Three challenges for spatiotemporal Hawkes modeling

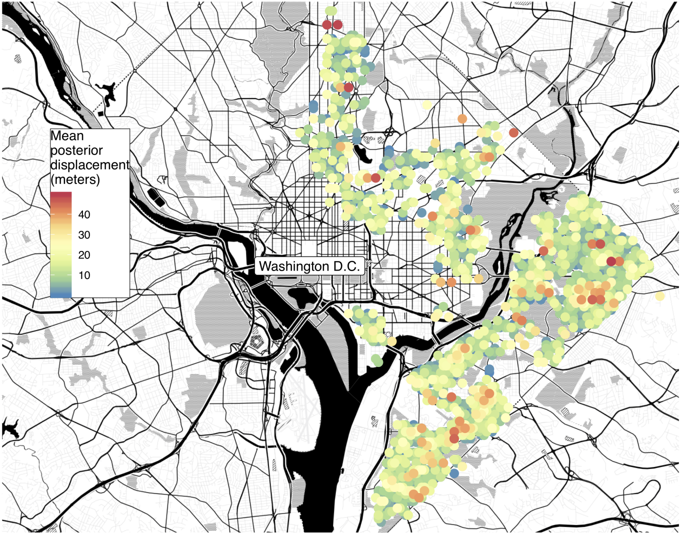
Andrew J. Holbrook

UCLA Biostatistics

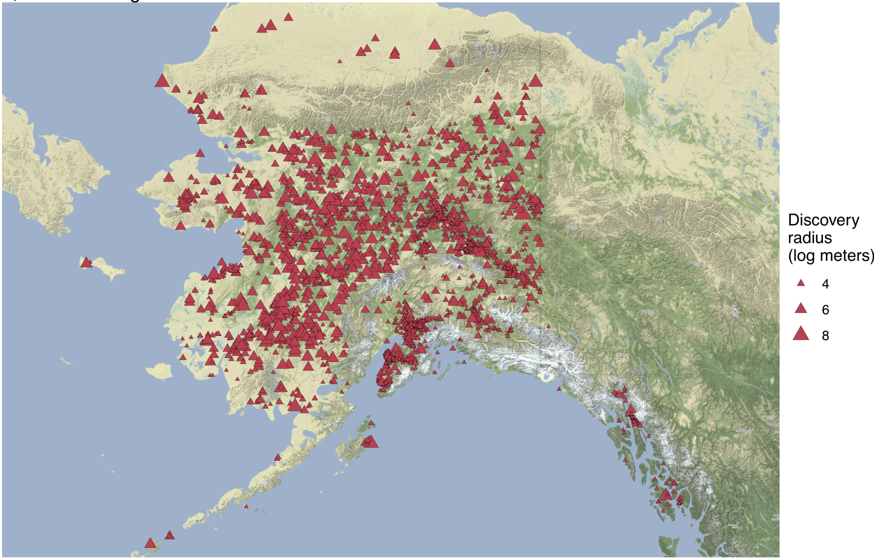
September 28, 2021

Spatiotemporal data in public health

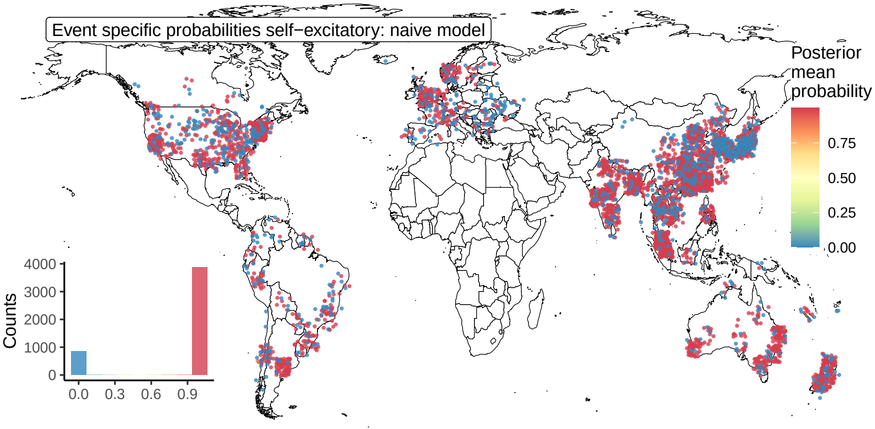
Washington D.C. gunshots (2018)



2,925 Wildfire ignition sites in Alaska: 2015–2019



Global influenza (2000-2012)



Poisson processes

A counting process $\{N(t),\ t > 0\}$ is a homogeneous Poisson point process with rate $\lambda > 0$ if

(i) $N(0) = 0$;

(ii) $\big(N(t_4) - N(t_3)\big) \perp \big(N(t_2) - N(t_1)\big)$ for $t_1 < t_2 \leq t_3 < t_4$;

(iii) $\big(N(t_2) - N(t_1)\big) \sim \mathit{Poisson}\big(\lambda(t_2 - t_1)\big)$ for $t_2 > t_1$.

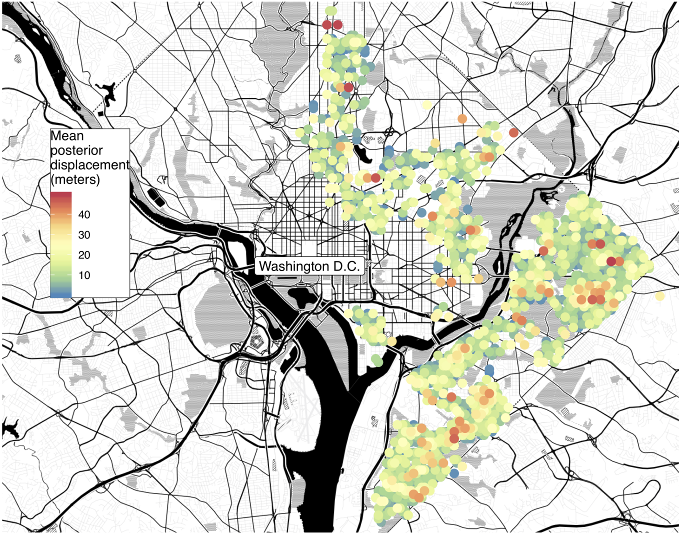
It is an inhomogeneous Poisson point process with rate $\lambda(t) > 0$ if

(i) $N(0) = 0$;

