COLUMBIA UNIVERSITY | MAILMAN SCHOOL of PUBLIC HEALTH
ENVIRONMENTAL HEALTH SCIENCES

unifying the epidemiological and evolutionary dynamics of PATHOGENS

Modeling spatial transmission of infectious diseases

Sen Pei, PhD

Assistant Professor
Department of Environmental Health Sciences
Mailman School of Public Health, Columbia University

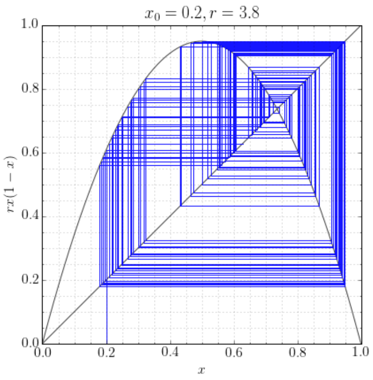
NORDITA, Stockholm, Sweden, 5/30/2023

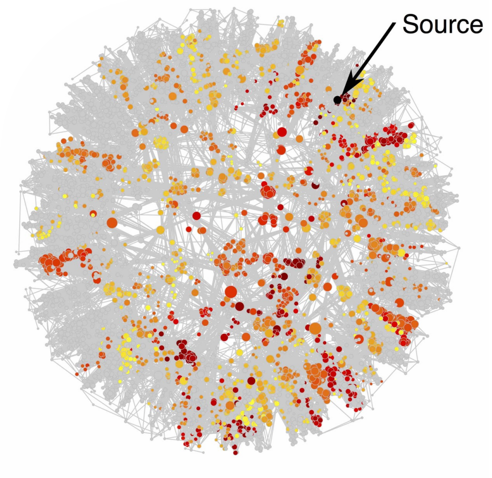
Content

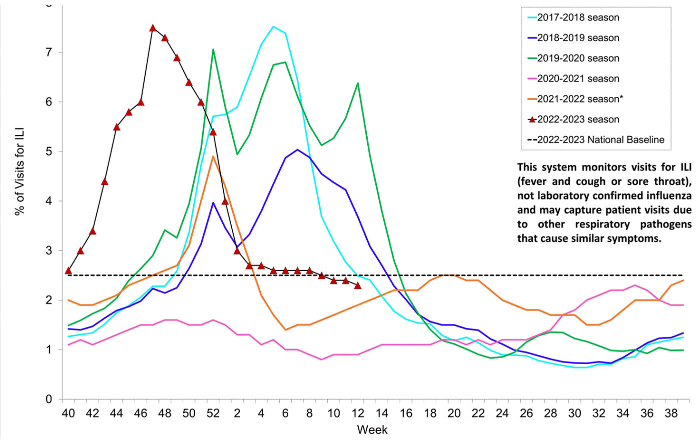
- Training and career path
- Forecasting spatial spread of influenza (2016-2019)
- Modeling and inference for COVID-19 (2020-2023)
- Use of foot-traffic data in epidemic models

Undergraduate | PhD | Postdoc

Dynamical systems | Networks | Infectious Disease

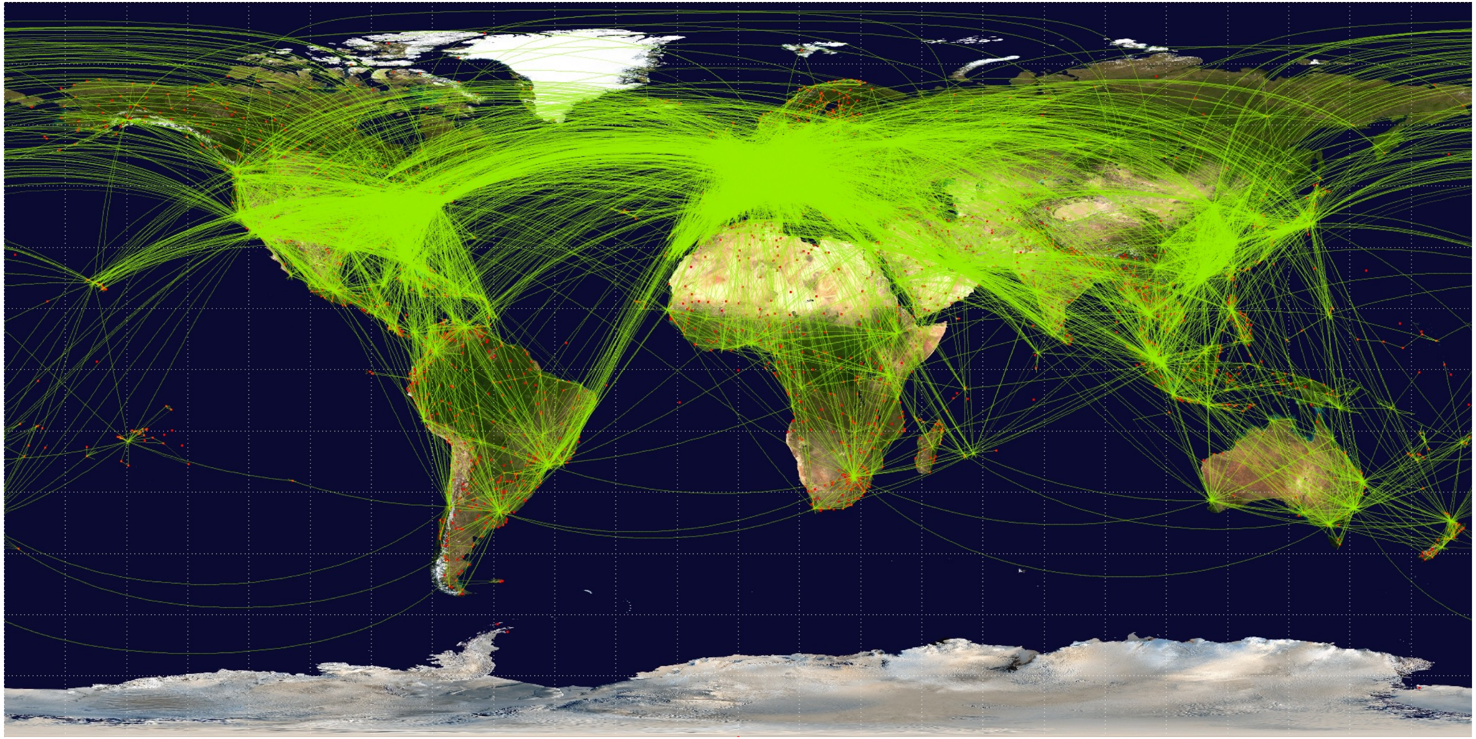






Pei 2020; CDC

Human mobility and spatial spread of infectious diseases

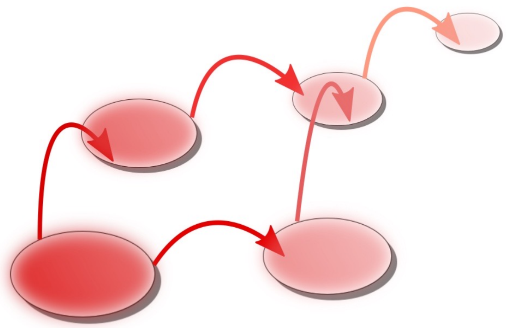


OpenFlights

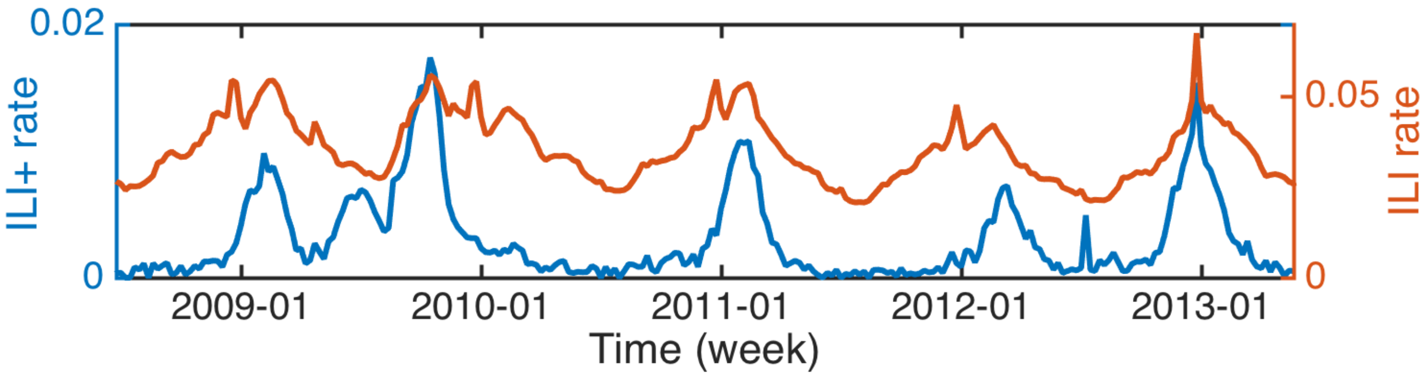
Can we predict the spatial spread of influenza in real time?

