Spatial-Temporal Hawkes Models of Infectious Disease Transmission with Covariates

Oral Exam

Sean Mulherin

February 9, 2026

UCLA, Department of Statistics and Data Science

Table of contents

Background & Motivation

Public Motivation

Modeling infectious disease transmission is important for the purposes of:

- Forecasting the spread of infections
- Anticipating hospital demand
- Determining public health policy

Statistical Motivation

Compartment models (e.g., SEIR) are commonly used to model infections, but not without drawbacks.

- Restrictive assumptions lead to poor fit to real data [1, 2, 3].
- Neglect pertinent information such as: infectivity behavior dynamics, proximity, mobility, viral strains, etc.

Multivariate Hawkes process (MHP) models have been shown to offer a substantially superior fit [1].

- Suitable for explaining **self-exciting events**: causally-linked events wherein the occurrence increases the likelihood of subsequent events.
- Impactful examples:
  - Earthquakes [4]
  - Crime [5, 6]
  - Finance [7]
  - Social media [8]

Basic Temporal Hawkes Model Demonstration

$$\lambda(t) = \mu + K \sum_{s<t} \phi(\cdot)$$

where $\mu \geq 0$ represents the exogenous rate and $K \geq 0$ describes productivity [9]. In the example below, an exponential decay function is used as the triggering kernel, $\phi(\beta; t; s) = e^{-\beta(t-s)}$.

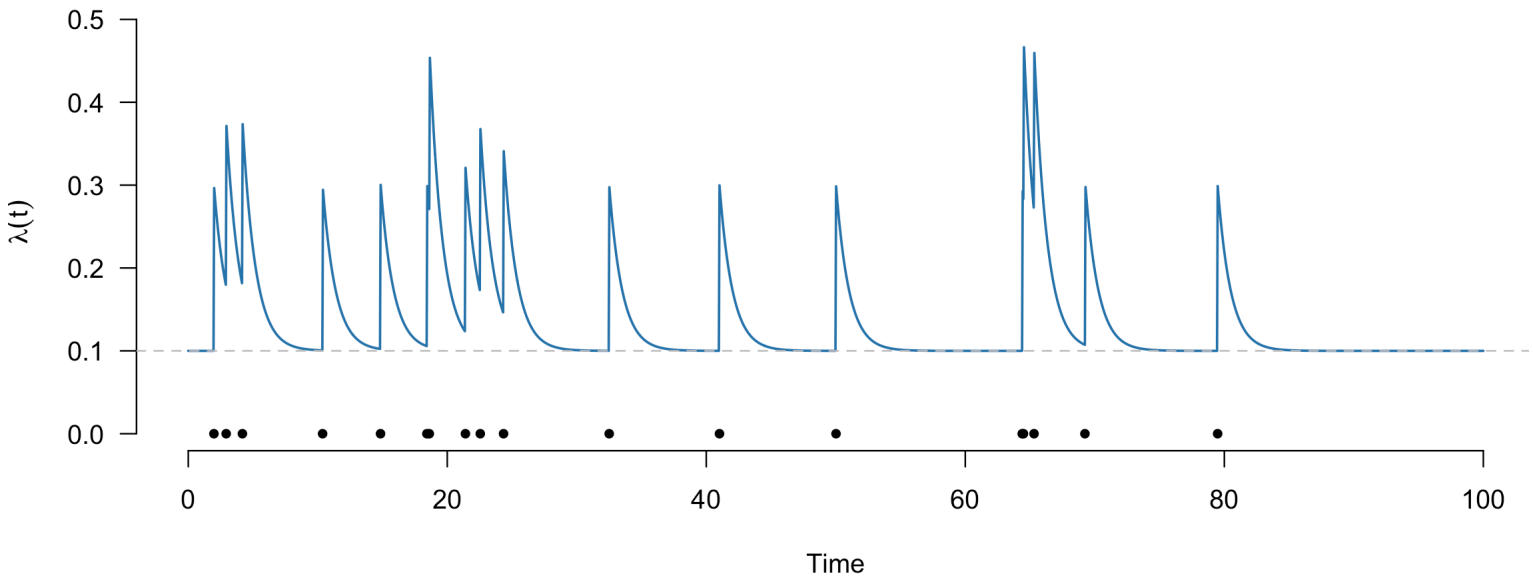


Figure 1: Simple Hawkes model demonstration, with $K = 0.2,\ \beta = 1.0,\ \mu = 0.1$.

Basic Spatial-Temporal Hawkes Model Demonstration

$$\lambda(t, x, y) = \mu + K \sum_{s<t} f(t-s)\, g(x - x_i, y - y_i). \tag{1}$$

where $f(\cdot)$ is the **temporal** triggering kernel and $g(\cdot)$ is the **spatial** triggering kernel.

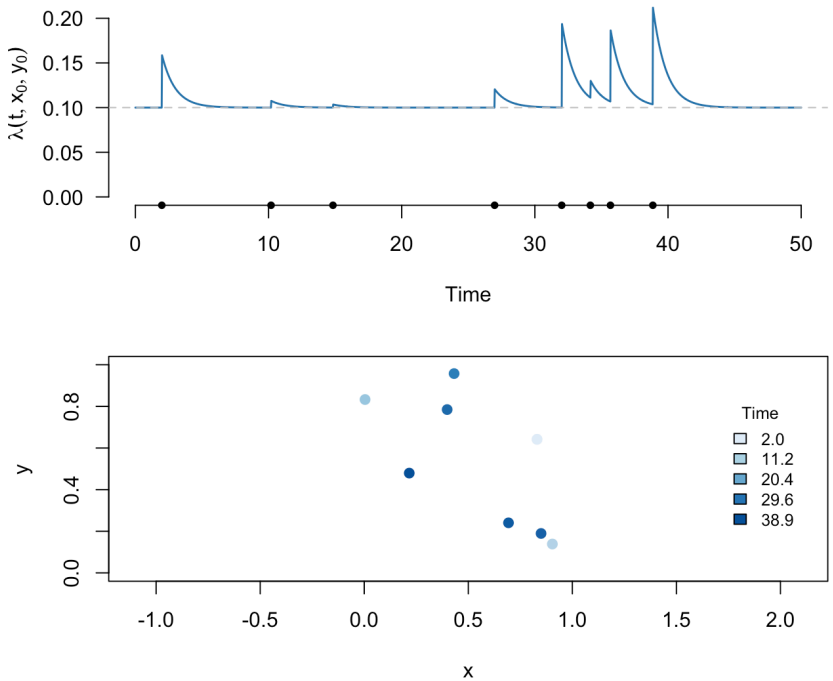
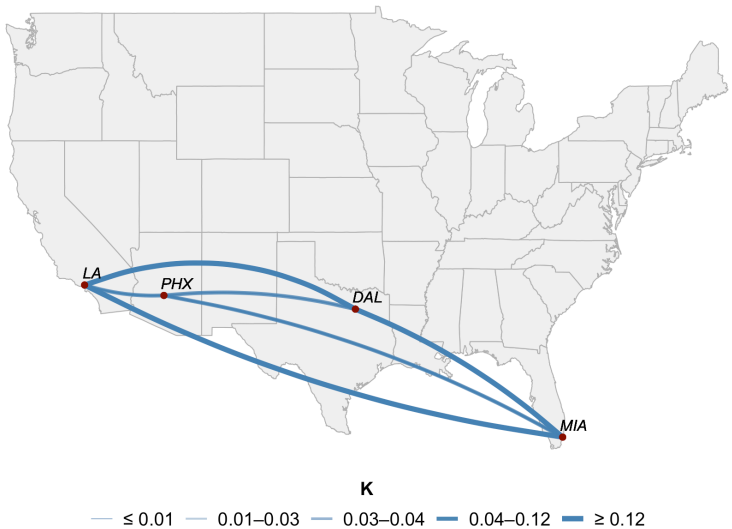


Figure 2: Simple spatial-temporal MHP model demonstration, with $\mu = 0.1,\ K = 0.2,\ \beta = 1.0$. $f(\cdot) = e^{-\beta(t-s)}$, $g(\cdot) = e^{-((x-x_i)^2+(y-y_i)^2)/2\sigma^2}$

Conventional MHP Problem

Fitting a purely statistical MHP model exclusively to incidence data leads to **inflated** cross-productivity estimates, particularly among locations that are far apart [10, 11].



