BAYESIAN MITIGATION OF SPATIAL COARSENING FOR A HAWKES MODEL APPLIED TO GUNFIRE, WILDFIRE AND VIRAL CONTAGION

BY ANDREW J. HOLBROOK[1,a], XIANG JI[2,b] AND MARC A. SUCHARD[3,c]

[1]*Department of Biostatistics, UCLA,* [a]*aholbroo@g.ucla.edu*

[2]*Department of Mathematics, Tulane University,* [b]*xji4@tulane.edu*

[3]*Departments of Biostatistics, Human Genetics and Computational Medicine, UCLA,* [c]*msuchard@ucla.edu*

Self-exciting spatiotemporal Hawkes processes have found increasing use in the study of large-scale public health threats, ranging from gun violence and earthquakes to wildfires and viral contagion. Whereas many such applications feature locational uncertainty, that is, the exact spatial positions of individual events are unknown, most Hawkes model analyses to date have ignored spatial coarsening present in the data. Three particular 21st century public health crises—urban gun violence, rural wildfires and global viral spread—present qualitatively and quantitatively varying uncertainty regimes that exhibit: (a) different collective magnitudes of spatial coarsening, (b) uniform and mixed magnitude coarsening, (c) differently shaped uncertainty regions and—less orthodox—(d) locational data distributed within the "wrong" effective space. We explicitly model such uncertainties in a Bayesian manner and jointly infer unknown locations together with all parameters of a reasonably flexible Hawkes model, obtaining results that are practically and statistically distinct from those obtained while ignoring spatial coarsening. This work also features two different secondary contributions: first, to facilitate Bayesian inference of locations and background rate parameters, we make a subtle yet crucial change to an established kernel-based rate model, and second, to facilitate the same Bayesian inference at scale, we develop a massively parallel implementation of the model's log-likelihood gradient with respect to locations and thus avoid its quadratic computational cost in the context of Hamiltonian Monte Carlo. Our examples involve thousands of observations and allow us to demonstrate practicality at moderate scales.

1. Introduction. Spatiotemporal Hawkes processes (Reinhart (2018)) are stochastic point processes that have found use in the modeling of various *self-excitatory* phenomena in space and time, such as earthquakes and their aftershocks (Hawkes (1973), Ogata (1988), Zhuang, Ogata and Vere-Jones (2004), Fox, Schoenberg and Gordon (2016)), retaliatory gun violence (Loeffler and Flaxman (2018), Park et al. (2019), Holbrook et al. (2021a)), wildfires (Schoenberg (2004)) and viral epidemics (Kim (2011), Meyer and Held (2014), Choi et al. (2015), Rizoiu et al. (2018), Kelly et al. (2019)). These applications all share at least one characteristic that, after observing an event, one expects to observe one or more events nearby and soon after. Because spatial proximity to an event increases the probability of observing another event, accurate model inference hinges on precise locational data.

Unfortunately, noisy, incomplete or otherwise coarsened spatial data seem to be the norm in many Hawkes process applications. Urban gunfire data sources may provide location data at city block precision or rounded to the nearest 100 meters (Holbrook et al. (2021a)). Scientists estimate the spatial position of an earthquake from noisy seismic wave energy arrival times at remote stations (Lomax, Michelini and Curtis (2009)). Thus, the recording of seismic locations amounts to an inverse problem arising from the use of remote sensing and complex

physical models. In the best scenario and due to privacy concerns, viral case registries provide the hospital or medical clinic that receives the sickened patient, an imprecise stand-in for the location at which the patient first contracts the virus. More often, epidemiological data arise from heterogeneous public health sources that make use of varying levels of spatial precision, be they on the national, provincial or municipal level (Park et al. (2018), Holbrook et al. (2021b)). We account for such spatial coarsening by directly incorporating locational uncertainty into our model in the form of prior distributions on spatial positions of individual events. The upshot is a Bayesian hierarchical model with global structure and event-specific prior distributions dictated by the weaknesses of the data at hand. Section 2.2 discusses these priors and how they relate to the general theoretical framework for coarsening established in Heitjan and Rubin (1991).

We demonstrate our approach with three distinct 21st century public health crises, each featuring its own particular spatial uncertainty and scope. We first consider Washington D.C. gunfire data generated throughout the span of 2018. Here, the Government of the District of Columbia has purposefully rounded each gunshot's latitudinal and longitudinal coordinates to an effective 100 meter precision. Because this data originate from a spatially precise acoustic gunshot location system (AGLS) (Loeffler and Flaxman (2018)), a reasonable prior on the spatial position of each gunshot is a uniform distribution on the 10,000 m^2 square centered at the observed data. Alaskan wildfire data from the years 2015 to 2019 feature a different kind of spatial uncertainty. Each observation features approximate spatial coordinates of the fire at the time of discovery as well as the fire's size, in acres, at the time of discovery. Because we do not know the direction of each fire's expansion at the time of discovery, the principle of indifference suggests that we model each wildfire's ignition location as taking position with equal probability within a circle centered at the given discovery coordinates but with area matching the discovery's acreage. In contrast to the Washington D.C. gunfire example, this application provides for differential spatial uncertainty between events.

A third application, the global spread of influenza from 2000 to 2012, presents a radically different flavor of spatial bias. Because spatial proximity to an event increases the probability of observing another event, the statistician that employs a spatiotemporal Hawkes process must take care to adequately define spatial relationships between locations in a way that takes the nature of the target phenomenon into account. Viruses spread across immensely complex human networks shaped by our relationships, institutions and economies. On the global scale, human air transportation networks capture the majority of viral transmission between geographic locations (Brockmann and Helbing (2013)). Thus, to model global viral spread, one must build information about these transportation networks into one's model. Failing to do so may lead to biased results that deliver incorrect insights into a crucial global public health challenge. This difficulty could be one of the reasons that the spatiotemporal Hawkes process has not found use for modeling global viral transmission. Another reason for such a hole in the literature is that global epidemiological data often arise from heterogeneous public health sources that make use of varying levels of locational precision. Since spatial nearness is a primary datum for the spatiotemporal Hawkes process, it is essential that our conception of nearness be coherent. How far is Beijing from China? How far is California from France? We would like to avoid such questions as well as mixed-methodological approaches, such as randomizing locations labels to, say, cities, according to some contrived weighting scheme prior to analysis.

We must, therefore, use an expressive prior to simultaneously account for these two sources of spatial bias. Bayesian multidimensional scaling (DeSarbo, Kim and Fong (1999), Oh and Raftery (2001), Oh and Raftery (2007), Holbrook et al. (2021b)) probabilistically maps from pairwise global air transportation distances between countries to random variables within a latent Euclidean space, while our spatiotemporal Hawkes model describes the

spread of viral cases, the locations of which are the very same low-dimensional latent variables. For viral case data arising from the same country, the temporal information provided by the Hawkes process efficiently informs the distribution of latent locations on a finer, domestic scale.

In meeting our goal of joint and fully Bayesian inference over location variables and model parameters, we must develop a model for the Hawkes process background rate that admits posterior inference for its individual parameters while retaining flexibility. As a secondary contribution, we develop just such a novel background rate model and use MCMC to compute posterior distributions for all spatiotemporal Hawkes process parameters, a first in the presence of a nontrivial background rate. As another secondary contribution, we have made significant additions to the HPHAWKES R package https://github.com/suchard-group/hawkes (Holbrook, Ji and Suchard (2022)) to facilitate high-performance computing for posterior distributions of Hawkes process locations. In particular, we have developed a fully parallelized implementation of the Hawkes log-likelihood gradient with respect to spatial locations (Appendix A).

2. Modeling. The strategy we use to model our three different target applications is to specify a single, adequately flexible data generative process in the form of a spatiotemporal Hawkes model (Section 2.1) and to design priors on event locations based on spatial biases encoded in each application's data (Section 2.2).

2.1. *Spatiotemporal self-excitation.* The spatiotemporal Hawkes process is an inhomogeneous Poisson point process (Daley and Vere-Jones (2003, 2008)) model for random variables $(\mathbf{x}, t) \in \mathbb{R}^D \times \mathbb{R}^+$ in space and time, where the intensity function

$$\lambda(\mathbf{x}, t) = \mu(\mathbf{x}, t) + \xi(\mathbf{x}, t) = \mu(\mathbf{x}, t) + \sum_{t_n < t} g(\mathbf{x} - \mathbf{x}_n, t - t_n)$$

describes the infinitesimal rate conditioned on all other observations $(\mathbf{x}_n, t_n)$ for $n = 1, \ldots, N$ and $\mathbf{x}_n = (x_{n1}, \ldots, x_{nD})$. Here, $\mu(\cdot, \cdot)$ is the background rate symmetric in time, $\xi(\mathbf{x}, t)$ the self-excitatory rate and $g(\cdot, \cdot)$ a triggering function determining the self-excitatory behavior of the process. As in Mohler (2014), Loeffler and Flaxman (2018), Holbrook et al. (2021a), we specify a triggering function that is exponential in time and Gaussian in space

$$\xi(\mathbf{x}, t) = \frac{\theta \omega}{h^D} \sum_{t_n < t} e^{-\omega(t - t_n)} \phi\left(\frac{\mathbf{x} - \mathbf{x}_n}{h}\right),$$

where ω, h and θ are strictly positive parameters. Similar to Holbrook et al. (2021a) but with the inclusion of the indicator function $\mathcal{I}_{[t \neq t_n]}$, a key difference, we use a flexible Gaussian kernel smoother for the endemic rate

$$\mu(\mathbf{x}, t) = \frac{\mu_0}{\tau_x^D \tau_t} \sum_{n=1}^{N} \phi\left(\frac{\mathbf{x} - \mathbf{x}_n}{\tau_x}\right) \phi\left(\frac{t - t_n}{\tau_t}\right) \mathcal{I}_{[t \neq t_n]},$$

where the indicator function efficiently ensures that events do not contribute to their own probability of occurrence (see joint probability density function equation (1), below), representing a novel and necessary departure from Holbrook et al. (2021a) if one wishes to infer background rate parameters and process locations in a Bayesian fashion. We call $1/\omega$ and h and τ_t and τ_x the self-excitatory and background lengthscales or bandwidths, respectively. We call μ_0 and θ the background and self-excitatory rate weights, and their relative magnitudes determine the amount of self-excitatory behavior exhibited by the process. With $\Theta = (\mu_0, \tau_x, \tau_t, \theta, \omega, h)$, the likelihood (Daley and Vere-Jones (2003)) for data

$(\mathbf{X}, \mathbf{t}) = ((\mathbf{x}_1, t_1), \ldots, (\mathbf{x}_N, t_N))$ is

$$\mathcal{L}(\mathbf{X}, \mathbf{t}|\Theta) = \exp\left(-\int_{\mathbb{R}^D} \int_0^{t_N} \lambda(\mathbf{x}, t)\, dt\, d\mathbf{x}\right) \prod_{n=1}^N \lambda(\mathbf{x}_n, t_n) := e^{-\Lambda(t_N)} \cdot \prod_{n=1}^N \lambda_n.$$

Although integrating over the entirety of $\mathbb{R}^D$ rather than a relevant subset is a popular and often necessary modeling decision, one must regard this choice as an approximation when measurement over $\mathbb{R}^D$ is incomplete (Schoenberg (2013)). This fact will provide an additional argument for our proposed modeling approach in Section 2.2.3. The background rate's indicator function does not change the integration term, so $\Lambda(t_N)$ is the same as in Holbrook et al. (2021a),

$$\Lambda(t_N) = \mu_0 \sum_{n=1}^N \left(\Phi\left(\frac{t_N - t_n}{\tau_t}\right) - \Phi\left(\frac{-t_n}{\tau_t}\right)\right) - \theta \sum_{n=1}^N \left(e^{-\omega(t_N - t_n)} - 1\right)$$

$$= \sum_{n=1}^N \left(\mu_0\left(\Phi\left(\frac{t_N - t_n}{\tau_t}\right) - \Phi\left(\frac{-t_n}{\tau_t}\right)\right) - \theta\left(e^{-\omega(t_N - t_n)} - 1\right)\right) := \sum_{n=1}^N \Lambda_n.$$

