8 shows the estimate of the smooth background rate in Figure 6(a) when applying the histogram estimator approach (Algorithm 1, step 2) to one simulated realization of ETAS. The histogram estimate shows substantial variability between the rates in nearby cells. The recovered mainshock hotspots are also much more discernible with the kernel estimate [Figure 6(b)]. This clearly demonstrates the advantage of using the kernel method (Algorithm 2) when it is believed that the true background seismicity varies smoothly over the observation region.

Figure 9 shows the histogram estimates of the triggering function from applying Algorithm 2 to 200 realizations of the specified ETAS model. The estimates in this figure appear to successfully recover the shape of each component of the triggering function governing the simulation. The most noticeable discrepancies seem to be in the right tail for estimates of $g(t)$ and $h(r)$; these are likely due to persistent boundary effects.

Tables 9–11 give the bias, standard deviation and RMSE of the triggering estimates as defined in equations (15)–(18). For reasons discussed in Section 4.1 and 4.2, we see a positive bias in the estimates of $g(t)$ over the left and mid-range of its support, and a negative bias in the estimates of the magnitude productivity. The bias in the estimates of $h(r)$ (Table 11) is quite reasonable, and contributes substantially less to the RMSE than the standard deviation for most bins. All tables show very reasonable standard deviations in the histogram estimates as well. Boundary correction values larger than $\varepsilon_r = 3$ and $\varepsilon_t = 3000$ may result in more accurate histogram estimates, however, the selected values seem sufficient for estimating the background intensity and recovering the shapes of the triggering components.

The results of this simulation study suggest that the nonparametric estimates appear to approximate both the true background and triggering components of the

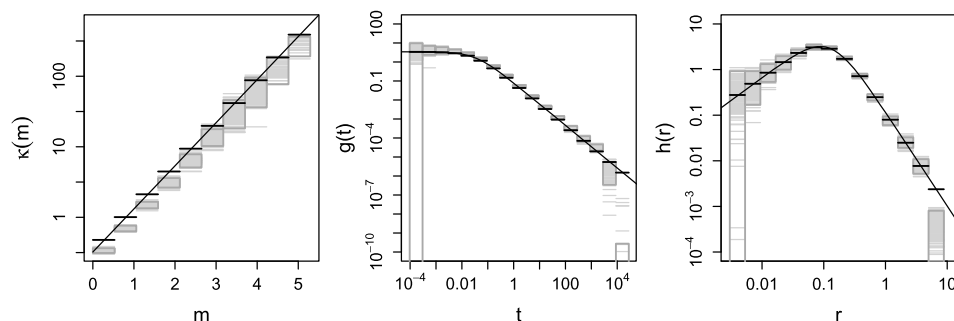


FIG. 9. Magnitude, temporal and distance components for the triggering function from the simulation study in Section 4.3. The black curves are the true triggering components used for the simulation, and the black horizontal lines are the binned average values of these functions. The gray horizontal lines in each bin are the histogram estimates from 200 simulations of ETAS; the dark gray boxes correspond to the pointwise 0.025 and 0.975 percentiles of the estimates in each bin.

process quite accurately. Note that, according to the Hawkes model, the properties (magnitude distribution, magnitude productivity, spatial and temporal distributions of triggered events) of mainshocks and aftershocks are identical, and thus we are only estimating one triggering function and one spatial background rate. Although the problem of nonidentifiability is often encountered when estimating both clustering and nonstationarity simultaneously, here this does not appear to be a problem since the form of the triggering function and spatial background rate are posited not to change over time. Hence, if the process is observed for an infinitely long time, one could recover both the true triggering function and spatial background rate up to any boundary effects.

TABLE 9

Average, bias, standard deviation and RMSE for the histogram estimates of $\kappa(m)$ from the simulation study in Section 4.3. Bias and RMSE are evaluated using the binned average values of $\kappa(m)$, which are provided in the column labeled Truth

