Bayes in the time of Big Data

Andrew J. Holbrook

Department of Human Genetics
University of California, Los Angeles

January 22, 2020

Overview

Part 1. One hundred years of deadly flu

Overview

Part 1. One hundred years of deadly flu

Part 2. A highly structured model for the spread of viruses along global transportation networks

Overview

Part 1. One hundred years of deadly flu

Part 2. A highly structured model for the spread of viruses along global transportation networks

Part 3. Modern Bayesian inference

Overview

Part 1. One hundred years of deadly flu

Part 2. A highly structured model for the spread of viruses along global transportation networks

Part 3. Modern Bayesian inference

Part 4. Massive parallelization

Overview

Part 1. One hundred years of deadly flu

Part 2. A highly structured model for the spread of viruses along global transportation networks

Part 3. Modern Bayesian inference

Part 4. Massive parallelization

Part 5. Global spread of influenza

Overview

Part 1. One hundred years of deadly flu

Part 2. A highly structured model for the spread of viruses along global transportation networks

Part 3. Modern Bayesian inference

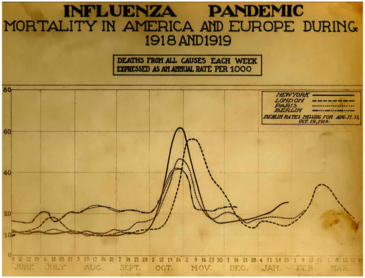
Part 4. Massive parallelization

Part 5. Global spread of influenza

Part 6. Big Bayes

Part 1. One hundred years of deadly flu

1918-1919 Influenza epidemic



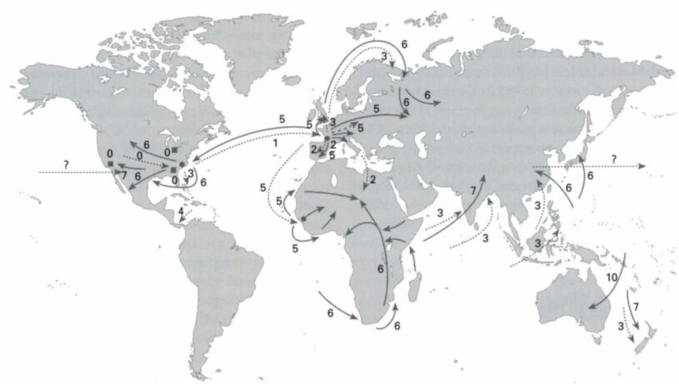
"Spanish flu" infected 500 million people worldwide and killed 50 million.

17 million in India; 675, 300 and 400 thousand in the U.S., Brazil and Japan...

A-H1N1 influenza, no more aggressive than previous strains.

Successful spread linked to First World War.

Global spread of Spanish flu



Nicholson et al. 1998

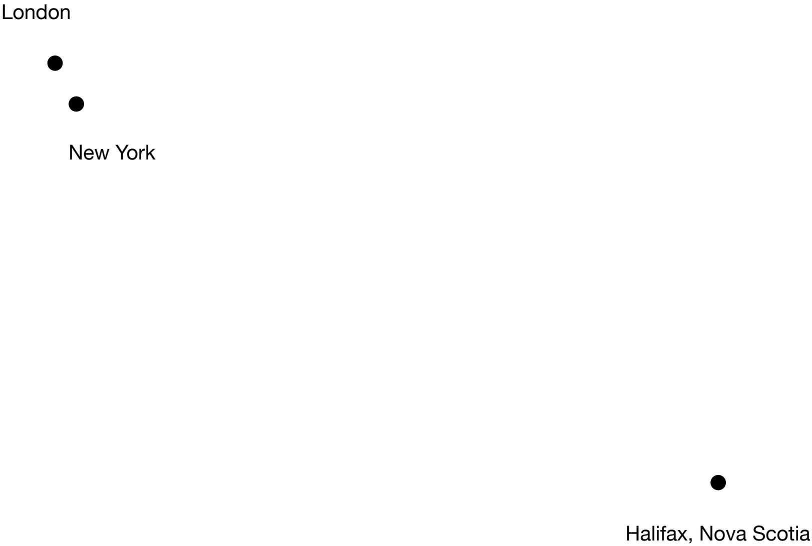
Naive spatial distances

London

Halifax, Nova Scotia

New York

Ocean traffic passenger distance

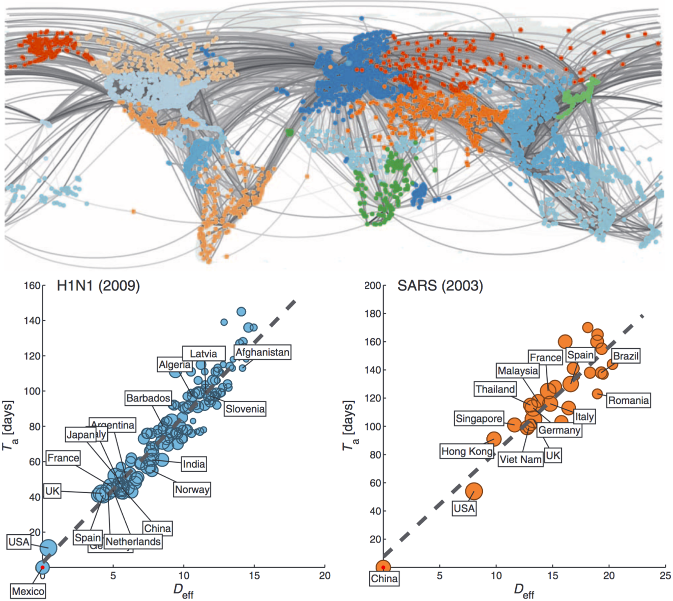


London

New York

Halifax, Nova Scotia

Contemporary networks of global travel

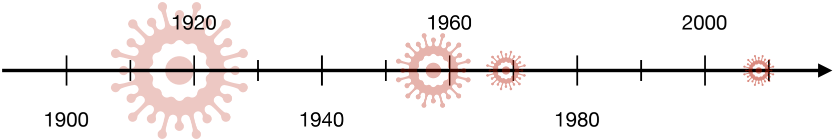


Brockmann and Helbing 2013

Deadly pandemics

“Spanish influenza”
A-H1N1
50-100 million deaths

“Hong Kong flu”
A-H3N2
1 million deaths



“Asian flu”
A-H2N2
1-4 million deaths

“Swine flu”
A-H1N1
150-580 thousand deaths

Some questions

Question 1. How should we characterize, quantify
and estimate rates of viral spread?