Taken together, the log-likelihood is

$$\ell(\mathbf{X}, \mathbf{t}|\Theta) = -\Lambda(t_N) + \sum_{n=1}^N \log \lambda_n$$

(1)

$$= \sum_{n=1}^N \left\{\log\left[\sum_{n'=1}^N \left(\frac{\mu_0 \mathcal{I}_{[t_n \neq t_{n'}]}}{\tau_x^D \tau_t} \phi\left(\frac{\mathbf{x}_n - \mathbf{x}_{n'}}{\tau_x}\right) \phi\left(\frac{t_n - t_{n'}}{\tau_t}\right)\right.\right.\right.$$

$$\left.\left.\left. + \frac{\theta \omega \mathcal{I}_{[t_{n'} < t_n]}}{h^D} e^{-\omega(t_n - t_{n'})} \phi\left(\frac{\mathbf{x}_n - \mathbf{x}_{n'}}{h}\right)\right)\right] - \Lambda_n\right\}$$

$$:= \sum_{n=1}^N \left[\log\left(\sum_{n'=1}^N \lambda_{nn'}\right) - \Lambda_n\right] := \sum_{n=1}^N \ell_n.$$

In all three applications we equip μ_0 and θ with standard normal priors truncated to be greater than 0. In contrast to Mohler (2014), Loeffler and Flaxman (2018), Holbrook et al. (2021a), we perform joint inference on all model lengthscales. To do so, we lend truncated normal priors to all model inverse lengthscales. We maintain constraints $0 < 1/\omega < \tau_t$ and $0 < h < \tau_x$. Finally, we set the prior standard deviations of the background inverse lengthscales to be 10 times those of their respective self-excitatory counterparts. In this way we encode our general expectation that self-excitation occurs at a finer scale than that of the background process. Regardless, we find that the thousands of observations present in each of our applications can easily overpower the soft prior constraints given by the prior standard deviations (Table 2).

Importantly, our model is similar to that of Mohler (2014), who finds that the spatial lengthscales h and τ_x may sometimes be exchanged with only a small change to the likelihood. For this reason, Mohler (2014) fixes the two parameters to be equal to avoid multimodality. Whereas our priors help ameliorate this issue, they do not solve it. Amazingly, we find that inferring locations, as discussed in the following section, can actually help solve this problem (Table 2). As an upshot, we are able to retain full model flexibility. This is all the more important because Reinhart and Greenhouse (2018) show that poorly estimated background processes contribute to biased estimates of self-excitation.

2.2. *Modeling spatial uncertainty.* Whereas we apply the same likelihood to each of our target applications, we must craft our priors on individual event locations in a way that respects the phenomenon being modeled and the spatial coarsening that gives rise to each specific dataset. Rather than obtaining observations that belong to the sample space of our random variable of interest, we observe *coarse* data (Heitjan and Rubin (1991), Heitjan (1993)) within the power set of that sample space. Rounded, heaped, truncated, censored and missing data are just a few common examples of coarsening. If one knows the coarsening mechanism and it is not stochastic, then the data are *grouped*. Rounding and truncation with fixed precision are prominent examples of grouping. Addressing rounding is as simple as integrating the originating likelihood over the uncertainty region the rounding induces. Adopting the established parlance of missing data, Heitjan and Rubin (1991) show that a stochastic coarsening mechanism is *ignorable* if: (a) it is coarsened at random (CAR), and (b) the parameters of the data generating and coarsening processes are distinct. If one assumes that ignorability holds, then modeling the data as grouped, that is, ignoring the stochasticity of the coarsening mechanism, is completely valid. In this paper we use the phrases spatial coarsening, spatial uncertainty and even spatial bias interchangeably.

In the remainder, $\mathbf{x}_n$ continues to denote an individual location that interfaces directly with the Hawkes model likelihood. This is a location variable. We denote its corresponding observed locational datum as $\mathfrak{x}_n$ and let $\mathfrak{X}$ be the collection of all N observed locations.

2.2.1. *Washington D.C. gun violence.* We first apply our Hawkes model to analyze gunfire in the American capital throughout the year 2018. The data feature 3982 gunshots obtained from a spatially precise AGLS (Section 1) but with latitudinal and longitudinal coordinates rounded to the nearest three decimal points for the purpose of privacy. This rounding amounts to recording observations within an approximate 100 meter precision in localized vertical and horizontal axes. Due to the precision of the original AGLS data, we are confident in specifying uniform priors over the 10000 m^2 square centered at each location for each location, that is, in local coordinates scaled to meters,

$$(2) \qquad p(\mathbf{x}_n) \propto 1, \qquad \mathfrak{x}_{nd} - 50 < \mathbf{x}_{nd} < \mathfrak{x}_{nd} + 50, \quad n = 1, \ldots, N, d = 1, 2.$$

Our uncertainty is uniform in both shape and magnitude throughout the sample. As stated above, rounding is an example of grouping, and we know that our prior specification is valid and corresponds directly to inference based on the *grouped-data likelihood* of Heitjan and Rubin ((1991), Example 1). In other words, our latent variable formulation accounts for grouping by integrating over the region of uncertainty induced by the grouping mechanism. Failing to account for this grouping leads to biased inference.

2.2.2. *Alaskan wildfire ignitions.* Next, we model the occurrence and spread of 2925 wildfires in Alaska through the years 2015 to 2019. Specifically, we would like to use the exact time and place of ignition for each wildfire as our data. Instead, we have the time, rough spatial coordinates $\mathfrak{x}_n$ and area A_n of the fire at discovery. Because we do not know the direction of each wildfire's extent, we invoke the principle of indifference (Marquis de Laplace (1925)) and assume equal extent in all directions, that is, that each uncertainty region is a circle centered at the given coordinates, assuming effects of geography are negligible. Here, we specify the radius r_n of each circular uncertainty region so that the circle's area matches the size of the wildfire at time of discovery,

$$(3) \qquad p(\mathbf{x}_n) \propto 1, \qquad \|\mathbf{x}_n - \mathfrak{x}_n\|_2 < r_n = \sqrt{A_n/\pi}, \quad n = 1, \ldots, N, D = 2.$$

This example, therefore, stands in contrast to the gun violence example, insofar as the shape of uncertainty regions are circular rather than square and the magnitudes of these circles vary

across all observations. The coarsening mechanism appears to be random, but it is impossible to capture the complicated processes that lead a passerby to discover a wildfire at any particular extent. Unlike the D.C. gunfire example, we do not know the exact circumstances that bring about the data's observed spatial coordinates, so we must simply assume that the CAR condition holds. Less problematic is the assumption that data generating and data coarsening mechanisms have distinct parameters. Having arrived at ignorability, our prior specification again corresponds to valid inference based on the grouped-data likelihood.

2.2.3. *Global influenza contagion.*

2.2.3.1. *Doubly debiased inference.*   Four thousand seven hundred thirty three influenza cases collected from 64 countries worldwide between 2000 and 2012 provide a much more difficult modeling task. Approximately, one-third of the observations bear labels for the city, another third for the province or state and the last third for the country in which the case occurred. We would like to proceed in a similar manner, as with the gun violence and wildfires analyses, but restricting location variables to the complicated borders of countries, provinces or cities is technically infeasible. Furthermore, naive spatial distances between locations on planet Earth fail to capture the way viruses propagate around the globe. The global human air transportation network, specifically, the number of humans traveling between locations, provides a much better tool for tracking the spread of viral strains (Brockmann and Helbing (2013)). We, therefore, propose to model the locations of each viral case in such a way that simultaneously accounts for the multiprecision nature of the data and the outsized role played by human air transport.

Classical multidimensional scaling (MDS) is a two-step method for mapping from pairwise distances or dissimilarities between objects to representations of these objects within a low-dimensional Euclidean space (Kruskal (1964)). In modeling the global spread of influenza, we let $\mathbf{Y} = \mathbf{Y}(\mathfrak{X})$ be an $N \times N$ matrix of pairwise distances $y_{nn'}$ generated by Brockmann and Helbing (2013) and inversely related to the number of air traffic passengers exchanged between the countries where cases n and n' occurred. Given any such matrix $\mathbf{Y}$, the centering transformation

$$\mathbf{Y} \longmapsto -\frac{1}{2}\Big(\mathbf{I} - \frac{1}{N}\mathbf{1}\mathbf{1}^T\Big)\mathbf{Y}^{\circ 2}\Big(\mathbf{I} - \frac{1}{N}\mathbf{1}\mathbf{1}^T\Big)$$

results in a positive semidefinite matrix corresponding to the sample covariance of N points existing in some D-dimensional subspace of N-dimensional Euclidean space. After obtaining this sample covariance, a simple application of principal component analysis (PCA) (Pearson (1901)) renders a low-dimensional representation of the N objects of interest. On the one hand, objects with smaller pairwise distances arrange themselves closer in L_2 distance within the low-dimensional space than objects with larger pairwise distances, leading to interpretable visualizations. On the other hand: (1) both the centering transformation and the eigendecomposition of PCA scale $O(N^3)$ in computational complexity, (2) low-dimensional representations fail to communicate uncertainty arising from randomness in the data generating mechanism and (3) secondary modeling of the low-dimensional representations results in difficult to quantify dependencies on the MDS process.

Bayesian MDS (BMDS) offers a way around these problems by positing that each object's latent location is a random variable, translating the MDS projection into a probability model on the observed pairwise distances conditioned on distances between latent locations (Ramsay (1982)) and specifying an appropriate prior distribution over these locations. For any two distinct objects n and n', we follow Oh and Raftery (2001) and model their observed pairwise distance as conditionally independent, truncated normal random variables

$$y_{nn'} \sim \mathrm{N}\big(\delta_{nn'}, \sigma^2\big)\mathcal{I}_{[y_{nn'}>0]} \quad \text{for } n > n',$$

where the centrality parameter $\delta_{nn'} = \|\mathbf{x}_n - \mathbf{x}_{n'}\|_2$ is the Euclidean distance between latent locations $\mathbf{x}_n$ and $\mathbf{x}_{n'}$ in $\mathbb{R}^D$. Conditioned on all latent locations $\mathbf{X}$, the probability density function of observed distance data $\mathbf{Y}$ becomes

(4)

$$p(\mathbf{Y} \mid \mathbf{X}, \sigma^2) \propto (\sigma^2)^{\frac{N(1-N)}{4}} \exp\left(-\sum_{n>n'} r_{nn'}\right),$$

$$r_{nn'} = \frac{(y_{nn'} - \delta_{nn'})^2}{2\sigma^2} + \log \Phi\left(\frac{\delta_{nn'}}{\sigma}\right),$$

for $\Phi(\cdot)$ the cumulative distribution function of a standard normal variate. Unlike classical MDS, BMDS uses the language of probability to describe the low-dimensional representations, thus: (1) exchanging the $O(N^3)$ computational complexity of classical MDS for the $O(N^2)$ complexity of evaluating the BMDS likelihood, (2) allowing for uncertainty quantification and (3) avoiding conceptual difficulties arising from the mixed-methodological application of probability models to the results of classical MDS. Indeed, in the BMDS framework, modeling the latent variables $\mathbf{x}$ is as straightforward as specifying the prior distribution within a hierarchical model. Examples of such an approach are the use of a mixture of D-dimensional normals in Oh and Raftery (2007) and Gaussian processes in Holbrook et al. (2021b), but there is no reason a priori to restrict the class of available priors to be Gaussian.