0.6571	0.0702	0.7581	0.3617	LA
0.0097	1.1489	0.0000	0.0000	PHX
0.0633	0.1184	0.4377	0.0000	DAL
0.0044	0.0608	0.1380	0.7950	MIA
LA	PHX	DAL	MIA	

Figure 3: Estimated cross-productivities ($\hat{K}$) under the joint estimation framework. Fitted on COVID-19 data from March 12–July 20, 2020

Conventional Hawkes Problem Cont.

Large cross-productivity estimates can result when distant locations have similar temporal patterns as a product of global factors such as:

- Lockdown procedures
- Reporting behaviors
- Evolution of new viral strains

This leads the model to overestimate the rate of contagion between locations [10].

Conventional MHP models make **no use** of the distance/mobility between locations $\rightarrow$ **unaware** of the increasingly improbable contagion imposed by geographic barriers and/or low human mobility rates.

Research Goal

1. Develop a **covariate-informed MHP framework** to mitigate bias in estimated cross-productivities by incorporating a covariate-weighted penalty term w_{ij}, governed by:

   i. Distance-based covariate D_{ij}

   ii. Mobility-based covariate M_{ij}

2. Conduct a **systematic comparison** to test if distance- and mobility-based penalties reduce bias in cross-excitation estimates more effectively than existing regularization methods, in terms of **predictive performance** and **parameter recovery**.

Literature Review

This research is an extension of a paper by Diouane, Schoenberg, and Mohler [10] wherein various MHP modeling methods were compared to estimate cross-excitation with no use of covariates.

- Proposed a **sequential technique**:
  - **Stage 1:** purely temporal (i.e. local) models are independently fitted to each location and their parameters fixed.
  - **Stage 2:** inter-county models are fitted to residuals from **Stage 1** to obtain more accurate cross-excitation parameter estimates, indicating inter-county transmission.

Novel distinction: human mobility and spatial distance are used to penalize cross-productivity rates.

Preliminary Research

Data - COVID-19 in the United States

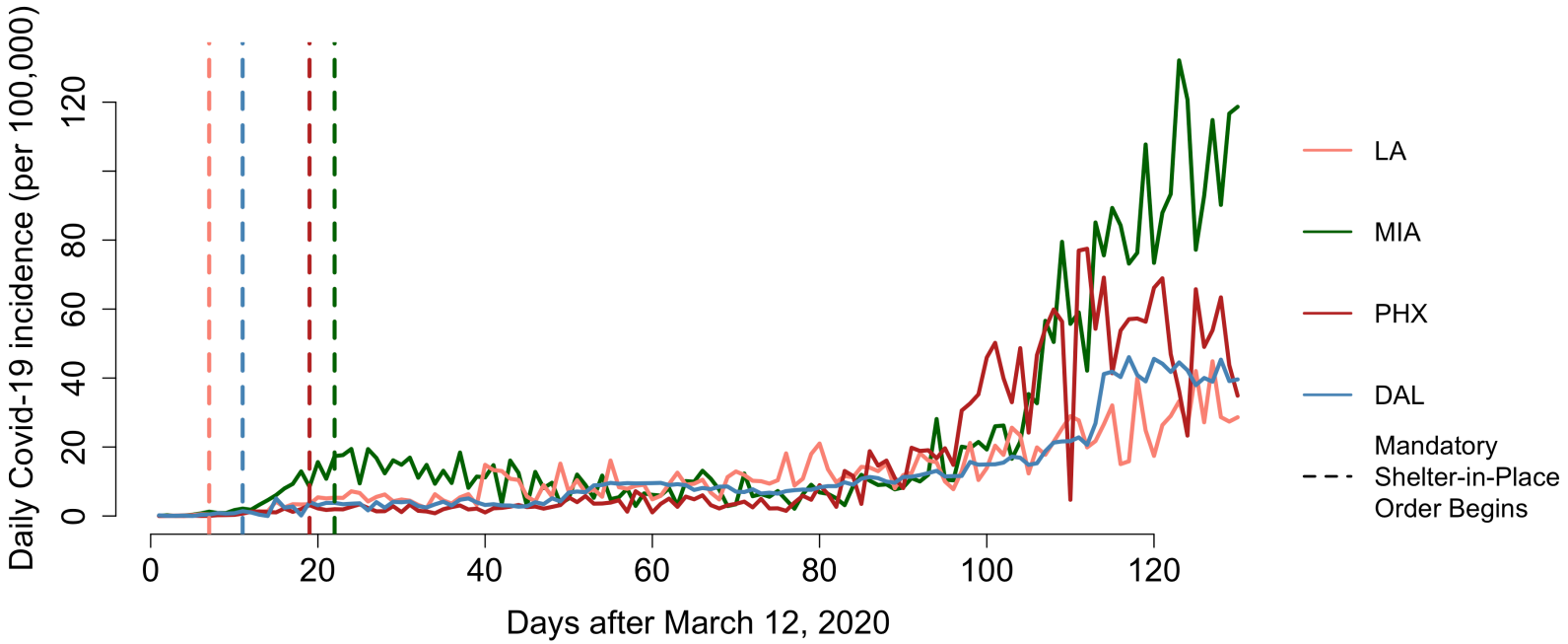


Figure 4: COVID-19 data were obtained from the NY Times Covid-19 GitHub repository, spanning 130 days from March 12–July 20, 2020 [12]. Los Angeles (LA), Miami-Dade (MIA), Maricopa (PHX), and Dallas (DAL) over 130 days starting after March 12, 2020. Data on mandatory shelter-in-place orders is obtained from U.S. Department of Health & Human Services [13].

Data - Distance Covariate

Locations are sourced from the R `maps` package. The Haversine formula is used to calculate the great-circle distance between locations:

$$D = 2 \cdot R \cdot \arcsin \sqrt{\sin^2\left(\frac{x_2 - x_1}{2}\right) + \cos x_1 \cdot \cos x_2 \cdot \sin^2\left(\frac{y_2 - y_1}{2}\right)}, \qquad (2)$$

where D denotes distance, $R = 6378$ km describes Earth's radius, and (x_i, y_i) are the latitude and longitude of location i [14].

	LA	PHX	DAL	MIA
LA	0	589	2011	3776
PHX	589	0	1426	3186
DAL	2011	1426	0	1784
MIA	3776	3186	1784	0

Table 1: Pairwise distances between counties, measured in kilometers.

Data - Mobility Covariate

Cellphone location data are obtained from Kang et al. [15] are used to estimate the proportion of travelers going from location to location, calculated as follows:

Mobility from i to j is defined here as the proportion of travelers who leave i and arrive in j, $M_{ij} = n_{ij} / \sum_j n_{ij}$, where n_{ij} denotes the number of travelers from i to j during March–July 2020.