Forecasting spatial spread is challenging

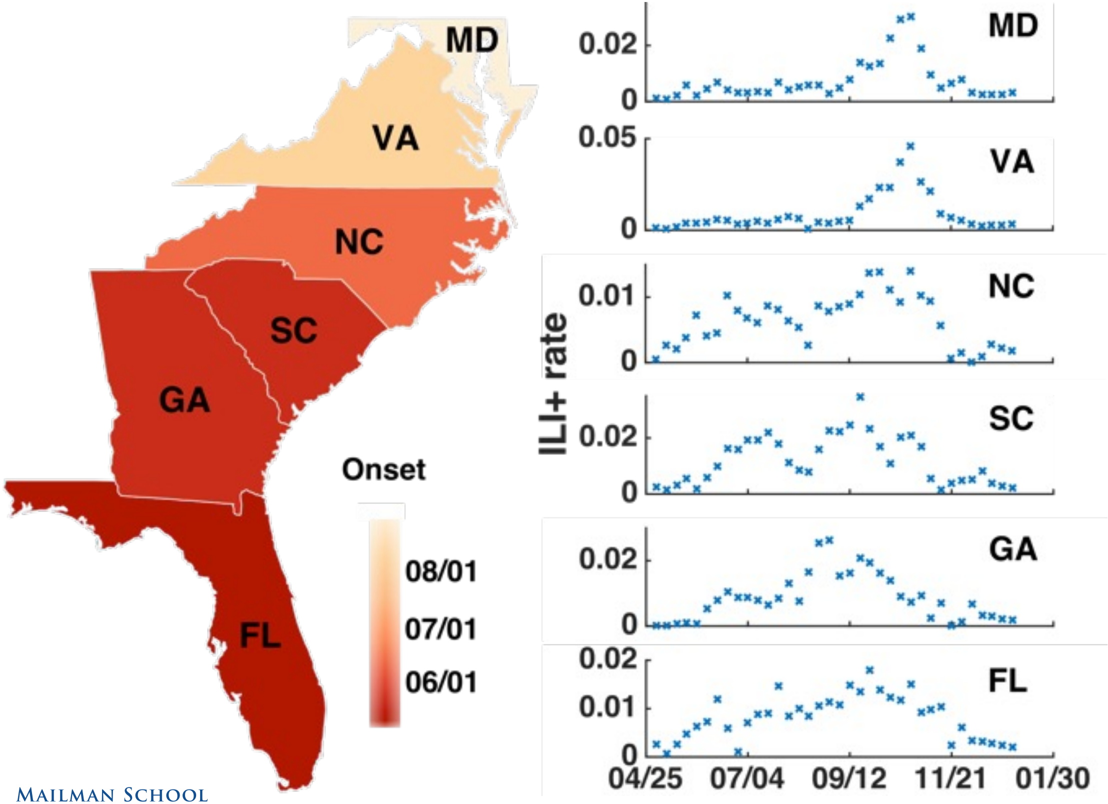
- Mobility data are not available in real-time
- Parameter may change over time
- Limited disease data to validate (in 2017)



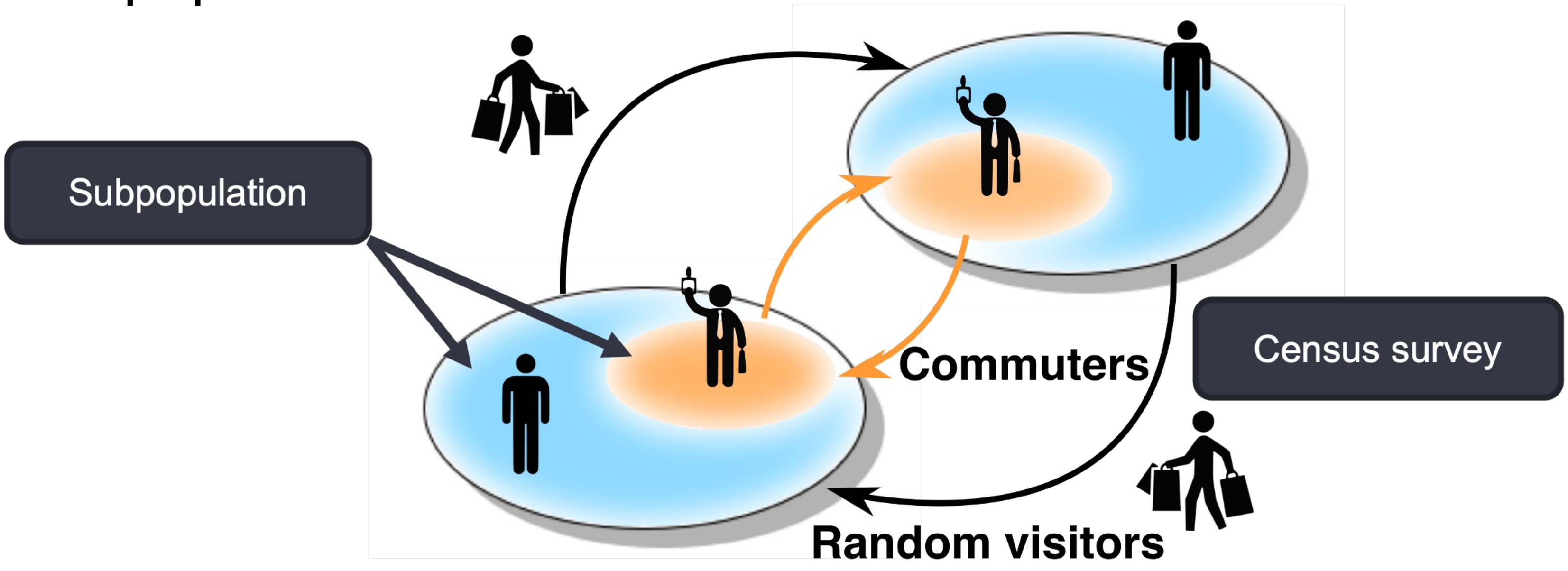
- A data-driven metapopulation model
- Influenza data from DoD healthcare system (35 states)



Spatial spread of 2009 H1N1 pandemic



A metapopulation structure



Random movement = θ × commute

Pei et al. 2018

A humidity-driven SIRS model

Day time transmission equations: SIRS model

Modulated by humidity data

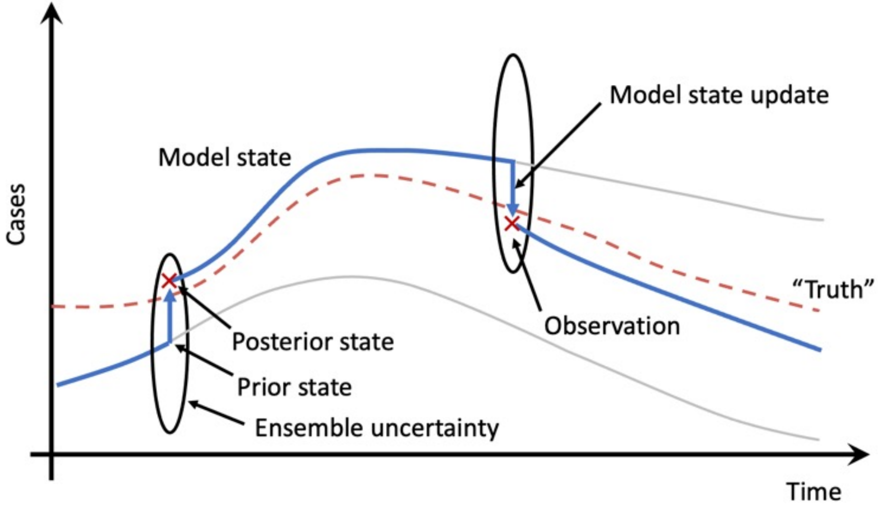
$$\frac{dI_n^k}{dt} = \frac{\beta_n(t) S_n^k I_n}{N_n} - \frac{I_n^k}{D} - \frac{\theta I_n^k}{N_n} \sum_{m \neq n} \bar{N}_m^n + \frac{\theta N_n^k}{N_n} \sum_{m \neq n} \bar{N}_n^m \frac{I_m}{N_m},$$