Some questions

Question 1. How should we characterize, quantify and estimate rates of viral spread?

Question 2. Do certain subtypes travel more effectively around the world? If so, which?

Some questions

Question 1. How should we characterize, quantify and estimate rates of viral spread?

Question 2. Do certain subtypes travel more effectively around the world? If so, which?

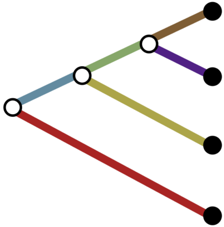
Question 3. How might we quantify our uncertainty?

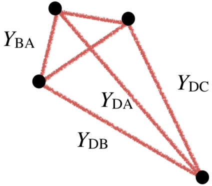
Part 2. A highly structured model for the spread of viruses along global transportation networks

The challenge

You are given

1. pairwise "air traffic" or "effective" distances y_n, $n = 1, \dots, N$, between N viral samples and
2. the evolutionary history of a specific viral strain.

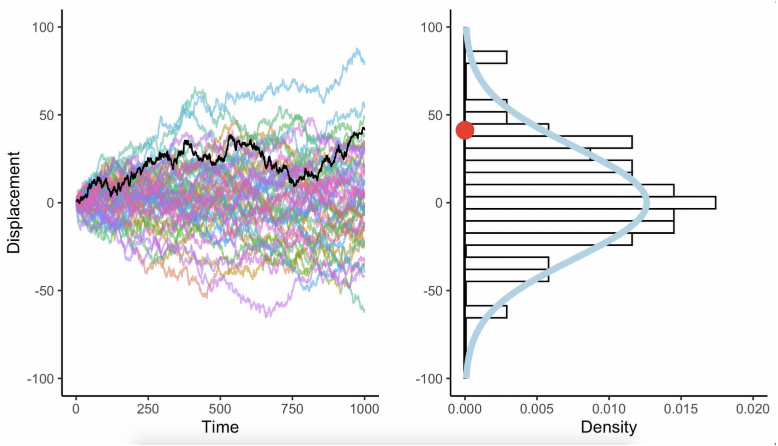




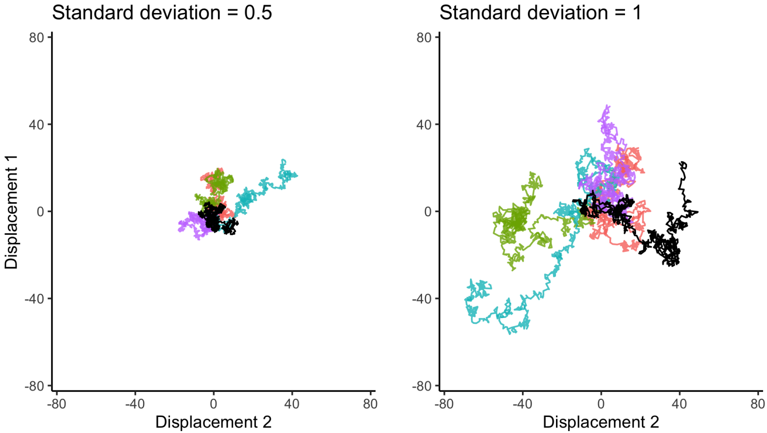
You are tasked to

1. use this data to discover how quickly the strain travels through global air traffic space and
2. quantify your uncertainty with respect to the relevant quantity.

Brownian motion



Brownian motion



Multidimensional Brownian diffusion

A standard Brownian motion w_t, $t > 0$ satisfies

(i) $w_0 = 0$

(ii) $(w_{t_4} - w_{t_3}) \perp (w_{t_2} - w_{t_1})$ for $t_1 < t_2 \leq t_3 < t_4$

(iii) $(w_{t_2} - w_{t_1}) \sim N(0, t_2 - t_1)$ for $t_2 > t_1$

(iv) w_t is continuous as a function of t

Stack independent $w_{d,t}$, $d = 1, \ldots, D$ and premultiply by infinitesimal rate matrix Σ to get general

$$\mathbf{x}_t = \sqrt{\Sigma}\, \mathbf{w}_t \,.$$

“Diffusivity”

In Fick’s laws of diffusion, the *diffusivity* or *diffusion coefficient* is proportional to the squared velocity of a diffusing particle (e.g. m^2/s).

For the stochastic differential equation

$$\mathrm{d}\mathbf{x}_t = \sqrt{\Sigma}\,\mathrm{d}\mathbf{w}_t$$

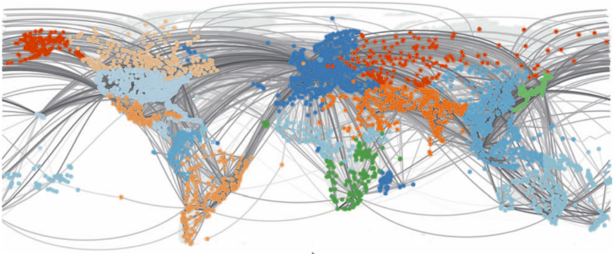
we use the identity

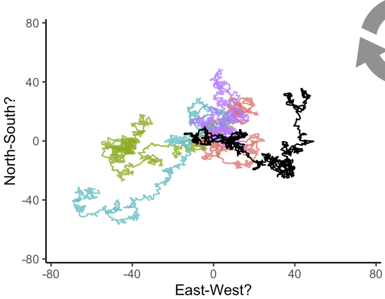
$$(\mathrm{d}w_t)^2 = \mathrm{d}t$$

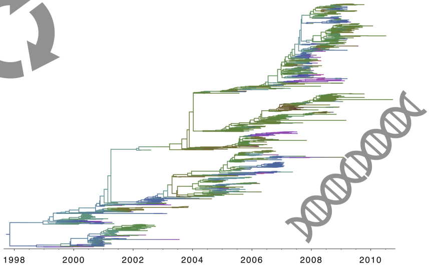
to get

$$\langle \mathrm{d}\mathbf{x}_t, \mathrm{d}\mathbf{x}_t \rangle = \mathrm{tr}\left(\Sigma\right)\,\mathrm{d}t \quad \text{or} \quad \mathrm{tr}\left(\Sigma\right) \text{ “} = \text{” } \frac{\langle \mathrm{d}\mathbf{x}_t, \mathrm{d}\mathbf{x}_t \rangle}{\mathrm{d}t}\,.$$