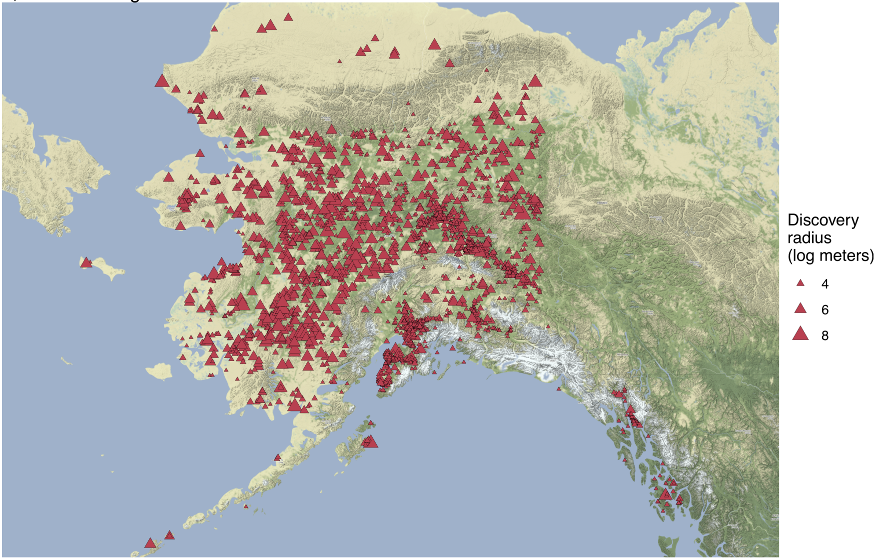
(ii) $\big(N(t_4) - N(t_3)\big) \perp \big(N(t_2) - N(t_1)\big)$ for $t_1 < t_2 \leq t_3 < t_4$;

(iii) $\big(N(t_2) - N(t_1)\big) \sim \mathit{Poisson}\big(\int_{t_1}^{t_2} \lambda(t)\mathrm{d}t\big)$ for $t_2 > t_1$.

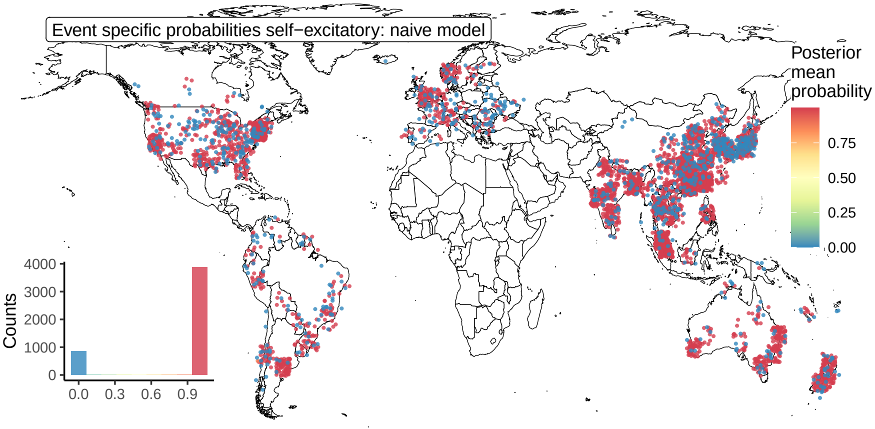
Washington D.C. gunshots (2018)



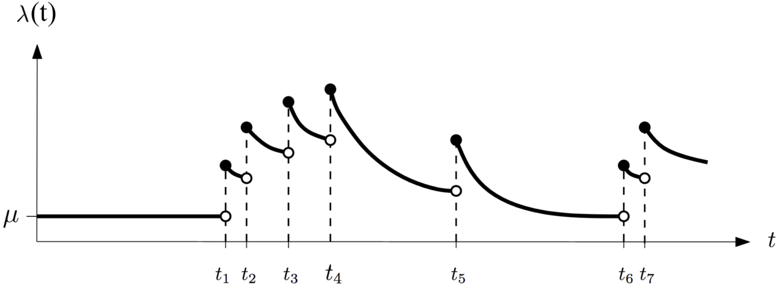
2,925 Wildfire ignition sites in Alaska: 2015–2019



Global influenza (2000-2012)



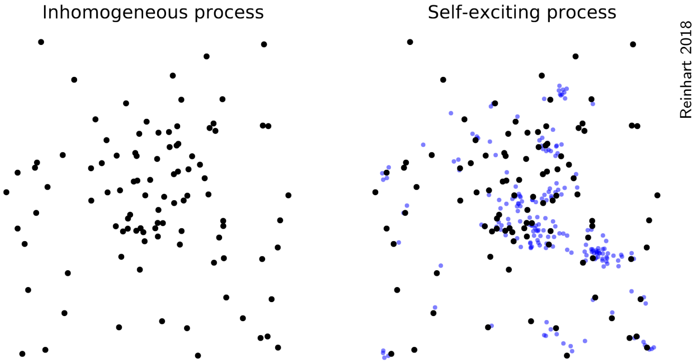
Hawkes process



Laub et al. 2015

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

Spatiotemporal Hawkes process



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

A simple model

We assume

1. an exponential decay triggering function,
2. Gaussian kernel spatial smoothers, and
3. separability in space/time:

$$\xi(\mathrm{x}, t) = \frac{\theta_0 \omega}{h^D} \sum_{t_n < t} e^{-\omega\,(t - t_n)} \phi\left(\frac{\mathrm{x} - \mathrm{x}_n}{h}\right)$$

$$\mu(\mathrm{x}, t) = \frac{\mu_0}{\tau_x^D\,\tau_t} \sum_{n=1}^{N} \phi\left(\frac{\mathrm{x} - \mathrm{x}_n}{\tau_x}\right) \cdot \phi\left(\frac{t - t_n}{\tau_t}\right) .$$

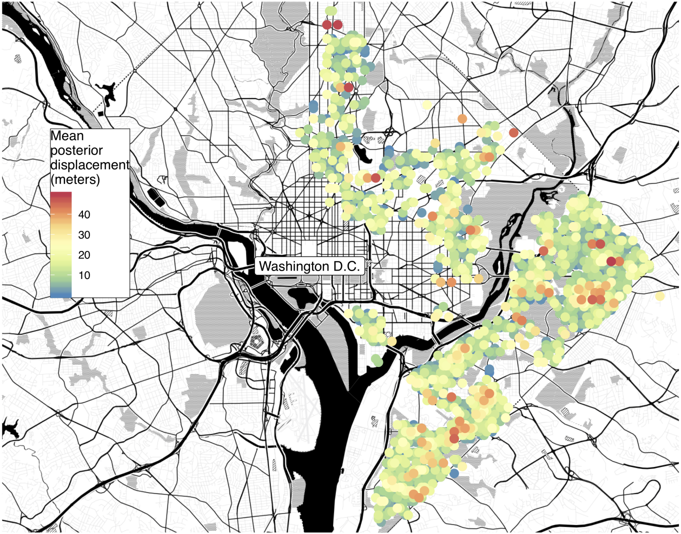
Three inferential challenges

Likelihood based inference encounters (at least) three challenges that are not independent from one another.

big data $\times$ spatial data precision $\times$ big model

Big data

Washington D.C. gunshots (2018)



D.C. gunshot data (2006-2018)

An acoustic gunshot location system recorded over 85k gunshots in Washington D.C. between 2006 and 2018.