2.2.3.2. *Choosing number of latent dimensions.* We would like to determine the optimal latent dimensionality D for the spatiotemporal Hawkes process; we use D to quantify the complexity of viral contagion through the global human air transportation network, and we let cross-validation (Geisser (1975)) dictate our choice of D. For BMDS our data are the distance matrix $\mathbf{Y}$ with off-diagonal elements $y_{nn'}$, the pairwise distances between objects n and n'. Within F-fold cross-validation, each fold f comprises held-out observations $\mathbf{Y}^f$ and the remaining observations $\mathbf{Y}^{-f}$. Let s index an MCMC state corresponding to a single draw from the posterior conditioned on $\mathbf{Y}^{-f}$, and denote the set of latent locations and model parameters $(\mathbf{X}, \Theta, \sigma^2)^{(s),-f}$ for $s = 1, \ldots, S$, the total number of MCMC states. We take the empirical *log pointwise predictive density* $(\widehat{lpd})$ as a measure of model fit and start with the log pointwise predictive densitiy *lpd* (Vehtari, Gelman and Gabry (2017)),

(5)

$$lpd = \sum_f \sum_{n<n'} \log p(y_{nn'}^f | \mathbf{Y}^{-f})$$

$$= \sum_f \sum_{n<n'} \log \int p(y_{nn'}^f | \mathbf{X}, \Theta, \sigma^2) p(\mathbf{X}, \Theta, \sigma^2 | \mathbf{Y}^{-f}) \, d(\mathbf{X}, \Theta, \sigma^2)$$

$$\approx \sum_f \sum_{n<n'} \log \frac{1}{S} \sum_{s=1}^{S} p(y_{nn'}^f | (\mathbf{X}, \Theta, \sigma^2)^{(s),-f}) = \widehat{lpd}.$$

Given competing models with different latent dimensionalities, we generally prefer the model with larger $\widehat{lpd}$.

3. Inference and implementation. We approach all three applications with an adaptive random scan Metropolis-within-Gibbs (Gilks, Best and Tan (1995)) scheme, building on Algorithm 1 of Holbrook et al. (2021a). For the first two applications the target posterior distribution takes the form

$$p(\Theta \mid \mathfrak{X}, \mathbf{t}) \propto p(\mathfrak{X}, \mathbf{t} \mid \Theta) p(\Theta) = \left(\int p(\mathfrak{X} \mid \mathbf{X}) \mathcal{L}(\mathbf{X}, \mathbf{t} | \Theta) \, d\mathbf{X}\right) p(\Theta),$$

where one obtains the uniform $p(\mathfrak{X} \mid \mathbf{X})$ by inverting the constraints of equations (2) and (3). We compute the high-dimensional integral over $\mathbf{X}$ using a Metropolis–Hastings kernel with

blockwise updates over sets of individual location variables. Satisfying the square constraints of equation (2) is straightforward using truncated normal proposals. Satisfying the circular constraints of equation (3) is less cut and dry. For an individual location variable $\mathbf{x}_n$ of the Markov chain's state s, we generate the $(s + 1)$th state according to the Metropolis kernel with proposal distribution

$$(6) \qquad \mathbf{x}_n^* \sim q(\mathbf{x}_n^* | \mathbf{x}_n) \propto 1, \qquad \|x_n^* - \mathfrak{x}_n\|_2 < r_n, \qquad \|x_n^* - \mathbf{x}_n\|_2 < r_n \epsilon, \quad \epsilon > 0,$$

where ϵ is an algorithmic parameter tuned with the help of diminishing adaptations (Roberts and Rosenthal (2007)). When $\|\mathbf{x}_n - \mathfrak{x}_n\|_2 + r_n \epsilon < r_n$, sampling is easy. Otherwise, we use a simple rejection sampler that satisfies the two circular constraints. In this case the Metropolis–Hastings accept-reject step requires calculating the area of the two circles' intersection (also referred to as the asymmetric lens), but one can easily obtain this quantity in closed-form and with negligible computational expense. Under our BMDS formulation for the third application, global viral contagion, the target posterior distribution is

$$p(\sigma^2, \Theta \mid \mathbf{Y}, \mathbf{t}) \propto p(\mathbf{Y} \mid \sigma^2) p(\sigma^2) p(\Theta) = \left(\int p(\mathbf{Y} \mid \mathbf{X}, \sigma^2) p(\mathbf{X}, \mathbf{t} \mid \Theta) \, d\mathbf{X} \right) p(\sigma^2) p(\Theta).$$

To compute the high-dimensional integral over values of $\mathbf{X}$, we use Hamiltonian Monte Carlo (HMC) (Neal (2011)) and again use simple Metropolis–Hastings proposals for the remaining parameters σ^2 and Θ. Hamiltonian Monte Carlo over $\mathbf{X}$ requires evaluation of both spatiotemporal Hawkes model and BMDS joint densities (equations (1) and (4)) and their gradients.

The Hawkes model likelihood (used in all three applications) and the BMDS likelihood and Hawkes likelihood gradients (used in the third application) all share the prohibitively burdensome computational complexity $O(N^2)$. We, therefore, use the OPENCL and C++ high-performance computing libraries MASSIVEMDS (Holbrook et al. (2021b)) https://github.com/suchard-group/MassiveMDS and HPHAWKES (Holbrook et al. (2021a)) https://github.com/suchard-group/hawkes (Holbrook, Ji and Suchard (2022)) to evaluate these functions and their gradients in parallel on either a graphics processing unit (GPU) or with a multicore central processing unit (CPU) with vectorization. In writing this paper, we have contributed GPU and CPU implementations of the Hawkes process log-likelihood gradient to the library HPHAWKES, and we detail the massively parallel Algorithms 1 and 2 and their resulting speedups in Appendix A. Finally, we access and embed the high performance implementations within the broader Metropolis-within-Gibbs scheme with the BEAST software package (Suchard et al. (2018)) using simple application programming interfaces.

4. Demonstrations. Besides the high-performance computing packages HPHAWKES and MASSIVEMDS we use for MCMC, we use the R programming language (R Core Team (2019)) and the R graphics packages GGPLOT2 (Wickham (2016)) and GGMAP (Kahle and Wickham (2013)) to produce and summarize results. The R package CODA (Plummer et al. (2006)) provides our effective sample size (ESS) measures, and we base reported 95% credible intervals on empirical posterior 0.025 and 0.975 quantiles. Finally, we make all analysis source files publicly available at https://github.com/andrewjholbrook/unknown_locs and https://github.com/andrewjholbrook/FluHawkes (Holbrook, Ji and Suchard (2022)).

4.1. *Washington D.C. gunfire in* 2018. We first apply our methodology to the analysis of 3982 gunshots occurring in Washington D.C. between January 2 and December 31, 2018. The Government of the District of Columbia makes gunfire data from the years 2014–2019 freely available at https://opendata.dc.gov. The data arise from ShotSpotter AGLS technology (Carr and Doleac (2016)) that has become increasingly accurate since first implemented in Washington D.C. in 2006. Loeffler and Flaxman (2018) use a spatiotemporal Hawkes process to
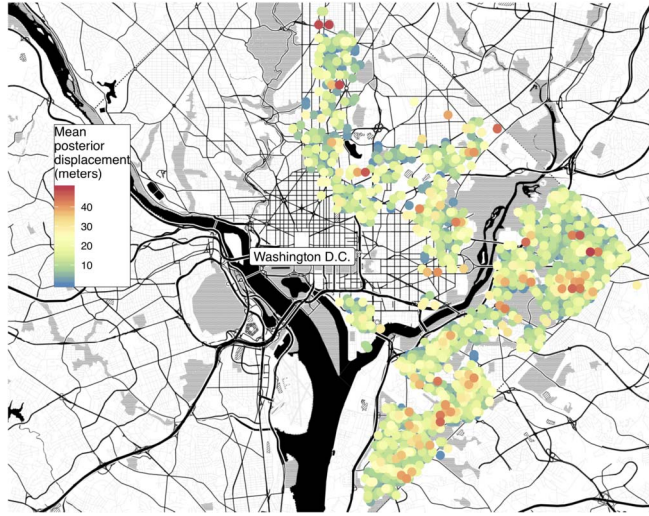
FIG. 1.   *We color the observed locations $\mathfrak{x}_n$ of 3982 gunshots occurring in the year 2018 by the magnitude of the mean posterior displacement of each event's inferred locations $\mathbf{x}_n$: $\|\sum_{s=1}^{S}(\mathbf{x}_n^{(s)} - \mathfrak{x}_n)/S\|_2$, where S is the total number of MCMC states. For each event this measure communicates the amount of posterior displacement in a single, general direction away from the observed location $\mathfrak{x}_n$.*

analyze a similar sample from the years 2010 through 2012, and Holbrook et al. (2021a) apply a related model to data from the years 2006 through 2019. The D.C. Government rounds all latitudinal and longitudinal coordinate data to three decimal places, a coarsening that corresponds to 100 meters precision. Because we wish to isolate this as the only source of spatial uncertainty and because of the gradual improvement of ShotSpotter technology, we choose to focus on a higher quality sample from 2018. We also remove all observations listed as potential firecrackers as well as all data from the first day of the year, again avoiding possible corruption due to misattribution to firecrackers. The result is 3982 events with locations plotted in Figure 1. The minimum, mean and maximum pairwise distances between raw locations are 0.0, 5.4 and 16.4 km. We compare these numbers to the data's spatial precision of 0.1 km.

To infer all six Hawkes model parameters and all 3982 location variables, we generate 30 million Markov chain states (requiring 48 hours on our Nvidia Quadro GP100 GPU) that provide minimum and mean ESS of 131 and 424 for latent locations and 401 and 571 for model parameters. First, we would like to know whether there is a spatial pattern to the posterior displacements away from raw locations for individual events, where we use the formula $\|\sum_{s=1}^{S}(\mathbf{x}_n^{(s)} - \mathfrak{x}_n)/S\|_2$ to quantify this displacement. Figure 1 shows that high posterior displacements occur both on the peripheries and at the centers of high activity areas. This fact suggests that more complex patterns underlie posterior displacements and that temporal relationships may play a role. Figure 2 presents three examples of event groups with larger posterior displacements. For the first group, posterior locations draw away from figure center despite relatively small temporal differences between events ranging from 11 to 120 hours. Here, temporal proximity appears to be overcome by the gunfire vacuum of a commercial shopping center at plot center. The second pair of events appear to have larger posterior displacements for a very different reason. Here, the two events are spatially isolated from other gunshots, so their posterior locations attract to each other, despite a larger temporal disparity of 55 days. Finally, the third cluster tells a much simpler story. The four events occupy the center of a large, high-activity area and gravitate toward the center of mass.
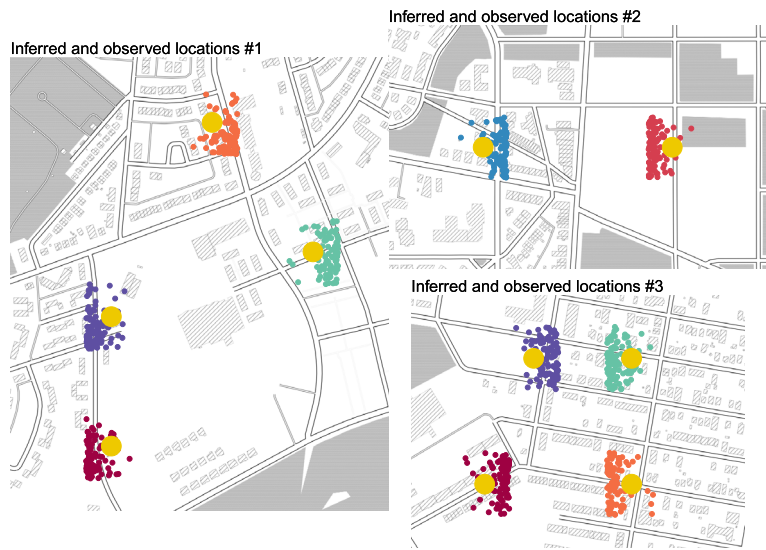
FIG. 2. *Visualizing the relationship between observed locations $\mathfrak{x}_n$ (yellow) and posterior sample locations $\mathbf{x}_n^{(s)}$ for 10 gunshots in the District of Columbia. Inferred locations may deviate from observed for multiple reasons. In the first plot, differences in date and time range from 11 to 120 hours; in the second, the two events differ by 55 days. On the one hand, the gunshots in the first plot occur in a gunfire dense area but separated by a low-activity shopping center. On the other hand, the gunshots in the second plot are spatially isolated from other events.*