A tale of two networks







Bayesian multidimensional scaling

Let $\mathbf{Y}$ be an $N \times N$ distance matrix with elements $y_{nn'}$ the distance between objects n and n'. Oh and Raftery (2001) model

$$y_{nn'} \sim N\left(||\mathbf{x}_n - \mathbf{x}_{n'}||, \sigma^2\right)\ I(y_{nn'} > 0)$$

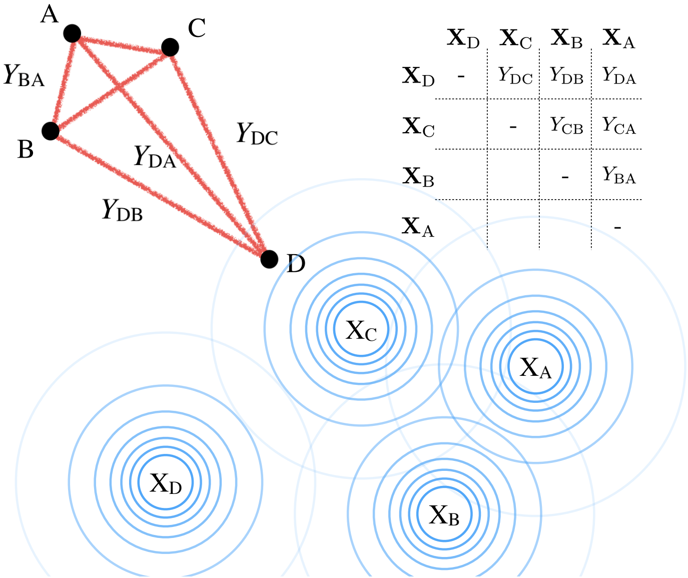
for random variables $\mathbf{x}_n, \mathbf{x}_{n'} \in \mathbb{R}^p$. Conditioned on latents $\mathbf{X} = (\mathbf{x}_1\,, \dots, \mathbf{x}_N)^T$, the *BMDS* likelihood is

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

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

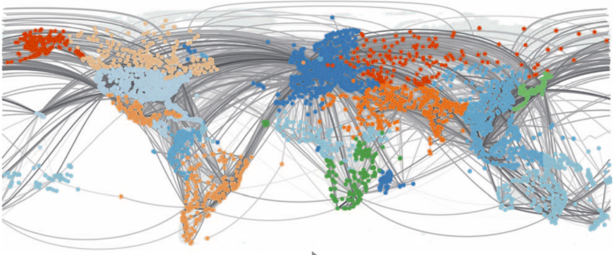
where $\delta_{nn'} = ||\mathbf{x}_n - \mathbf{x}_{n'}||$.

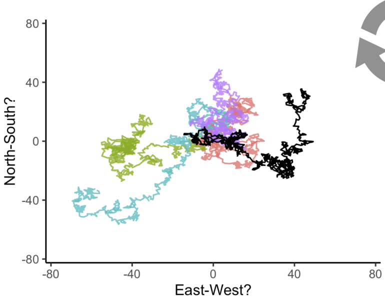
Bayesian multidimensional scaling

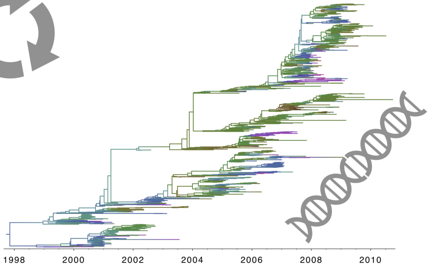

A
C
Y_{BA}
B
Y_{DA}
Y_{DC}
Y_{DB}
D
X_{C}
X_{A}
X_{D}
X_{B}

	$\mathbf{X}_{\mathrm{D}}$	$\mathbf{X}_{\mathrm{C}}$	$\mathbf{X}_{\mathrm{B}}$	$\mathbf{X}_{\mathrm{A}}$
$\mathbf{X}_{\mathrm{D}}$	-	Y_{DC}	Y_{DB}	Y_{DA}
$\mathbf{X}_{\mathrm{C}}$		-	Y_{CB}	Y_{CA}
$\mathbf{X}_{\mathrm{B}}$			-	Y_{BA}
$\mathbf{X}_{\mathrm{A}}$				-

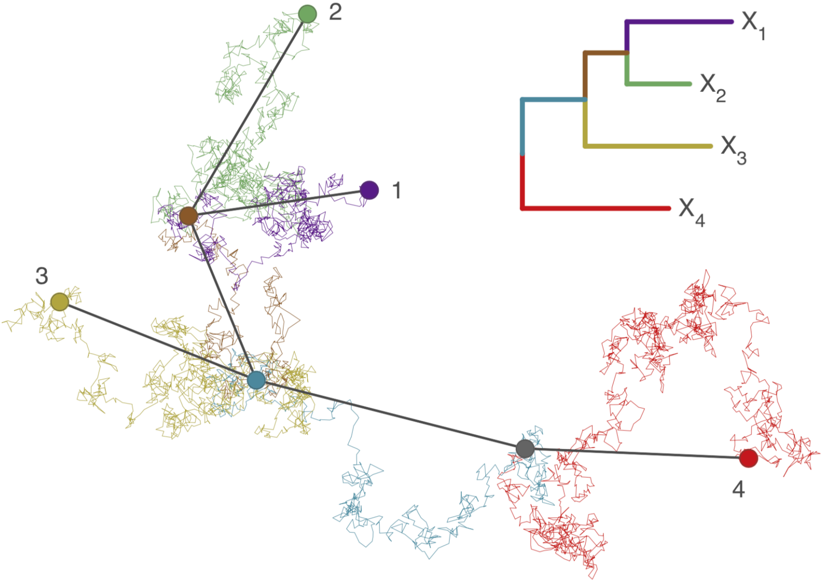
A tale of two networks







Brownian phylogenetic diffusion



Brownian phylogenetic diffusion

Associate to each tip n of a rooted, N-tipped, binary tree a Brownian motion $\mathbf{x}_n$, centered at its parent node $\mathbf{x}_{pa(n)}$. Then

$$\mathbf{x}_n | \mathbf{x}_{pa(n)} \sim N_p(\mathbf{x}_{pa(n)}, t_n \Sigma)\,,$$

for t_n the branch length of node n to its parent.