Data cover approximately 10% of the U.S. population, derived from anonymized cellphone location tracking, thereby accounting for all modes of transportation.

	LA	PHX	DAL	MIA
LA	0	0.82	0.10	0.08
PHX	0.86	0	0.08	0.06
DAL	0.47	0.32	0	0.21
MIA	0.51	0.22	0.27	0

Table 2: Pairwise mobility rates. Columns indicate origin while rows indicate destination.

Staged Hawkes Model

For each location $i = 1, \ldots, Q$ on day $t = 1, \ldots, T$, incidence is modeled with a discrete-time MHP,

$$\lambda_i(t|\mathcal{H}_{t-}) = \mu_i + \underbrace{K_{ii} \sum_{s<t} \phi_{\text{local}}(t, s; \beta_i, p_i) N_i(s)}_{\text{Stage 1 (local)}} + \underbrace{\sum_{j \neq i} K_{ij} \sum_{s<t} \phi_{\text{cross}}(t, s; \bar{\beta}_i, \bar{p}_i) N_j(s)}_{\text{Stage 2 (foreign)}} \tag{3}$$

- $\mathcal{H}_{t-}$ is the event history up to time t.
- $N_i(s) \in \{0, 1, 2, \ldots\}$ is the recorded case count on day s
- $\phi_{\text{local}}(\cdot)$ and $\phi_{\text{cross}}(\cdot)$ are discretized Weibull triggering kernels given by

$$\phi(t, s; \beta, p) = (1-p)^{(t-s-1)^\beta} - (1-p)^{(t-s)^\beta}, 0 < p < 1, \beta > 0 \tag{4}$$

Estimate 1 - Joint Baseline

We first consider completely unconstrained parameters where $\theta = \{\mu_i, K_{ij}, \beta_i, p_i\}$, for $i, j = 1, 2, 3, 4$ is estimated jointly by MLE with no regularization and no constraints on the parameters other than $K_{ij} \geq 0, 0 < p < 1, \beta > 0$, and the subcriticality condition that specifies $\rho(K) < 1$.

$$\ell(\theta) = \sum_{i=1}^{Q} \sum_{t=1}^{T} \log\left(\frac{\lambda_i(t;\theta)^{N_i(t)}}{N_i(t)!} e^{-\lambda_i(t;\theta)} \right) \tag{5}$$

Such estimates tend to vastly overestimate cross-excitations, especially among distant locations [10].

Estimate 2 - Purely Univariate (Stage 1)

Here, we set $K_{ij} = 0$ for $i \neq j$ to force estimates to represent **purely local transmission**.

Local parameter estimates $\theta_{\text{local}} = \{\mu_i, K_{ii}, \beta_i, p_i\}_{i=1}^{Q}$ are obtained by MLE with elastic-net regularization on parameters K and β.

$$\max_{\theta_{\text{local}}} \ell_1(\theta_{\text{local}}) - \underbrace{\left(\frac{\alpha_1}{2} \sum_{i=1}^{Q} \left[\eta_1 K_{ii}^2 + \eta_2 \beta_i^2 \right] + (1 - \alpha_1) \sum_{i=1}^{Q} \left[\eta_1 |K_{ii}| + \eta_2 |\beta_i| \right] \right)}_{\text{Elastic-Net Regularization}} \quad (6)$$

Estimate 3 - Staged with no Covariate Constraints

Two-stage procedure: local parameters $\hat{\theta}_{\text{local}}$ are estimated in **Stage 1** and held fixed, while cross-excitation parameters are estimated in **Stage 2**.

This prioritizes the local component:

$\rightarrow$ Most incidence is attributed locally.

$\rightarrow$ Reserves K_{ij} to capture only the residual clustering not explained by the local component.

Hence, with $\hat{\theta}_{\text{local}}$ fixed, cross-productivities $K_{i \neq j}$ are estimated by

$$\max_{K_{\text{cross}}} \ \ell_2(K_{\text{cross}} | \hat{\theta}_{\text{local}}) - \eta_3 \left[\frac{\alpha_2}{2} \sum_{i=1}^{Q} \sum_{j \neq i} K_{ij}^2 + (1 - \alpha_2) \sum_{i=1}^{Q} \sum_{j \neq i} |K_{ij}| \right] \tag{7}$$

Estimate 4 - Staged with Covariate Constraints

Here, we modulate $K_{i \neq j}$ by a covariate-derived weight w_{ij} determined by (i) spatial distance or (ii) human mobility.

$$\max_{K_{\text{cross}}} \ \ell_2(K_{\text{cross}} | \hat{\theta}_{\text{local}}) - \eta_3 \left[\frac{\alpha_2}{2} \sum_{i=1}^{Q} \sum_{j \neq i} w_{ij} K_{ij}^2 + (1 - \alpha_2) \sum_{i=1}^{Q} \sum_{j \neq i} w_{ij} |K_{ij}| \right]. \quad (8)$$

In (i) $w_{ij} \propto D_{ij}$ according to $w_{ij} = \exp(\delta D_{ij}^*)$ where D_{ij}^* is the standardized distance matrix, $D_{ij}^* = D_{ij} / \max_{i \neq j} D_{ij}$.

In (ii) $w_{ij} \propto 1/M_{ij}$ according to $w_{ij} = \exp(\gamma(1 - M_{ij}))$.

Both (i) and (ii) are structured in such a way as to shrink implausible long-range connections while preserving short-range transmission.

Optimization

Elastic-net mixing parameters (α_1, α_2), strength parameters (η_1, η_2, η_3), and covariate scaling parameters (γ, δ) are tuned via grid search on a 104-day (80%) training window (March 12 - June 24) with a 26-day (20%) testing window (June 25–July 20).

Each stage is initialized with the **Genetic Algorithm** (GA) [16] followed by `optim` to refine the solution and compute standard errors.

Simulation

Conducted a **Monte Carlo simulation** study for each covariate regime: (i) distance and (ii) mobility. For each, 100 independent datasets were generated to report mean parameter estimates and evaluation metrics.

Covariate matrices **D** and **M** are used to initiate the ground-truth system, allowing us to test the degree to which the model captures the effects of cross-excitation caused by trends in distance and mobility.