| Bin | Truth | Avg | Bias | SD | RMSE |
|---------------|-------|-------|--------|--------|-------|
| (0, 0.529] | 0.478 | 0.34 | −0.138 | 0.0138 | 0.139 |
| (0.529, 1.06] | 1.01 | 0.705 | −0.302 | 0.0319 | 0.303 |
| (1.06, 1.59] | 2.12 | 1.49 | −0.628 | 0.0876 | 0.634 |
| (1.59, 2.12] | 4.46 | 3.12 | −1.34 | 0.255 | 1.37 |
| (2.12, 2.65] | 9.4 | 6.47 | −2.92 | 0.7 | 3.01 |
| (2.65, 3.17] | 19.8 | 13.7 | −6.09 | 2.12 | 6.45 |
| (3.17, 3.7] | 41.6 | 28.6 | −13.1 | 6.66 | 14.7 |
| (3.7, 4.23] | 87.7 | 61.7 | −26 | 16.4 | 30.8 |
| (4.23, 4.76] | 185 | 123 | −62.1 | 30.2 | 69.1 |
| (4.76, 5.29] | 389 | 288 | −101 | 52.5 | 114 |

TABLE 10
Average, bias, standard deviation and RMSE for the histogram estimates of $g(t)$ from the simulation study in Section 4.3. Bias and RMSE are evaluated using the binned average values of $g(t)$, which are provided in the column labeled Truth

| Bin | Truth | Avg | Bias | SD | RMSE |
|--------------------|----------|----------|-----------|----------|----------|
| (0.000096, 0.0003] | 3.41 | 4.36 | 0.958 | 2.56 | 2.73 |
| (0.0003, 0.00095] | 3.36 | 4.51 | 1.15 | 1.33 | 1.76 |
| (0.00095, 0.003] | 3.23 | 4.18 | 0.955 | 0.848 | 1.28 |
| (0.003, 0.0094] | 2.86 | 3.76 | 0.893 | 0.416 | 0.986 |
| (0.0094, 0.03] | 2.12 | 2.76 | 0.645 | 0.219 | 0.681 |
| (0.03, 0.093] | 1.16 | 1.52 | 0.361 | 0.0907 | 0.373 |
| (0.093, 0.29] | 0.458 | 0.599 | 0.141 | 0.0321 | 0.145 |
| (0.29, 0.93] | 0.147 | 0.192 | 0.0447 | 0.0106 | 0.046 |
| (0.93, 2.9] | 0.0428 | 0.0554 | 0.0126 | 0.00322 | 0.013 |
| (2.9, 9.2] | 0.0121 | 0.0155 | 0.00346 | 0.00104 | 0.00362 |
| (9.2, 29] | 0.00335 | 0.00432 | 0.000965 | 0.000342 | 0.00102 |
| (29, 91] | 0.000929 | 0.00118 | 0.000256 | 0.000118 | 0.000282 |
| (91, 286] | 0.000257 | 0.000335 | 7.74E-05 | 4.66E-05 | 9.04E-05 |
| (286, 898] | 7.11E-05 | 9.43E-05 | 2.32E-05 | 1.56E-05 | 2.8E-05 |
| (898, 2827] | 1.97E-05 | 3.29E-05 | 1.32E-05 | 6.35E-06 | 1.46E-05 |
| (2827, 8898] | 5.44E-06 | 2.87E-06 | −2.58E-06 | 1.41E-06 | 2.94E-06 |
| (8898, 28,000] | 1.51E-06 | 8.03E-10 | −1.5E-06 | 5.99E-09 | 1.5E-06 |

TABLE 11
Average, bias, standard deviation and RMSE for the histogram estimates of $h(r)$ from the simulation study in Section 4.3. Bias and RMSE are evaluated using the binned average values of $h(r)$, which are provided in the column labeled Truth

| Bin | Truth | Avg | Bias | SD | RMSE |
|------------------|---------|----------|-----------|----------|---------|
| (0.003, 0.0053] | 0.277 | 0.298 | 0.021 | 0.249 | 0.25 |
| (0.0053, 0.0094] | 0.488 | 0.518 | 0.0294 | 0.202 | 0.204 |
| (0.0094, 0.017] | 0.854 | 0.896 | 0.0422 | 0.209 | 0.214 |
| (0.017, 0.029] | 1.46 | 1.46 | 0.00725 | 0.238 | 0.238 |
| (0.029, 0.052] | 2.31 | 2.33 | 0.0181 | 0.203 | 0.203 |
| (0.052, 0.092] | 3.07 | 3.13 | 0.0574 | 0.194 | 0.202 |
| (0.092, 0.16] | 2.87 | 2.94 | 0.0719 | 0.125 | 0.144 |
| (0.16, 0.29] | 1.72 | 1.76 | 0.0443 | 0.0775 | 0.0893 |
| (0.29, 0.51] | 0.718 | 0.731 | 0.0135 | 0.0379 | 0.0403 |
| (0.51, 0.9] | 0.248 | 0.251 | 0.0028 | 0.0196 | 0.0198 |
| (0.9, 1.6] | 0.0796 | 0.0781 | −0.00158 | 0.0089 | 0.00903 |
| (1.6, 2.8] | 0.0249 | 0.0263 | 0.00138 | 0.0039 | 0.00414 |
| (2.8, 5] | 0.0077 | 0.00767 | −3.55E-05 | 0.00146 | 0.00146 |
| (5, 8.8] | 0.00238 | 0.000403 | −0.00198 | 0.000217 | 0.00199 |