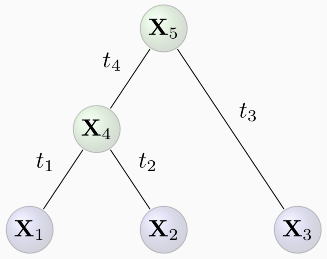
Write the joint distribution as

$$\mathbf{X} \sim MN_{N\times p}(\mathbf{0}, \mathbf{V}, \Sigma)$$

for

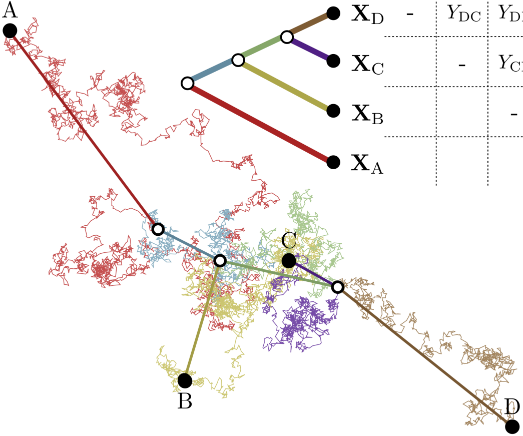
$$[\mathbf{V}]_n = t_n + t_{pa(n)} + t_{pa(pa(n))+\dots}$$

$$[\mathbf{V}]_{nn'} = \begin{cases} [\mathbf{V}]_n - t_n, & pa(n) = pa(n') \\ 0, & o/w \end{cases}$$

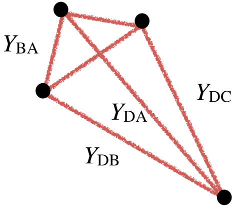


Zhang et al. 2019

Bayesian phylogenetic multidimensional scaling



	$\mathbf{X}_{\mathrm{D}}$	$\mathbf{X}_{\mathrm{C}}$	$\mathbf{X}_{\mathrm{B}}$	$\mathbf{X}_{\mathrm{A}}$
$\mathbf{X}_{\mathrm{D}}$	-	Y_{DC}	Y_{DB}	Y_{DA}
$\mathbf{X}_{\mathrm{C}}$		-	Y_{CB}	Y_{CA}
$\mathbf{X}_{\mathrm{B}}$			-	Y_{BA}
$\mathbf{X}_{\mathrm{A}}$				-



Part 3. Modern Bayesian inference

Likelihood based inference and Bayes

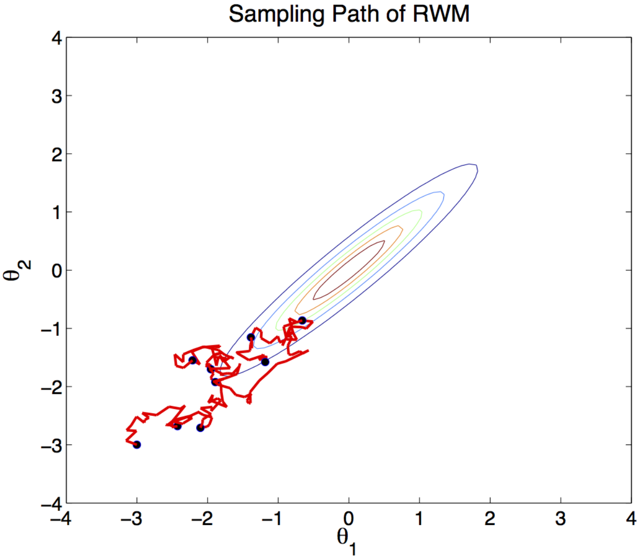
Assume data generated according to $\mathbf{y}_n \overset{\perp}{\sim} f(\mathbf{y}_n|\boldsymbol{\theta}, \mathbf{x}_n)$ with prior distributions $\boldsymbol{\theta} \sim p_\theta(\boldsymbol{\theta})$ and $(\mathbf{x}_1, \ldots, \mathbf{x}_N) = \mathbf{X} \sim p_x(\mathbf{X})$.

Bayes' theorem says:

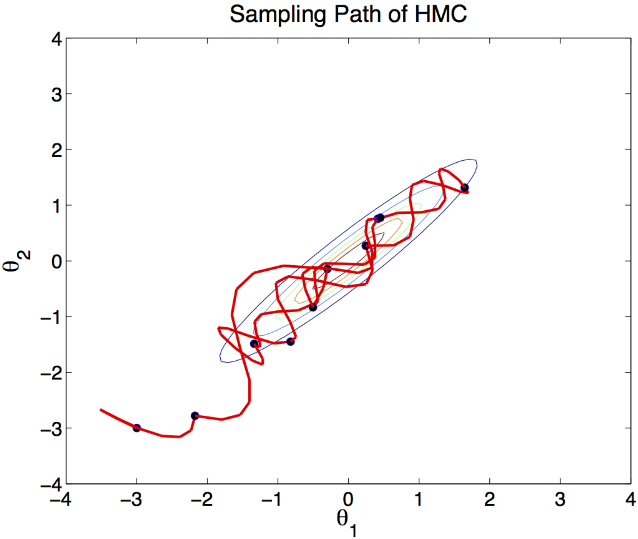
$$p(\boldsymbol{\theta}|\mathbf{Y}) = \frac{f(\mathbf{Y}|\boldsymbol{\theta})\, p_\theta(\boldsymbol{\theta})}{f(\mathbf{Y})} = \frac{\int_{\mathbf{X}} f(\mathbf{Y}|\mathbf{X}, \boldsymbol{\theta}) p_x(\mathbf{X}) \mathrm{d}\mathbf{X}\ p_\theta(\boldsymbol{\theta})}{\int_\Theta \left(\int_{\mathbf{X}} f(\mathbf{Y}|\mathbf{X}, \boldsymbol{\theta}) p_x(\mathbf{X}) \mathrm{d}\mathbf{X} \right) p_\theta(\boldsymbol{\theta}) \mathrm{d}\boldsymbol{\theta}}\,,$$

where $f(\mathbf{Y}|\boldsymbol{\theta}, \mathbf{X}) = \prod_n^N f(\mathbf{y}_n|\boldsymbol{\theta}, \mathbf{x}_n)$ is the *likelihood* function and $f(\mathbf{Y}|\boldsymbol{\theta})$ is the *marginal likelihood*.

