Annals of the Institute of Statistical Mathematics (2019) 71:1271–1287
https://doi.org/10.1007/s10463-018-0690-9

A recursive point process model for infectious diseases

Frederic Paik Schoenberg[1] · **Marc Hoffmann**[2] · **Ryan J. Harrigan**[3]

Received: 23 May 2017 / Revised: 28 May 2018 / Published online: 9 October 2018

Abstract
We introduce a new type of point process model to describe the incidence of contagious diseases. The model incorporates the premise that when a disease occurs at low frequency in the population, such as in the primary stages of an outbreak, then anyone with the disease is likely to have a high rate of transmission to others, whereas when the disease is prevalent, the transmission rate is lower due to prevention measures and a relatively high percentage of previous exposure in the population. The model is said to be recursive, in the sense that the conditional intensity at any time depends on the productivity associated with previous points, and this productivity in turn depends on the conditional intensity at those points. Basic properties of the model are derived, estimation and simulation are discussed, and the recursive model is shown to fit well to California Rocky Mountain Spotted Fever data.

Keywords Conditional intensity · Contagious diseases · Hawkes process · Productivity

1 Introduction

The *Hawkes* self-exciting point process (Hawkes 1971) is a type of branching point process model that has become very commonly used in modeling clustered phenomena. For example, versions of Hawkes models are used to model seismicity (Ogata 1988, 1998), crimes (Mohler et al. 2011), invasive plants (Balderama et al. 2012), terrorist strikes (Porter and White 2012), and perturbations in financial markets (Bacry et al. 2013, 2015).

✉ Frederic Paik Schoenberg
frederic@stat.ucla.edu

[1] Department of Statistics, University of California, 8142 Math-Science Building, Los Angeles, CA 90095–1554, USA

[2] Ceremade, CNRS, PSL Research University, Université Paris-Dauphine, 75016 Paris, France

[3] Center for Tropical Research, Institute of the Environment and Sustainability, University of California, Los Angeles, CA 90095-1554, USA

Although Hawkes models have some features making them amenable to modeling incidence of infectious diseases, consideration of the nature of the spread of disease may suggest a somewhat different type of model. For instance, Hawkes processes have the property that the productivity (the expected number of secondary events triggered directly by the given event, or in the case of infectious disease, the expected number of transmissions from one individual to another) is static. In the case of Hawkes models applied to earthquakes (e.g., Ogata 1988, 1998), the basic Hawkes model was extended to allow the productivity of an earthquake to depend on its magnitude, but still not to depend on the time or location of the event, nor on the number of previously occurring events. When considering infectious diseases, however, this assumption of static productivity seems questionable. Early in the onset of an epidemic, when prevalence of the disease is still low, one would expect the rate of transmission to be much higher than when the prevalence of the disease is higher, because of human efforts at containment and intervention of the disease, and because some potential hosts of the disease may have already been exposed. This is a feature common to grid-based contagion models, but as noted by Law et al. (2009), relative to such models, where the surface of study is divided into an array of pixels on a grid, point processes can enable greater precision of forecasts, and can offer a more detailed and precise account of heterogeneity and clustering. Thus, we introduce a new type of point process model where the productivity (expected number of transmissions) for a subject infected at time t is a function of the conditional intensity at t. Since the conditional intensity in turn depends critically on this productivity, we call the model *recursive*.

Here we present this extension of Hawkes point process models (as they apply to infectious diseases) in the following format. After a brief review of point processes in general and Hawkes models in particular in Sect. 2, we introduce the recursive model in Sect. 3, followed by the derivation of some basic properties of the model in Sect. 4. Simulation and estimation are discussed in Sects. 5 and 6, respectively, and in Sect. 7, we fit the model to data on recorded cases of Rocky Mountain Spotted Fever in California from Jan 1, 1960, to Dec 31, 2011. Section 8 contains some concluding remarks.

2 Hawkes point processes

A point process (Daley and Vere-Jones 2003, 2007) is a σ-finite collection of points $\{\tau_1, \tau_2, \dots\}$ occurring in some metric space S. While the definitions and results below can be extended quite readily to other spaces, we will assume for simplicity throughout that the metric space S is a bounded interval $[0, T]$ in time, and the results here extend easily to the case of a space–time point process on a bounded region $B \times [0, T]$ in space–time, where B is a subset of some complete separable metric space equipped with Lebesgue measure, ℓ, and assuming for convenience that the spatial region is scaled so that $\ell(B) = 1$. A point process is *simple* if, with probability 1, none of the points overlap exactly.

A temporal point process is typically modeled via its conditional intensity, $\lambda(t)$, which represents the infinitesimal rate at which points are accumulating at time t, given information on all points occurring prior to time t. Simple point processes are

uniquely characterized by their conditional intensity (Daley and Vere-Jones 2007); for models for non-simple point processes, see Schoenberg (2006).

For a simple Hawkes process (Hawkes 1971), the conditional rate of events at time t can be written

$$\lambda(t) = \mu + K \int_0^t g\left(t - t'\right) \mathrm{d}N\left(t'\right), \tag{1}$$

where $\mu > 0$, is the background rate, $g(v) \geq 0$ is the *triggering density* satisfying $\int_0^\infty g(u)\mathrm{d}u = 1$ which describes the conductivity of events, and the constant K is the productivity, which is typically required to satisfy $0 \leq K < 1$ in order to ensure stationarity and subcriticality (Hawkes 1971).

Ogata (1988) extended the Hawkes model in order for earthquakes of larger magnitudes to have greater productivity. Hawkes models and their extension to the temporal-magnitude case were called *epidemic* by Ogata (1988), since they posit that an earthquake can produce aftershocks which in turn produce their own aftershocks, etc. Several forms of the triggering function g have been posited for describing seismological data, such as $g(u) = \frac{1}{(u+c)^p}$, where u is the time elapsed since a previous event (Ogata 1988).

Hawkes processes have been extended to describe the space–time–magnitude distribution of seismic events. A version suggested by Ogata (1998) called the epidemic-type aftershock sequence (ETAS) model uses a spatially inhomogeneous background rate and circular aftershock regions where the squared distance between an aftershock and its triggering event follows a Pareto distribution. The model may be written

$$\lambda(s,t) = \mu(s) + K \int_m \int_B \int_0^t g\left(s - s', t - t', m'\right) \mathrm{d}N\left(s', t', m'\right),$$

with triggering function

$$g(u,v,m) = \left(||u||^2 + d\right)^{-q} \exp\left\{a(m - M_0)\right\} (v + c)^{-p}, \tag{2}$$

where $||s_i - s_j||^2$ represents the squared distance between the epicenters or hypocenters s_i and s_j of earthquakes i and j, respectively, and $d > 0$ and $q > 0$ are parameters describing the spatial distribution of triggered earthquakes about their respective mainshocks.