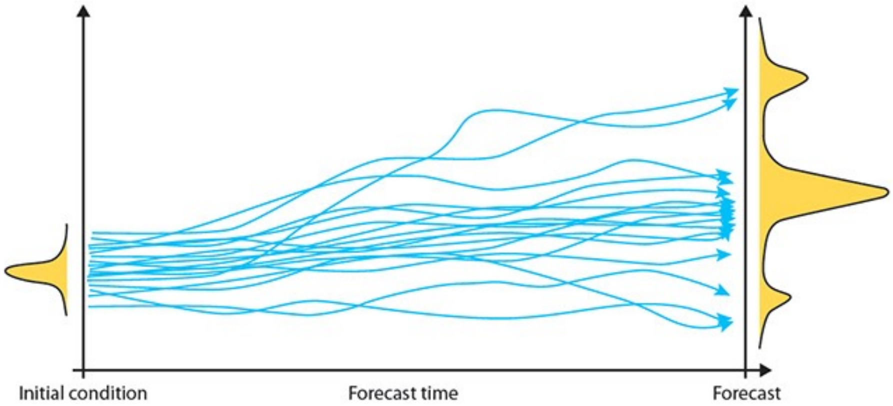
$$\frac{dS_n^k}{dt} = \frac{N_n^k - S_n^k - I_n^k}{L} - \frac{\beta_n(t) S_n^k I_n}{N_n} - \frac{\theta S_n^k}{N_n} \sum_{m \neq n} \bar{N}_m^n + \frac{\theta N_n^k}{N_n} \sum_{m \neq n} \bar{N}_n^m \frac{S_m}{N_m}.$$

Within location transmission — Population exchange

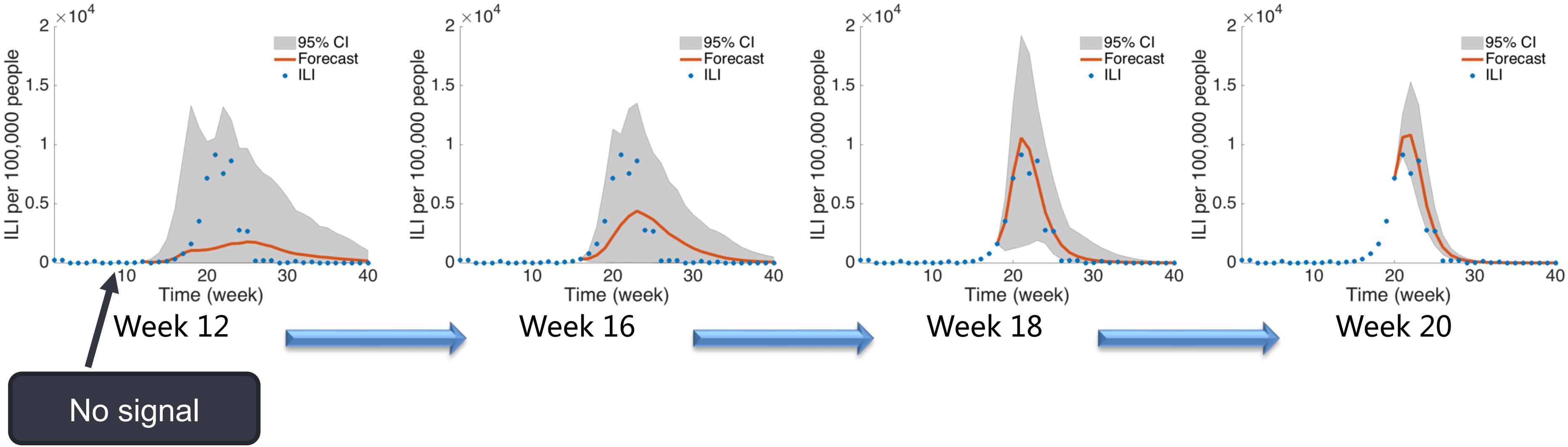
Model calibration and forecasting

- A high-dimensional system
- Ensemble Kalman filter (numerical weather prediction)