Loeffler and Flaxman (2018) used a subset of 9k gunshots in the paper titled *Is gun violence contagious? A spatiotemporal test*.

They answered ‘yes’, but did the results hold for a complete data analysis?

Likelihood based inference

The likelihood for data (x_1, t_1), ..., (x_N, t_N) is

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

The log-likelihood involves the term

$$\sum_{n=1}^{N}\log\lambda_n = \sum_{n=1}^{N}\log\left(\mu_n + \frac{\theta_0\omega}{h^D}\sum_{t_{n'}<t_n} e^{-\omega\,(t_n - t_{n'})}\phi\left(\frac{\mathrm{x}_n - \mathrm{x}_{n'}}{h}\right)\right)$$

The gradient w.r.t. Θ also features a double summation.

Parallelization methods

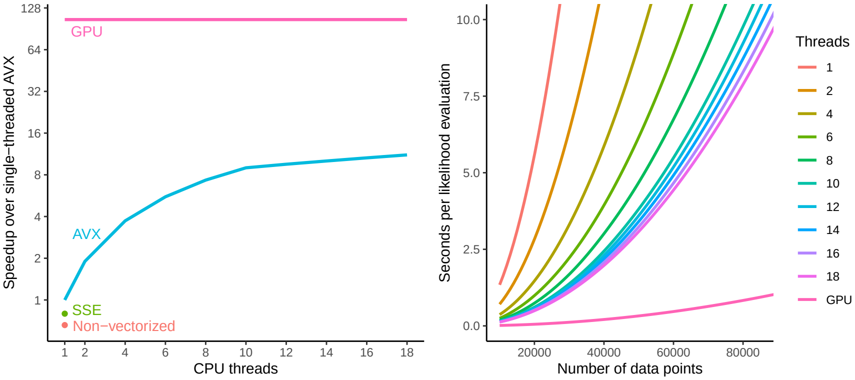
Central processing unit (CPU):

1. Global parallelization: 2 to hundreds of cores (multi-core)
2. Local parallelization: single instruction multiple data (SIMD)

Graphics processing unit (GPU):

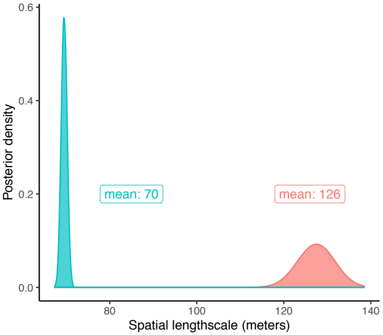
1. Thousands of cores (many-core)
2. Single instruction multiple threads (SIMT)
3. High memory bandwidth

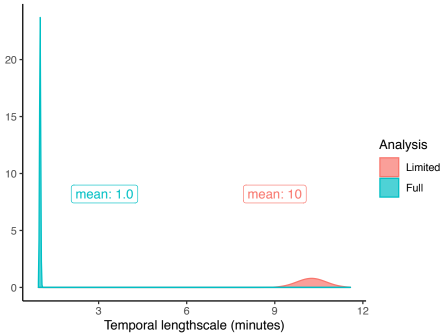
Significant speedups

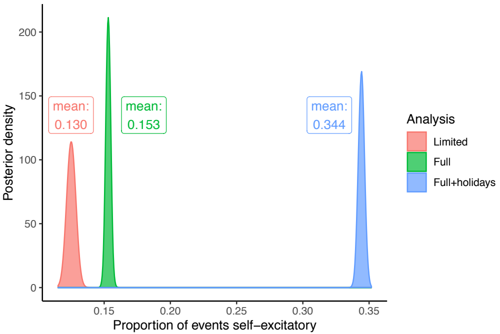


Significant speedups

	Seconds per evaluation			Relative speedup		
N	GPU	C++	R	GPU	C++	R
5000	0.004	0.80	5.02	1255.00	6.27	1
10000	0.01	2.66	18.74	1338.57	7.05	1
20000	0.05	10.10	105.54	1991.32	10.45	1
30000	0.12	21.10	232.51	1970.42	11.02	1

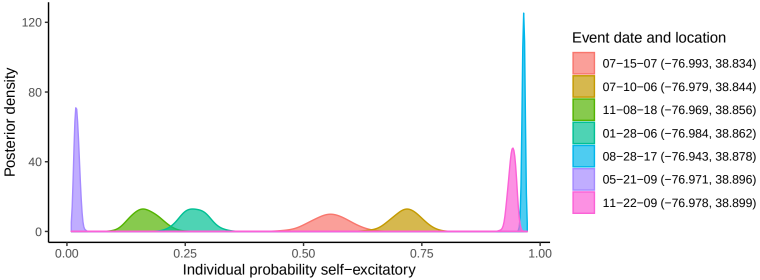

Posterior density
0.6
0.4
0.2
0.0
mean: 70
mean: 126
80
100
120
140
Spatial lengthscale (meters)


20
15
10
5
0
mean: 1.0
mean: 10
3
6
9
12
Temporal lengthscale (minutes)
Analysis
Limited
Full


Posterior density
200
150
100
50
0
mean: 0.130
mean: 0.153
mean: 0.344
0.15
0.20
0.25
0.30
0.35
Proportion of events self−excitatory
Analysis
Limited
Full
Full+holidays

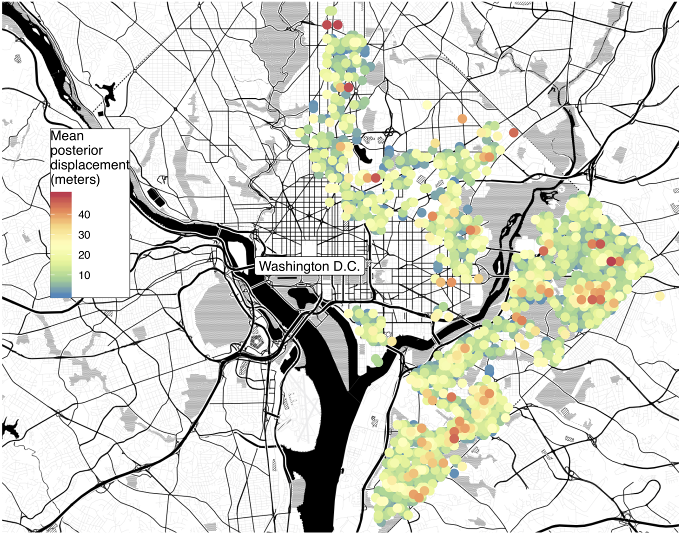
Postprocessing is expensive too

We can also consider the posterior distribution for the probability an event comes from self-excitation: $\xi_n/(\xi_n + \mu_n)$.