The ETAS model has been extended by allowing the parameters to vary spatially and temporally. For example, the HIST-ETAS model (Ogata et al. 2003; Ogata 2004) assumes the parameters in the ETAS model are locally constant within small spatial–temporal cells. Similarly, Harte (2014) allows the ETAS model's productivity parameter to vary smoothly in space and time. In the following section, we extend the model in a different way, allowing the productivity to vary as a function of λ.

3 Proposed recursive model

Considering what we know from both specific disease epidemics (Wallinga and Teunis 2004; Cauchemez et al. 2016) and general modeling of outbreaks (Lloyd-Smith et al. 2015), there is good reason to question the usual assumption in Hawkes models of static productivity. For instance, when the prevalence of the disease is low or zero in a region, as is the case when the epidemic has never struck before or has not struck in considerable time, then the conditional intensity λ is small and one would expect the rate of transmission for each infected person to be quite high, as a majority of hosts are likely immunologically naive, and a carrier of the disease may be expected to infect many others. When the epidemic is at its peak and many subjects have contracted the disease, on the other hand, λ is large and one might expect the rate of transmission to be lower due to human efforts at containment and intervention of the disease, and because many subjects may have already been exposed and thus might be recovered and immune to further infection, or deceased (in either case no longer part of a susceptible pool). These considerations suggest a point process model where the productivity for a subject infected at time t is inversely related to the conditional intensity at time t. Since the conditional intensity in turn depends critically on this productivity, we call the model *recursive*.

We may write this model

$$\lambda(t) = \mu + \int_0^t H\left(\lambda_{t'}\right) g\left(t - t'\right) \mathrm{d}N\left(t'\right), \tag{3}$$

where $\mu > 0$, $g > 0$ is a density function, and $\lambda_{t'}$ means $\lambda(t')$. The productivity function H should typically be decreasing in light of the considerations above regarding the transmission of disease, and we focus in particular in what follows on the case where $H(x) = \kappa x^{-\alpha}$, with $\kappa > 0$, so that

$$\lambda(t) = \mu + \kappa \int_0^t \lambda_{t'}^{-\alpha} g\left(t - t'\right) \mathrm{d}N\left(t'\right). \tag{4}$$

The triggering density g may be given, e.g., by an exponential density,

$$g(u) = \beta \exp(\beta u). \tag{5}$$

When $\alpha = 0$, (4) reduces to a Hawkes process. We will refer to the special case where $\alpha = 1$, i.e., where

$$\lambda(t) = \mu + \kappa \int_0^t \frac{g\left(t - t'\right)}{\lambda_{t'}} \mathrm{d}N\left(t'\right) \tag{6}$$

as *standard*. The standard recursive model has especially simple and attractive features, some of which are described in Sect. 4.

4 Basic properties of the recursion model

We prove the existence of a simple point process with conditional intensity (4), and find the mean, variance, and certain large sample properties of the process.

4.1 Existence

Proposition 1 *Given a complete probability space, a recursive model with conditional intensity satisfying* (4) *can be constructed with* $H(x) = \kappa x^{-\alpha}$, *for any* $\alpha, \kappa > 0$.

Proof Let $(e_k)_{k\geq 1}$ be a sequence of independent exponential random variables with unit mean. Set $T_0 = 0$ and

$$T_{k+1} = \inf\left\{ t > T_k,\ \int_{T_k}^{t-} \left(\mu + \kappa \sum_{i=1}^{k} \lambda_{T_i}^{-\alpha} g(s - T_i)\right) \mathrm{d}s \geq e_{k+1} \right\}.$$

Define, for $k \geq 1$ the sequence of processes $N_t^{(k)} = \sum_{i=1}^{k} \mathbf{1}_{\{T_i \leq t\}}$. For a given integer k, $N^{(k)}$ only has points at $T_1, T_2, \ldots T_k$ by construction. The intensity of $N^{(k)}$ can be verified directly or by appealing to the random time change theorem (Meyer 1971, or Daley and Vere-Jones 2003, Theorem 7.4.1) such that

$$\lambda_t^{(k)} = \mu + \kappa \int_0^{t-} \left(\lambda_u^{(k)}\right)^{-\alpha} g(t-u)\mathrm{d}N_u^{(k)},$$

for $N_{t^-}^{(k)} < k$ and $\lambda_t^{(k)} = 0$ when $N_{t^-}^{(k)} = k$, where $N_{t^-}^{(k)} = \lim_{u \nearrow t} N_u^{(k)}$. Let $N_t = \lim_{k\to\infty} N_t^{(k)}$. Let us show that N_t is well-defined, *i.e.,* has no accumulation of jumps. We have

$$\begin{aligned}
E\left[N_t^{(k)}\right] &= E\left[\int_0^t \lambda_{s-}^{(k)} \mathrm{d}s\right] \\
&\leq \mu t + \kappa E\left[\int_0^t \int_0^{(s-)} \left(\lambda_u^{(k)}\right)^{-\alpha} g(s-u)\mathrm{d}N_u^{(k)}\mathrm{d}s\right] \\
&\leq \mu t + \kappa \mu^{-\alpha} E\left[\int_0^t \int_0^{(s-)} g(s-u)\mathrm{d}N_u^{(k)}\mathrm{d}s\right] \\
&= \mu t + \kappa \mu^{-\alpha} E\left[\int_0^t g(t-s)N_s^{(k)}\mathrm{d}s\right]
\end{aligned}$$

where the last line can be obtained for instance by Lemma 22 in Delattre et al. (2016). Hence

$$E\left[N_t^{(k)}\right] \leq \mu t + \kappa \mu^{-\alpha} \int_0^t g(t-s) E\left[N_s^{(k)}\right] \mathrm{d}s$$

and the function $G_k(t) = E[N_t^{(k)}]$ satisfies $G_k(t) \leq \mu t + \kappa\mu^{-\alpha} \int_0^t g(t-s)G_k(s)\mathrm{d}s$, for which Gronwall lemma implies $\sup_k G_k(t) \leq \mu t C(t, g, \mu, \alpha)$ for some constant C depending on t, g, μ and α, provided g is locally integrable (see, for instance Lemma 23(i) in Delattre et al. 2016). Letting $k \to \infty$, we infer by monotone convergence that $E\left[N_t\right] < \infty$ and thus $N_t < \infty$ P-almost surely follows. From this, one can observe that the stochastic intensity λ of N satisfies the desired equation. □