Figure 3 and Table 1 present inferential results for the Hawkes model parameters, where the normalized self-excitatory weight $\theta/(\theta + \mu_0)$ communicates the proportion of all events arising from self-excitation rather than the background process. Here, we see generally con-
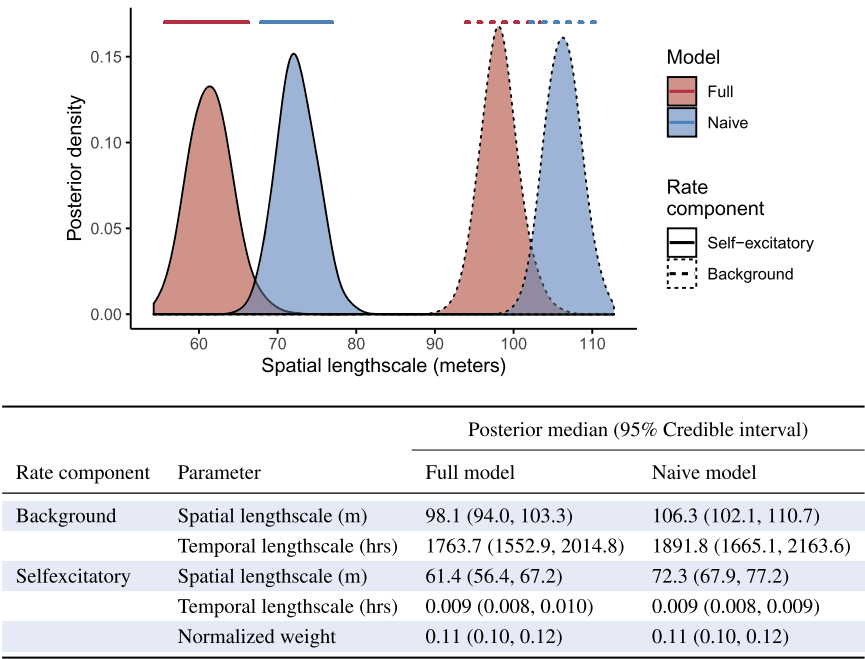

Rate component	Parameter	Posterior median (95% Credible interval)	
		Full model	Naive model
Background	Spatial lengthscale (m)	98.1 (94.0, 103.3)	106.3 (102.1, 110.7)
	Temporal lengthscale (hrs)	1763.7 (1552.9, 2014.8)	1891.8 (1665.1, 2163.6)
Selfexcitatory	Spatial lengthscale (m)	61.4 (56.4, 67.2)	72.3 (67.9, 77.2)
	Temporal lengthscale (hrs)	0.009 (0.008, 0.010)	0.009 (0.008, 0.009)
	Normalized weight	0.11 (0.10, 0.12)	0.11 (0.10, 0.12)

FIG. 3 AND TABLE 1. *Posterior densities and 95% credible intervals of Hawkes model background and self--excitatory lengthscales for "full" (locations inferred) and "naive" models of gun violence in the District of Columbia. Credible intervals for self-excitatory lengthscales do not overlap, while those of the background component display marginal overlap.*

sistent results between the model that incorporates spatial uncertainty and that which does not. Perhaps unsurprisingly, the major discrepancies in posterior inference between these models are for the two spatial lengthscales. The self-excitatory spatial lengthscales for the full and naive models are 61.4 m (56.4, 67.2) and 72.3 m (67.9, 77.2), and the background spatial lengthscales are 98.1 m (94.0, 103.3) and 106.3 m (102.1, 110.7). Smaller spatial lengthscales make sense, insofar as inferred locations may attract to each other, but one might find these statistically significant and marginally statistically significant differences surprising, given the relatively small spatial uncertainty (0.1 km) precision relative to a mean pairwise distance in the data of roughly five km. Despite statistically significant differences, we judge the practical differences to be small, and this is good news for practitioners who want to avoid integrating over latent locations. Nonetheless, this good news only seems to apply when spatial uncertainty is: (a) relatively small and (b) uniform across observations a priori. In Section C we use simulated data to test this hypothesis and find that the naive model indeed fails under moderate coarsening.

4.2. *Alaskan wildfire ignitions*: 2015–2019.   The Alaska Interagency Coordination Center makes various wildfire data resources freely accessible at https://fire.ak.blm.gov/predsvcs/maps.php. In particular, we apply our methodology to data consisting of wildfire geographic coordinates, date and time of fire discovery and size in acres at time of discovery. Figure 4 displays the raw locations for all 2925 wildfires, plotting each with size proportional to its radius on the log scale. The minimum, mean and maximum pairwise distances between raw locations are 0.0, 500.3 and 2373.8 km. The empirical distribution of wildfire discovery site radii roughly resembles a power law, with minimum, median, mean and maximum of 0.01, 0.01, 0.08 and 4.42 km. In this way the spatial uncertainty relative to the scale of locational spread is much smaller for this application than for the Washington D.C. gunfire example. This difference allows us to partially isolate the effects of differential uncertainty across the observed sample on posterior inference.

Using the prior of equation (3) on locations, we model the spread of wildfire ignitions with our spatiotemporal Hawkes model and, simultaneously, infer ignition locations. Again, we generate 30 million Markov chain states that provide a minimum and mean ESS of 120 and 398 for inferred locations and 510 and 586 for the six Hawkes' model parameters. Despite the smaller sample size, compared to the gun violence data, MCMC for this example
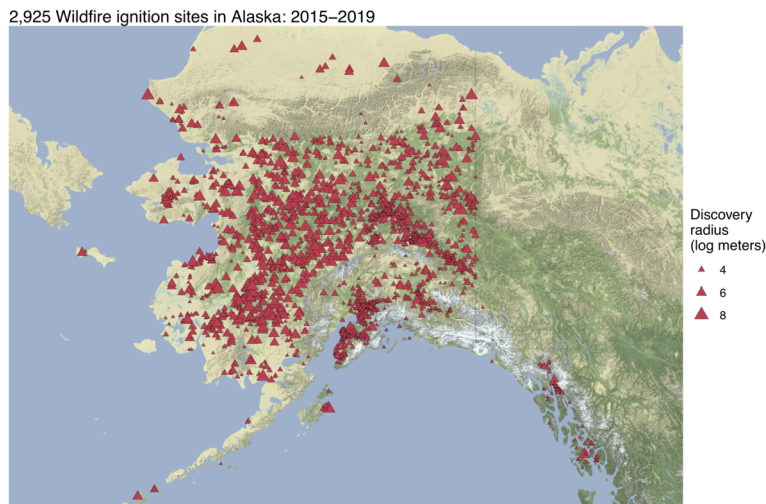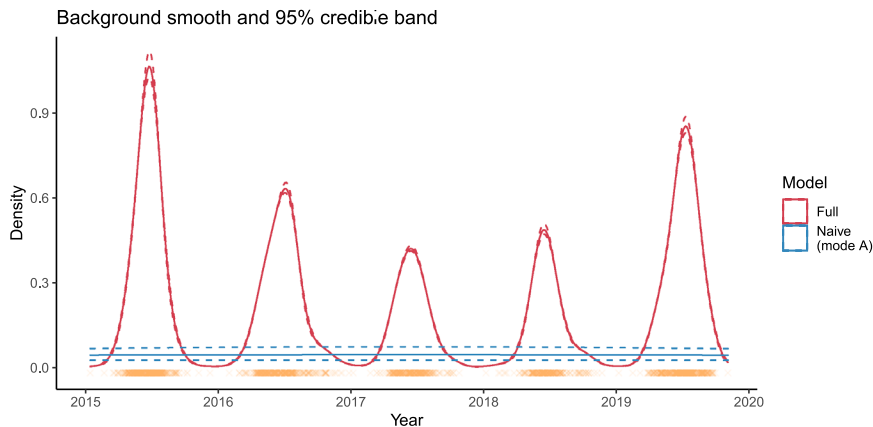

FIG. 4.   *Data include time of fire discovery as well as size* (*in acres*) *and location of each wildfire at time of discovery.*

Background smooth and 95% credible band

		Posterior median (95% Credible interval)		
Rate component	Parameter	Full model	Naive model *A*	Naive model *B*
Background	Spatial lengthscale (km)	34.8 (32.9, 37.6)	**23.5 (22.3, 24.6)**	63.0 (58.7, 68.7)
	Temporal lengthscale (days)	25.9 (23.8, 27.9)	3244.0 (1929.7, 5803.5)	**10.2 (9.4, 11.1)**
Self-excitatory	Spatial lengthscale (km)	11.1 (10.1, 12.0)	**23.3 (22.2, 24.4)**	6.5 (5.9, 7.2)
	Temporal lengthscale (days)	1.1 (0.9, 1.4)	2.2 (1.9, 2.5)	**10.0 (9.2, 10.8)**
	Normalized weight	0.34 (0.31, 0.37)	0.27 (0.17, 0.36)	**0.44 (0.41, 0.47)**

FIG. 5 AND TABLE 2. *Spatiotemporal Hawkes model posterior inference for* 2015–2019 *wildfire ignitions in Alaska with "Full" model* (*locations inferred*) *and of "Naive" model* (*locations not inferred*) *modes* A *and* B: *inferring locations may help avoid modes at near-equal lengthscales* (**bold**). *Mode* A *provides an unreasonably large background temporal lengthscale that fails to incorporate seasonal trends, and the normalized self-excitatory weight of mode* B *may be considered too large to be realistic.*

requires 50% more time (totaling roughly 72 hours), due to the rejection sampler we use to generate from (6). Figure 5 and Table 2 present inferential results from the model with locations inferred as well as the naive model. As discussed in Section 2.1, simultaneous inference for self-excitatory and background lengthscales can sometimes lead to multimodality. This issue is so problematic that Mohler (2014) fixes the problem by setting spatial lengthscales to be equal, that is, by removing flexibility from the model. Here, we find that inferring locations may actually help mitigate such multimodality: while the naive model gets stuck in two different modes (*A* and *B*), the full model does not. As shown in Figure 2, multimodality can lead to poor model fits and unreasonable results. With a posterior mean of 3244.0 days and 95% credible interval of (1929.7, 5803.5), inference for the background temporal lengthscale of the naive model's mode A suggests no seasonal wildfire trend whatsoever *in Alaska*. In contrast, inference for the full model fully captures seasonal trends with a posterior mean of 25.9 days (23.8, 27.9) for the temporal lengthscale of the background rate.