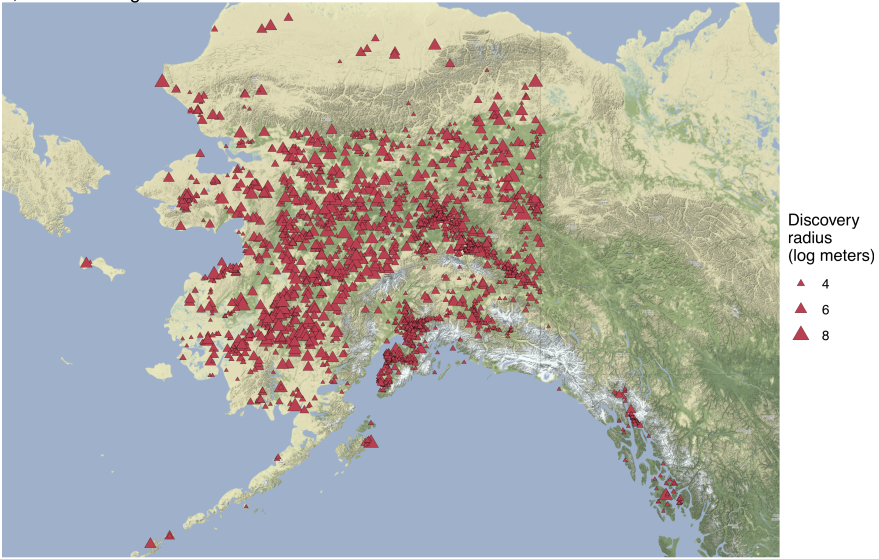


Spatial data precision

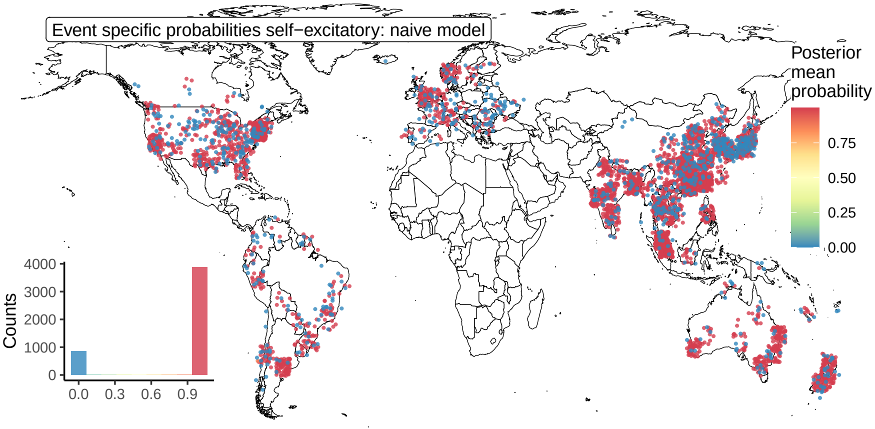
Washington D.C. gunshots (2018)



2,925 Wildfire ignition sites in Alaska: 2015–2019



Global influenza (2000-2012)



Simultaneously inferring gunshot locations

The P.D. rounds the data to the nearest 100 meters. A uniform prior over the 10k m^2 square centered at each observation $\mathfrak{x}_n$

$$p(\mathrm{x}_n) \propto 1\,, \quad \mathfrak{x}_{nd} - 50 < \mathrm{x}_{nd} < \mathfrak{x}_{nd} + 50\,,\ d = 1, 2$$

corresponds to using the *grouped data likelihood* of Heitjan and Rubin (1991).

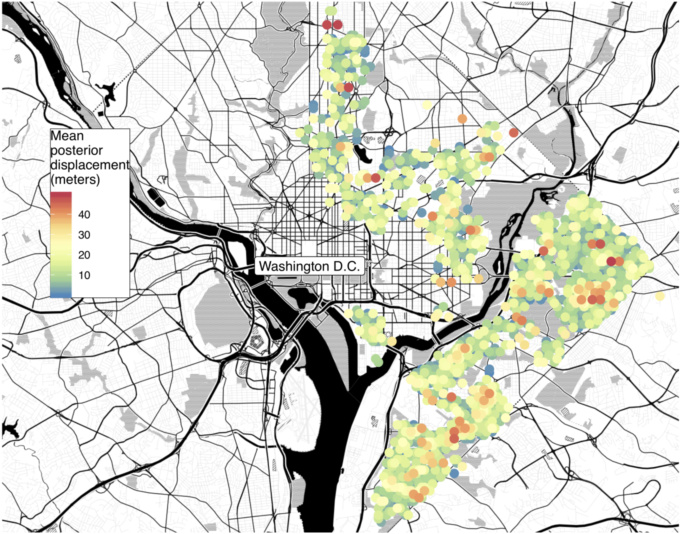
Simultaneously inferring gunshot locations

Inferred and observed locations #1

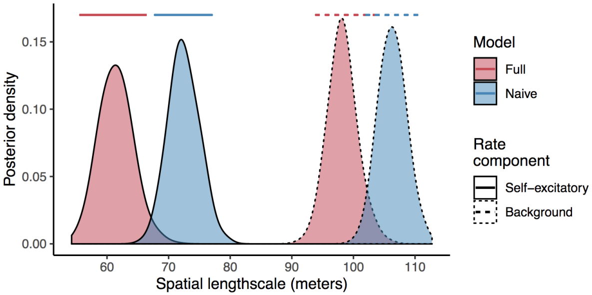
Inferred and observed locations #2

Inferred and observed locations #3

Simultaneously inferring gunshot locations

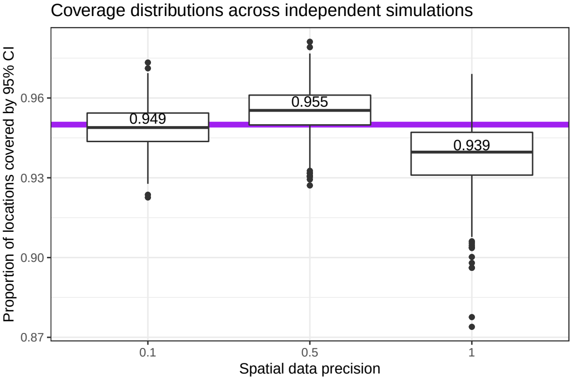


Good news!