4.2 Mean and variance

The mean of the recursive process (4) can be obtained simply by using the martingale property for point processes (Daley and Vere-Jones 2003; Zhuang 2015). For $S = [0, T]$,

$$\begin{aligned}
\frac{1}{T}EN(S) &= \frac{1}{T}E\int_S \mathrm{d}N \\
&= \frac{1}{T}E\int_0^T \lambda_t \,\mathrm{d}t \\
&= \frac{1}{T}E\int_0^T \left\{\mu + \kappa\int_0^t \lambda_{t'}^{-\alpha}\, g\left(t-t'\right)\mathrm{d}N_{t'}\right\}\mathrm{d}t \\
&= \mu + \frac{\kappa}{T}E\int_0^T\int_0^t \lambda_{t'}^{1-\alpha}\, g\left(t-t'\right)\mathrm{d}t\mathrm{d}t' \\
&= \mu + \frac{\kappa}{T}E\int_0^T \lambda_{t'}^{1-\alpha}\left\{\int_0^{T-t'} g\left(t-t'\right)\mathrm{d}t\right\}\mathrm{d}t' \\
&\to \mu + \frac{\kappa}{T}E\int_0^T \lambda_{t'}^{1-\alpha}\,\mathrm{d}t', \qquad (7)
\end{aligned}$$

as $T \to \infty$, provided

$$\lim_{T\to\infty}\int_0^{T-t'} g\left(t-t'\right)\mathrm{d}t = 1,\ \forall t'. \qquad (8)$$

If assumption (8) is violated, then Eq. (7) is merely an approximation. Impacts of violations to assumption (8) are investigated in Schoenberg (2016).

Note that for the standard recursive model, $\alpha = 1$, and (7) reduces simply to

$$\mu + \kappa. \qquad (9)$$

This highlights a major difference between Hawkes models and recursive models. For a Hawkes process, doubling the background rate amounts to doubling the total expected number of points, but this is far from true for the recursive process. As an example, in the rather realistic simulations in Fig. 1a where $\mu = 0.1$ and $\kappa = 2$, doubling μ would only increase the total expected number of points by less than 5%,

and in the case of the process simulated in Fig. 1c where $\mu = 0.01$ and $\kappa = 2$, doubling μ would increase the total expected number of points by less than 0.5%.

4.3 Law of large numbers

We specialize in this section to the case $\alpha = 1$ and show that $T^{-1}N_T$ converges to $\mu + \kappa$ as $T \to \infty$ with rate of convergence $\sqrt{T}$ in L^2.

Proposition 2 *Assume* $\limsup_{T\to\infty} T^{1/2} \int_T^\infty g(t)\mathrm{d}t < \infty$. *Then*

$$\sup_T T E\left[\left(T^{-1}N_T - (\mu+\kappa)\right)^2\right] < \infty.$$

Proof Write $T^{-1}N_T - (\mu+\kappa) = A_T + B_T$, with

$$A_T = T^{-1}N_T - \frac{1}{T}\int_0^T \lambda_s \mathrm{d}s \quad \text{and} \quad B_T = \frac{1}{T}\int_0^T \lambda_s \mathrm{d}s - (\mu+\kappa).$$

We claim that both $\sup_T T E[A_T^2] < \infty$ and $\sup_T T E[B_T^2] < \infty$, from which the proposition readily follows. Let us first consider the term involving B_T. We have

$$\begin{aligned} B_T &= \mu + \kappa \frac{1}{T}\int_0^T g(T-s)\int_0^{s-} \frac{\mathrm{d}N_u}{\lambda_u}\mathrm{d}s - (\mu+\kappa) \\ &= \kappa\left\{\frac{1}{T}\int_0^T g(T-s)\widetilde{N}_s \mathrm{d}s - 1\right\} \\ &= \kappa\frac{1}{T}\int_0^T g(T-s)\left(\widetilde{N}_s - s\right)\mathrm{d}s + \kappa\left\{\frac{1}{T}\int_0^T g(T-s)s\mathrm{d}s - 1\right\} \\ &= B_T^{(1)} + B_T^{(2)}, \end{aligned}$$

say, where $\widetilde{N}_s = \int_0^{s-} \frac{\mathrm{d}N_u}{\lambda_u}\mathrm{d}s$. Clearly

$$\begin{aligned} \frac{1}{T}\int_0^T g(T-s)s\mathrm{d}s - 1 &= -\frac{1}{T}\int_0^T g(s)\mathrm{d}s + \int_0^T g(s)\mathrm{d}s - 1 \\ &= -\frac{1}{T}\int_0^T g(s)\mathrm{d}s + \int_T^\infty g(s)\mathrm{d}s \end{aligned}$$

and this (deterministic) term is $O(T^{-1/2})$ by assumption and thus $B_T^{(2)}$ has the right order. As for $B_T^{(1)}$, since $s \mapsto g(T-s)$ is a probability density, we successively use Jensen's inequality, Fubini, the fact that $\widetilde{N}_s$ is a martingale with predictable compensator s, hence $(\widetilde{N}_s - s)^2$ itself also a martingale with predictable compensator s to obtain

$$E\left[\left(B_T^{(1)}\right)^2\right] \le \kappa^2 \frac{1}{T^2}\int_0^T g(T-s)E\left[\left(\widetilde{N}_s - s\right)^2\right]\mathrm{d}s$$

$$= \kappa^2 \frac{1}{T^2} \int_0^T g(T-s) E\left[\langle \widetilde{N}_\cdot - \cdot \rangle_s\right] \mathrm{d}s$$
$$= \kappa^2 \frac{1}{T^2} \int_0^T g(T-s) s \mathrm{d}s$$

and this term multiplied by T is negligible, as for the term $B_T^{(2)}$. We finally turn to the important term A_T. Since $N_t - \int_0^t \lambda_s \mathrm{d}s$ is a martingale, its predictable compensator is also $\int_0^t \lambda_s \mathrm{d}s$. It follows that

$$E\left[A_T^2\right] = T^{-2} E\left\{\left(N_T - \int^T \lambda_s \mathrm{d}s\right)^2\right\}$$
$$= T^{-2} E\left\{\left\langle N_\cdot - \int_0^\cdot \lambda_s \mathrm{d}s \right\rangle_T\right\}$$
$$= T^{-2} \int_0^T E\left[\lambda_s\right] \mathrm{d}s.$$

The remainder of the proof consists of showing that $\sup_{s>0} E[\lambda_s] < \infty$. This follows in the same line as for the proof of non-accumulation of jumps in Proposition 1. □

4.4 Productivity

The productivity of a point τ_i is typically defined in the context of Hawkes or ETAS processes as the expected number of first generation offspring of the point τ_i. For a Hawkes process, the productivity of each point is simply K.