4.3. *Global influenza cases*: 2000–2012. We analyze the worldwide influenza spread using 4733 cases recorded between the years of 2000 and 2012. Of the 4733 cases, 1161 are H1N1 subtype, 1341 H3N2 subtype, 1195 Victoria lineage (VIC) and 1036 Yamagata lineage (YAM). H1N1 and H3N2 are influenza type A and, generally, more prevalent than Victoria and Yamagata which are both type B and contribute to significantly less infections annually. Between H1N1 and H3N2, H1N1 is responsible for two major pandemics, the Spanish flu of 1918–1919 and the swine flu of 2009, while H3N2 has contributed to one, the Hong Kong flu of 1968–1969. Bedford et al. (2015) relate the greater epidemiological success of type A influenza to higher rates of antigenic drift, leading to different age groups becoming infected at different rates. In particular, adults are more susceptible to H1N1 and, being more likely frequent fliers than children, help the subtype travel more quickly through global air

travel networks (Bedford et al. (2014)) than competing strains. Combining BMDS with a phylogenetic diffusion model that conditions on each subtype and lineage's evolutionary history, Holbrook et al. (2021b) confirms that the rate of diffusion through the global air traffic network is significantly greater for H1N1 than for H3N2, YAM and VIC. Here, we are interested in whether inference based on our BMDS-Hawkes model renders similar results and how greater efficiency of H1N1 might express itself for individual Hawkes model parameters, for example, shorter lengthscales or greater rates of self-excitation.

The data we consider here are a subset of the 5392 analyzed in Holbrook et al. (2021b), where we have removed those cases that lack a precise date. Moreover, we use the exact same matrix of pairwise air traffic distances between countries $\mathbf{Y}$. Brockmann and Helbing (2013) create these distances from a network for which nodes are 4096 airports worldwide and edges (when they exist) between nodes inversely relate to the total number of passengers traveling between the two airports each year. Motivated by the multiprecision nature of the influenza case data—spatial labels are approximately one-third cities, one-third provinces and one-third countries—Holbrook et al. (2021b) then collapse across airports to obtain effective distances between countries on this global transportation network. We use the Hawkes model to infill the relationships between latent locations coming from the same country. Through the spatiotemporal Hawkes likelihood that interfaces with temporal data $\mathbf{t}$, the BMDS-Hawkes model further informs latent positions $\mathbf{X}$. Thus, we efficiently and simultaneously: (a) adapt our data to the realities of global air transport and (b) exploit all data despite its multiprecision nature.

Before producing the full analysis, we use the $\widehat{lpd}$ of equation (5) as measure of model fit and perform five-fold cross-validation to select the latent dimensionality of our BMDS-Hawkes model. Dimensions 2 through 8 provide $\widehat{lpd}$s of -13.2, -8.1, -6.2, -5.5, -5.2, -5.1 and -5.0 million. Noting a lack of relative improvement for further dimensions, we judge the six-dimensional model to be sufficiently complex. Next, we use HMC (Section 3) to generate 80 million Markov chain states. Employing Algorithm 2 for massively parallel Hawkes log-likelihood gradient calculations, this requires roughly 10 days on our Nvidia Titan V GPU.

The top two plots of Figure 6 show the naive global distribution of the influenza case data colored by the posterior mean probability that each event arises from self-excitation, that is,

$$\frac{1}{S} \sum_{s=1}^{S} \xi^{(s)}\big(\mathbf{x}_n^{(s)}, t_n\big) / \big(\xi^{(s)}\big(\mathbf{x}_n^{(s)}, t_n\big) + \mu^{(s)}\big(\mathbf{x}_n^{(s)}, t_n\big)\big)$$

for $\mathbf{x}_n^{(s)}$ a location in the six-dimensional latent air traffic network space and $\xi^{(s)}(\cdot, \cdot)$ and $\mu^{(s)}(\cdot, \cdot)$ the self-excitatory and background rates parameterized by parameters $\Theta^{(s)}$. Posterior concentration around this posterior mean is extremely tight for all observations, so the models strongly believe that blue cases arise from the background process while red arise from self-excitation. For the naive model there are as many blue cases as there are locations: the model regards the earliest case in every location as coming from the background process and every case thereafter as arising from this earliest case. Reflecting this fact, the posterior distributions of the naive model's self-excitatory spatial lengthscale concentrate below three km (Figure 9). Thus, naive model inference communicates no information beyond what one would garner from a simple exploratory data analysis. On the other hand, the six-dimensional BMDS-Hawkes model reveals significantly less background activity and a model more in tune with the self-excitatory reality of viral spread. Still, we can interpret background activity as arising from relatively large and fast traversals of the global air traffic network. The third plot is similar but shows the arrangement of latent locations for
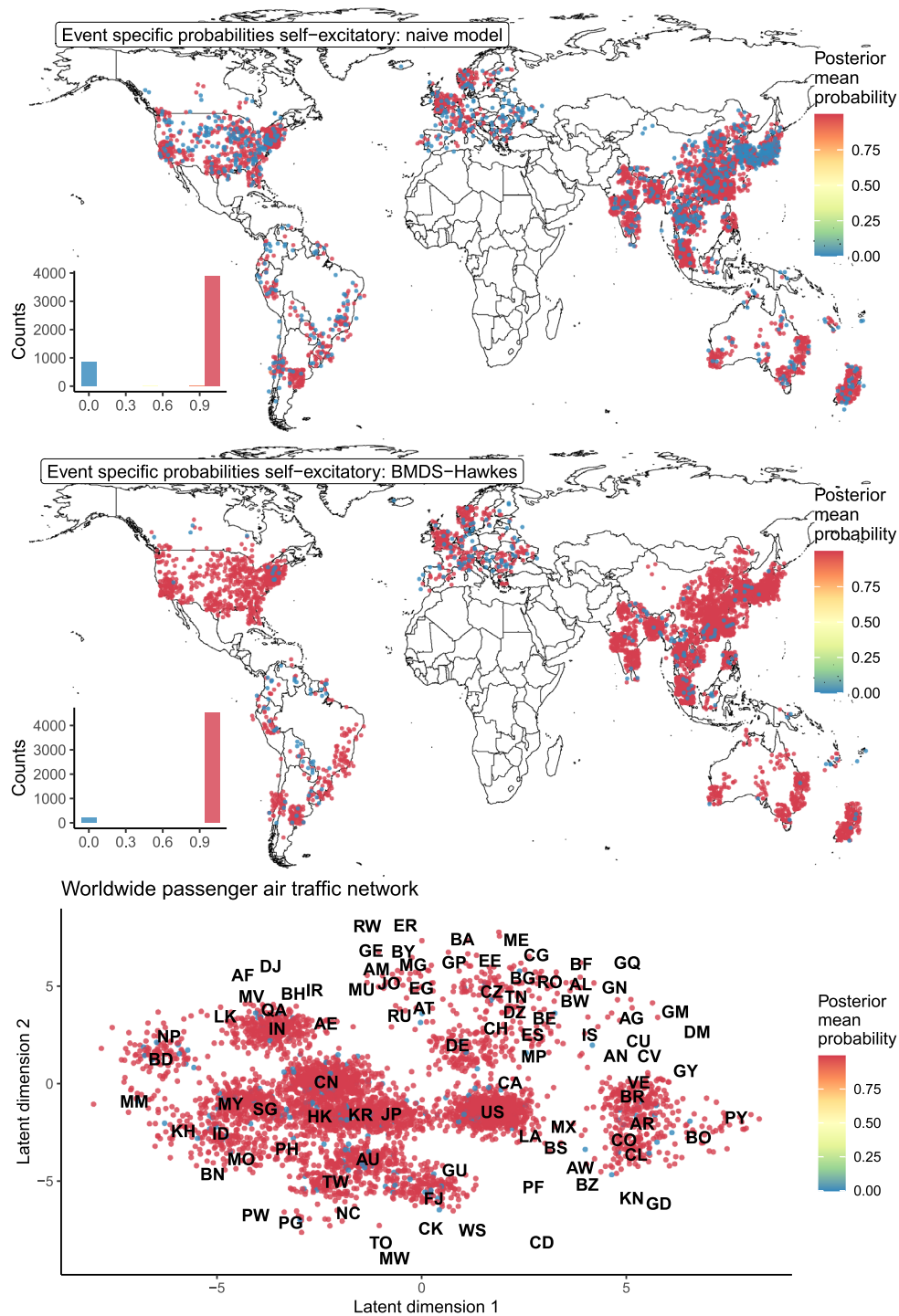
FIG. 6. *Geographic and network positions of* 4733 *influenza cases, each colored by posterior mean probability the case originates from another "parent" case. Top figure shows results from spatially naive model; bottom two figures from the six- and two-dimensional combined Bayesian multidimensional scaling and spatiotemporal Hawkes model (BMDS-Hawkes). In the latter, proximity of Laos (LA) to the United States (U.S.) portends poor cross-validation results.*
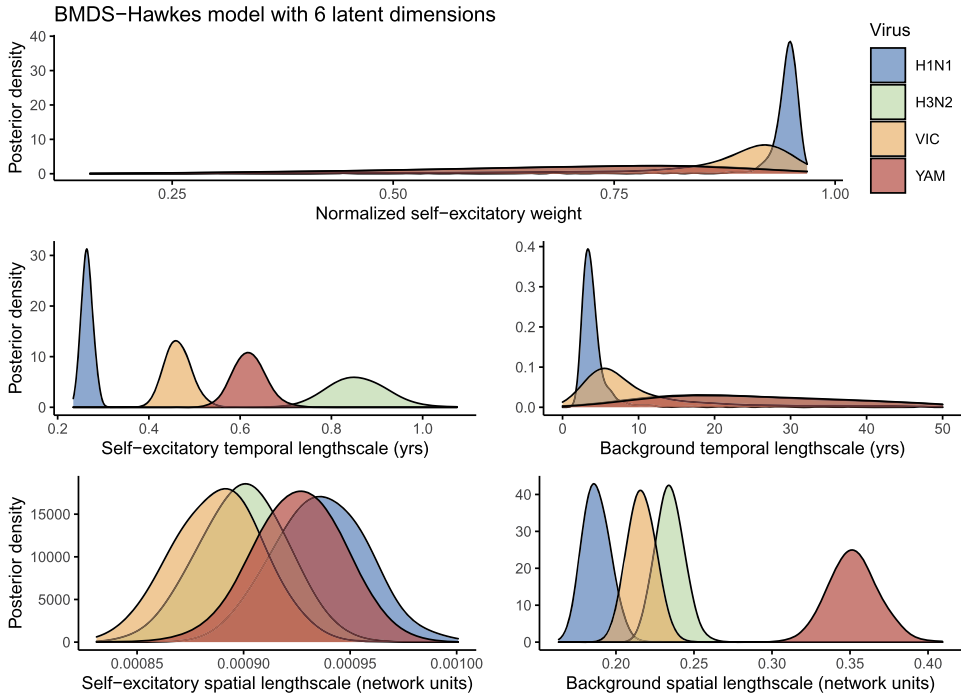
FIG. 7.   *Posterior distributions for Hawkes model parameters based on* 1161 *H1N1 subtype,* 1341 *H3N2 subtype,* 1195 *Victoria lineage and* 1036 *Yamagata lineage influenza cases. In general, H1N1 is more prevalent and infects adults in greater numbers than it does children.*

a single posterior sample for the two-dimensional BMDS-Hawkes model. In general, the world economic powers gravitate toward the middle while smaller countries tend toward the outside. These arrangements are largely as one might hope, but there are hints that the two-dimensional model is insufficient. For example, Laos (LA) is much closer to the United States (U.S.) than it is the rest of Asia, in general, and China in particular. This suggests that a higher dimensionality might be more appropriate, a fact which cross-validation results bear out.