Random walk Metropolis



Hamiltonian Monte Carlo



Hamiltonian Monte Carlo

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

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

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

$$\begin{aligned} \frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} &= \frac{\partial H}{\partial \mathbf{p}} = \mathbf{p} \\ \frac{\mathrm{d}\mathbf{p}}{\mathrm{d}t} &= -\frac{\partial H}{\partial \mathbf{x}} = \nabla \log \pi(\mathbf{x})\ . \end{aligned}$$

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

Hamiltonian Monte Carlo

Benefits. HMC computes high-dimensional integrals;
scales to 30,000+ parameters.

Challenges. HMC necessitates repeated computation of
log-likelihood and its gradient (best case $\mathcal{O}(N)$).

HMC for BMDS?

The BMDS likelihood scales $\mathcal{O}(N^2)$:

$$-\log p(\mathbf{Y}|\mathbf{X},\sigma^2) \propto \sum_{n>n'} \frac{(y_{nn'}-\delta_{nn'})^2}{2\sigma^2} + \log\Phi\left(\frac{\delta_{nn'}}{\sigma}\right) .$$

The gradient also scales $\mathcal{O}(N^2)$:

$$\begin{aligned}\frac{\partial}{\partial \mathbf{x}_n}\log p(\mathbf{Y}|\mathbf{X},\sigma^2) &= \frac{\partial}{\partial \delta_{nn'}}\log p(\mathbf{Y}|\mathbf{X},\sigma^2)\,\frac{\partial \delta_{nn'}}{\partial \mathbf{x}_n} \\ &= -\sum_{n'\neq n}\left(\frac{(\delta_{nn'}-y_{nn'})}{\sigma^2} + \frac{\phi(\delta_{nn'}/\sigma)}{\sigma\Phi(\delta_{nn'}/\sigma)}\right)\frac{(\mathbf{x}_n-\mathbf{x}_{n'})}{\delta_{nn'}} \\ &:= -\sum_{n'\neq n}\mathbf{r}_{nn'} .\end{aligned}$$

Recap

Goal: quantify and infer the diffusion rate of global contagion.

1. Brownian diffusion is a useful model (flexible/tractable).
   Brownian diffusion does not account for network structures.
2. Model adapts Brownian diffusion to network realities.
   Model inference is hard: integral dimension grows $\mathcal{O}(N)$.
3. HMC scales inference to tens of thousands of dimensions.
   HMC for model costs $\mathcal{O}(N^2)$.

HMC for BMDS?

The BMDS likelihood scales $\mathcal{O}(N^2)$:

$$-\log p(\mathbf{Y}|\mathbf{X},\sigma^2) \propto \sum_{n>n'} \frac{(y_{nn'}-\delta_{nn'})^2}{2\sigma^2} + \log \Phi\left(\frac{\delta_{nn'}}{\sigma}\right) .$$

The gradient also scales $\mathcal{O}(N^2)$:

$$\begin{aligned}\frac{\partial}{\partial \mathbf{x}_n} \log p(\mathbf{Y}|\mathbf{X},\sigma^2) &= \frac{\partial}{\partial \delta_{nn'}} \log p(\mathbf{Y}|\mathbf{X},\sigma^2) \frac{\partial \delta_{nn'}}{\partial \mathbf{x}_n} \\ &= -\sum_{n'\neq n} \left(\frac{(\delta_{nn'}-y_{nn'})}{\sigma^2} + \frac{\phi(\delta_{nn'}/\sigma)}{\sigma\Phi(\delta_{nn'}/\sigma)} \right) \frac{(\mathbf{x}_n - \mathbf{x}_{n'})}{\delta_{nn'}} \\ &:= -\sum_{n'\neq n} \mathbf{r}_{nn'} .\end{aligned}$$

Part 4. Massive parallelization

Parallelization methods

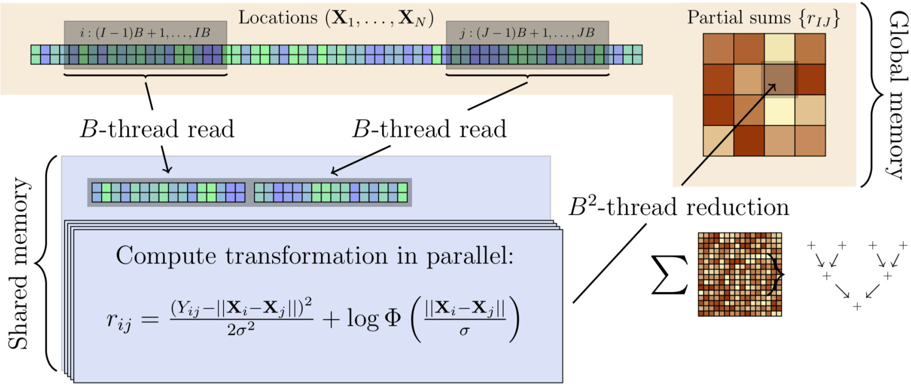
Central processing unit (CPU):

1. Global parallelization: 2 to 60 cores (multi-core)
2. Local parallelization: single instruction multiple data (SIMD)

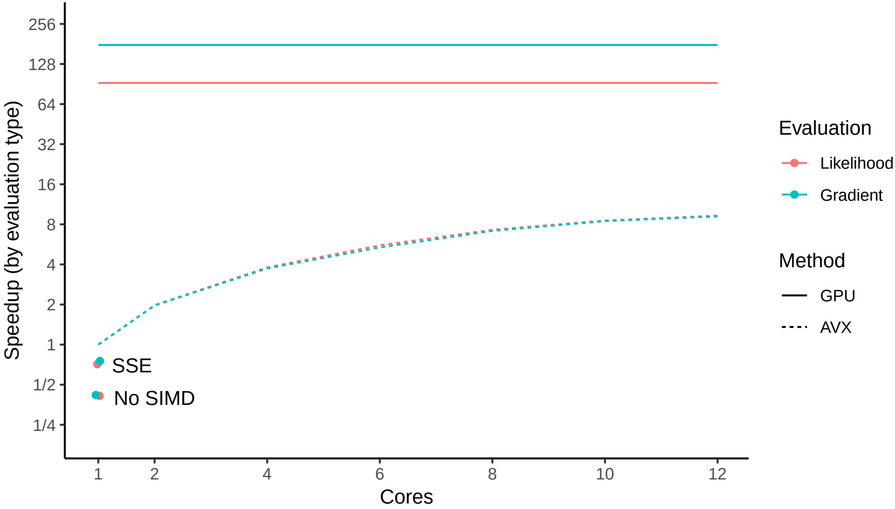
Graphics processing unit (GPU):