In the case of the recursive model (3), the productivity of any point τ_i is given by $H\{\lambda(\tau_i)\}$. Thus the total productivity, for n points $\tau_1, \tau_2, \ldots, \tau_n$, is $\sum_{i=1}^n H\{\lambda(\tau_i)\}$, and for the special case of the standard recursive model (6), the expected value of the total productivity is

$$\kappa E \int_0^T \frac{1}{\lambda_t} \mathrm{d}N_t = \kappa E \int_0^T \frac{1}{\lambda_t} \lambda_t \, \mathrm{d}t = \kappa T.$$

Thus under assumption (8) the average productivity for the standard recursive model is $\kappa T / N(S) \to \kappa/(\mu + \kappa)$ a.s., since $N(S)/T \to \mu + \kappa$ a.s. This highlights another difference between the recursive and Hawkes models. For a Hawkes process, the points $\tau_1, \tau_2, \ldots$ all have constant productivity, K. For a standard recursive process, the productivity for the first point τ_1 is κ/μ, which is larger than the productivity for any subsequent point. The productivity of the second point is $\kappa / \{\mu + \kappa g(\tau_2 - \tau_1)/\mu\}$, the productivity of the third point is $\kappa / \{\mu + \kappa g(\tau_3 - \tau_1)/\mu + \kappa g(\tau_3 - \tau_2)/(\mu + \kappa g(\tau_2 - \tau_1)/\mu)\}$, etc., and the productivity of the points ultimately averages $\kappa/(\mu + \kappa)$.

4.5 Declustering

In the seismological context, one is often interested in *mainshocks*, and it can occasionally be desirable to remove the earthquakes that could be considered *aftershocks* from a catalog. Zhuang et al. (2002) proposed a method of *stochastic declustering* for Hawkes or ETAS processes whereby one may assign to each observed point τ_i a probability that it was a mainshock, attributable to the background rate μ, and to each pair of points (τ_i, τ_j) one may compute the probability that earthquake j was *triggered* by earthquake i, and may thus be considered an *aftershock* of event i.

Similarly, when discussing the spread of a contagious disease in a given spatial region, one may consider the probability that events generated by the recursive model (3) are new outbreak points, attributable to the background rate μ, or whether point τ_j was infected by point τ_i. Such triggering or infection probabilities would be extremely relevant to a statistical analysis of epidemic data.

Fortunately, these background and infection probabilities are very easy to calculate for the recursive model. Whereas in a subcritical Hawkes process, the expected proportion of background points is $1/(1-K)$, for the standard recursive process, this proportion is $\mu/(\mu+\kappa)$. This follows directly from the formula (9) for the mean of the recursive process. Referring to the form of the recursive model in (3), for any points $\tau_i < \tau_j$, the probability that subject j was infected by subject i is given by

$$\frac{H(\lambda_{\tau_i})g(\tau_j - \tau_i)}{\mu + \int_0^{\tau_j} H(\lambda_{t'})\, g\left(\tau_j - t'\right) \mathrm{d}N_{t'}}, \tag{10}$$

which can readily be computed.

5 Simulation

One way to simulate a recursive point process is using the thinning technique of Lewis and Shedler (1979). Specifically, one sets b to some large value, generates a homogeneous Poisson process of *candidate* points with rate b on the domain S, sorts the candidate points in order of time, and for each candidate point τ_i, for $i = 1, 2, \ldots,$, one keeps the point independently of the others with probability $\lambda(\tau_i)/b$. Here, $\lambda(\tau_i)$ is computed using Eq. (3), where in calculating

$$\lambda(t) = \mu + \int_0^t H\left(\lambda_{t'}\right)\, g\left(t - t'\right) \mathrm{d}N\left(t'\right) = \mu + \sum_{i:t_i<t} H\left(\lambda_{t_i}\right)\, g(t - t_i),$$

the sum is taken over only the *kept* points t_i. Hawkes processes may be simulated in a similar manner.

Figure 1 shows simulations of a recursive process and a Hawkes process over the same domain, with the same exponential triggering density, and the same background rate μ. In the top panels, $\mu = 0.05$ and $\beta_t = 0.8$, and in the bottom panels, $\mu = 0.1$ and $\beta = 1$. In each case, the parameter K of the Hawkes process was selected as

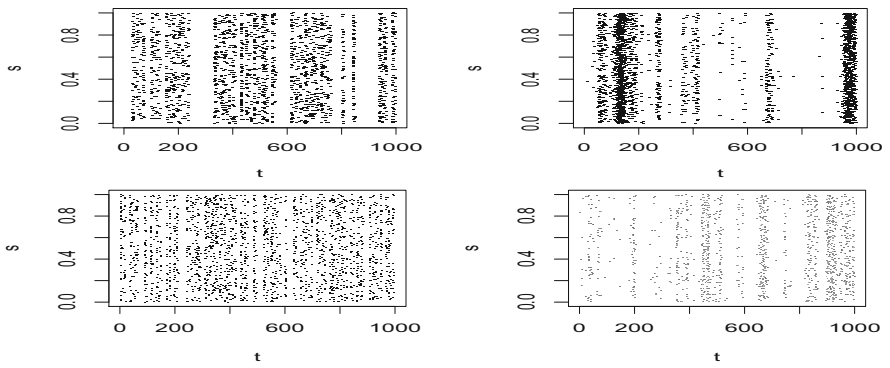

Fig. 1 (Top left) simulation of a standard recursive model (6) with $\mu = 0.05$, $\kappa = 2$, and g as in (5) with $\beta_t = 0.8$. (Top right) Simulation of a Hawkes model (1) with the same g and μ as in, and with $K = \mu/(\mu + \kappa)$ so that the processes the top row have the same expected number of points. (bottom left) Simulation of a standard recursive model (6) with $\mu = 0.1$, $\kappa = 2$, and g as in (5) with $\beta_t = 1$. (Bottom right) Simulation of a Hawkes model (1) with the same g and μ and with $K = \mu/(\mu + \kappa)$ so that the processes in the bottom row have the same expected number of points. All four simulations are over the same temporal domain [0, 1000]. The points are spread uniformly over the y axis for ease of visualization