Illustratively, Figure S3 and Table S3 in Appendix C of the supplement [Fox, Schoenberg and Gordon (2016)] show the bootstrap intervals (Section 3.3) for one nonparametric estimate of the specified ETAS model with a smoothly varying background rate.

4.4. Sensitivity to initial conditions. The nonparametric estimation methods described in Section 3 require an initialization of the mainshock-aftershock probabilities (8) in matrix P . For the simulation studies we used the initialization

$$(20) \quad p_{ij} = \begin{cases} 1/i, & j \leq i; i = 1, \dots, N, \\ 0, & \text{otherwise.} \end{cases}$$

We can consider other initializations, for instance,

$$(21) \quad p_{ij} = \begin{cases} 1, & i = j = 1, \\ \alpha, & i = j; i > 1, \\ \frac{1-\alpha}{i-1}, & j < i, \\ 0, & \text{otherwise,} \end{cases}$$

where $0 < \alpha < 1$. Setting $\alpha = 0.5$, for example, means that at initialization we assign a 0.5 mainshock probability to each event $i = 2, \dots, N$.

Figure 10 shows the convergence of the nonparametric estimates of the total mainshock rate when using different initial values for probability matrix P . The

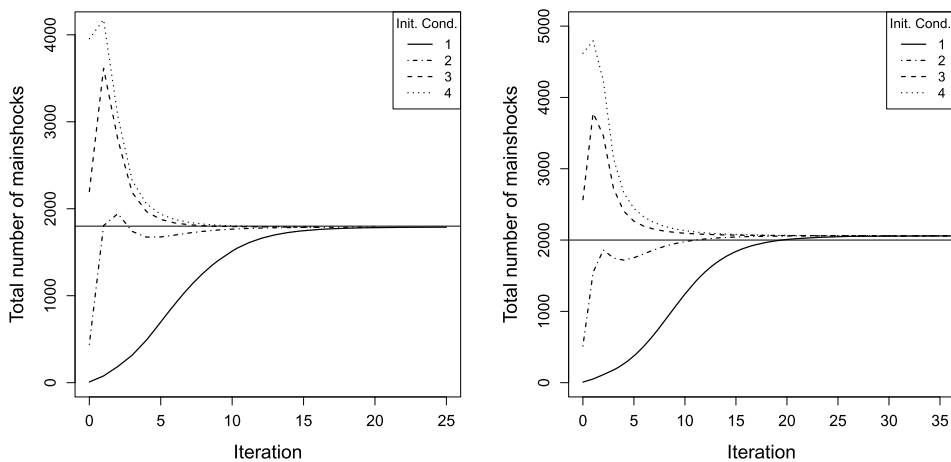


FIG. 10. Convergence of estimates of the total mainshock rate using different initial conditions. The left panel shows the convergence of Algorithm 1 applied to one realization of the ETAS model specified in Section 4.1. The right panel shows the convergence of Algorithm 2 applied to one realization of the ETAS model specified in Section 4.3. The horizontal line in each panel is the true total number of mainshocks for each simulation. Initial conditions 1–4 are given by (20) and (21) for $\alpha = 0.1, 0.5, 0.9$, respectively.