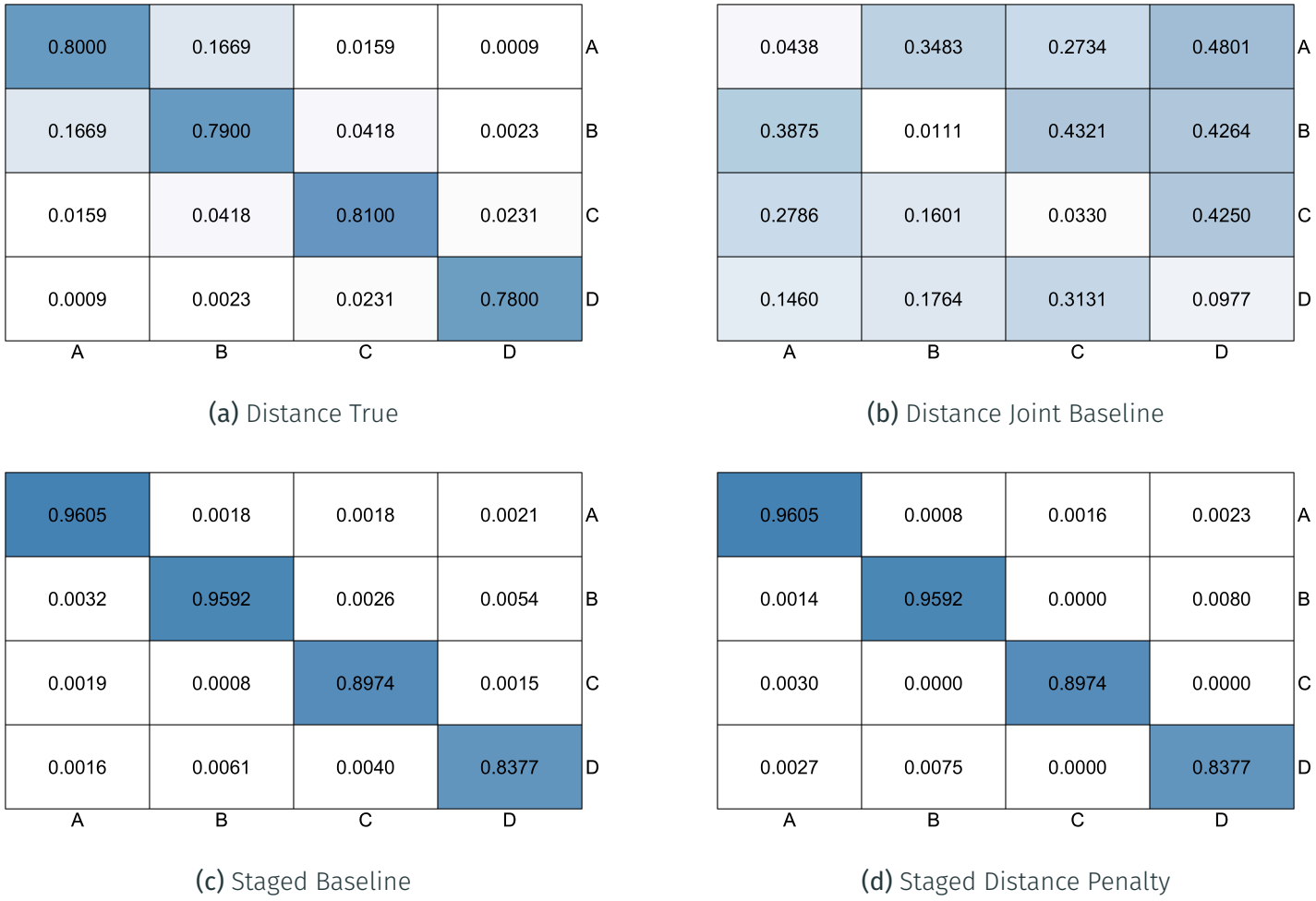
Focus is on recovery of reproduction matrix K and accuracy of out-of-sample predictions.

Simulation Results - Distance

Estimate	NLL	RMSE	$K_{i=j}$	$K_{i\neq j}$
True			0.795	0.042
Joint Baseline	6575.6 (389.2) [5812.7, 7338.6]	69.2 (2.8) [63.6, 74.7]	0.046 (0.009) [0.029, 0.064]	0.321 (0.006) [0.310, 0.333]
Univariate	1802.4 (30.9) [1741.9, 1862.9]	17.5 (0.6) [16.3, 18.8]	0.914 (0.005) [0.903, 0.924]	0.000 (0.000) [0.000, 0.000]
Staged Baseline	**1781.9** (17.0) [1748.6, 1815.3]	**17.3** (0.4) [16.5, 18.1]	0.914 (0.005) [0.903, 0.924]	0.003 (0.000) [0.002, 0.003]
Staged Distance Penalty	**1781.9** (17.0) [1748.6, 1815.1]	**17.3** (0.4) [16.5, 18.1]	0.914 (0.005) [0.903, 0.924]	0.002 (0.000) [0.002, 0.003]

Table 3: Distance setting. Pooled negative log-likelihood (NLL), pooled root mean squared error (RMSE), fitted diagonal (local) productivity ($\hat{K}_{i=j}$) and off-diagonal (cross) productivity ($\hat{K}_{i\neq j}$) across four estimation methods (mean with standard errors; 95% confidence intervals in brackets). Results from 100 simulations; true productivities shown for reference.

Simulation Results



(a) Distance True

(b) Distance Joint Baseline

(c) Staged Baseline

(d) Staged Distance Penalty

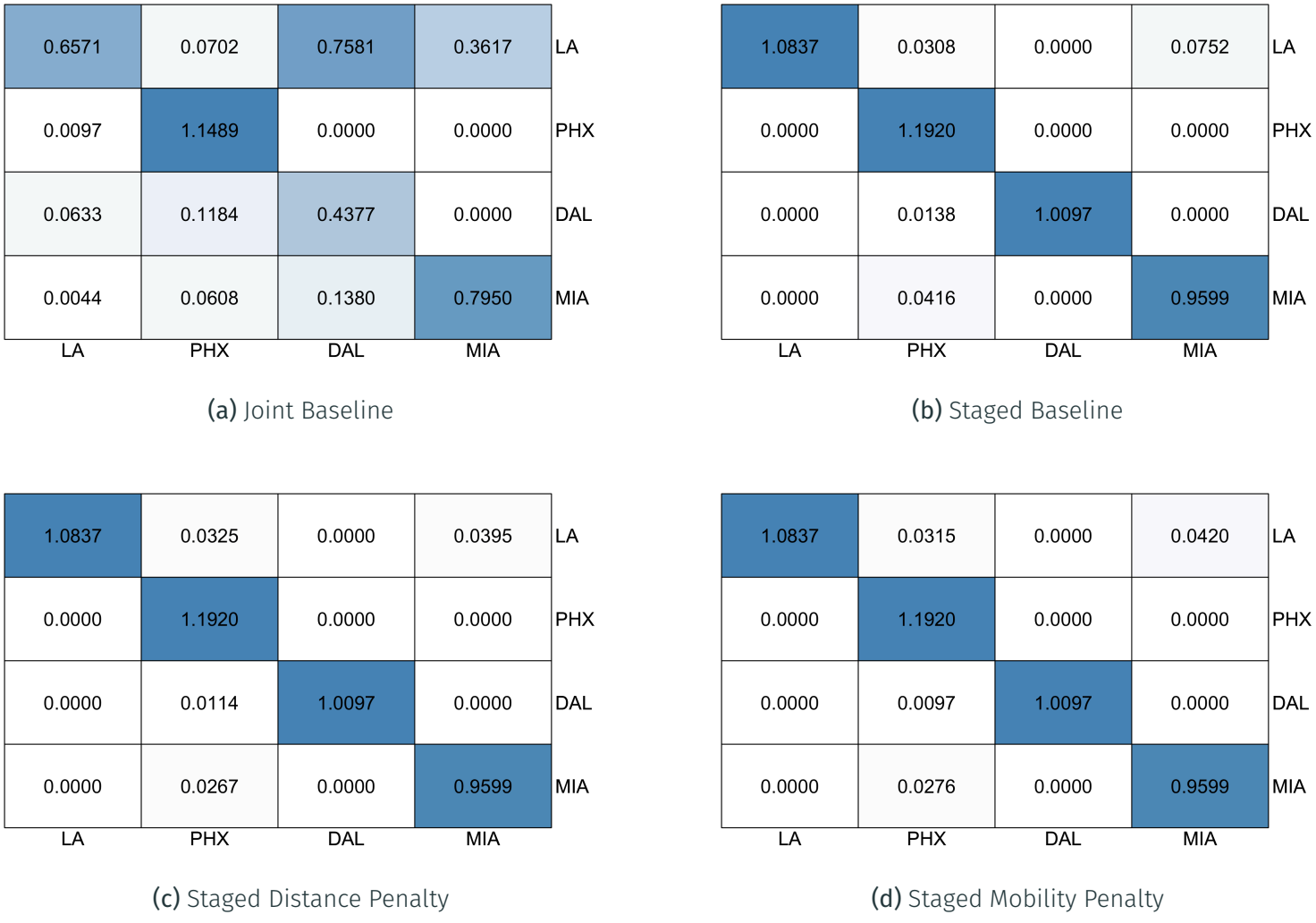
Figure 5: Fitted productivity estimates for the distance-based simulation under three penalty regimes (i.e. joint baseline, staged baseline, staged penalty). True parameter values are included for comparison.