Figure 7 displays posterior densities for the Hawkes model parameters for each influenza subtype and lineage. We immediately notice that the H1N1 model attributes more influenza activity to self-excitation than do the other models. The posterior means and 95% credible intervals of the normalized self-excitatory weight for H1N1, H3N2, VIC and YAM are 0.95 (0.82, 0.96), 0.72 (0.31, 0.91), 0.91 (0.47, 0.95), 0.72 (0.32, 0.90). We note that overlapping credible intervals indicate uncertainty as to this specific ordering. On the other hand, there is very little posterior uncertainty with regards to the ordering of the self-excitatory temporal lengthscales. In order, the same posterior measures for H1N1, VIC, YAM and H3N2 are 0.26 years (0.24, 0.29), 0.46 years (0.42, 0.52), 0.62 years (0.56, 0.69) and 0.86 years (0.74, 0.98). This result suggests that the self-excitatory temporal lengthscale of our Hawkes model and the rate of diffusion of the phylogenetic diffusion model of Holbrook et al. (2021b) are similar, insofar as they both capture the greater efficiency with which H1N1 uses passenger air traffic networks to quickly travel the globe. Moreover, H1N1's posterior mean self-excitatory temporal lengthscale is small enough to fully capture seasonal trends. Finally, the large posterior mean for the same parameter of the H3N2 model and its inability to capture seasonal trends suggest that a network build solely of air transportation may be insufficient for modeling the spread of H3N2.

5. Discussion.   The spatiotemporal Hawkes process is a powerful tool for modeling the complex spatial dynamics of many real-world phenomena. Although its use is growing increasingly widespread, previous applications of the model have glossed over the presence of spatial coarsening and uncertainty in the problems analyzed and the role played by the same uncertainty in biasing model inference. By considering three diverse applications, we have demonstrated: (a) the prevalence of spatial uncertainty in processes commonly regarded as self-exciting, (b) the practicality of integrating over such uncertainty in the manners proposed and (c) the statistically and practically significant differences between full and naive approaches. Furthermore, we have shown that our strategies may also be useful in mitigating multimodality, a problem that makes model fitting, diagnostics and interpretation more cumbersome. Indeed, we have demonstrated that one can reap these benefits without having to sacrifice model flexibility.

That said, there are a few meaningful changes to our proposed approach that may lead to improved inference. First, this approach seeks to account for spatial coarsening but fails to adapt to temporal coarsening. For this reason we removed hundreds of viral cases that lack full temporal precision from our analysis of global influenza. One could make retainment of these observations possible by directly modeling the coarsened data in a similar way to how we have modeled locations in our first example. Unfortunately, implementing the MCMC to integrate over latent times would be especially difficult, due to the self-excitatory rate function's reliance only on past events. The required combinatorial integration would necessitate careful bookkeeping, and it is difficult to predict the empirical mixing of such a Markov chain. Second, our priors over locations in both the gunfire and wildfire examples could take further advantage of geographical information. In the former example, one might use the fact that gun violence is more likely to occur indoors. In the latter, one could use rivers and lakes to further reduce the support of wildfire ignitions. We are not sure how one might generate and respect such priors at nontrivial scale. Third, we use Bayesian multidimensional scaling to model viral spread through Euclidean space instead of a complex network. It is plausible that other continuous spaces might be more appropriate. For example, a low-dimensional torus with its varying curvature might be better able to capture pairwise relationships more efficiently than Euclidean space of the same dimension.

There are usually many valid ways to approach a challenge, and other approaches may prove more tractable over time. In modeling global viral spread, for instance, it might also prove useful to equip the self-excitatory rate with a dramatically heavy-tailed triggering function, such as a continuous scale-mixture of truncated Gaussian functions (Polson, Scott and Windle (2014), Nishimura and Suchard (2018)). This would allow the same triggering function to account for both domestic and international transmission. That said, this would ameliorate the issue of adapting to transportation dynamics but would not account for spatially coarsened and multiprecision data.

To conclude, we hope that this paper's results will prove instructive to scientists interested in spatiotemporal modeling, regardless of the model or statistical paradigm they choose to employ. At the very least, our results suggest that sensitivity testing by perturbing spatial data is a good idea. In an exact analogy to the way we have selected priors on spatial locations, one might generate perturbations in a way that reflects the a reasonable model of the spatial uncertainty at hand. Finally, our proposed approach appears to be easily translatable into the frequentist paradigm in the form of marginal maximum likelihood (Geyer (1991)). MCMC similar to that performed here would provide for integration over locations, although it is not immediately clear how such integration would influence the already complex consistency arguments involved in maximum likelihood estimation for spatiotemporal model parameters (Schoenberg (2016)).

Algorithm 1 Parallel computation of Hawkes process log-likelihood gradient: *uses multiple central processing unit (CPU) cores along with loop vectorization to compute log-likelihood gradient. For double-precision floating point, the algorithm uses either SSE or AVX vectorization to make $J = 2$ or 4 long jumps and cut loop iterations by one-half or three-fourths, respectively. Here, B is the number of CPU threads available. Symbols ℓ, λ and Λ appear in equation* (1)

1: Compute rates $\lambda_1, \ldots, \lambda_N$:
   a: **parfor** $b \in \{1, \ldots, B\}$ **do**
   b: **if** $b \neq B$ **then**
   c: $Upper \leftarrow b\lfloor N/B \rfloor$
   d: **else**
   e: $Upper \leftarrow \lceil N/B \rceil$
   f: **end if**
   g: **for** $n \in \{(b-1)\lfloor N/B \rfloor + 1, \ldots, Upper\}$ **do**
   h: copy $\mathbf{x}_n, t_n$ to cache
   i: $\lambda_n \leftarrow \mathbf{0}$ $\triangleright$ vector of length J
   j: $n' \leftarrow 1$
   k: **while** $n' < N$ **do**
   l: $J \leftarrow \min(J, N - n')$
   m: copy $\mathbf{x}_{n':(n'+J)}, t_{n':(n'+J)}$ to cache
   n: $\boldsymbol{\Delta}_{nn'} : \boldsymbol{\Delta}_{nn':(n'+J-1)} \leftarrow (\mathbf{x}_n - \mathbf{x}_{n'}) : (\mathbf{x}_n - \mathbf{x}_{n'+J-1})$ $\triangleright$ vectorized subtraction
   o: calculate $\delta_{nn'} : \delta_{n(n'+J-1)}$ $\triangleright$ vectorized multiplication
   p: calculate $\lambda_{nn'} : \lambda_{n(n'+J-1)}$ $\triangleright$ vectorized evaluation
   q: $\lambda_n \leftarrow \lambda_n + \lambda_{nn'} : \lambda_{n(n'+J-1)}$ $\triangleright$ vectorized addition
   r: $n' \leftarrow n' + J$
   s: **end while**
   t: **end for**
   u: **end parfor**

2: Compute N D-dimensional gradients $\frac{\partial \ell}{\partial \mathbf{x}_n}$:
   a: **parfor** $b \in \{1, \ldots, B\}$ **do**
   b: **if** $b \neq B$ **then**
   c: $Upper \leftarrow b\lfloor N/B \rfloor$
   d: **else**
   e: $Upper \leftarrow \lceil N/B \rceil$
   f: **end if**
   g: **for** $n \in \{(b-1)\lfloor N/B \rfloor + 1, \ldots, Upper\}$ **do**
   h: copy $\mathbf{x}_n, t_n$ to cache
   i: $\frac{\partial \ell}{\partial \mathbf{x}_n} \leftarrow \mathbf{0}$ $\triangleright$ vector of length J
   j: $n' \leftarrow 1$
   k: **while** $n' < N$ **do**
   l: $J \leftarrow \min(J, N - n')$
   m: copy $\mathbf{x}_{n':(n'+J)}, t_{n':(n'+J)}$ to cache
   n: $\boldsymbol{\Delta}_{nn'} : \boldsymbol{\Delta}_{nn':(n'+J-1)} \leftarrow (\mathbf{x}_n - \mathbf{x}_{n'}) : (\mathbf{x}_n - \mathbf{x}_{n'+J-1})$ $\triangleright$ vectorized subtraction
   o: calculate $\delta_{nn'} : \delta_{n(n'+J-1)}$ $\triangleright$ vectorized multiplication
   p: calculate $\mu_{nn'} : \mu_{n(n'+J-1)}$ $\triangleright$ vectorized evaluation
   q: calculate $\xi_{nn'} : \xi_{n(n'+J-1)}$ $\triangleright$ vectorized evaluation
   r: calculate $\xi_{n'n} : \xi_{(n'+J-1)n}$ $\triangleright$ vectorized evaluation
   s: **for** $j \in n', \ldots, n' + J - 1$ **do**
   t: $\frac{\partial \ell}{\partial \mathbf{x}_n} \leftarrow \frac{\partial \ell}{\partial \mathbf{x}_n} + (\frac{\mu_{nj}}{\lambda_n} + \frac{\mu_{jn}}{\lambda_j})\frac{\boldsymbol{\Delta}_{jn}}{\tau_x^2} + (\frac{\xi_{nj}}{\lambda_n} + \frac{\xi_{jn}}{\lambda_j})\frac{\boldsymbol{\Delta}_{jn}}{\sigma_x^2}$
   u: **end for**
   v: $n' \leftarrow n' + J$
   w: **end while**
   x: **end for**
   y: **end parfor**

Algorithm 2 Parallel computation of Hawkes process log-likelihood gradient: *calculates the log-likelihood gradient with multiple levels of parallelization on graphics processing unit (GPU). In practice, we specify B = 128 to be the the size of the GPU work groups. Symbols ℓ, λ and Λ appear in equation* (1)

1: Compute rates $\lambda_1, \ldots, \lambda_N$:

  a: **parfor** $n \in \{1, \ldots, N\}$ **do**

  b: copy $\mathbf{x}_n, t_n$ to local $\triangleright$ *B* threads

  c: **parfor** $N' \in \{1, \ldots, \lfloor N/B \rfloor\}$ **do**

  d: $n' \leftarrow N'$

  e: $\lambda_{nN'} \leftarrow 0$

  f: **while** $n' < N$ **do**

  g: copy $\mathbf{x}_{n'}, t_{n'}$ to local $\triangleright$ *B* threads

  h: $\Delta_{nn'} \leftarrow \mathbf{x}_n - \mathbf{x}_{n'}$ $\triangleright$ vectorized subtraction

  i: calculate $\delta_{nn'} = \sqrt{\sum \Delta_{nn'} \circ \Delta_{nn'}}$ $\triangleright$ vectorized multiplication

  j: $\lambda_{nN'} \leftarrow \lambda_{nN'} + \lambda_{nn'}$ $\triangleright$ $\lambda_{nn'}$ a function of $\delta_{nn'}, t_n$ and $t_{n'}$

  k: $n' \leftarrow n' + B$

  l: **end while**

  m: **end parfor**

  n: $\lambda_n \leftarrow \sum_{N'} \lambda_{nN'}$ $\triangleright$ binary tree reduction on chip

  o: **end parfor**

2: Compute N D-dimensional gradients $\frac{\partial \ell}{\partial \mathbf{x}_n}$:

  a: **parfor** $n \in \{1, \ldots, N\}$ **do**

  b: copy $\mathbf{x}_n, t_n$ to local $\triangleright$ *B* threads

  c: **parfor** $N' \in \{1, \ldots, \lfloor N/B \rfloor\}$ **do**

  d: $n' \leftarrow N'$

  e: $(\frac{\partial \ell}{\partial \mathbf{x}_n})_{N'} \leftarrow \mathbf{0}$

  f: **while** $n' < N$ **do**

  g: copy $\mathbf{x}_{n'}, t_{n'}$ to local $\triangleright$ *B* threads

  h: $\Delta_{nn'} \leftarrow \mathbf{x}_n - \mathbf{x}_{n'}$ $\triangleright$ vectorized subtraction