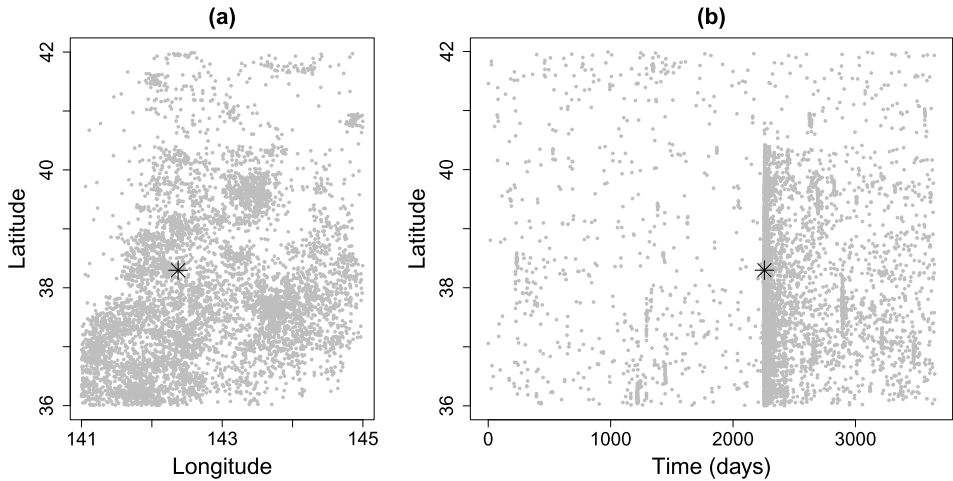


FIG. 11. *Epicentral locations (a) and space–time plot (b) of earthquakes, magnitude 4.0 or greater, occurring off the east coast of the Tohoku District, Japan. The asterisk corresponds to the 2011 Tohoku earthquake of magnitude 9.0.*

first panel shows the estimated total number of mainshocks at each iteration of Algorithm 1 applied to one simulated realization of the ETAS model specified in Section 4.1 (using boundary correction for aftershocks $\varepsilon_t = 10^6$ and $\varepsilon_r = 1000$). The second panel shows the estimated total number of mainshocks at each iteration of Algorithm 2 applied to one simulated realization of the ETAS model specified in Section 4.3. We considered the initialization in (20), as well as (21) for $\alpha = 0.1, 0.5, 0.9$. These initializations are labeled 1–4, respectively, in Figure 10.

Figure 10 demonstrates that Algorithms 1 and 2 are robust to choice of initialization of the probability matrix P since the algorithms converge to the same total mainshock rate for a wide variety of initial values.

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 window is a $141^{\circ}\sim 145^{\circ}$ E longitude and $36^{\circ}\sim 42^{\circ}$ 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 shown in Figure 11, 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 12. Here we chose the smoothing parameter $n_p = 50$, corresponding to the best choice for the simulation study in Section 4.3. Figure 12 is an important plot for assessing seismic risk since it shows the estimate of the

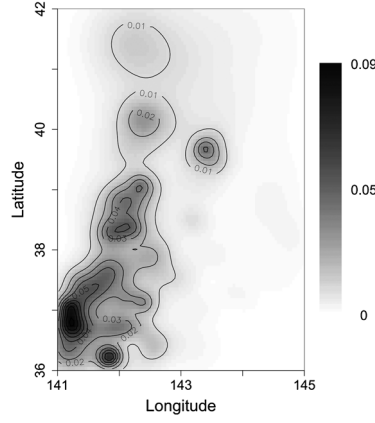


FIG. 12. Estimate of background rate (Algorithm 2, step 2) for Japan earthquake dataset (Section 5). Rate values are in events/day/degree².

underlying spatially inhomogeneous Poisson processes $\mu(x, y)$ for mainshock activity which persists over time in the region. In total, the algorithm estimated there to have been 801 mainshocks, or 13.2% of the total seismicity; this suggests that most of the events in the dataset were aftershocks, temporally and spatially linked to previously occurring earthquakes. Moreover, a comparison of Figures 11(a) and 12 reveal that most of the earthquakes observed in the lower right-hand quadrant were aftershocks.

The histogram estimate of each triggering function component is shown in Figure 13. The bootstrap error bars (Section 3.3) were formed by taking the 0.025 and 0.975 percentiles from 200 bootstrap replicates of the fitted nonparametric model.