$\kappa/(\mu + \kappa)$ so that the Hawkes and recursive processes would have the same expected number of points. One sees also how the parameter β influences the degree of clustering in the processes. One main difference between the Hawkes and recursive models is that the former exhibits occasional small clusters with just a few or even just one isolated point, whereas the latter produces almost exclusively large clusters. The form for $H(\lambda)$ implies that events occurring following a period of no events or relatively few events will trigger more offspring. In models for earthquakes such as ETAS, rare events trigger more offspring only in the sense that larger magnitude events are rare. However, in ETAS and other Hawkes-type models for earthquakes, two events of the same magnitude have the same productivity regardless of when they occur. In the epidemic disease context, this may be violated since diseases may spread more rapidly when the disease has been dormant for a while and human attempts to curb contagion are not in place. In the earthquake context, it has been noted by some authors, including notably Ogata (1988), that large earthquakes have frequently followed periods of relative quiescence, but this effect has typically not been incorporated into models such as ETAS for forecasting earthquakes.

6 Estimation

As with most temporal or space–time point process models including Hawkes and ETAS processes, the parameters in recursive point processes can be estimated by maximizing the log-likelihood,

$$L(\theta) = \int_S \log \lambda(s) \mathrm{d}N(s) - \int_S \lambda(s) \mathrm{d}t, \tag{11}$$

where θ is the vector of parameters to be estimated. Maximum likelihood estimates (MLEs) of the parameters in such point process models are consistent, asymptotically normal, and efficient (Ogata 1978).

Despite the recursive nature of the model, the log-likelihood of a recursive point process can be computed quite directly. For any given realization of points $\{\tau_1, \tau_2, \ldots, \tau_n\}$, given a particular value of the parameter vector θ, $\lambda(\tau_1) = \mu$ so one can immediately compute $H\{\lambda(\tau_1)\} = H(\mu)$, and thus $\lambda(\tau_2) = \mu + H(\mu)g(\tau_2 - \tau_1)$. One therefore has $H\{\lambda(\tau_2)\} = H(\mu + H(\mu)g(\tau_2 - \tau_1))$, and one can continue in this fashion to compute $\lambda(\tau_i)$ for $i = 1, 2, \ldots, n$.

The integral term $\int_S \lambda(s)\mathrm{d}t$ may readily be approximated in the standard way (see, e.g., Schoenberg 2013). Assuming $g(t)$ has negligible mass for $t > T - \tau_i$, one may invoke the approximation

$$\begin{aligned}\int_S \lambda(s)\mathrm{d}t &= \int_0^T \left\{\mu + \int_0^t H(\lambda(s))g(t-s)\mathrm{d}N(s)\right\}\mathrm{d}t \\ &= \mu T + \int_0^T H(\lambda(s)) \int_0^{T-s} g(u)\mathrm{d}N(s)\mathrm{d}u \\ &\approx \mu T + \int_0^T H(\lambda(s))\mathrm{d}N(s) \\ &= \mu T + \sum_i H(\lambda(\tau_i)),\end{aligned}$$

which is trivial to compute. The parameter vector θ maximizing the approximation of (11) can then be estimated by standard Newton–Raphson optimization routines. In what follows, we use the function optim() in *R*. Approximate standard errors can be derived via the diagonal elements of the inverse of the Hessian of the log-likelihood (Ogata 1978), or by repeatedly simulating and re-estimating by MLE as suggested by Harte (2010) and taking an outlier-resistant measure of standard error as in Wang et al. (2010), such as the root-median-squared errors for each parameter.

7 Application to Rocky Mountain Spotted Fever cases in California

Recorded cases of Rocky Mountain Spotted Fever in California from Jan 1, 1960, to Dec 31, 2011, were obtained from Project Tycho, www.tycho.pitt.edu (Panhuis et al. 2013). The data consist of weekly counts of confirmed cases of Rocky Mountain Spotted Fever in California published by the United States Centers of Disease Control (CDC) in its open access weekly Morbidity and Mortality Weekly Reports. Weeks with no data over this period were treated as having zero confirmed cases. Since the temporal resolution of the data is by week, the onset time for each individual case was drawn uniformly within each 7-day time interval, as, e.g., in Althaus (2014) and Chaffee (2017). Figure 2 shows a histogram of the cases, along with the estimated rate of the recursive model (4) with exponential triggering fit to the data by maximum likelihood. The estimated parameters are $(\hat{\mu}, \hat{\kappa}, \hat{\beta}, \hat{\alpha})$ = (0.000139 points/day, 0.00205 triggered points/observed point, 0.00151 points/day, 1.09), with corresponding

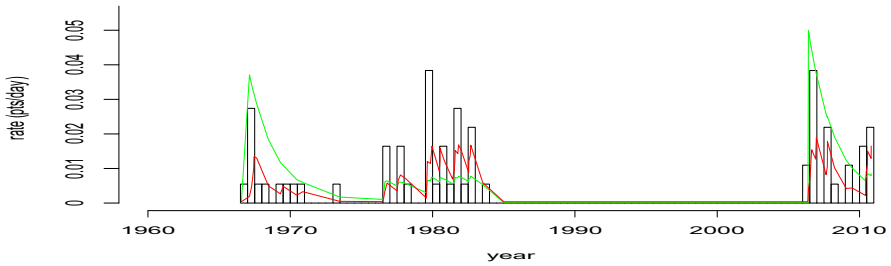

Fig. 2 Histogram (black) of Rocky Mountain Spotted Fever cases in California from Jan 1, 1960, to Dec 31, 2011, along with the estimated rate of the recursive model (green) and Hawkes model (red), each with exponential triggering function and fit by maximum likelihood (color figure online)

standard error estimates (0.000144, 0.0403, 0.00630, 0.0731). The estimated Hessian of the log-likelihood was singular in this case, so as suggested in Harte (2010), the standard errors were estimated by 100 repeated simulations and re-estimation by MLE. Note in particular the statistical significance of the estimated coefficient α, indicating significant disagreement between the data and the Hawkes model.