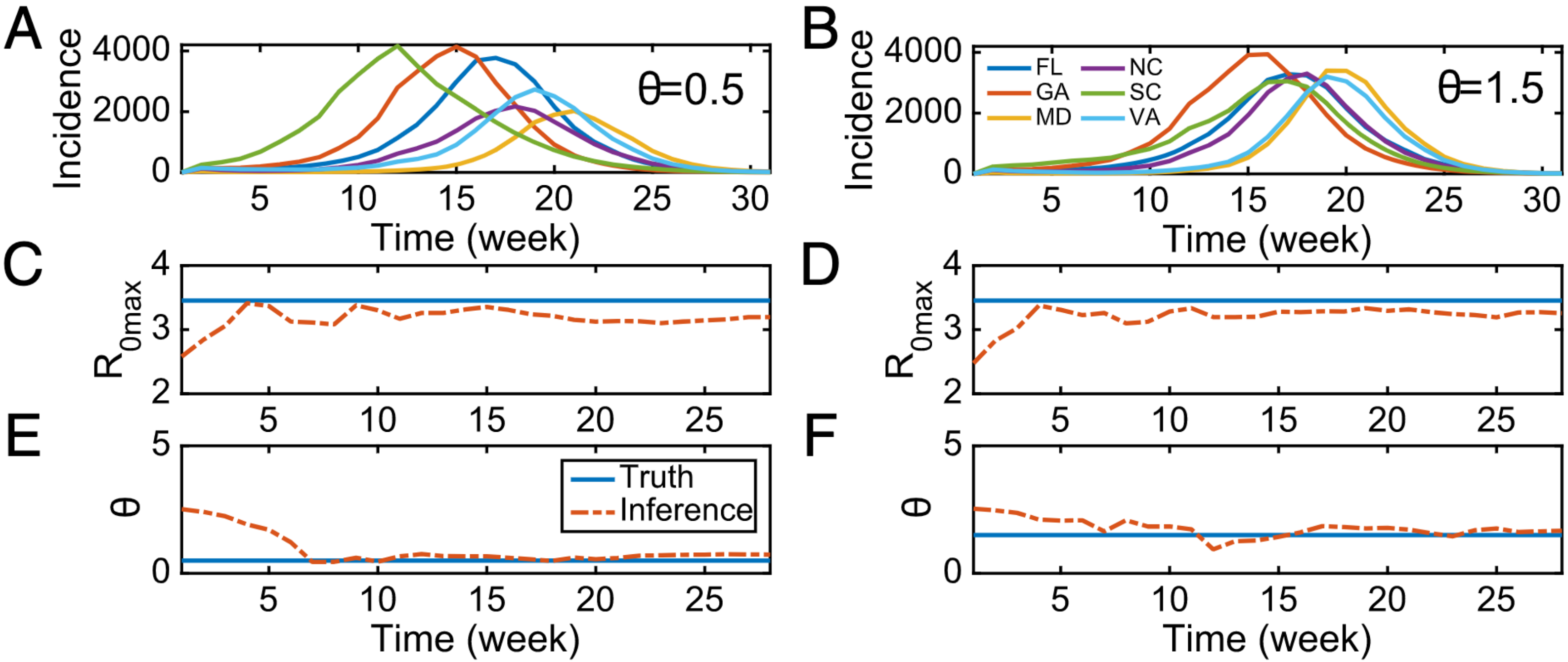




An example of ensemble forecasting using a local model



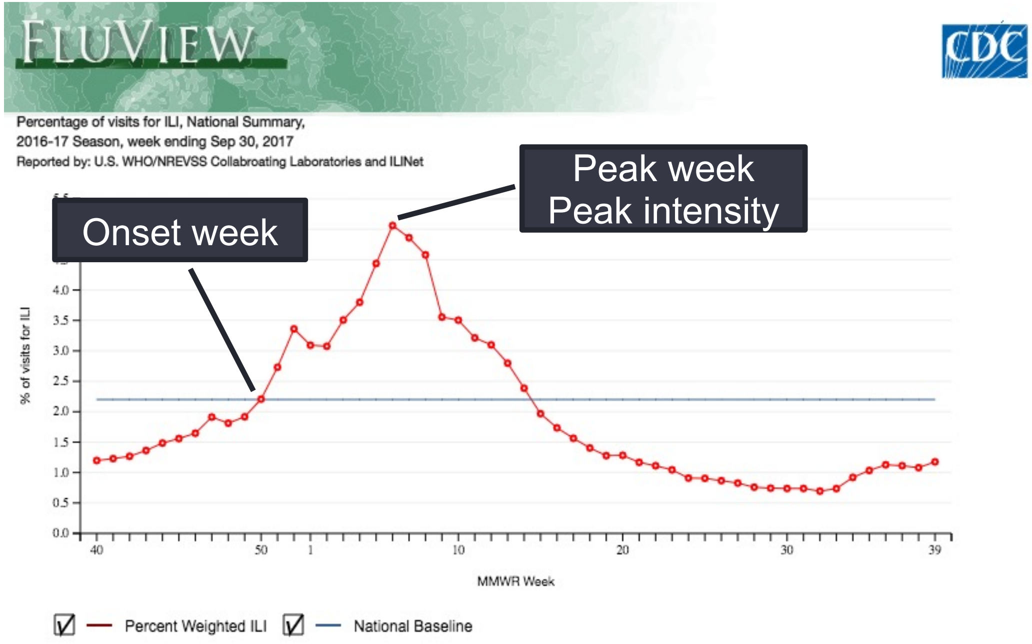
Parameter inference



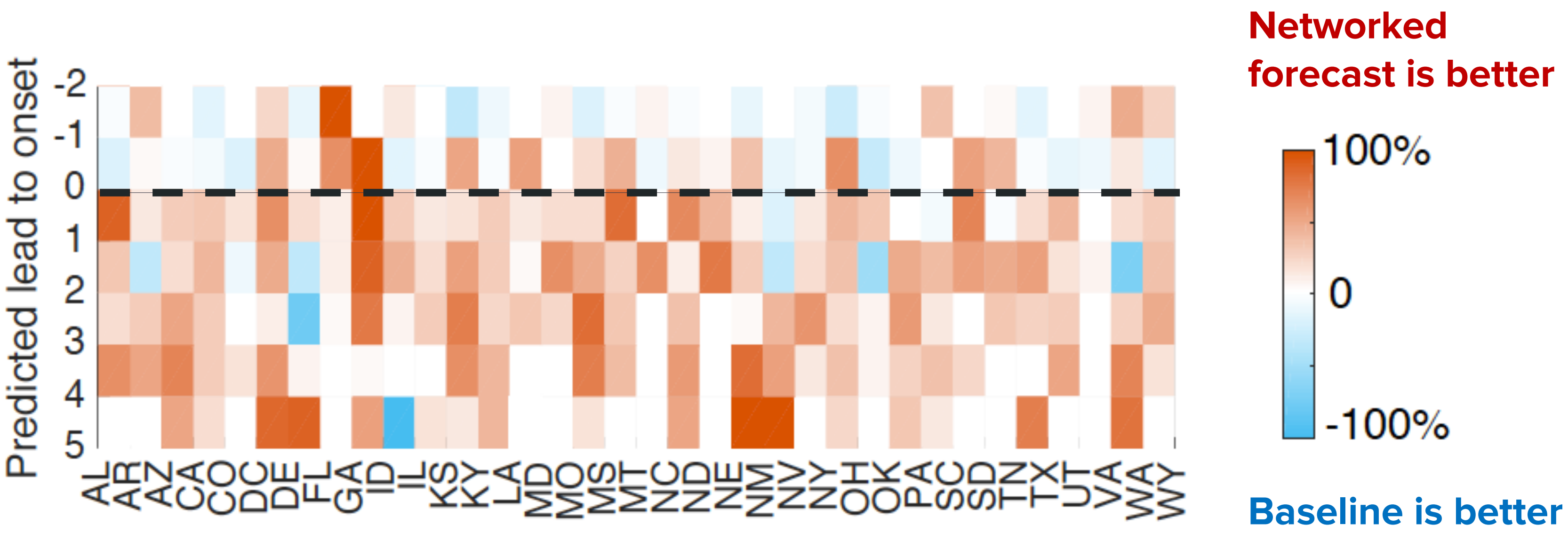
Retrospective forecasting

- 2008-2012, 35 US states

- Weekly forecasting
  - Metapopulation model
  - Local model (Baseline)



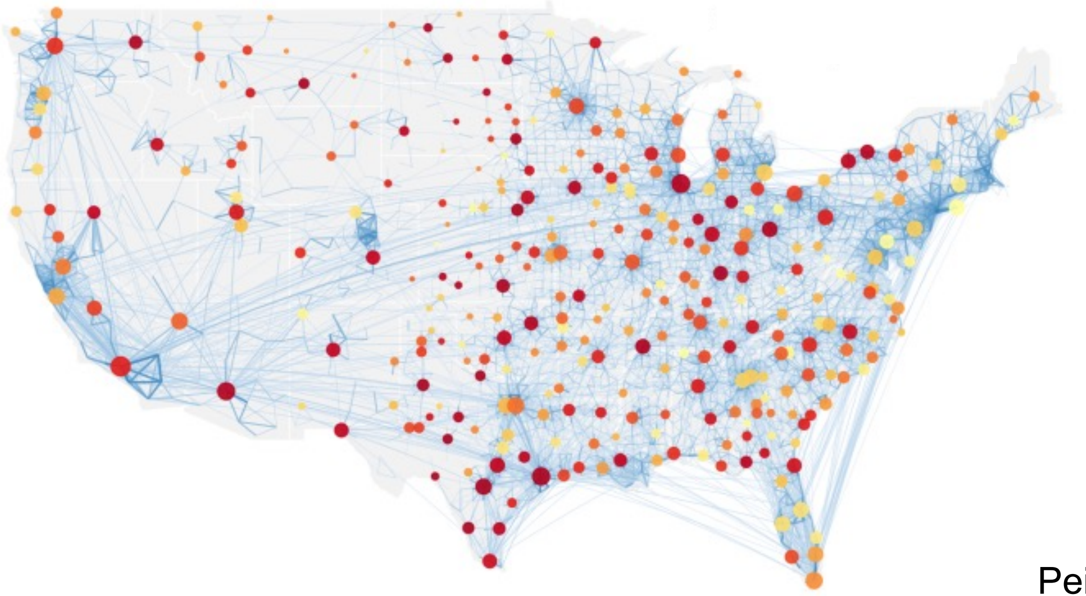
Improvement of onset prediction for each individual state