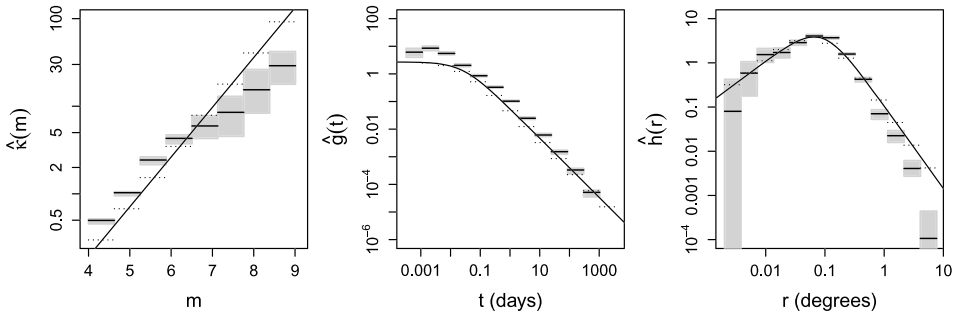


FIG. 13. Magnitude, temporal and distance components for the triggering function estimated from the Japan earthquake dataset. The histogram estimate of each component is plotted with black horizontal lines and gray bootstrap error bars. The black curves are the parametric estimates from Ogata (1998) in the same region, and the dashed horizontal lines are the binned average values of these functions.

The density estimates of $g(t)$ and $h(r)$ both exhibit power-law type behavior. The bootstrap error bars are centered around both density estimates and indicate very reasonable sampling variability for $t \in (0.001, 1000)$ days and $r \in (0.01, 4)$ degrees. Note that the estimates at the right-tail ends of these distributions ($t > 1000$ days and $r > 4$ 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. In the dataset there are only 3 events of magnitude 7.4 or above, and hence greater 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; this may be due to boundary effects since many of the aftershocks may have occurred outside the observation region. Moreover, a bias in the completeness magnitude (i.e., minimum magnitude above which all earthquakes are reliably detected) has been reported in the first few hours following the Tohoku earthquake [Lengliné et al. (2012)]. This could also explain the underestimation of the magnitude productivity since many events may not have been detected.

Superimposed on Figure 13 are the parametric estimates of the ETAS model given by equations (4)–(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 13 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.

The bootstrap technique described in Section 3.3 can also be used for uncertainty quantification for the total mainshock rate by taking the 0.025 and 0.975 percentiles of the replicates of the number of mainshock events estimated from the bootstrap datasets. However, the bootstrap interval (797, 1003) obtained for the Japan data is not centered around the estimated value of 801, and demonstrates a tendency for the bootstrap procedure to overestimate the number of mainshocks. This is perhaps not surprising since the estimates of the total mainshock rate in the Section 4.3 simulation study also showed a positive bias. Thus, the estimated value of 801 is mostly likely larger than the true total for this region. Moreover, when generating bootstrap replicates from the fitted nonparametric model, the re-estimates of the total mainshock rate are affected by similar biases, which lead to an asymmetrical percentile interval.

6. Discussion.

6.1. *Additional remarks about boundary issues.* The results from the simulation studies revealed that the nonparametric estimates are sensitive to boundary effects when only considering events that occur within the observation region. In Section 4.2, we demonstrated that these boundary effects become negligible once aftershocks occurring outside the observation region are included in the estimation. However, it is important to note that this proposed boundary correction is only applicable to simulated data and not to real earthquake data. In the simulation, mainshocks are constrained to occur within the observation region, but may trigger aftershocks outside of this region. For the boundary correction we are using the additional information that events occurring outside the observation region must be aftershocks. For real earthquake data the boundary correction cannot be applied since it is unknown whether events outside the observation region are either mainshocks or aftershocks. Designing boundary correction schemes that can be applied to real data would be an important avenue for future research.

The bootstrap intervals discussed in Section 3.3 only account for the sampling variability of the estimates, and do not account for any biases caused by boundary effects. Thus, the bootstrap error bars for the histogram estimates of the Hawkes process model can only be interpreted as approximate confidence intervals. As an illustration, Appendix C of the supplement [Fox, Schoenberg and Gordon (2016)] provides a realization of the bootstrap intervals for an estimated model from each of the simulation studies discussed in Section 4.