1. Thousands of cores (many-core)
2. Single instruction multiple threads (SIMT)
3. High memory bandwidth (not strictly maths anymore)

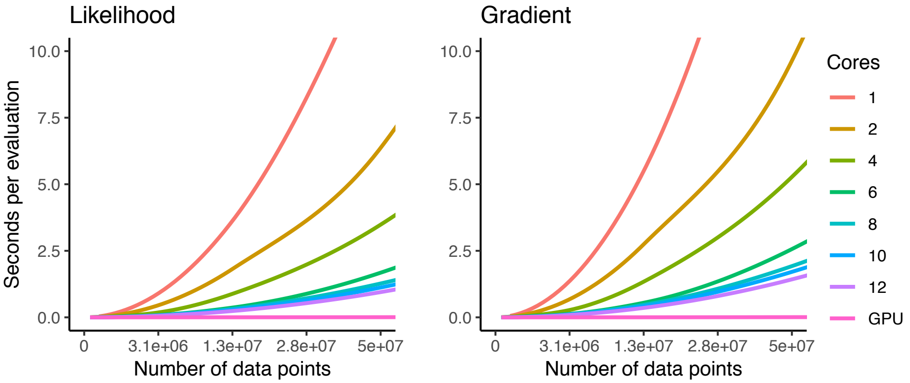
Exploiting likelihood parallelism


Locations $(\mathbf{X}_1, \dots, \mathbf{X}_N)$
$i : (I-1)B+1, \dots, IB$
$j : (J-1)B+1, \dots, JB$
Partial sums $\{r_{IJ}\}$
Global memory
B-thread read
B-thread read
B^2-thread reduction
Shared memory
Compute transformation in parallel:
$r_{ij} = \frac{(Y_{ij} - ||\mathbf{X}_i - \mathbf{X}_j||)^2}{2\sigma^2} + \log \Phi \left(\frac{||\mathbf{X}_i - \mathbf{X}_j||}{\sigma} \right)$
$\sum$

Significant speedups



Significant speedups



Open-source software

1. MassiveMDS: C++ library and R package;
   `https://github.com/suchard-group/MassiveMDS`

2. RcppXsimd: R wrapper package for Xsimd;
   `https://cran.r-project.org/web/packages/RcppXsimd`

Part 5. Global spread of influenza

Influenza data

Data consist of spatial locations and RNA sequences of 5,392 viral samples from 189 countries between 2001 and 2013.

Influenza type A:

1. H1N1: 1,370
2. H3N2: 1,389

Influenza type B:

1. Victoria: 1,393
2. Yamagata: 1,240

We convert locations data into

$$\binom{5,392}{2} = 14,534,136$$

pairwise "air traffic" distances.

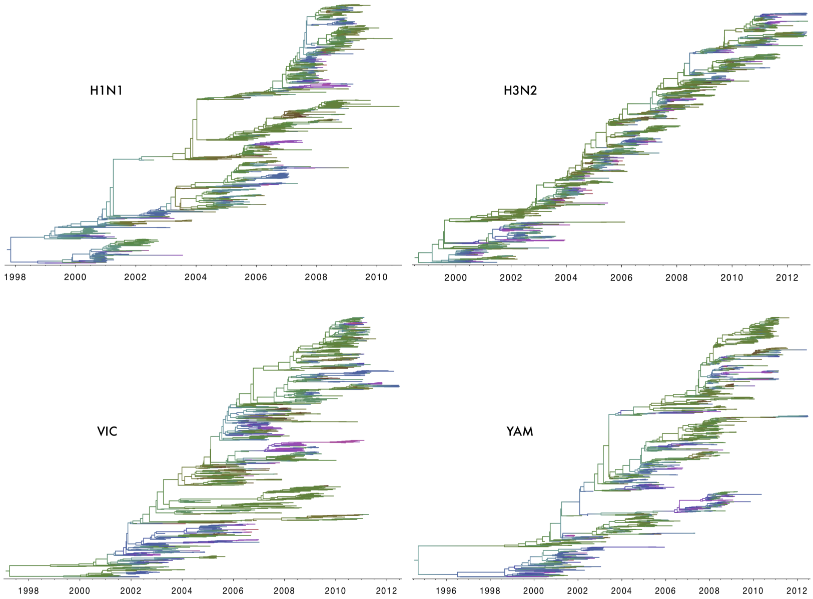
Model selection and inference

5-fold cross validation with log pointwise predictive density ($\widehat{lpd}$) from 10,000 MCMC samples dictates dimension count of 6.

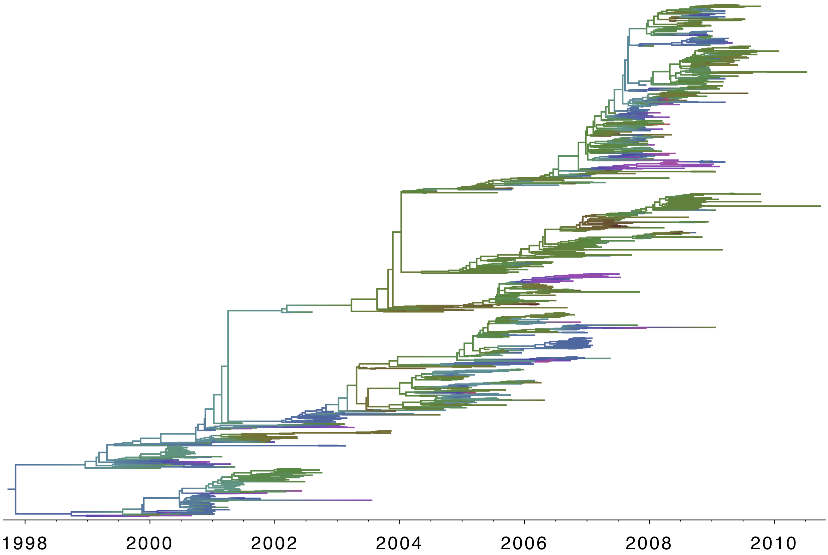
Dimension	2	3	4	5	6	7
-$\widehat{lpd}$ ($\times 10^6$)	7.1	4.2	3.4	3.5	2.8	7.0

We then use HMC-within-Gibbs to generate 2 million states for all $\mathbf{X}_{N\times 6}$ and strain-specific Σ and $\mathcal{T}$.