Forecasts for peak week and peak intensity are also improved

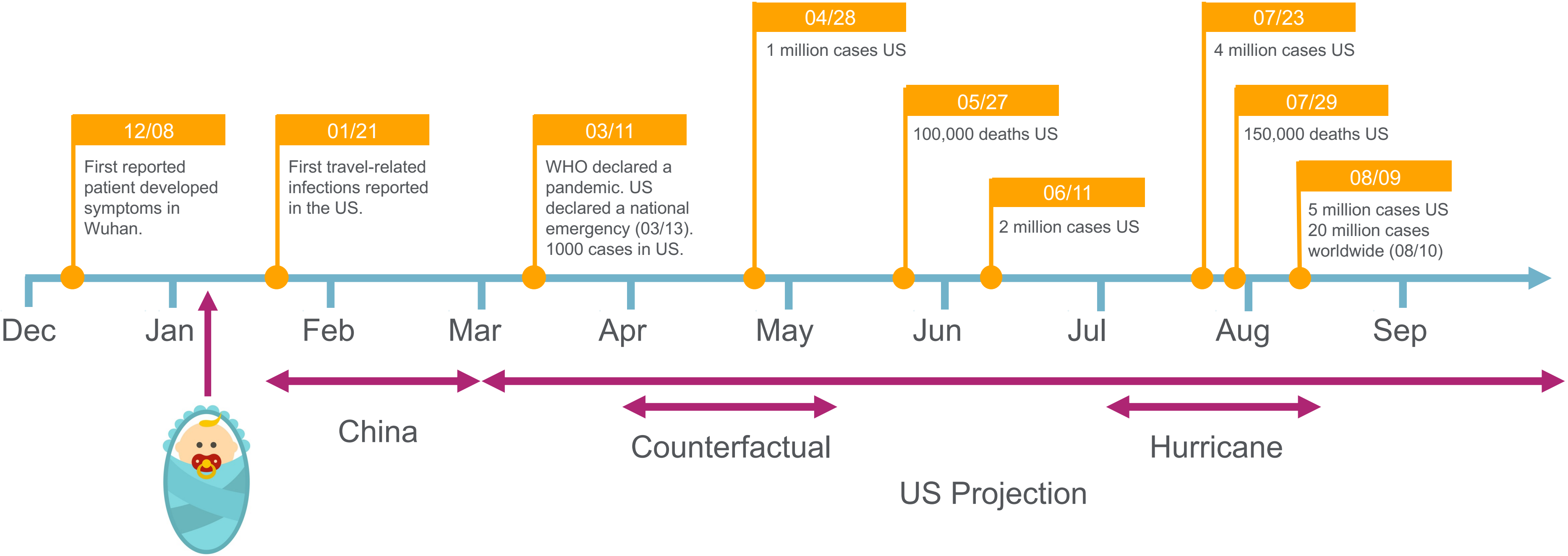
A county-level model for respiratory viruses

- Optimal selection of surveillance sites (2018-2019)
- Influenza, human metapneumovirus, **endemic coronavirus**



Pei et al. (2021)

COVID-19 in 2020

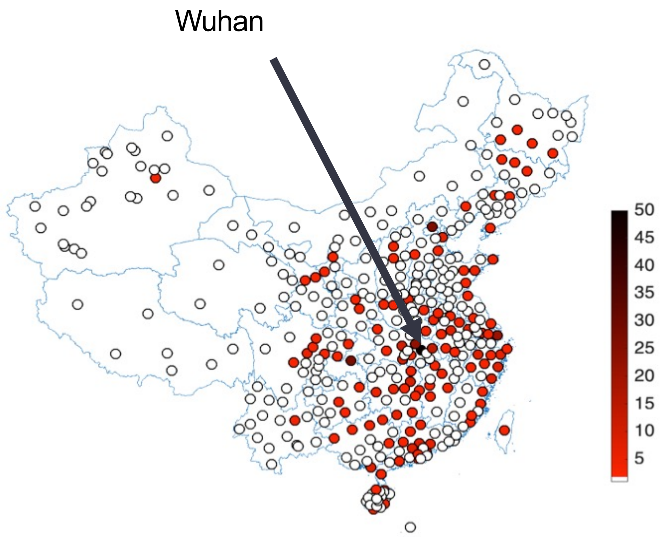


A pandemic?

- Fast spatial spread, only ~900 cases
- Endemic coronavirus (229E, NL63, OC43, HKU1)
  - Infections with no/mild symptoms

How many infections were not detected?
How infectious are the undetected infections?

- Potential for a pandemic?



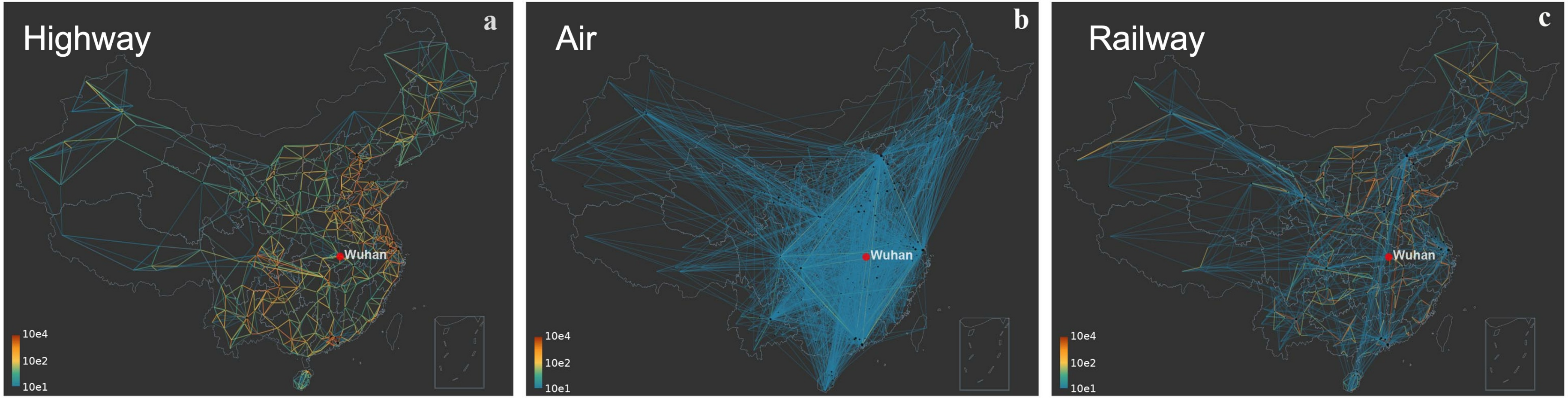
Cumulative number of reported cases in 375 cities as of Jan 23. Wuhan has 454 cases.

Li et al 2020

What is the role of undetected infections?

Modeling SARS-CoV-2 spread in China

- Chunyun: travel during lunar New Year (3 billion trips in 40 days)
- Tencent location-based service data in 2018
- 375 cities, Jan 10 – Jan 23, before Wuhan lockdown



Xu et al. 2023

Metapopulation SEIR model

- Susceptible (S), exposed (E), documented infection (Ir), undocumented infection (Iu), removed (R)
- **Fraction of undocumented infection, relative contagiousness**

$$
\begin{aligned}
\frac{dS_i}{dt} &= -\frac{\beta S_i I_i^r}{N_i} - \frac{\mu \beta S_i I_i^u}{N_i} + \theta \sum_j \frac{M_{ij} S_j}{N_j - I_j^r} - \theta \sum_j \frac{M_{ji} S_i}{N_i - I_i^r} \\
\frac{dE_i}{dt} &= \frac{\beta S_i I_i^r}{N_i} + \frac{\mu \beta S_i I_i^u}{N_i} - \frac{E_i}{Z} + \theta \sum_j \frac{M_{ij} E_j}{N_j - I_j^r} - \theta \sum_j \frac{M_{ji} E_i}{N_i - I_i^r} \\
\frac{dI_i^r}{dt} &= \alpha \frac{E_i}{Z} - \frac{I_i^r}{D} \\
\frac{dI_i^u}{dt} &= (1-\alpha) \frac{E_i}{Z} - \frac{I_i^u}{D} + \theta \sum_j \frac{M_{ij} I_j^u}{N_j - I_j^r} - \theta \sum_j \frac{M_{ji} I_i^u}{N_i - I_i^r}
\end{aligned}
$$

Within city transmission

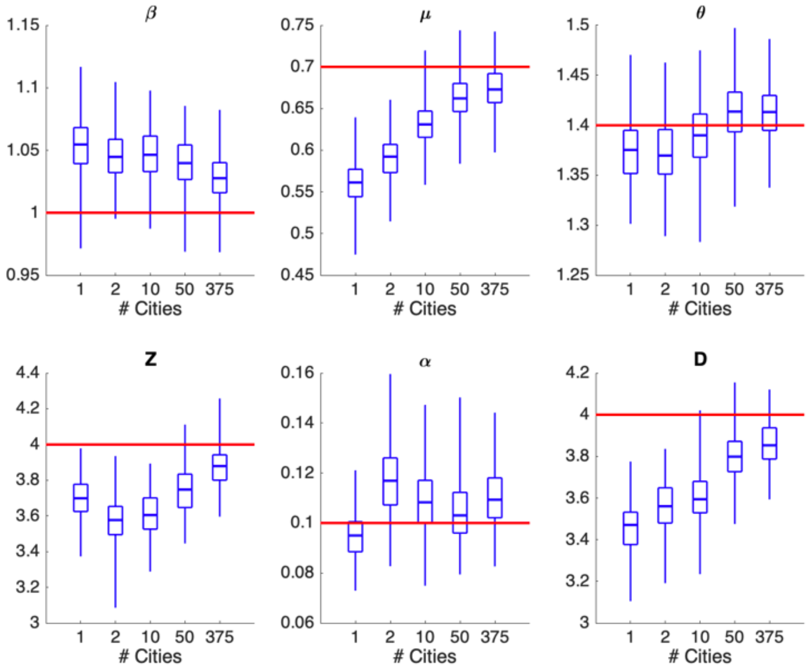
Cross city migration

Parameter
Transmission rate (β, days^{-1})
Relative transmission rate (μ)
Latency period (Z, days)
Infectious period (D, days)
Reporting rate (α)
Basic reproductive number (R_e)
Mobility factor (θ)