  i: calculate $\delta_{nn'} = \sqrt{\sum \Delta_{nn'} \circ \Delta_{nn'}}$ $\triangleright$ vectorized multiplication

  j: $(\frac{\partial \ell}{\partial \mathbf{x}_n})_{N'} \leftarrow (\frac{\partial \ell}{\partial \mathbf{x}_n})_{N'} + (\frac{\mu_{nn'}}{\lambda_n} + \frac{\mu_{n'n}}{\lambda_j})\frac{\Delta n'_n}{\tau_x^2} + (\frac{\xi_{nn'}}{\lambda_n} + \frac{\xi_{n'n}}{\lambda_{n'}})\frac{\Delta n'_n}{\sigma_x^2}$

  k: $n' \leftarrow n' + B$

  l: **end while**

  m: **end parfor**

  n: $\frac{\partial \ell}{\partial \mathbf{x}_n} \leftarrow \sum_{N'} (\frac{\partial \ell}{\partial \mathbf{x}_n})_{N'}$ $\triangleright$ binary tree reduction on chip
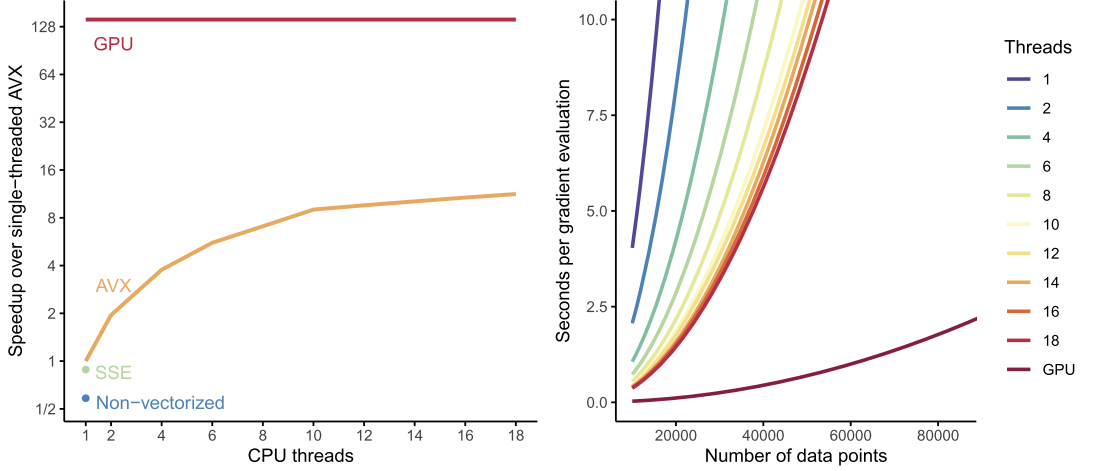
  o: **end parfor**

FIG. 8.   *Computing the spatiotemporal Hawkes process log-likelihood gradient with respect to locations using central and graphics processing units* (*CPU and GPU*). [*Left*] *Relative speedups over single-core advanced vector extensions* (*AVX*) *vectorization for single-core nonvectorized and streaming SIMD extensions* (*SSE*), *multicore AVX and many-core GPU implementations for* 75,000 *simulated observations.* [*Right*] *Absolute time to perform gradient evaluation for single- and multicore AVX processing and GPU as a function of the number of simulated data points.*

APPENDIX A: PARALLELIZATION

In writing this paper, we have developed massively parallel implementations of the gradient of the log-likelihood with respect to spatial locations for our spatiotemporal Hawkes model. We have also added this code to HPHAWKES, a C++ library and R package for high-performance computing for Bayesian inference under the spatiotemporal Hawkes process. We have made this open-source software freely available at https://github.com/suchard-group/hawkes (Holbrook, Ji and Suchard (2022)).

Letting $\lambda_{nn'}$ be as defined in equation (1), define $\lambda_n := \sum_{n'}^{N} \lambda_{nn'}$. Then, the gradient of the Hawkes likelihood with respect to locations is

$$\frac{\partial \ell}{\partial \mathbf{x}_n} = \sum_{n'=1}^{N} \left(\frac{\mu_{nn'}}{\lambda_n} + \frac{\mu_{n'n}}{\lambda_{n'}} \right) \left(\frac{\mathbf{x}_{n'} - \mathbf{x}_n}{\tau_x^2} \right) + \left(\frac{\xi_{nn'}}{\lambda_n} + \frac{\xi_{n'n}}{\lambda_{n'}} \right) \left(\frac{\mathbf{x}_{n'} - \mathbf{x}_n}{\sigma_x^2} \right)$$

$$:= \sum_{n'=1}^{N} \left(\frac{\partial \ell}{\partial \mathbf{x}_n} \right)_{n'}.$$

Algorithms 1 and 2 describe our massively parallel implementation of this gradient on a CPU and GPU, respectively, and Figure 8 illustrates some of the speedups achieved by these implementations.

APPENDIX B: ADDITIONAL RESULTS

This section contains posterior inferential results for the naive model from Section 4.3 in the form of Figure 9. In this figure we are particularly interested in the bottom-left plot of posterior densities for self-excitatory spatial lengthscales. No matter the influenza strain, these posterior distributions concentrate below three km. Such small lengthscales amount to zero self-excitation between even closely neighboring cities, let alone any transmission on the provincial, national or global scales. This way the naive model attributes one influenza case (the earliest) to the background process for each distinct location and all successive cases
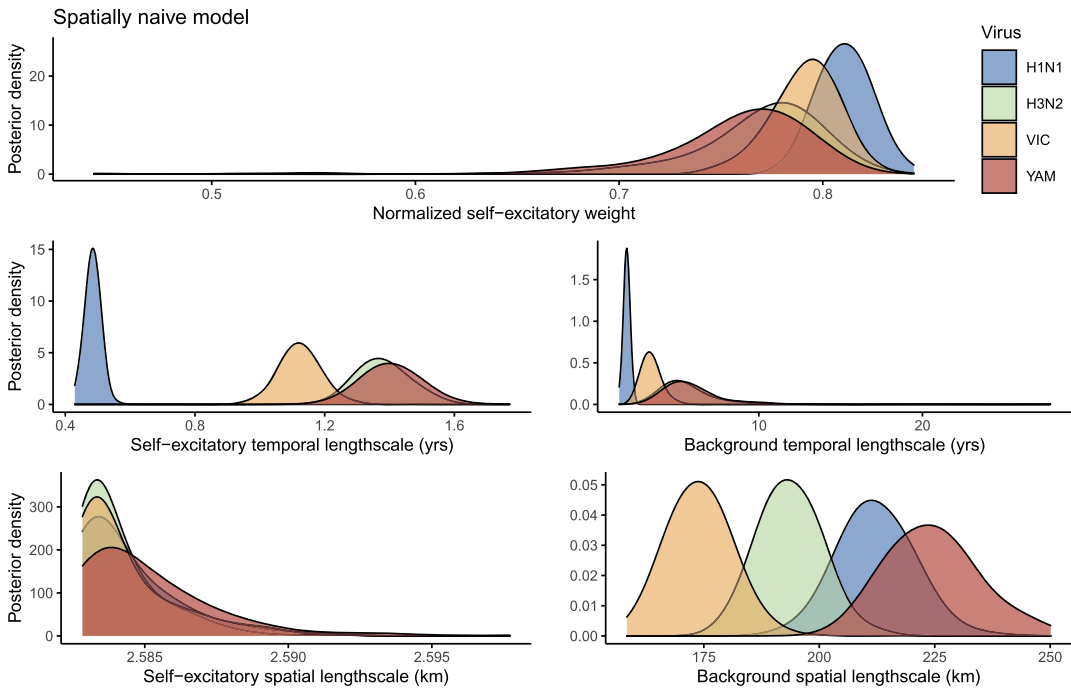
FIG. 9. *Strain specific posterior inference for a naive model.*

to self-excitation. One might address this issue by spatially perturbing all cases within, say, the same city, but this would not address the multiprecision nature of the data we consider and certainly would not take any transportation network into account. In addition, it seems that the resulting spatial lengthscale estimates would depend on the amount of perturbation applied in ways that are difficult to quantify or motivate a priori.

APPENDIX C: SIMULATION STUDY

For 800 independent instances, we: (a) simulate a spatiotemporal Hawkes process, (b) induce three different levels of coarsening and (c) apply our model with both fixed coarsened locations data and inferred locations. We then compare the coverage of resulting credible intervals for both the self-excitatory spatial lengthscale h and the unobserved locations. Although there is no expectation for Bayesian credible intervals to have perfect frequentist operating characteristics, the model with inferred locations performs reasonably well and outperforms the naive model, especially when coarsening becomes more severe.

To simulate a Hawkes process we use the clustering based algorithm of Zhuang, Ogata and Vere-Jones (2004) which first draws events from an inhomogeneous background process and then simulates successive generations according to the self-excitatory triggering function and

TABLE 3
Empirical coverage from 800 independent simulations for 50%, 80% and 95% credible intervals (CIs) of the self-excitatory spatial lengthscale h

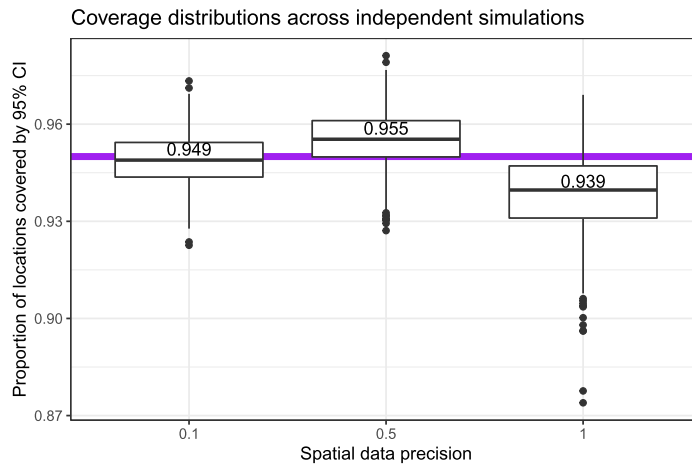|	50% CIs			80% CIs			95% CIs		
Spatial precision	1.0	0.5	0.1	1.0	0.5	0.1	1.0	0.5	0.1
Fixed locations	0.00	0.19	0.52	0.00	0.42	0.81	0.00	0.68	0.96
Sampled locations	0.53	0.49	0.53	0.84	0.81	0.81	0.98	0.95	0.96

FIG. 10. *Boxplots and mean proportions of locations covered by* 95% *credible intervals* (*CI*) *across* 800 *independent simulations. The purple line represents* 0.95.

its weight. For the background process we use a rate function with three Gaussian modes and draw an average of 200 points. Conditional on these points, we iteratively simulate new generations with an expected number of children of 0.5, and spatial lengthscales of 0.5. We then coarsen the locations data by rounding to 0.1, 0.5 and 1.0 to investigate performance when the magnitude of coarsening is less than, equal to and greater than the spatial lengthscale. We then generate 30,000 MCMC samples, both inferring the latent locations within bounded margins (0.1, 0.5, 1.0), as in Section 2.2.1, and keeping them fixed. In total, this simulation requires the generation of $30,000 \times 4 \times 800 = 96,000,000$ MCMC samples for which we use our Nvidia Quadro GP100 GPU and the algorithms of Appendix A. Table 3 shows credible interval empirical coverage of the true spatial lengthscale h. In general, we find that coverage is reasonably close to nominal values when inferring locations but deteriorates as a function of the data's spatial precision for the naive model. Figure 10 shows the distribution of proportions of locations covered by 95% credible intervals across the independent simulations. In general, coverage is reasonably close to 0.95 but deteriorates for data rounded to nearest integer (precision $= 1$). Part of this behavior, including the increased number of outliers, plausibly arises from the need to generate longer MCMC chains.