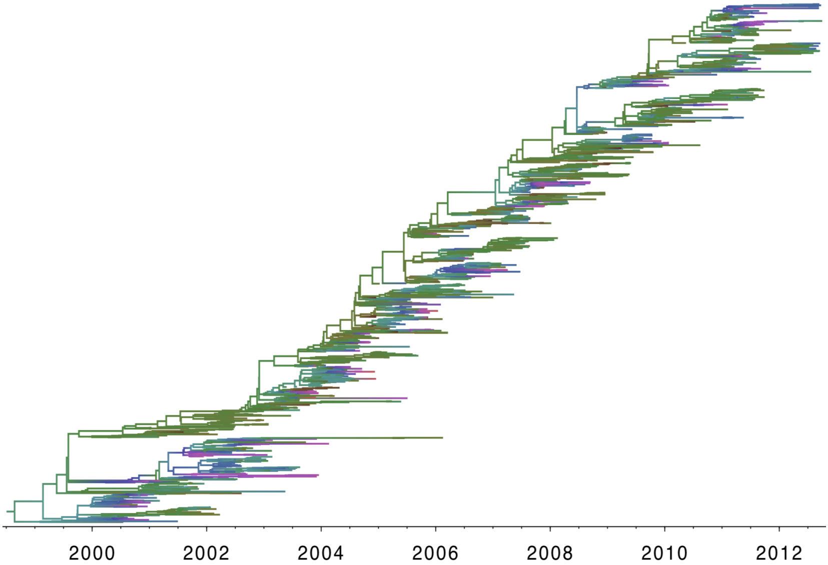
Learning phylogenies



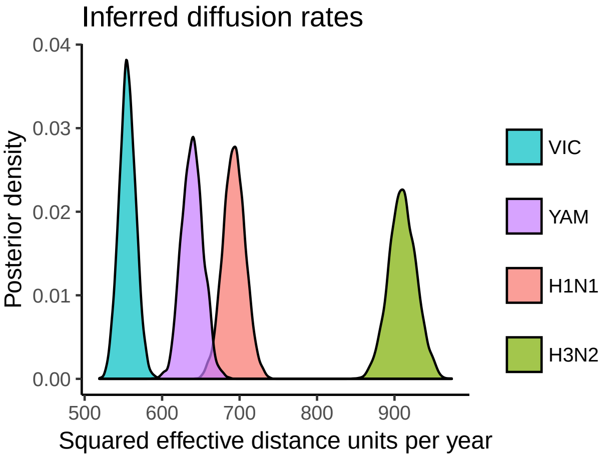
H1N1



H3N2

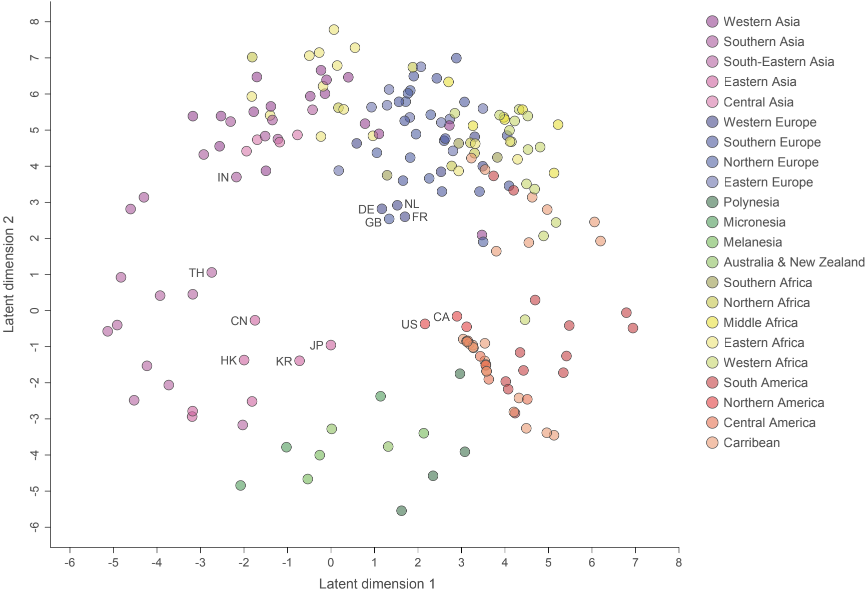


Learning diffusivities

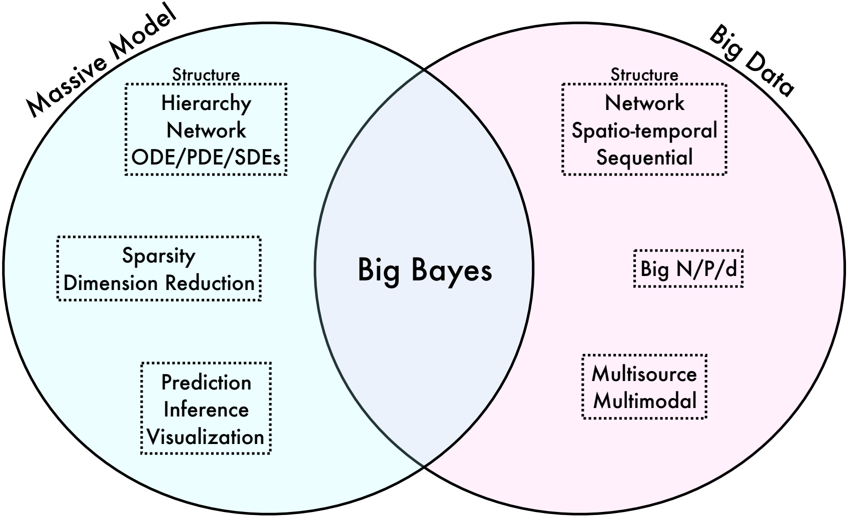


Recall *diffusivity* takes units of squared distance over time. We take diffusivity $\mathrm{tr}(\Sigma)$ as object of parametric interest.