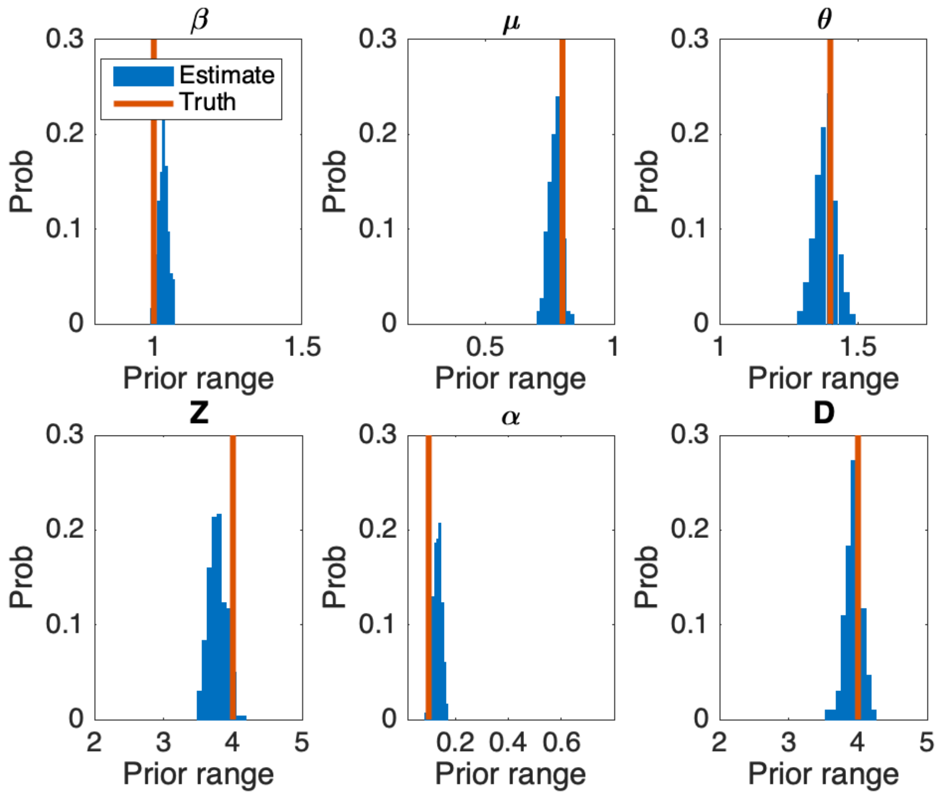
A gamma-distributed reporting delay estimated using line-list data

Parameter inference

- Iterated filtering with ensemble adjustment Kalman filter
- Parameter identifiability?
- Connectivity improves identifiability!

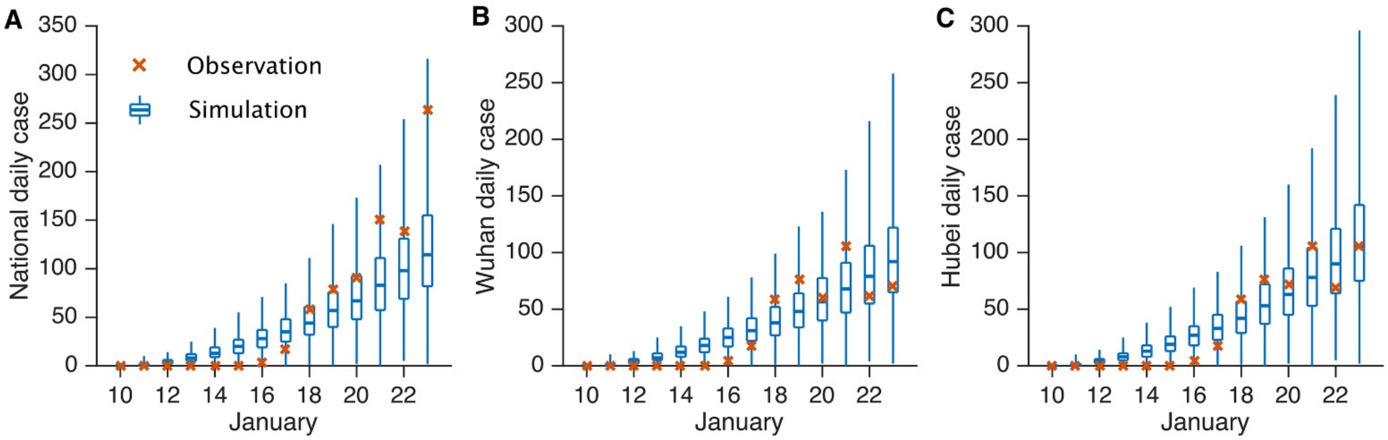


Validation using synthetic outbreaks



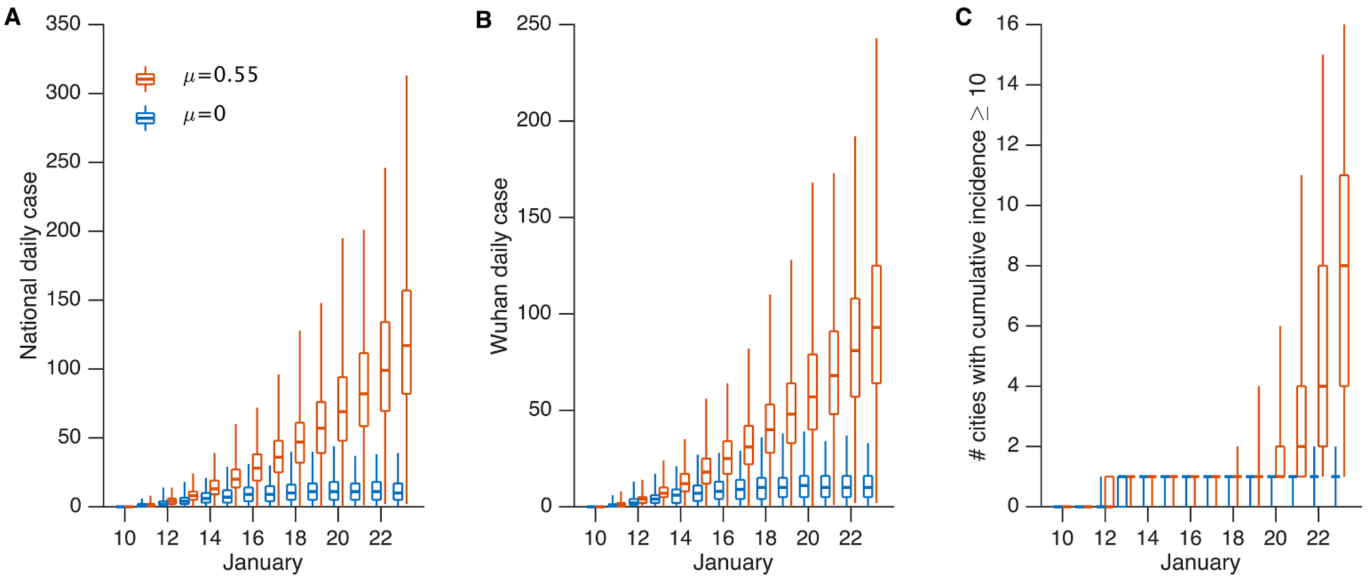
Inference results

- 86% [82%, 90%] infections were undocumented before travel restrictions
- The transmission rate of undocumented infections was 55% [46%, 62%] of documented infections