For the simulation studies presented in this article we only examined the case where mainshocks occurring inside the observation region can trigger aftershocks outside the observation region. However, with real observed data mainshocks occurring outside of the observation region can trigger aftershocks inside the observation region. To address the latter case, we performed additional simulation studies where the observation region is a subset of a larger spatial region over which earthquakes are generated. Estimation is then only performed on earthquakes occurring within the subset observation region. The main result is that the boundary effects on the background rate become more severe when allowing mainshocks to occur outside the observation region. The details of this additional simulation study are provided in Appendix D.1 of the supplement [Fox, Schoenberg and Gordon (2016)].

Another important question is how altering the size of the observation region affects the nonparametric estimates. In Appendix D.2 of the supplement [Fox, Schoenberg and Gordon (2016)] we implement the same simulation study discussed in Section 4.1, except on a larger observation region. Interestingly, we find that the biases in the background rate estimates do not change much when enlarging the observation region. This perhaps suggests that the severity of boundary effects on the background rate depends on whether most of the aftershock activity is accounted for and not the size of observation region.

6.2. *Future work.* There are many directions for future methodological and applied research on nonparametric estimators for Hawkes point processes. Below we list several important possibilities:

1. Due to computational limitations, we did not implement a simulation study to assess the empirical coverage of the bootstrap confidence intervals discussed in Section 3.3. Such a study would provide a quantitative evaluation of the performance of the proposed intervals. It would be necessary to first improve the computational efficiency of the estimation algorithm since an assessment of coverage requires running the estimation algorithm many thousands of times.
2. In the application to real data, the error bars for the histogram estimates of the triggering function can potentially be improved by investigating ways to incorporate measurement uncertainty in the times and locations of earthquakes.
3. In this paper we only considered modeling the conditional intensity over a two-dimensional spatial domain. However, earthquakes occur in the crust in a three-dimensional volume. A possible extension is to consider a nonparametric space–time Hawkes model with the depth of an earthquake incorporated as a mark.
4. In the simulation studies we discovered that boundary effects result in a negative bias in the right-tail ends of the estimates $g(t)$ and $h(r)$. For future applications of the method it would be useful to quantify, as a function of the size of the observation region, up to which point t and r the histogram density estimates remain reliable.
5. We can compare the weighted kernel approach described in this paper with other methods for estimating a smoothly varying background rate such as the B-spline method described in Ogata and Katsura (1988).

6.3. *Concluding remarks.* This paper investigated two new ways to incorporate an inhomogeneous background rate estimator into the EM-type method of Marsan and Lengliné (2008) for estimating a Hawkes process nonparametrically. The first approach uses a two-dimensional histogram estimator, and the second approach uses kernel smoothing. Both types of background rate estimators are weighted by the estimated probabilities that each event in the catalog is a main-shock. In this paper, we also proposed a bootstrap procedure for adding error bars to the histogram estimates of the triggering function. The error bars quantify the variability of the estimates over multiple realizations of the seismic process.

The results of the simulation studies demonstrated the ability of our nonparametric methods (Algorithms 1 and 2) to recover a spatially varying background rate and separable triggering function with components for magnitude productivity, time and distance. Some boundary effects were noticeable, such as bias in the background rate estimates and the right-tail ends of the histogram estimates of the triggering densities. Further simulations demonstrated that these boundary effects are attributable to neglecting aftershock activity occurring outside the observation

region. Once boundary effects were taken into account in the simulation study the biases in the resulting estimates were substantially less severe.

The new methodology was applied to an earthquake dataset from the Tohoku District in Japan. A striking result was that the nonparametric estimate agreed closely with a previously estimated parametric form of the ETAS model for this region. This further justifies ETAS as an adequate model of seismicity.

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 the space–time Hawkes process model may be considered. The nonparametric methods discussed in this paper can serve as a diagnostic to assess which parameterization is a good fit to the data. In other applications of self-exciting point processes, such as crime or ecology, there is a less established literature on suitable parametric models. In such applications, nonparametric estimation can be a powerful exploratory tool in determining a suitable parameterization of the triggering function. The bootstrap error bars on the histogram 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.

Acknowledgment. We thank the reviewers for very helpful comments which greatly improved this paper.

SUPPLEMENTARY MATERIAL

Supplement to “Spatially inhomogeneous background rate estimators and uncertainty quantification for nonparametric Hawkes point process models of earthquake occurrences” (DOI: [10.1214/16-AOAS957SUPP](https://doi.org/10.1214/16-AOAS957SUPP); .pdf). This supplementary document contains the following: (1) a derivation of Algorithm 1 as iterations of an EM-type algorithm; (2) the simulation algorithm for generating bootstrap datasets from a fitted nonparametric Hawkes process model; (3) additional figures and tables for Section 4; (4) additional simulation studies about boundary effects; (5) an example of nonparametrically estimating a Hawkes process model using distance to the fault.