Worldwide air traffic space



Part 6. Big Bayes


Massive Model
Big Data
Structure
Hierarchy
Network
ODE/PDE/SDEs
Sparsity
Dimension Reduction
Prediction
Inference
Visualization
Big Bayes
Structure
Network
Spatio-temporal
Sequential
Big N/P/d
Multisource
Multimodal

Applications in neural decoding

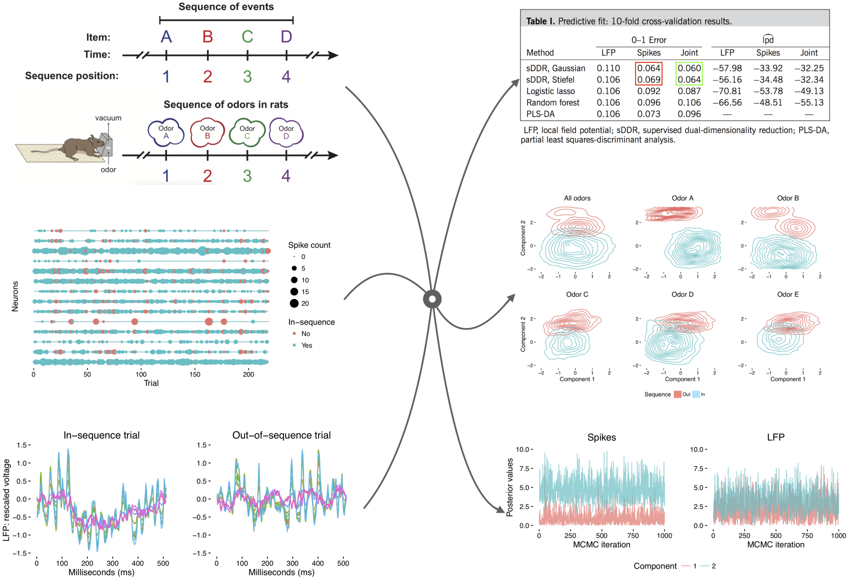


Table I. Predictive fit: 10-fold cross-validation results.

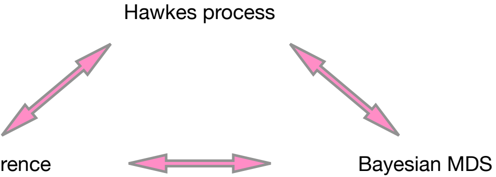
	0–1 Error			$\widehat{\text{lpd}}$		
Method	LFP	Spikes	Joint	LFP	Spikes	Joint
sDDR, Gaussian	0.110	0.064	0.060	−57.98	−33.92	−32.25
sDDR, Stiefel	0.106	0.069	0.064	−56.16	−34.48	−32.34
Logistic lasso	0.106	0.092	0.087	−70.81	−53.78	−49.13
Random forest	0.106	0.096	0.106	−66.56	−48.51	−55.13
PLS-DA	0.106	0.073	0.096	—	—	—

LFP, local field potential; sDDR, supervised dual-dimensionality reduction; PLS-DA, partial least squares-discriminant analysis.

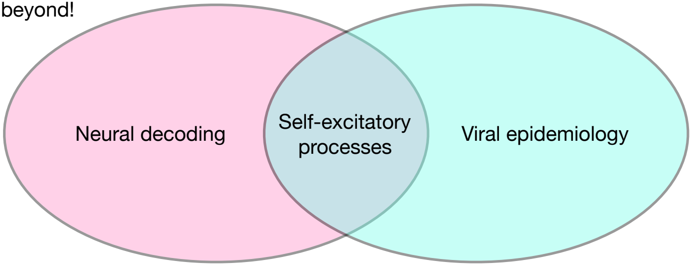
Holbrook et al. (2017); Holbrook et al. (2018); Lan, Holbrook et al. (2019)

Future work I, modeling

1. Predictive phylogenetics
   (K25 under review)



2. And beyond!



Future work II, scalability

1. Sampling developments [computational statistics] (Holbrook et al. 2018)
2. Massive parallelization [statistical computing] (Holbrook et al. 2019b)
3. Neural network aided MCMC [just plain fun] (Li, Holbrook et al. 2019)

Thank you!

Publications I

1. **Holbrook A**, Lemey P, Baele G, Dellicour S, Brockmann D, Rambaut A, Suchard M. *Massive parallelization boosts big Bayesian multidimensional scaling*. Submitted to Journal of Computational and Graphical Statistics, 2019.
2. Ji X, Zhang Z, **Holbrook A**, Nishimura A, Baele G, Rambaut A, Lemey P, Suchard M. *Gradients do grow on trees: a linear-time O(N)-dimensional gradient for statistical phylogenetics*. Submitted to Annals of Applied Statistics, 2019.
3. **Holbrook A**, Tustison N, Marquez F, Roberts J, Yassa M, Gillen D. *Anterolateral entorhinal cortex thickness as a biomarker for early detection of Alzheimer's disease*. Submitted to Alzheimer's and Dementia: The Journal of the Alzheimer's Association, 2019.
4. Lan S, **Holbrook A**, Elias G, Fortin N, Ombao H, Shahbaba B. *Flexible Bayesian Dynamic Modeling of Correlation and Covariance Matrices*. Bayesian Analysis, In Press, 2019.
5. **Holbrook A**, Lumley T, Gillen D. *Estimating prediction error for complex samples*. Canadian Journal of Statistics, In Press, 2019.
6. Tustison N, **Holbrook A**, Avants B, Roberts J, Cook P, Reagh Z, Stone J, Gillen D, Yassa M. *Longitudinal mapping of cortical thickness measurements: an Alzheimer's Disease Neuroimaging Initiative-based evaluation study*. Journal of Alzheimer's Disease, vol. 71, no. 1, pp. 165-183, 2019.

Publications II

7. Li L, **Holbrook A**, Shahbaba B, Baldi P. *Neural network gradient Hamiltonian Monte Carlo*. Computational Statistics, vol. 34, no. 1, pp. 281-299, 2019.
8. **Holbrook A**. *Differentiating the pseudo determinant*. Linear Algebra and its Applications, vol. 548, pp. 293-304, 2018.
9. **Holbrook A**, Lan S, Vandenberg-Rodes A, Shahbaba B. *Geodesic Lagrangian Monte Carlo over the space of positive definite matrices: with application to Bayesian spectral density estimation.* Journal of Statistical Computation and Simulation, vol. 88, no. 5, pp. 982-1002, 2018.
10. **Holbrook A**, Vandenberg-Rodes A, Fortin N, Shahbaba B. *A Bayesian supervised dual-dimensionality reduction model for simultaneous decoding of LFP and spike train signals.* Stat Journal, vol. 6, no. 1, pp. 53-67, 2017.
11. Grill J, **Holbrook A**, Pierce A, Hoang D, Gillen D. *Attitudes toward Potential Participant Registries.* Journal of Alzheimer's Disease, vol. 56, no. 3, pp. 939-946, 2017.