Counterfactual simulations

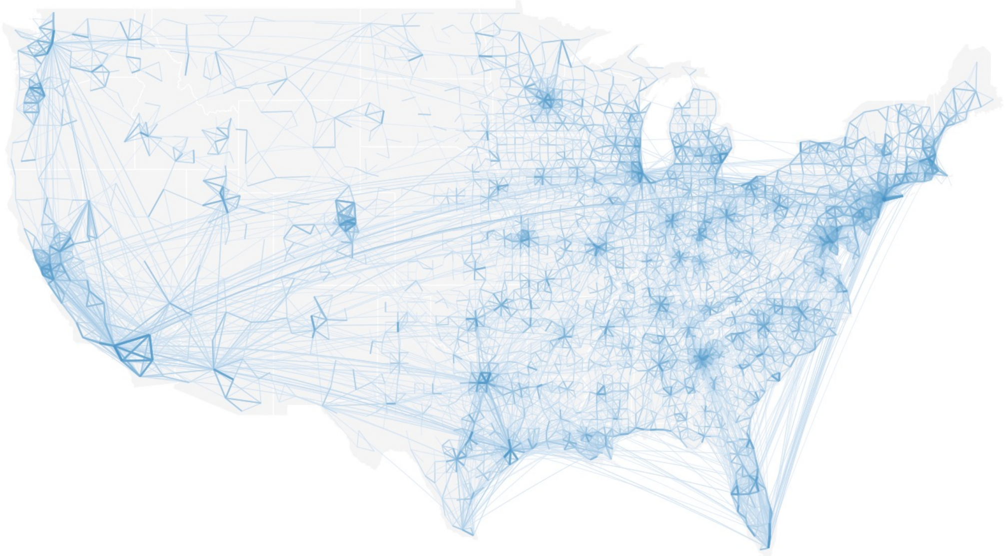
- Assume undocumented infections are not contagious



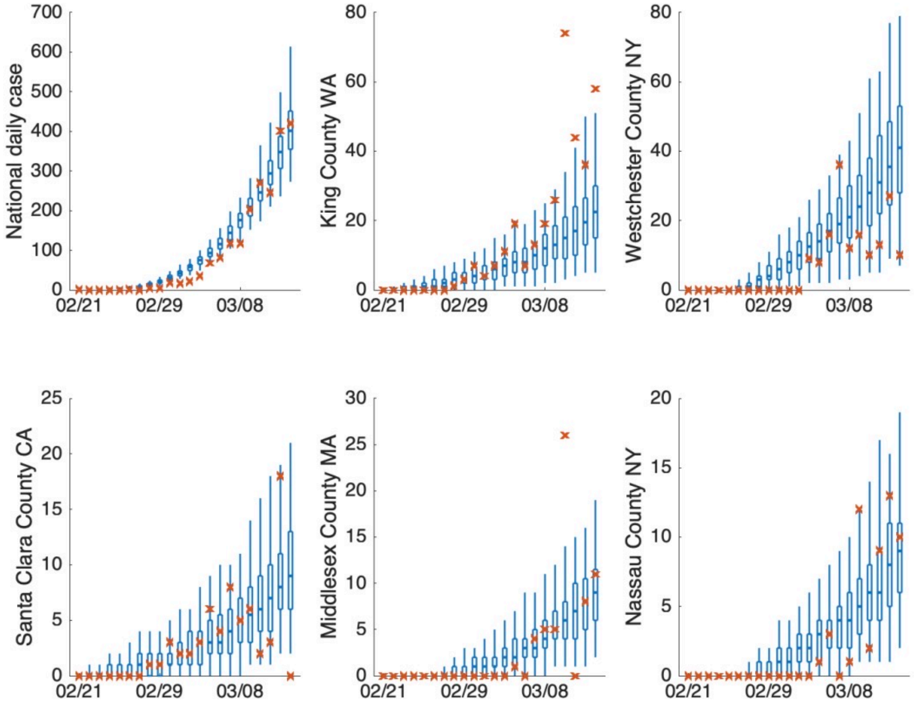
“These findings explain the rapid geographic spread of SARS-CoV-2 and indicate that containment of this virus will be particularly challenging.”

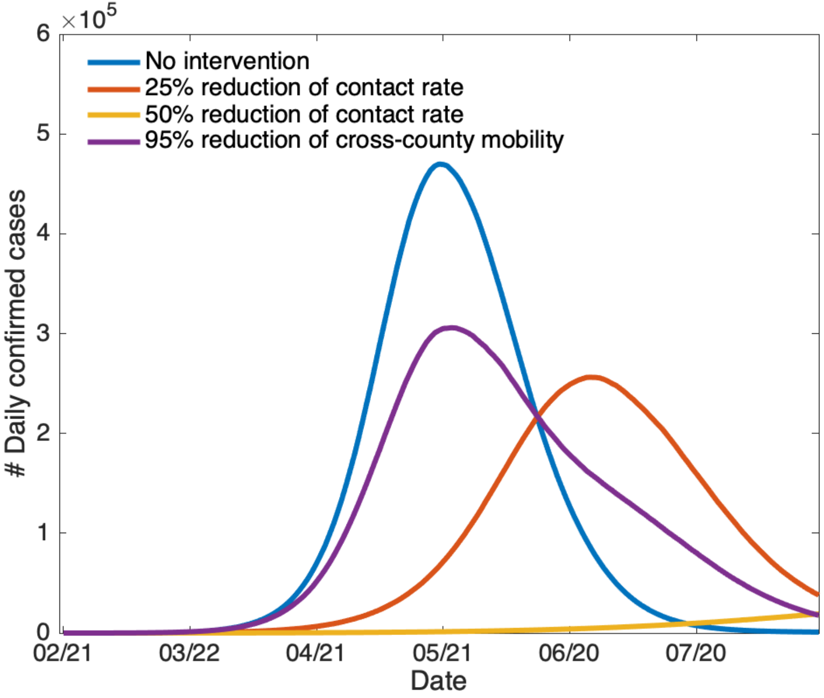
SARS-CoV-2 spread in the US

- A metapopulation model for 3142 US counties
- Commuting data from census
- Fit to county-level data from NYT



Early simulation in the US





Pei & Shaman, 2020

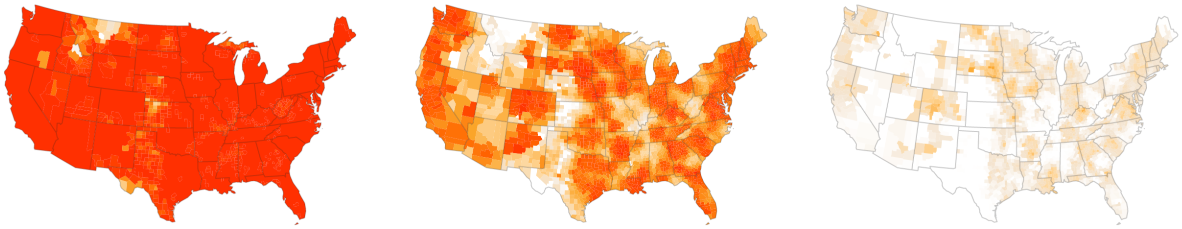
Communications with the public

The New York Times

PLAY THE CROSSWORD

Coronavirus Could Overwhelm U.S. Without Urgent Action, Estimates Say

By James Glanz, Lauren Leatherby, Matthew Bloch, Mitch Smith, Larry Buchanan, Jin Wu and Nicholas Bogel-Burroughs March 20, 2020



Three scenarios for how the outbreak could spread.