It is important to check that the estimates obtained by MLE are reasonable and not merely local rather than global optima, and one common way to check is to compute the ratio $\int_0^T \hat{\lambda}(t)\mathrm{d}t/N(0,T)$, as suggested in Harte (2015), which should be close to 1 since $E\int_0^T \lambda(t)\mathrm{d}t = E\int_0^T \mathrm{d}N = N(0,T)$. For the Rocky Mountain Spotted Fever data, the ratio is 0.978, and the log-likelihood is -364.2.

To compare with the recursive model (4), a Hawkes model (1) was also fit to the Rocky Mountain Spotted Fever dataset. The Hawkes model is identical to (4) but with the restriction $\alpha = 0$, and the MLEs of the other parameters are $(\hat{\mu}, \hat{\kappa}, \hat{\beta}) =$ (0.000368 points/day, 0.895 triggered points/observed point, 0.00422 points/day), with corresponding standard error estimates (0.000665, 0.0221, 0.0340), again obtained by simulating and re-estimating the model 100 times. The estimated conditional intensity of the fitted Hawkes model is also shown in Fig. 2. The log-likelihood for the fitted Hawkes model is -385.1, or 19.9 less than the log-likelihood of the recursive model. As these are nested models, the difference in log-likelihoods is approximately χ^2-distributed, and based on the Akaike Information Criterion (Akaike 1974), the improvement in fit using the recursive model is statistically significant.

In order to assess the fit of the model, we used super-thinned residuals (Clements et al. 2013). In super-thinning, one selects a constant b, thins the observations by keeping each observed point τ_i independently with probability $b/\hat{\lambda}(\tau_i)$ if $\hat{\lambda}(\tau_i) > b$, and superposes points from a Poisson process with rate $(b - \hat{\lambda})\mathbf{1}_{\hat{\lambda} \le b}$, where $\mathbf{1}$ denotes the indicator function. A default choice for b is the mean of the $\hat{\lambda}$ at the observed points, as suggested in Gordon et al. (2015). The resulting super-thinned residuals form a homogeneous Poisson process with rate b iff. $\hat{\lambda}$ is the true conditional rate of the observed point process (Clements et al. 2013). If t_i are the times of the super-thinned points, one may consider the interevent times, $r_i = t_i - t_{i-1}$ (with the convention $t_0 = 0$), which should be exponential with mean $1/b$ if the fitted model $\hat{\lambda}$ is correct, and it is natural therefore to inspect the uniformity of the standardized interevent times

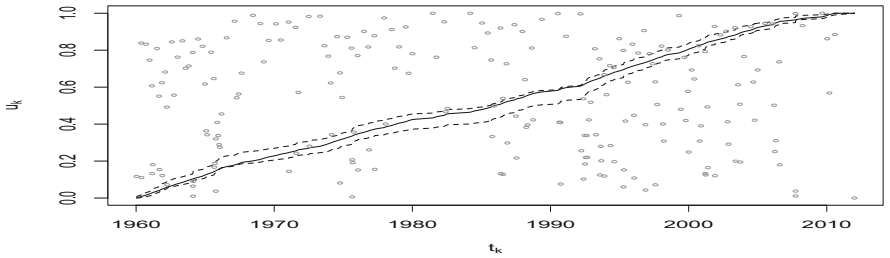

Fig. 3 Super-thinned residuals t_k using $b = 100$ points/year and their corresponding standardized interevent times u_k. The solid line shows, for each value of t_k, the normalized cumulative sum $\sum_{i=1}^{k} u_i / \sum_{i=1}^{m} u_i$, where m is the number of super-thinned residuals. Dotted lines show lower and upper simultaneous 95% confidence bounds based on 1000 simulations of the normalized cumulative sums of m uniform random variables

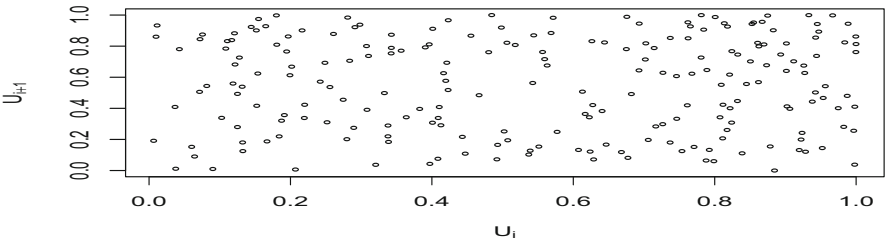

Fig. 4 Lag plot of the standardized interevent times u_i of the super-thinned residuals using $b = 100$ points/year

$u_i = F^{-1}(r_i)$, where F is the cumulative distribution function of the exponential with mean $1/b$. Figure 3 shows the super-thinned residuals t_i along with their corresponding standardized interevent times u_i, as well as the cumulative sum of the standardized interevent times, and the individual 95% confidence bounds based on 1000 simulations of an equivalent number of uniform random variables. The super-thinned residuals appear to be well scattered, though the model does not fit perfectly; there are fewer small interevent times than expected, especially between 1979 and 1985. Despite the noticeable absence of small interevent times around 1970 and 1979–1985, in general the interevent times appear to be largely well scattered, as confirmed in the lag plot of the standardized interevent times in Fig. 4.

The proposed recursive model can also be compared with the Hawkes model in terms of predictive efficacy. We separated the California Rocky Mountain Spotted Fever data into a training set (1/1/1960–12/31/2006) on which the models would be fit by MLE, and saved the last 5 years of this dataset for testing. For the fitted recursive model, the log-likelihood over the test data was 65.8, and the log-likelihood for the fitted Hawkes model was -128.8, for a difference of 194.6. To make this difference concrete, we imagine using as a threshold the 95th percentile of $\hat{\lambda}(t_i)$, evaluated at the observed points t_i in the training set. Again, this intensity $\hat{\lambda}(t)$ was fit only on the data up to 12/31/2006, and then evaluated on the test data from 2007 to 2011. With this threshold, for the recursive model, 13 out of 23 observed points (56.5%) in the test set occurred when the estimated value of λ exceeded the threshold, and only 9.6% of