		Posterior median (95% Credible interval)	
Rate component	Parameter	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)
Self-excitatory	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)

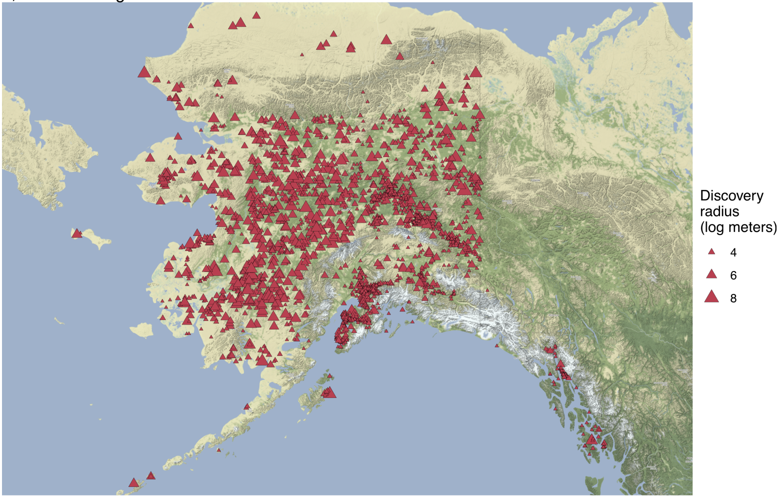
Breaking the model: decreasing precision



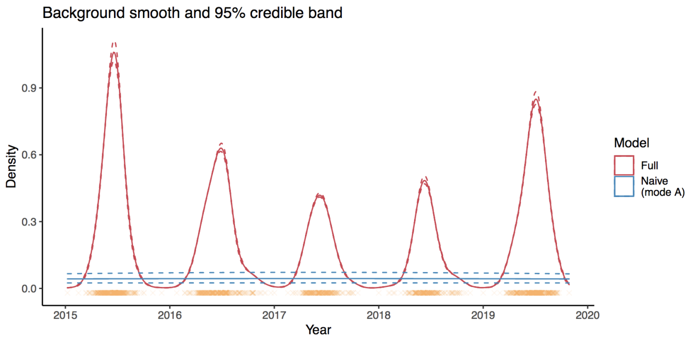
	50% CIs			80% CIs			95% CIs		
Spatial precision	1.0	0.5	0.1	1.0	0.5	0.1	1.0	0.5	0.1
Fixed locations	0.00	0.19	0.52	0.00	0.42	0.81	0.00	0.68	0.96
Sampled locations	0.53	0.49	0.53	0.84	0.81	0.81	0.98	0.95	0.96

Breaking the model: variable precision

2,925 Wildfire ignition sites in Alaska: 2015–2019



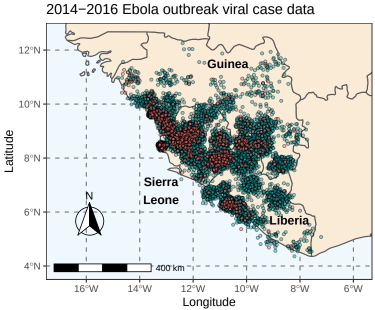
Breaking the model: variable precision

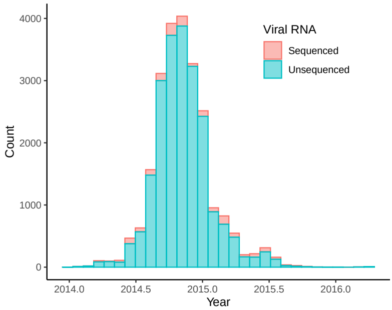


		Posterior median (95% Credible interval)		
Rate component	Parameter	Full model	Naive model *A*	Naive model *B*
Background	Spatial lengthscale (km)	34.8 (32.9, 37.6)	**23.5 (22.3, 24.6)**	63.0 (58.7, 68.7)
	Temporal lengthscale (days)	25.9 (23.8, 27.9)	3244.0 (1929.7, 5803.5)	**10.2 (9.4, 11.1)**
Self-excitatory	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)

Big model

2014-2016 Ebola virus outbreak in West Africa





- 1,610 sequenced viruses (1,367 of which have locations data)
- 21,811 unsequenced cases

Variable degrees of contagion

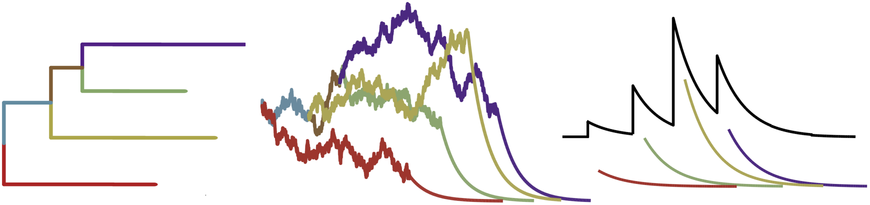
One can tailor the triggering function to change for each observation (Schoenberg et al., 2019):

$$\lambda(\mathrm{x}, t) = \mu(\mathrm{x}) + \sum_{t_n < t} g_n(\mathrm{x} - \mathrm{x}_n, t - t_n)\,.$$

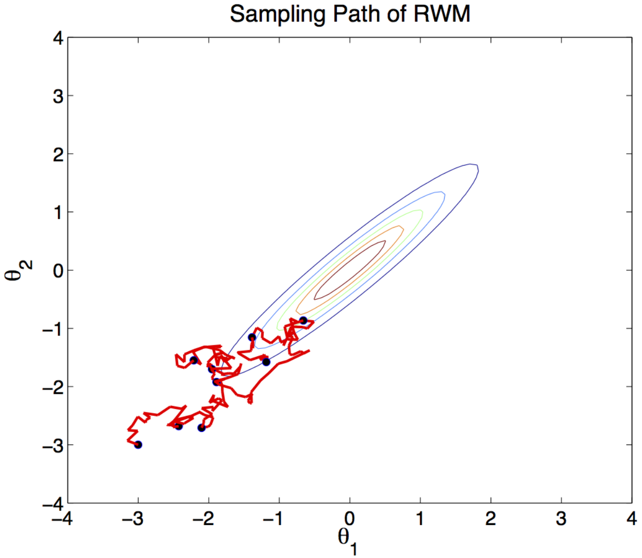
In the following, I specify

$$\xi(\mathrm{x}, t) = \frac{\theta_0 \omega}{h^D} \sum_{t_n < t} \theta_n \, e^{-\omega\,(t - t_n)} \phi\left(\frac{\mathrm{x} - \mathrm{x}_n}{h}\right)\,.$$

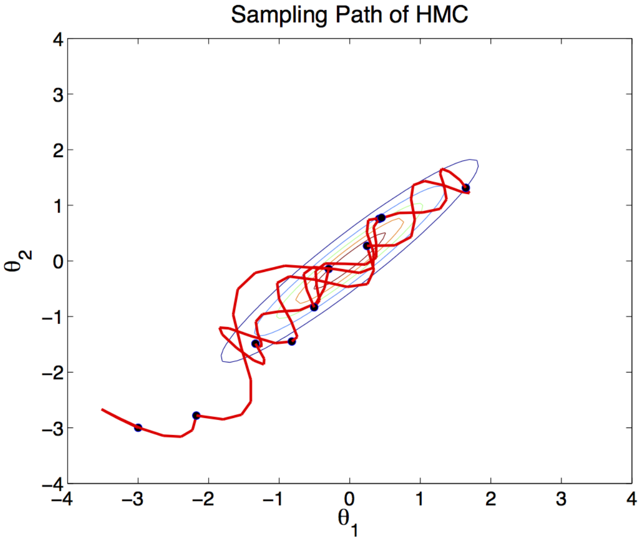
Phylogenetic Hawkes process



Random walk Metropolis



Hamiltonian Monte Carlo



Hamiltonian Monte Carlo

Augment parameter space with auxiliary Gaussian variable p and construct a Hamiltonian energy function:

$$
\begin{aligned}
H(\mathsf{z}, \mathsf{p}) &= -\log(\pi(\mathsf{z}) \times \phi(\mathsf{p})) \\
&\propto -\log \pi(\mathsf{z}) + \frac{1}{2}\mathsf{p}^T \mathsf{M}^{-1}\mathsf{p}\,.
\end{aligned}
$$