Application to COVID-19 - Local Parameter Recovery

Parameter	LA	PHX	DAL	MIA
μ	49.06 (2.856) [43.47, 54.66]	1.98 (0.576) [0.85, 3.11]	8.09 (1.042) [6.05, 10.14]	20.57 (1.874) [16.90, 24.25]
K	1.08 (0.009) [1.07, 1.10]	1.19 (0.010) [1.17, 1.21]	1.01 (0.010) [0.99, 1.03]	0.96 (0.009) [0.94, 0.98]
β	0.61 (0.019) [0.57, 0.65]	1.18 (0.042) [1.10, 1.26]	1.30 (0.096) [1.11, 1.49]	1.63 (0.074) [1.48, 1.77]
p	0.33 (0.009) [0.31, 0.34]	0.27 (0.013) [0.25, 0.30]	0.30 (0.030) [0.24, 0.36]	0.22 (0.017) [0.19, 0.25]

Table 4: Fitted Stage 1 local parameters with (standard errors) and [95% confidence intervals] for COVID-19 data trained on 104 days from March 12, 2020 to June 24, 2020.

Application to COVID-19 - Productivity Recovery



(a) Joint Baseline

(b) Staged Baseline

(c) Staged Distance Penalty

(d) Staged Mobility Penalty

Figure 6: Fitted productivity estimates under different penalty regimes for observed COVID-19 data. Columns indicate origin while rows indicate destination.

Application to COVID-19 - Productivity Recovery

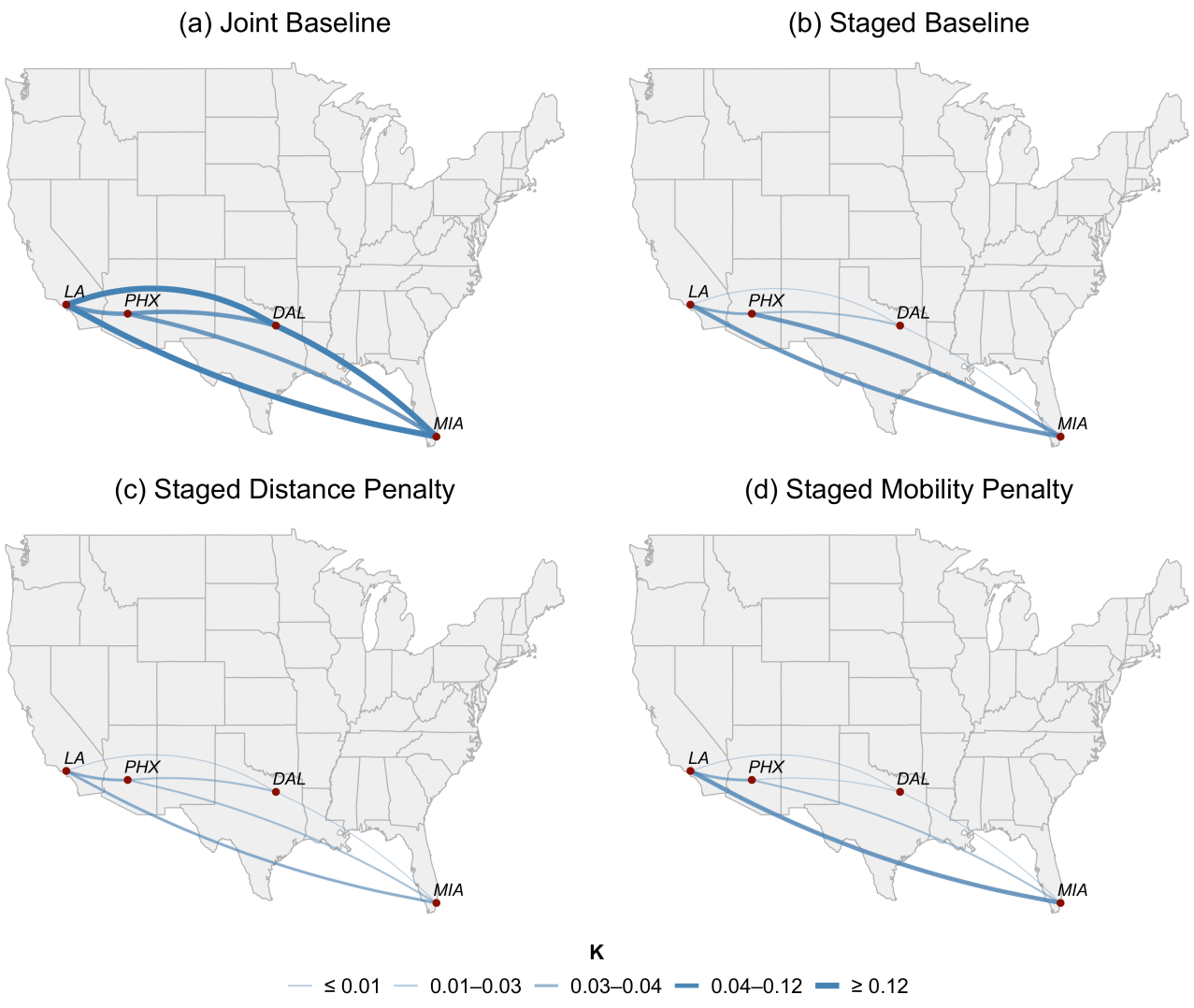


Figure 7: Fitted inter-county productivity rates under four different estimation regimes.