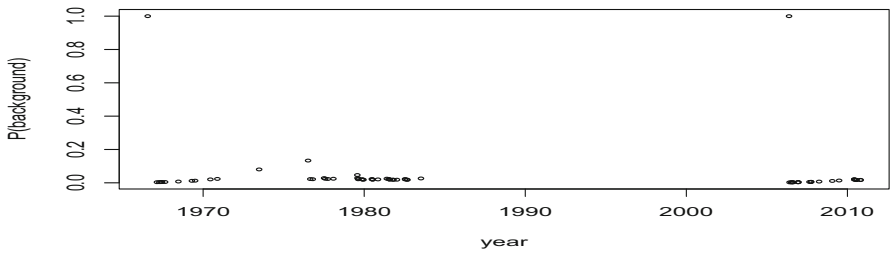

Fig. 5 Stochastic declustering of the Rocky Mountain Spotted Fever cases in CA based on the fitted model (4). For each observed point t_i, the y-coordinate, $\mu/\lambda(t_i)$, is the probability, based on model (4), that the point is attributed to the background rate (μ) as opposed to contagion from previous points

days in the test set were false alarms, i.e., days when the threshold was exceeded yet no points occurred. By contrast, for the fitted Hawkes model, the corresponding 95th percentile threshold was lower and as a result 26.2% of days in the test set would have been false alarms, yet only 7 out of 23 observed points (30.4%) in the test set occurred when $\hat{\lambda}$ exceeded the threshold.

Figure 5 shows the stochastic declustering of the Rocky Mountain Spotted Fever data in California using the fitted model (4). The y-axis shows the probability, based on the fitted model (4), that the point is attributed to the background rate (μ) as opposed to contagion from previous points. The vast majority of points are attributed to contagion rather than novel outbreaks. Two particular points in 1966 and 2006 are assigned near certainty of being attributed to new outbreaks, and two points in 1973 and 1976 are assigned substantially higher probability of being attributable to new outbreaks rather than contagion from one of the other points in the dataset, according to the fitted model.

8 Concluding remarks

The recursive point process model proposed here and fit to infectious disease data seems to be an improvement over Hawkes models because of its more general form and its flexibility, enabling it to account for changes in the rate of contagion over the course of an epidemic. We should note that although Hawkes models are widely used in seismology and are occasionally called *epidemic-type* models, and although the processes by which humans spread contagious diseases seem naturally to lend themselves to such models, the use of Hawkes models in describing the spread of infections has been scant. Exceptions are Becker (1977), who proposed purely temporal self-exciting point process models to describe the temporal spread of smallpox in Brazil, Farrington et al. (2003), who describe the effect of vaccinations on the spread of Rocky Mountain Spotted Fever in the USA using self-exciting point process models, and Balderama et al. (2012), who model invasive red banana plant locations and times using a parametric space–time Hawkes point process model. Diggle (2006) investigated inhomogeneity in foot-and-mouth disease using spatial–temporal point process models estimated by partial likelihood methods, and Diggle (2014) discusses some

successful uses of spatial–temporal point process modeling in describing in detail ecological phenomena such as the locations of Japanese black pine saplings as well as public health data such as liver cirrhosis in Northeastern England, but these efforts currently do not appear to have been widely replicated. Perhaps the added flexibility of the recursive model proposed here will facilitate the more frequent use of point process models for such epidemic data.

The improvement in fit from the recursive model relative to the Hawkes model is significant and cannot be explained as overfitting, as even when the models were fit using a training dataset (1/1/1960 to 12/31/2006) and then assessed on a separate testing time period (1/1/2007–12/31/2011), the recursive model significantly outperformed the Hawkes model using this data on Rocky Mountain Spotted Fever in California. Note that there may be cases missing from the CDC reports from which these data were obtained, as well as cases of Rocky Mountain Spotted Fever that were misdiagnosed as other diseases (Centers for Disease Control and Prevention 2015). In addition, the parameter estimates are based on when the Rocky Mountain Spotted Fever cases were reported. This may differ from the actual times when subjects contracted the disease. In this analysis, the onset times of the disease for each subject were uniformly distributed within a 7-day period, and if more detailed temporal data were available, perhaps more small-scale clustering of the data would be observed. There may be numerous covariates, such as climate, geographical and geological variables for instance, that are omitted here yet may influence the relationship observed here between previously observed points and the rate of future points. The conditional intensity may nevertheless be consistently estimated in the absence of such information provided the impact of the missing covariates is suitable small, as shown in Schoenberg (2016).

We have presented an extension of the Hawkes point process model, a recursive model, that allows for previous disease status to inform a flexible component describing the time intervals between contagious events. In the special case where the productivity is inversely proportional to the conditional intensity (i.e., when $\alpha = 1$), we have shown that this standard recursive model is computationally trivial to estimate, and does not require estimates of more complex parameters typically needed for accurate estimations of transmission events. We have demonstrated that these recursive models perform well on historical disease datasets, and can lead to insights into the transmission dynamics of particularly contagious diseases. These advances are particularly relevant, given the increase in emerging infectious diseases and their effects on human populations (Jones et al. 2008), and will hopefully encourage informed strategies as to how best prevent and mitigate future outbreaks. While we have used the term *alarms* in the context of false alarm rates and assessment of the hypothetical predictive capacity of the recursive and Hawkes models, we must emphasize that we are in no way proposing the issuance of actual alarms based on these models. Much further research is needed to determine how best to warn the public of the dangers of disease epidemics, to determine how best to decide if differences in the conditional intensity between the recursive model and alternatives such as the Hawkes model are statistically significant, and to allow the productivity to depend not only on the conditional intensity but on covariates as well, such as population density, climate, and other biotic and abiotic factors.

Acknowledgements This material is based upon work supported by the National Science Foundation under grant number DMS 1513657. Computations were performed in *R* (https://www.r-project.org). Thanks to the US CDC for supplying the data and to Project Tycho for making it so easily available. Thanks to J. Lloyd-Smith for helpful discussions on disease theory. Thanks to UCLA and Paris for allowing Professor Schoenberg a 1 year sabbatical during which time this research was performed.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*, 716–723.

Althaus, C. L. (2014). Estimating the reproduction number of Ebola virus (EBOV) during the 2014 outbreak in West Africa. *PLOS Current Outbreaks*, *6*, 1–6.

Bacry, E., Delattre, S., Hoffmann, M., Muzy, J. F. (2013). Modeling microstructure noise with mutually exciting point processes. *Quantitative Finance*, *13*, 65–77.

Bacry, E., Mastromatteo, I., Muzy, J.-F. (2015). Hawkes processes in finance. *Market Microstructure and Liquidity*, *01*(01), 1550005.

Balderama, E., Schoenberg, F. P., Murray, E., Rundel, P. W. (2012). Application of branching models in the study of invasive species. *Journal of the American Statistical Association*, *107*(498), 467–476.