New states of the Markov chain are proposed by forward integrating Hamilton's equations:

$$
\begin{aligned}
\frac{\mathrm{d}\mathsf{z}}{\mathrm{d}t} &= \frac{\partial H}{\partial \mathsf{p}} = \mathsf{M}^{-1}\mathsf{p} \\
\frac{\mathrm{d}\mathsf{p}}{\mathrm{d}t} &= -\frac{\partial H}{\partial \mathsf{z}} = \nabla \log \pi(\mathsf{z})\ .
\end{aligned}
$$

Numerical simulation induces discretization error, which we correct with a Metropolis accept-reject step.

Hamiltonian Monte Carlo

Benefits:

- HMC scales to tens-of-thousands of parameters.

Challenges:

- HMC necessitates repeated computation of log-likelihood and its gradient (best case $\mathcal{O}(N)$).
- Preconditioning required for ill-conditioned posteriors. May involve *additional* expensive Hessian evaluations.

HMC for variable rates?

- The Hawkes likelihood scales $\mathcal{O}(N^2)$ ✓

- But the Hawkes log-likelihood gradient also scales $\mathcal{O}(N^2)$

$$\frac{\partial \ell}{\partial \theta_n} = -\frac{\partial \Lambda_n}{\partial \theta_n} + \sum_{t_n < t_{n'}} \frac{1}{\lambda_{n'}} \frac{\partial \lambda_{n'n}}{\partial \theta_n}$$

$$= \theta_0 \left(e^{-\omega\,(t_N - t_n)} - 1 \right) + \sum_{t_n < t_{n'}} \frac{1}{\lambda_{n'}} \frac{\theta_0 \omega}{h^D} e^{-\omega\,(t_{n'} - t_n)} \phi\left(\frac{\mathrm{x}_{n'} - \mathrm{x}_n}{h} \right)$$

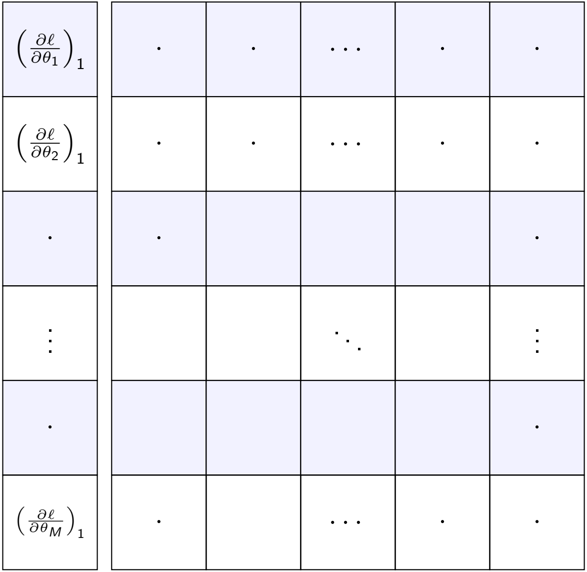
- And the Hawkes log-likelihood Hessian diagonal also scales $\mathcal{O}(N^2)$!

$$\mathrm{M}_{mm}^{-1} \approx -\frac{\partial^2 \ell}{\partial \theta_m^2} = \sum_{t_m < t_n} \frac{1}{\lambda_n^2} \frac{\theta_0^2 \omega^2}{h^{2D}} e^{-2\omega\,(t_n - t_m)} \phi^2 \left(\frac{\mathrm{x}_n - \mathrm{x}_m}{h} \right) .$$

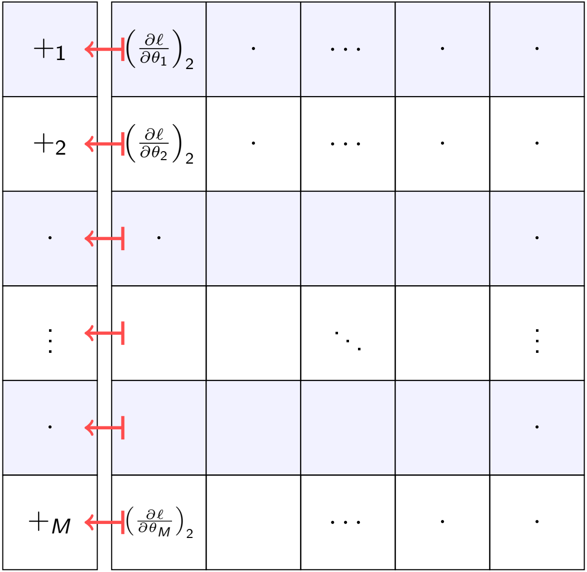
Parallel gradient calculations

$\left(\frac{\partial \ell}{\partial \theta_1}\right)_1$	$\left(\frac{\partial \ell}{\partial \theta_1}\right)_2$	$\cdot$	$\cdots$	$\cdot$	$\left(\frac{\partial \ell}{\partial \theta_1}\right)_N$	$\frac{\partial \ell}{\partial \theta_1}$
$\left(\frac{\partial \ell}{\partial \theta_2}\right)_1$	$\left(\frac{\partial \ell}{\partial \theta_2}\right)_2$	$\cdot$	$\cdots$	$\cdot$	$\left(\frac{\partial \ell}{\partial \theta_2}\right)_N$	$\frac{\partial \ell}{\partial \theta_2}$
$\cdot$	$\cdot$				$\cdot$	$\cdot$
$\vdots$			$\ddots$		$\vdots$	$\vdots$
$\cdot$					$\cdot$	$\cdot$
$\left(\frac{\partial \ell}{\partial \theta_M}\right)_1$	$\cdot$		$\cdots$	$\cdot$	$\left(\frac{\partial \ell}{\partial \theta_M}\right)_N$	$\frac{\partial \ell}{\partial \theta_M}$