SUPPLEMENTARY MATERIAL

hpHawkes source (DOI: 10.1214/21-AOAS1517SUPPA; .zip). A zip file 'hpHawkes.zip' containing the source code for R package HPHAWKES.

Global influenza source (DOI: 10.1214/21-AOAS1517SUPPB; .zip). A zip file 'Flu Hawkes.zip' containing the source code for global influenza analysis.

Gunfire and wildfire source (DOI: 10.1214/21-AOAS1517SUPPC; .zip). A zip file 'unknownLocs.zip' containing the source code for Washington D.C. gunfire and Alaskan wildfire analyses.

REFERENCES

BEDFORD, T., SUCHARD, M. A., LEMEY, P., DUDAS, G., GREGORY, V., HAY, A. J., McCAULEY, J. W., RUSSELL, C. A., SMITH, D. J. et al. (2014). Integrating influenza antigenic dynamics with molecular evolution. *eLife* **3** e01914. https://doi.org/10.7554/eLife.01914

BEDFORD, T., RILEY, S., BARR, I. G., BROOR, S., CHADHA, M., COX, N. J., DANIELS, R. S., GUNASEKARAN, C. P., HURT, A. C. et al. (2015). Global circulation patterns of seasonal influenza viruses vary with antigenic drift. *Nature* **523** 217–220. PMCID: PMC4499780.

BROCKMANN, D. and HELBING, D. (2013). The hidden geometry of complex, network-driven contagion phenomena. *Science* **342** 1337–1342.

CARR, J. and DOLEAC, J. L. (2016). The geography, incidence, and underreporting of gun violence: New evidence using ShotSpotter data. *Brookings* (April 26, 2016).

CHOI, E., DU, N., CHEN, R., SONG, L. and SUN, J. (2015). Constructing disease network and temporal progression model via context-sensitive Hawkes process. In 2015 *IEEE International Conference on Data Mining* 721–726. IEEE, New York.

DALEY, D. J. and VERE-JONES, D. (2003). *An Introduction to the Theory of Point Processes*. *Vol. I*: *Elementary Theory and Methods*, 2nd ed. *Probability and Its Applications* (*New York*). Springer, New York. MR1950431

DALEY, D. J. and VERE-JONES, D. (2008). *An Introduction to the Theory of Point Processes*. *Vol. II*: *General Theory and Structure*, 2nd ed. *Probability and Its Applications* (*New York*). Springer, New York. MR2371524 https://doi.org/10.1007/978-0-387-49835-5

DESARBO, W. S., KIM, Y. and FONG, D. (1999). A Bayesian multidimensional scaling procedure for the spatial analysis of revealed choice data. *J. Econometrics* **89** 79–108. MR1681137 https://doi.org/10.1016/S0304-4076(98)00056-6

FOX, E. W., SCHOENBERG, F. P. and GORDON, J. S. (2016). Spatially inhomogeneous background rate estimators and uncertainty quantification for nonparametric Hawkes point process models of earthquake occurrences. *Ann. Appl. Stat.* **10** 1725–1756. MR3553242 https://doi.org/10.1214/16-AOAS957

GEISSER, S. (1975). The predictive sample reuse method with applications. *J. Amer. Statist. Assoc.* **70** 320–328.

GEYER, C. J. (1991). Markov chain Monte Carlo maximum likelihood.

GILKS, W. R., BEST, N. G. and TAN, K. (1995). Adaptive rejection Metropolis sampling within Gibbs sampling. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **44** 455–472.

HAWKES, A. (1973). Cluster models for earthquakes-regional comparisons. *Bull. Int. Stat. Inst.* **45** 454–461.

HEITJAN, D. F. (1993). Ignorability and coarse data: Some biomedical examples. *Biometrics* **49** 1099–1109.

HEITJAN, D. F. and RUBIN, D. B. (1991). Ignorability and coarse data. *Ann. Statist.* **19** 2244–2253. MR1135174 https://doi.org/10.1214/aos/1176348396

HOLBROOK, A. J., JI, X. and SUCHARD, M. A. (2022). Supplement to "Bayesian mitigation of spatial coarsening for a Hawkes model applied to gunfire, wildfire and viral contagion." https://doi.org/10.1214/21-AOAS1517SUPPA, https://doi.org/10.1214/21-AOAS1517SUPPB, https://doi.org/10.1214/21-AOAS1517SUPPC

HOLBROOK, A. J., LOEFFLER, C. E., FLAXMAN, S. R. and SUCHARD, M. A. (2021a). Scalable Bayesian inference for self-excitatory stochastic processes applied to big American gunfire data. *Stat. Comput.* **31** Paper No. 4, 15 pp. MR4199464 https://doi.org/10.1007/s11222-020-09980-4

HOLBROOK, A. J., LEMEY, P., BAELE, G., DELLICOUR, S., BROCKMANN, D., RAMBAUT, A. and SUCHARD, M. A. (2021b). Massive parallelization boosts big Bayesian multidimensional scaling. *J. Comput. Graph. Statist.* **30** 11–24. MR4235961 https://doi.org/10.1080/10618600.2020.1754226

KAHLE, D. and WICKHAM, H. (2013). ggmap: Spatial visualization with ggplot2. *R J.* **5** 144–161.

KELLY, J. D., PARK, J., HARRIGAN, R. J., HOFF, N. A., LEE, S. D., WANNIER, R., SELO, B., MOSSOKO, M., NJOLOKO, B. et al. (2019). Real-time predictions of the 2018–2019 Ebola virus disease outbreak in the Democratic Republic of the Congo using Hawkes point process models. *Epidemics* **28** 100354.

KIM, H. (2011). Spatio-temporal point process models for the spread of avian influenza virus (H5N1). Ph.D. thesis, UC Berkeley. MR2926851

KRUSKAL, J. B. (1964). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika* **29** 1–27. MR0169712 https://doi.org/10.1007/BF02289565

LOEFFLER, C. and FLAXMAN, S. (2018). Is gun violence contagious? A spatiotemporal test. *J. Quant. Criminol.* **34** 999–1017.

LOMAX, A., MICHELINI, A. and CURTIS, A. (2009). Earthquake location, direct, global-search methods. *Encycl. Complex. Syst. Sci.* **5** 2449–2473.

MARQUIS DE LAPLACE, P. S. (1825). *Essai philosophique sur les probabilités*. Bachelier.

MEYER, S. and HELD, L. (2014). Power-law models for infectious disease spread. *Ann. Appl. Stat.* **8** 1612–1639. MR3271346 https://doi.org/10.1214/14-AOAS743

MOHLER, G. (2014). Marked point process hotspot maps for homicide and gun crime prediction in Chicago. *Int. J. Forecast.* **30** 491–497.

NEAL, R. M. (2011). MCMC using Hamiltonian dynamics. In *Handbook of Markov Chain Monte Carlo*. *Chapman & Hall/CRC Handb. Mod. Stat. Methods* 113–162. CRC Press, Boca Raton, FL. MR2858447

NISHIMURA, A. and SUCHARD, M. A. (2018). Prior-preconditioned conjugate gradient method for accelerated Gibbs sampling in "large *n* & large *p*" sparse Bayesian regression. Preprint. Available at arXiv:1810.12437.

OGATA, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes. *J. Amer. Statist. Assoc.* **83** 9–27.

OH, M.-S. and RAFTERY, A. E. (2001). Bayesian multidimensional scaling and choice of dimension. *J. Amer. Statist. Assoc.* **96** 1031–1044. MR1947251 https://doi.org/10.1198/016214501753208690

OH, M.-S. and RAFTERY, A. E. (2007). Model-based clustering with dissimilarities: A Bayesian approach. *J. Comput. Graph. Statist.* **16** 559–585. MR2351080 https://doi.org/10.1198/106186007X236127

PARK, J., CHAFFEE, A. W., HARRIGAN, R. J. and SCHOENBERG, F. P. (2018). A non-parametric Hawkes model of the spread of Ebola in West Africa. *J. Appl. Stat*. To appear. https://doi.org/10.1080/02664763.2020.1825646

PARK, J., SCHOENBERG, F. P., BERTOZZI, A. L. and BRANTINGHAM, P. J. (2019). Investigating clustering and violence interruption in gang-related violent crime data using spatial-temporal point processes with covariates.

PEARSON, K. (1901). LIII. On lines and planes of closest fit to systems of points in space. *Lond. Edinb. Dublin Philos. Mag. J. Sci.* **2** 559–572.

PLUMMER, M., BEST, N., COWLES, K. and VINES, K. (2006). CODA: Convergence diagnosis and output analysis for MCMC. *R News* **6** 7–11.

POLSON, N. G., SCOTT, J. G. and WINDLE, J. (2014). The Bayesian bridge. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 713–733. MR3248673 https://doi.org/10.1111/rssb.12042

R CORE TEAM (2019). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

RAMSAY, J. O. (1982). Some statistical approaches to multidimensional scaling data. *J. Roy. Statist. Soc. Ser. A* **145** 285–312. MR0678529 https://doi.org/10.2307/2981865

REINHART, A. (2018). A review of self-exciting spatio-temporal point processes and their applications. *Statist. Sci.* **33** 299–318. MR3843374 https://doi.org/10.1214/17-STS629

REINHART, A. and GREENHOUSE, J. (2018). Self-exciting point processes with spatial covariates: Modelling the dynamics of crime. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **67** 1305–1329. MR3873709 https://doi.org/10.1111/rssc.12277

RIZOIU, M.-A., MISHRA, S., KONG, Q., CARMAN, M. and XIE, L. (2018). SIR-Hawkes: Linking epidemic models and Hawkes processes to model diffusions in finite populations. In *Proceedings of the* 2018 *World Wide Web Conference on World Wide Web* 419–428. International World Wide Web Conferences Steering Committee.

ROBERTS, G. O. and ROSENTHAL, J. S. (2007). Coupling and ergodicity of adaptive Markov chain Monte Carlo algorithms. *J. Appl. Probab.* **44** 458–475. MR2340211 https://doi.org/10.1239/jap/1183667414

SCHOENBERG, F. P. (2004). Testing separability in spatial-temporal marked point processes. *Biometrics* **60** 471–481. MR2066282 https://doi.org/10.1111/j.0006-341X.2004.00192.x

SCHOENBERG, F. P. (2013). Facilitated estimation of ETAS. *Bull. Seismol. Soc. Amer.* **103** 601–605.

SCHOENBERG, F. P. (2016). A note on the consistent estimation of spatial-temporal point process parameters. *Statist. Sinica* **26** 861–879. MR3497774

SUCHARD, M. A., LEMEY, P., BAELE, G., AYRES, D. L., DRUMMOND, A. J. and RAMBAUT, A. (2018). Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol.* **4** vey016. https://doi.org/10.1093/ve/vey016

VEHTARI, A., GELMAN, A. and GABRY, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Stat. Comput.* **27** 1413–1432. MR3647105 https://doi.org/10.1007/s11222-016-9696-4

WICKHAM, H. (2016). *Ggplot2: Elegant Graphics for Data Analysis*. Springer, New York.

ZHUANG, J., OGATA, Y. and VERE-JONES, D. (2004). Analyzing earthquake clustering features by using stochastic reconstruction. *J. Geophys. Res.*, *Solid Earth* **109** B05301.