Becker, N. (1977). Estimation for discrete time branching processes with application to epidemics. *Biometrics*, *33*(3), 515–522.

Cauchemez, S., Nouvellet, P., Cori, A., Jombart, T., Garske, T., Clapham, H., et al. (2016). Unraveling the drivers of MERS-CoV transmission. *Proceedings of the National Academy of Sciences*, *113*(32), 9081–9086.

Centers for Disease Control and Prevention. (2015). Chapter 13. In J. Hamborsky, A. Kroger, S. Wolfe (Eds.), *Epidemiology and prevention of vaccine-preventable diseases* (13th ed., pp. 209–230). Washington D.C.: Public Health Foundation.

Chaffee, A. W. (2017). *Comparative Analysis of SEIR and Hawkes Models for the 2014 West Africa Ebola Outbreak*. MS Thesis, UCLA Statistics, University of California, Los Angeles (pp. 1–32).

Clements, R. A., Schoenberg, F. P., Veen, A. (2013). Evaluation of space–time point process models using super-thinning. *Environmetrics*, *23*(7), 606–616.

Daley, D., Vere-Jones, D. (2003). *An introduction to the theory of point processes, volume 1*: *Elementary theory and methods* (2nd ed.). New York: Springer.

Daley, D., Vere-Jones, D. (2007). *An introduction to the theory of point processes, volume 2*: *General theory and structure* (2nd ed.). New York: Springer.

Delattre, S., Fournier, N., Hoffmann, M. (2016). Hawkes processes on large networks. *Annals of Applied Probability*, *26*, 216–261.

Diggle, P. J. (2006). Spatio-temporal point processes, partial likelihood, foot-and-mouth. *Statistical Methods in Medical Research*, *15*, 325–336.

Diggle, P. J. (2014). *Statistical analysis of spatial and spatio-temporal point patterns* (3rd ed.). Boca Raton: CRC Press.

Farrington, C. P., Kanaan, M. N., Gay, N. J. (2003). Branching process models for surveillance of infectious diseases controlled by mass vaccination. *Biostatistics*, *4*(2), 279–295.

Gordon, J. S., Clements, R. A., Schoenberg, F. P., Schorlemmer, D. (2015). Voronoi residuals and other residual analyses applied to CSEP earthquake forecasts. *Spatial Statistics*, *14B*, 133–150.

Harte, D. S. (2010). PtProcess: An R package for modelling marked point processes indexed by time. *Journal of Statistical Software*, *35*(8), 1–32.

Harte, D. S. (2014). An ETAS model with varying productivity rates. *Geophysical Journal International*, *198*(1), 270–284.

Harte, D. S. (2015). Log-likelihood of earthquake models: Evaluation of models and forecasts. *Geophysical Journal International*, *201*(2), 711–723.

Hawkes, A. G. (1971). Point spectra of some mutually exciting point processes. *Journal of the Royal Staistical Society*, *B33*, 438–443.

Jones, K. E., Patel, N. G., Levy, M. A., Storeygard, A., Balk, D., Gittleman, J. L., et al. (2008). Global trends in emerging infectious diseases. *Nature*, *451*, 990–993.

Law, R., Illian, J., Burslem, D. F. R. P., Gratzer, G., Gunatilleke, C. V. S., Gunatilleke, I. A. U. N. (2009). Ecological information from spatial patterns of plants: Insights from point process theory. *Journal of Ecology*, *97*(4), 616–628.

Lewis, P. A. W., Shedler, G. S. (1979). Simulation of non-homogeneous Poisson processes by thinning. *Naval Research Logistics Quarterly*, *26*(3), 403–413.

Lloyd-Smith, J. O., Funk, S., McLean, A. R., Riley, S., Wood, J. L. N. (2015). Nine challenges in modelling the emergence of novel pathogens. *Epidemics*, *10*, 35–39.

Meyer, P. (1971). Demonstration simplifiée d'un théorème de Knight. In *Séminaire de Probabilités V, Université Strasbourg, Lecture notes in mathematics* (Vol. 191, pp. 191–195).

Mohler, G. O., Short, M. B., Brantingham, P. J., Schoenberg, F. P., Tita, G. E. (2011). Self-exciting point process modeling of crime. *Journal of the American Statistical Association*, *106*(493), 100–108.

Ogata, Y. (1978). The asymptotic behavious of maximum likelihood estimators for stationary point processes. *Annals of the Institute of Statistical Mathematics*, *30*, 243–261.

Ogata, Y. (1988). Statistical models for earthquake occurrence 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*(2), 379–402.

Ogata, Y. (2004). Space–time model for regional seismicity and detection of crustal stress changes. *Journal of Geophysical Research*, *109*(B3), B03308, 1–16.

Ogata, Y., Katsura, K., Tanemura, M. (2003). Modelling heterogeneous spacetime occurrences of earthquakes and its residual analysis. *Applied Statistics*, *52*, 499–509.

Porter, M. D., White, G. (2012). Self-exciting hurdle models for terrorist activity. *Annals of Applied Statistics*, *6*(1), 106–124.

Schoenberg, F. P. (2006). On non-simple marked point processes. *Annals of the Institute of Statistical Mathematics*, *58*(2), 223–233.

Schoenberg, F. P. (2013). Facilitated estimation of ETAS. *Bulletin of the Seismological Society of America*, *103*(1), 601–605.

Schoenberg, F. P. (2016). A note on the consistent estimation of spatial-temporal point process parameters. *Statistica Sinica*, *26*, 861–789.

van Panhuis, W. G., Grefenstette, J., Jung, S. Y., Chok, N. S., Cross, A., Eng, H., et al. (2013). Contagious diseases in the United States from 1888 to the present. *New England Journal of Medicine*, *369*(22), 2152–2158.

Wallinga, J., Teunis, P. (2004). Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. *American Journal of Epidemiology*, *160*(6), 509–516.

Wang, Q., Schoenberg, F. P., Jackson, D. D., Kagan, Y. Y. (2010). Standard errors of parameter estimates in the ETAS model. *Bulletin of the Seismological Society of America*, *100*(5a), 1989–2001.

Zhuang, J. (2015). Weighted likelihood estimators for point processes. *Spatial Statistics*, *14*, 166–178.

Zhuang, J., Ogata, Y., Vere-Jones, D. (2002). Stochastic declustering of space–time earthquake occurrences. *Journal of the American Statistical Association*, *97*(458), 369–380.