REFERENCES

- ADELFO, G. and CHIODI, M. (2013). Mixed estimation technique in semi-parametric space–time point processes for earthquake description. In *28th International Workshop on Statistical Modeling* **1** 65–70. Palermo, Italy.
- ADELFO, 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 *The Thirty-First Asilomar Conference on Signals, Systems and Computers* **2** 1087–1091. IEEE Computer Society, Los Alamitos, CA.

- DALEY, D. J. and VERE-JONES, D. (2003). *An Introduction to the Theory of Point Processes. Vol. I: Elementary Theory and Methods*, 2nd ed. Springer, New York. [MR1950431](#)
- DAVISON, A. C. and HINKLEY, D. V. (1997). *Bootstrap Methods and Their Application. Cambridge Series in Statistical and Probabilistic Mathematics 1*. Cambridge Univ. Press, Cambridge. [MR1478673](#)
- FOX, E. W., SCHOENBERG, F. P. and GORDON, J. S. (2016). Supplement to “Spatially inhomogeneous background rate estimators and uncertainty quantification for nonparametric Hawkes point process models of earthquake occurrences.” DOI:[10.1214/16-AOAS957SUPP](#).
- GUTENBERG, B. and RICHTER, C. F. (1944). Frequency of earthquakes in California. *Bull. Seismol. Soc. Amer.* **34** 185–188.
- HAWKES, A. G. (1971). Spectra of some self-exciting and mutually exciting point processes. *Biometrika* **58** 83–90. [MR0278410](#)
- LENGLINÉ, O., ENESCU, B., PENG, Z. and SHIOMI, K. (2012). Decay and expansion of the early aftershock activity following the 2011, Mw9.0 Tohoku earthquake. *Geophysical Research Letters* **39** L18309.
- LEWIS, P. A. W. and SHEDLER, G. S. (1979). Simulation of nonhomogeneous Poisson processes by thinning. *Naval Res. Logist. Quart.* **26** 403–413. [MR0546120](#)
- 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. *Journal of Geophysical Research: Solid Earth* **115** B09302.
- MOHLER, G. O., SHORT, M. B., BRANTINGHAM, P. J., SCHOENBERG, F. P. and TITA, G. E. (2011). Self-exciting point process modeling of crime. *J. Amer. Statist. Assoc.* **106** 100–108. [MR2816705](#)
- MUSMECI, F. and VERE-JONES, D. (1992). A space–time clustering model for historical earthquakes. *Ann. Inst. Statist. Math.* **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. [MR3200305](#)
- OGATA, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes. *J. Amer. Statist. Assoc.* **83** 9–27.
- OGATA, Y. (1998). Space–time point-process models for earthquake occurrences. *Ann. Inst. Statist. Math.* **50** 379–402.
- OGATA, Y. and KATSURA, K. (1988). Likelihood analysis of spatial inhomogeneity for marked point patterns. *Ann. Inst. Statist. Math.* **40** 29–39. [MR0946013](#)
- SCHOENBERG, F. P. (2013). Facilitated estimation of ETAS. *Bull. Seismol. Soc. Amer.* **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. *J. Amer. Statist. Assoc.* **103** 614–624. [MR2523998](#)
- ZHUANG, J., OGATA, Y. and VERE-JONES, D. (2002). Stochastic declustering of space–time earthquake occurrences. *J. Amer. Statist. Assoc.* **97** 369–380. [MR1941459](#)
- ZHUANG, J., OGATA, Y. and VERE-JONES, D. (2004). Analyzing earthquake clustering features by using stochastic reconstruction. *Journal of Geophysical Research: Solid Earth* **109** B05301.

DEPARTMENT OF STATISTICS
 UNIVERSITY OF CALIFORNIA
 8125 MATH SCIENCES BUILDING
 LOS ANGELES, CALIFORNIA 90095-1554
 USA
 E-MAIL: eric.fox@stat.ucla.edu
frederic@stat.ucla.edu
joshuagordon@g.ucla.edu