Application to COVID-19 - Predictive Performance

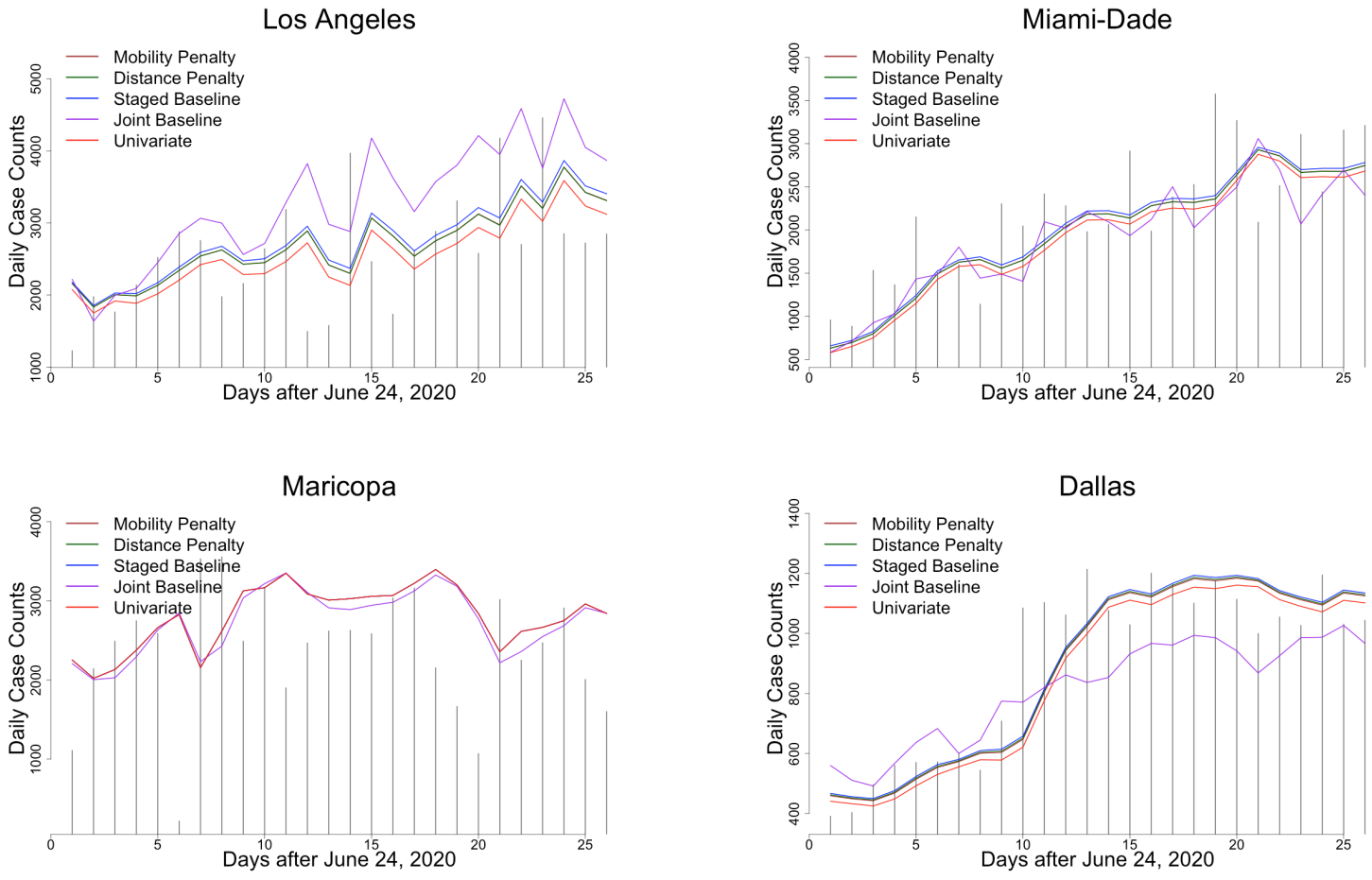


Figure 8: Daily recorded COVID-19 case counts (bars) with five estimation method predictions (curves) overlaid for each location from June 25, 2020 to July 20, 2020. All plots are fitted to observed COVID-19 data [12].

Application Results - Evaluation Metrics

Estimate	NLL	RMSE	$\hat{K}_{i=j}$	$\hat{K}_{i\neq j}$
Joint Baseline	13642	795	0.76	0.132
Univariate	11458	682	1.06	0
Staged Baseline	**10961**	**676**	1.06	0.013
Staged Distance Penalty	11060	676	1.06	0.009
Staged Mobility Penalty	11057	676	1.06	0.009

Table 5: Pooled negative log-likelihood (NLL), pooled Root Mean Squared Error (RMSE), mean diagonal (local) productivity estimates ($\hat{K}_{i=j}$), and mean off-diagonal (cross) productivity estimates ($\hat{K}_{i\neq j}$) under five estimation frameworks for observed COVID-19 data.

Future Research Plan

Scaling in Time - Time-Varying Parameters

Current Limitation: Currently, parameters are time-invariant, statically structured to explain transmission over the course of a dynamic pandemic.

Proposed Enhancement: Allow parameters to vary piecewise by statistically identifying significant change-points and partitioning the time series accordingly.

- Scale to the entire course of a pandemic, not merely 130 days.
- Determine change-points by detecting trend changes:
  - Smooth case counts (7-day)
  - Estimate change-points m
  - Estimate $\mu_i^{(m)}, K_{ij}^m$ for all i, j, m

Scaling in Time Cont.

Change-points are likely to exist as a consequence of:

- Policy changes (e.g., quarantines, mandatory mask mandates)
- Emergence of variants (e.g., Omicron)
- Seasonality (e.g., winter vs. summer trends)
- Mobility shifts (e.g., virtual work adoption, airport travel patterns)

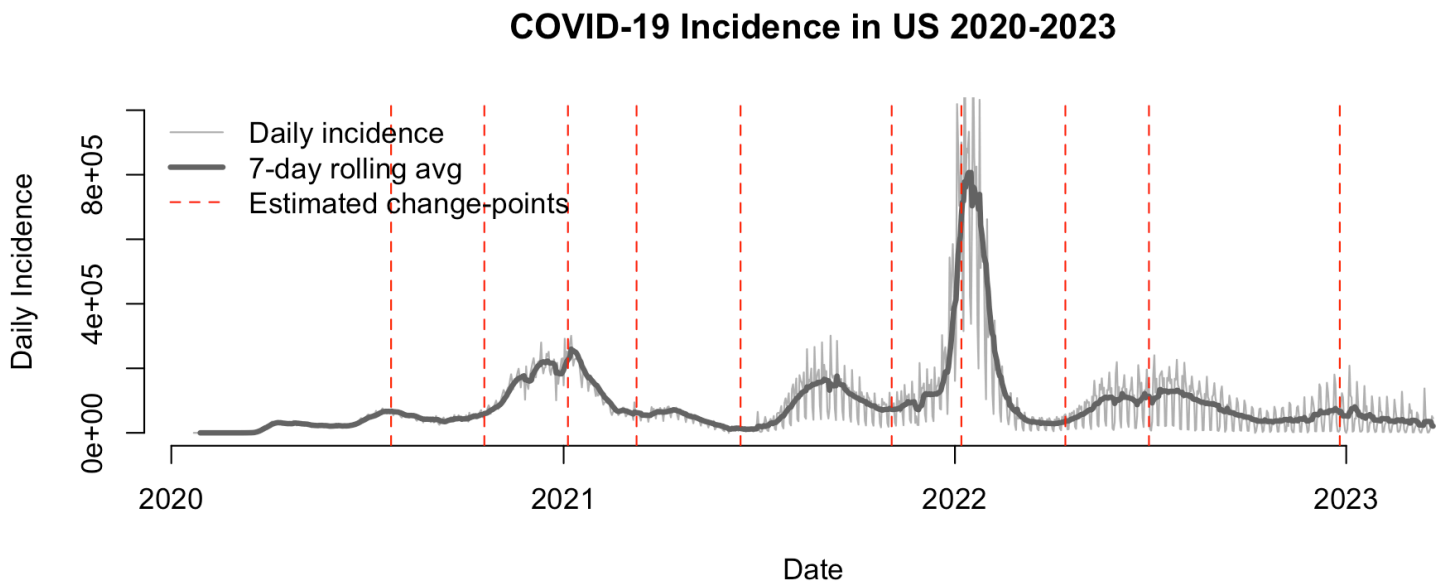


Figure 9: COVID-19 data of total daily incidence in the USA were obtained from the NY Times COVID-19 GitHub repository [12].

Scaling in Time Cont.