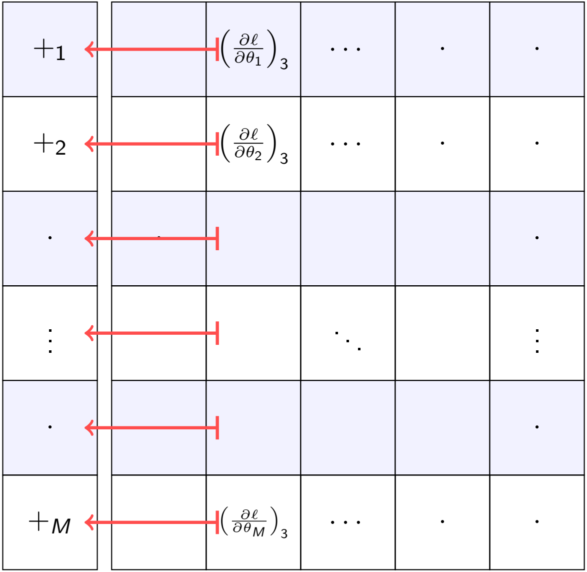
Parallel gradient calculations



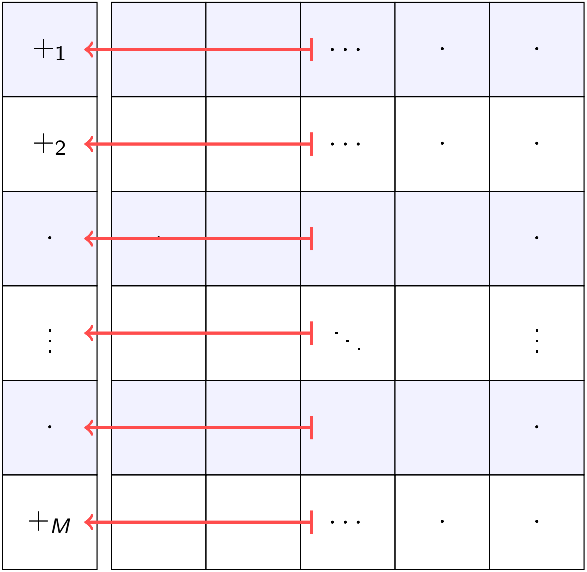
Parallel gradient calculations



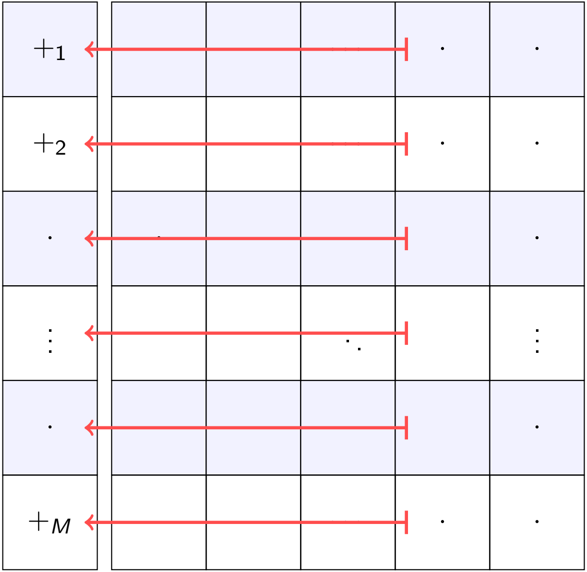
Parallel gradient calculations



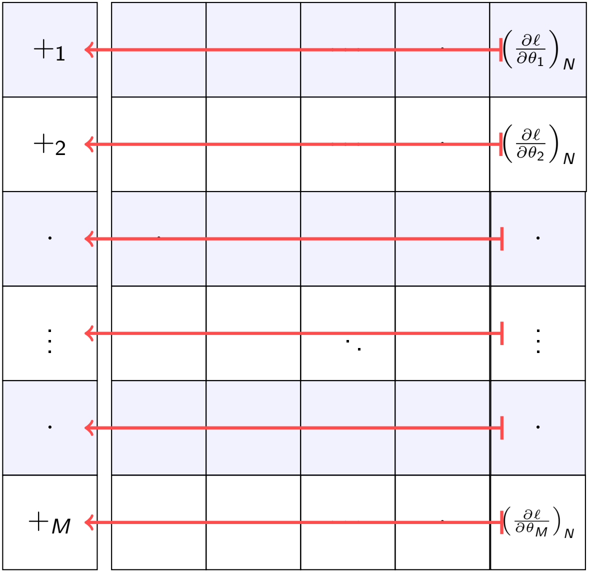
Parallel gradient calculations



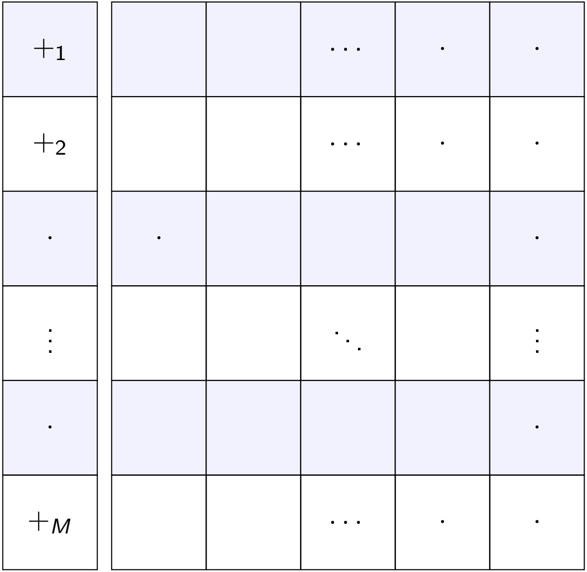
Parallel gradient calculations



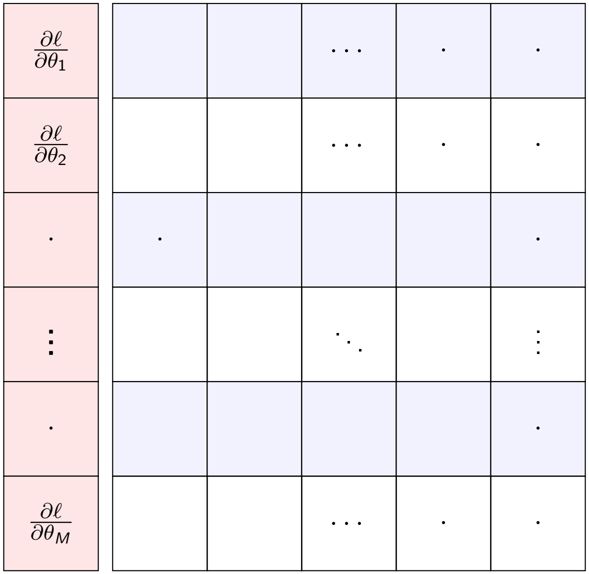
Parallel gradient calculations

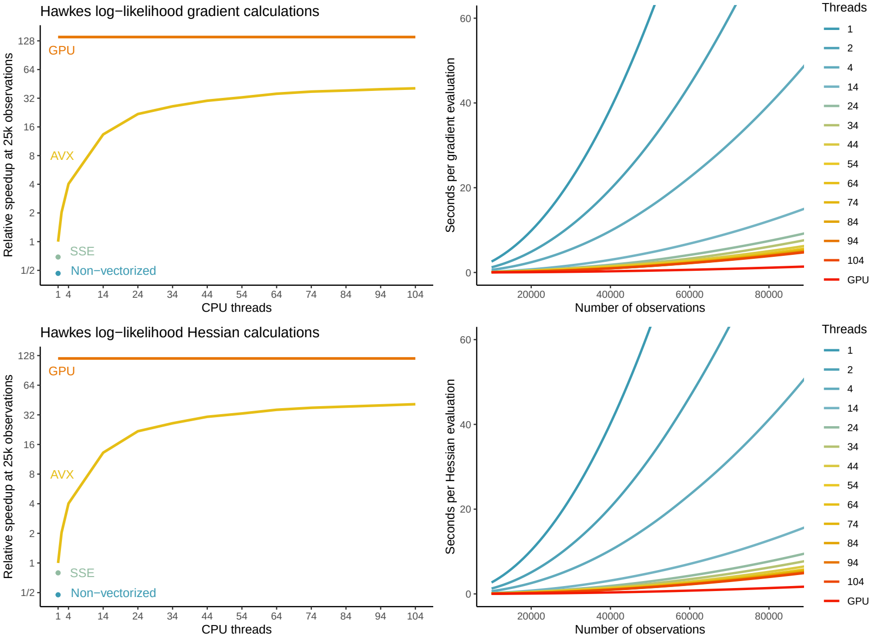


Parallel gradient calculations



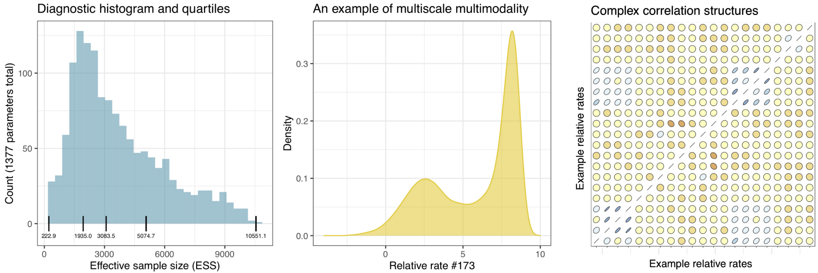
Parallel gradient calculations



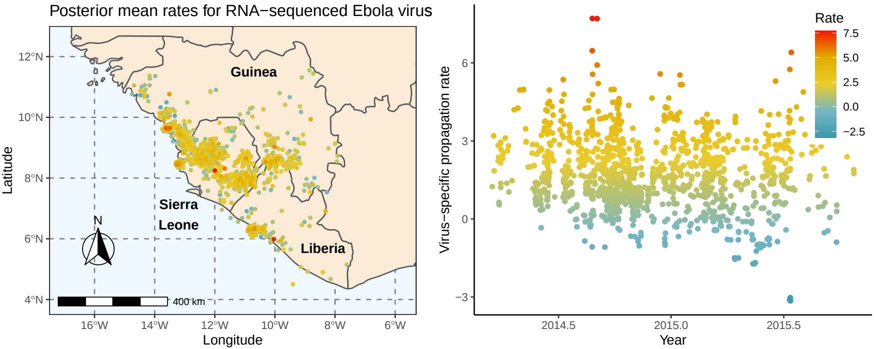

Hawkes log−likelihood gradient calculations
Relative speedup at 25k observations
128
64
32
16
8
4
2
1
1/2
GPU
AVX
SSE
Non−vectorized
1 4 14 24 34 44 54 64 74 84 94 104
CPU threads
Seconds per gradient evaluation
60
40
20
0
20000 40000 60000 80000
Number of observations
Threads
1
2
4
14
24