Breakpoints are estimated by fitting a piecewise-linear regression model to 7-day smoothed incidence.

$$\lambda_i(t|\mathcal{H}_{t-}) = \mu_i^{(m)} + K_{ii}^{(m)} \sum_{s<t} \phi_{\text{local}}\left(t, s; \beta_i, p_i\right) N_i(s) + \sum_{j \neq i} K_{ij}^{(m)} \sum_{s<t} \phi_{\text{cross}}\left(t, s; \bar{\beta}_i, \bar{p}_i\right) N_j(s) \tag{9}$$

where $m = 1, \ldots, M$ denotes disjoint time segments and $m - 1$ is used as **priors** to initialize segment m. Change-points are estimated globally and shared across locations rather than locally for the purposes of scalability, though local segmentation would be a natural extension. Event history persists across segments via triggering kernels, in alignment with infectious disease dynamics.

Scaling in Space

Current Limitation: A mere four-node network has been studied up to this point, neglecting many locations in order to mitigate computational expense.

Proposed Enhancement: Scale the model to efficiently handle many nodes, either state- or country-wide.

- CA has 58 counties.
- USA has approximately $3,200$ counties.

To reduce computational load, the following methods are proposed:

- Directional covariate gates.
- Parallelization across locations.
- Modulating μ, K by a population density scaling factor.
- Residual-based gating.

Population Density

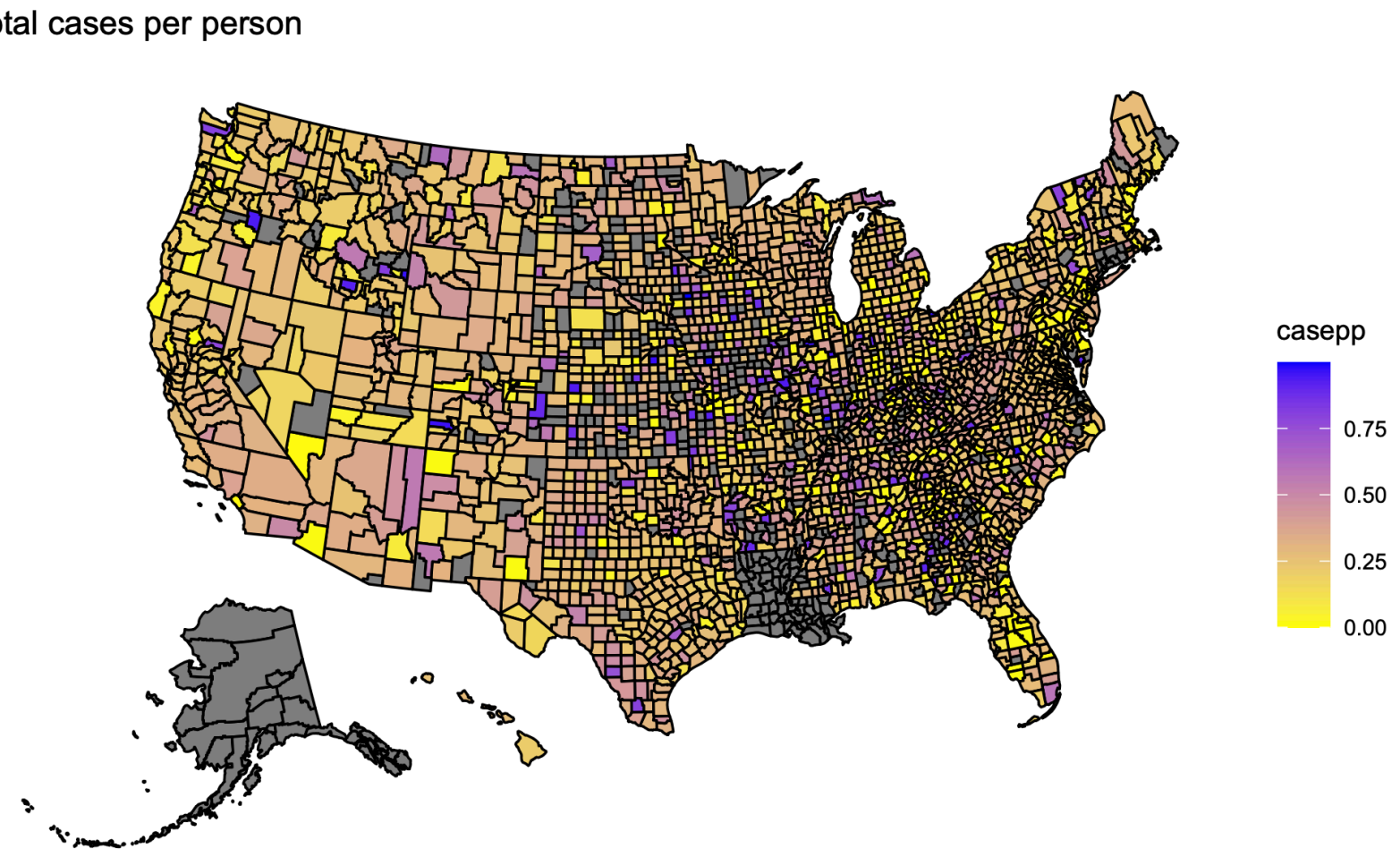


Figure 10: US map of COVID-19 cases standardized by county population. COVID-19 cases are cumulated from 2020-2023, while county population is set to 2021 rates.

Theoretical Properties of the Staged Estimator

Proposition 1 (Stability): The multivariate Hawkes process is stable.

Proposition 2 (Consistency): Assuming stability and Stage 1 estimators are consistent, Stage 2 estimators are consistent up to regularization bias.

Proposition 3 (Error): The estimation error of the cross-excitation parameter K_{ij} decomposes into

$$\hat{K}_{ij} - K_{ij} = \underbrace{\left(K_{ij}^{\star} - K_{ij}\right)}_{\text{Stage 1 error}} + \underbrace{\mathbb{E}\left[\hat{K}_{ij} \mid \text{Stage 1}\right] - K_{ij}^{\star}}_{\text{regularization bias}} + \underbrace{\hat{K}_{ij} - \mathbb{E}\left[\hat{K}_{ij} \mid \text{Stage 1}\right]}_{\text{sampling noise}} \tag{10}$$