34
44
54
64
74
84
94
104
GPU
Hawkes log−likelihood Hessian calculations
Relative speedup at 25k observations
GPU
AVX
SSE
Non−vectorized
CPU threads
Seconds per Hessian evaluation
Number of observations
Threads

Posterior inference

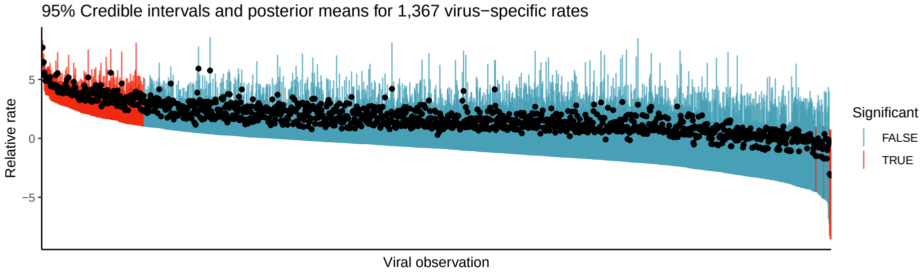
Generate 100 million Markov chain states ($\sim$ 3.5 million samples/day on Nvidia GV100) in 1 month.



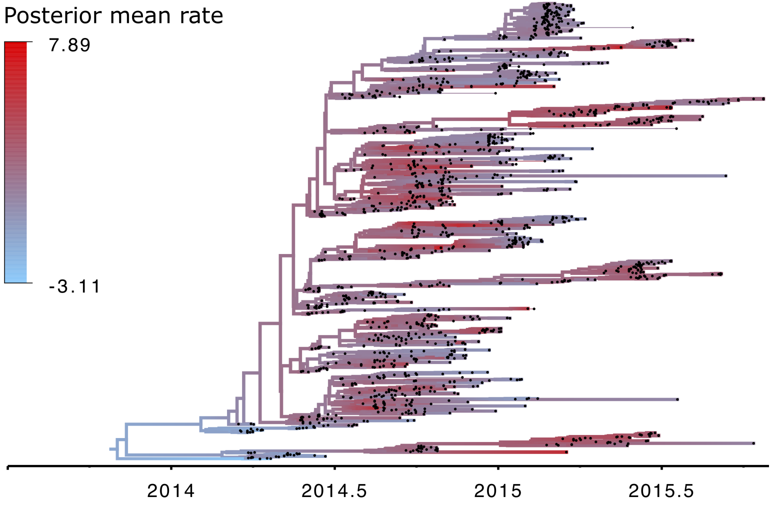
Inferred rates of contagion



Inferred rates of contagion



Biologically modulated rates



Big data exacerbates challenges

There are more challenges:

- Flexible models (the irony of model based nonparametrics)
- Boundary issues (censoring and truncation)
- Differential sampling

Acknowledgements

Joint work with

- Marc Suchard (UCLA)
- Xiang Ji (Tulane)

Funded by K-Award *Big Data Predictive Phylogenetics with Bayesian Learning*.

NIH National Institute of Allergy and Infectious Diseases

Thank you!