where $K_{ij}^{\star}$ is the oracle estimator

Monkeypox

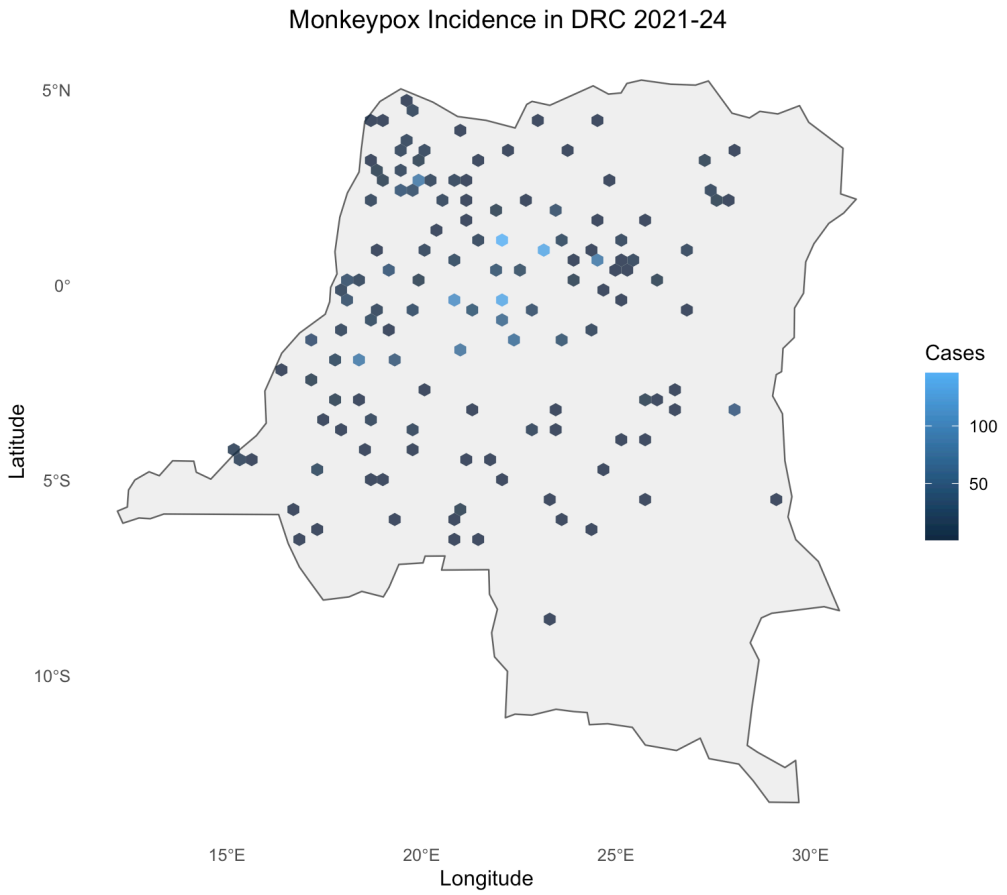


Figure 11: Monkeypox data in the Democratic Republic of the Congo 2022-2024.

Conclusion

Conclusion

Current problem:

- Overly zealous and biased cross-productivities as a result of naturally similar temporal patterns.

Proposed solution:

- Staged estimation
- Covariate-informed penalties
- Time-varying parameters
- Modulate with population density
- Smart spatial scaling

Anticipated result:

- Improved parameter recovery and predictive performance.
- Extend conventional MHPs to be covariate-informed & scalable to large networks.

Roadmap

Milestones:

1. Scale in time
    - Statistically identify notable change points over the course of the COVID-19 pandemic (March 2020 - May 2023)
    - Estimate parameters for each disjoint time segment m
2. Scale in space
    - Directional covariate gates
    - Parallelization
    - Modulate μ, K by population density
    - Residual-based gating
3. Staged estimator
    - Stability
    - Consistency
    - Error decomposition

References

[1] Conor Kresin, Frederic Paik Schoenberg, and George Mohler. **“Comparison of the Hawkes and SEIR models for the spread of COVID-19”.** In: *Advances and Applications in Statistics* 74 (2022), pp. 83–106. DOI: DOI:10.17654/0972361722019.

[2] Sen Pei and Jeffrey Shaman. **“Initial Simulation of SARS-CoV2 Spread and Intervention Effects in the Continental US”.** In: *medRxiv* (2020). DOI: 10.1101/2020.03.21.20040303. eprint: https://www.medrxiv.org/content/early/2020/03/27/2020.03.21.20040303.full.pdf.

[3] E. O. Nsoesie et al. **“A systematic review of studies on forecasting the dynamics of influenza outbreaks”.** In: *Influenza and other respiratory viruses* 8.3 (2014), pp. 309–316. URL: https://doi.org/10.1111/irv.12226.

[4] Frederic Paik Schoenberg. **“A NOTE ON THE CONSISTENT ESTIMATION OF SPATIAL-TEMPORAL POINT PROCESS PARAMETERS”.** In: *Statistica Sinica* 26.2 (2016), pp. 861–879. ISSN: 10170405, 19968507. URL: http://www.jstor.org/stable/24721302 (visited on 05/30/2025).

[5] George Mohler and Michael D. Porter. **“Rotational grid, PAI-maximizing crime forecasts”.** In: *Statistical Analysis and Data Mining: An ASA Data Science Journal* 11.5 (2018), pp. 227–236. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/sam.11389. URL: https://doi.org/10.1002/sam.11389.

[6] Alex Reinhart and Joel Greenhouse. **“Self-Exciting Point Processes with Spatial Covariates: Modelling the Dynamics of Crime”.** In: *Journal of the Royal Statistical Society Series C: Applied Statistics* 67.5 (Mar. 2018), pp. 1305–1329. ISSN: 0035-9254. DOI: 10.1111/rssc.12277. eprint: https://academic.oup.com/jrsssc/article-pdf/67/5/1305/49337113/jrsssc_67_5_1305.pdf. URL: https://doi.org/10.1111/rssc.12277.

[7] Alessio Sancetta. **“ESTIMATION FOR THE PREDICTION OF POINT PROCESSES WITH MANY COVARIATES”.** In: *Econometric Theory* 34.3 (2018), pp. 598–627. DOI: 10.1017/S026646661700017 2.

[8] Yichen Jiang and Michael Porter. **“Discovering Influence of Yelp Reviews Using Hawkes Point Processes”.** In: *Intelligent Systems and Applications*. Ed. by Kohei Arai. Cham: Springer International Publishing, 2022, pp. 84–104. ISBN: 978-3-030-82199-9.

[9] Frederic Paik Schoenberg, Marc Hoffmann, and Ryan J. Harrigan. **“A recursive point process model for infectious diseases”.** In: *Annals of the Institute of Statistical Mathematics* 71.5 (Oct. 1, 2019), pp. 1271–1287. ISSN: 1572-9052. DOI: 10.1007/s10463-018-0690-9. URL: https://doi.org/10.1007/s10463-018-0690-9.

[10] Youness Diouane, Frederic Schoenberg, and George Mohler. **“Accurate Estimation of Cross-Excitation in Multivariate Hawkes Process Models of Infectious Diseases”.** In: (2024). DOI: 10.1109/DSAA61799.2024.10722768. URL: https://par.nsf.gov/biblio/10549612.

[11] Baichuan Yuan, Frederic P. Schoenberg, and Andrea L. Bertozzi. **“Fast estimation of multivariate spatiotemporal Hawkes processes and network reconstruction”.** In: *Annals of the Institute of Statistical Mathematics* 73.6 (Dec. 1, 2021), pp. 1127–1152. ISSN: 1572-9052. DOI: 10.1007/s10463-020-00780-1. URL: https://doi.org/10.1007/s10463-020-00780-1.

[12] The New York Times. ***Coronavirus (Covid-19) Data in the United States.*** 2021. URL: https://github.com/nytimes/covid-19-data.

[13] U.S. Department of Health & Human Services. ***U.S. State, Territorial, and County Stay-At-Home Orders: March 15–May 5 by County by Day.*** Metadata updated July 26, 2021. July 26, 2021. URL: https://healthdata.gov/dataset/U-S-State-Territorial-and-County-Stay-At-Home-Orde/e9zu-hjtk/about_data (visited on 06/24/2025).

[14] R. W. Sinnott. **“Virtues of the Haversine”.** In: *Sky and Telescope* 68.2 (1984), p. 159.

[15] Yuhao Kang et al. **“Multiscale Dynamic Human Mobility Flow Dataset in the U.S. during the COVID-19 Epidemic”.** In: *Scientific Data* (390 2020), pp. 1–13. URL: https://doi.org/10.1038/s41597-020-00734-5.

[16] Melanie Mitchell. **“Genetic Algorithms: An Overview”.** In: *Complexity* 1.1 (1995), pp. 31–39. DOI: 10.1002/cplx.6130010108.