Percent infected by **July 1** 25 50 75%


"All the News That's Fit to Print"

The New York Times

Late Edition

VOL. CLXIX ... No. 58,639 NEW YORK, SATURDAY, MARCH 21, 2020 $3.00

PRESSURE ON TRUMP AS MILLIONS ARE KEPT HOME

Mixed Signals From President Sow Confusion

How the Outbreak Could Spread Across U.S. Counties Under Three Social Control Scenarios

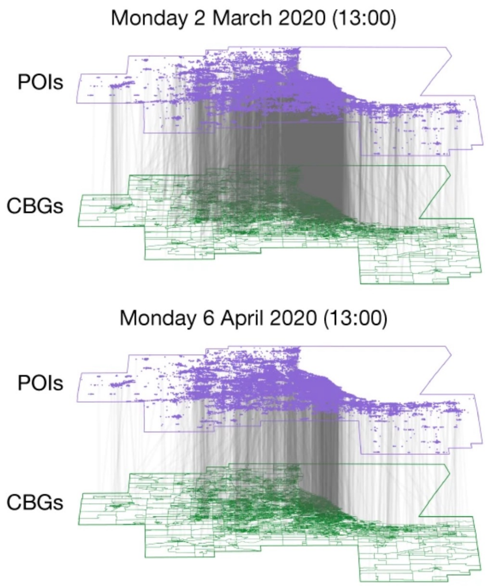
Decrees From New York and Illinois — Virus Tightens Grip on Nation

Short on Beds and Ventilators, New York Hospitals Face Surge

$1 Trillion Stimulus Spurs Lobbying Gold Rush

Skipping Meals So Her Children Can Have Theirs

New data streams on human mobility



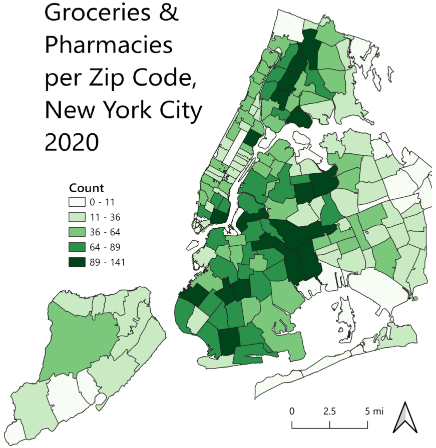
- Aggregated mobility data shared by private companies
  - SafeGraph, Cuebiq, Google, GPS, etc.
- Unprecedented data for infectious disease modeling
- High-resolution foot-traffic data to understand contact patterns and mobility
  - SafeGraph: points of interest (POIs) and mobility
- Potentials and challenges

Chang et al. 2020

POIs and mobility in NYC

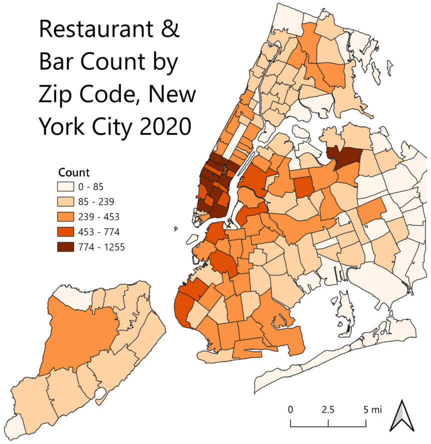
- Place categories: Grocery & Pharmacy, Other Retails, Art & Entertainment, Restaurant & Bar, Education, Healthcare, Other Places.

Groceries & Pharmacies per Zip Code, New York City 2020

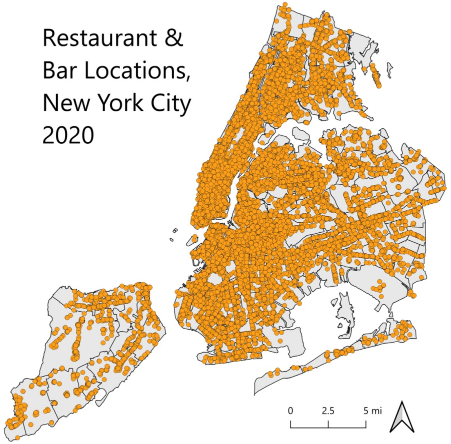


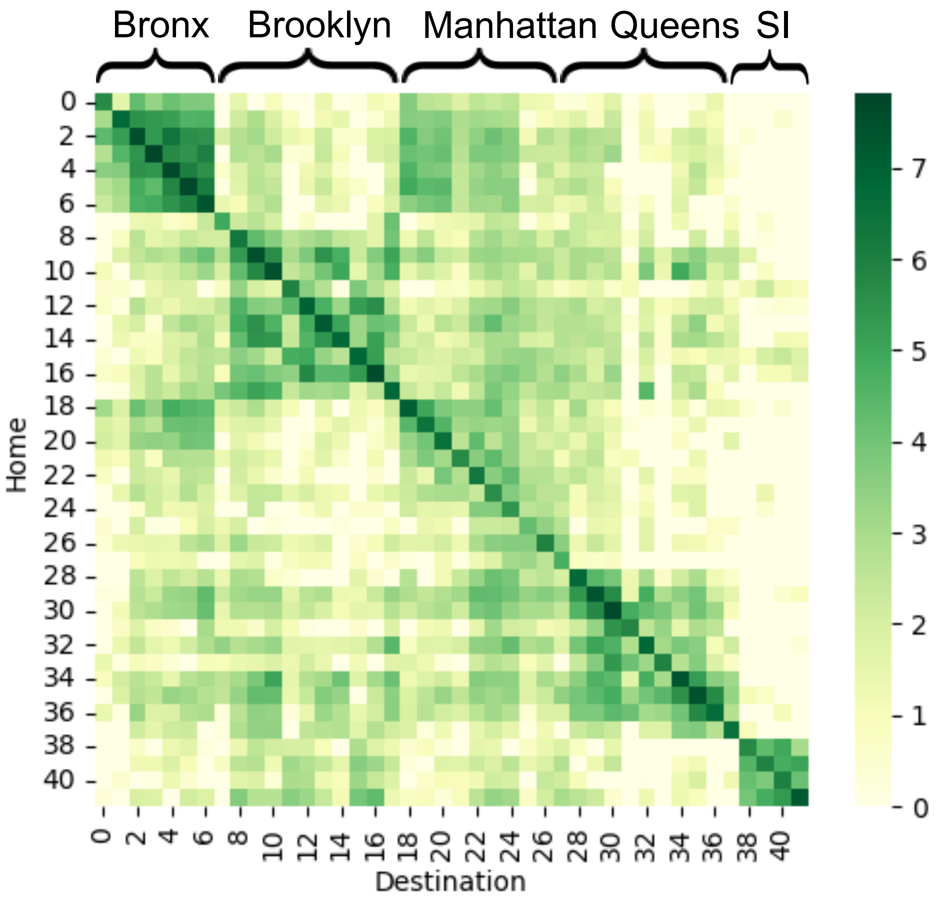
Grocery & Pharmacy Locations, New York City 2020

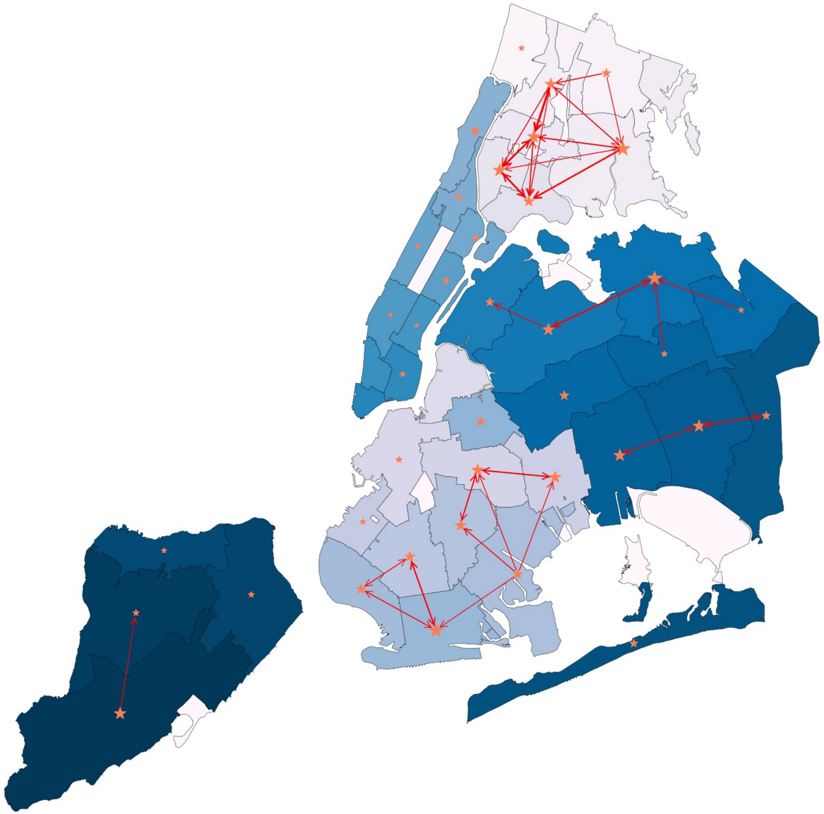
Restaurant & Bar Count by Zip Code, New York City 2020



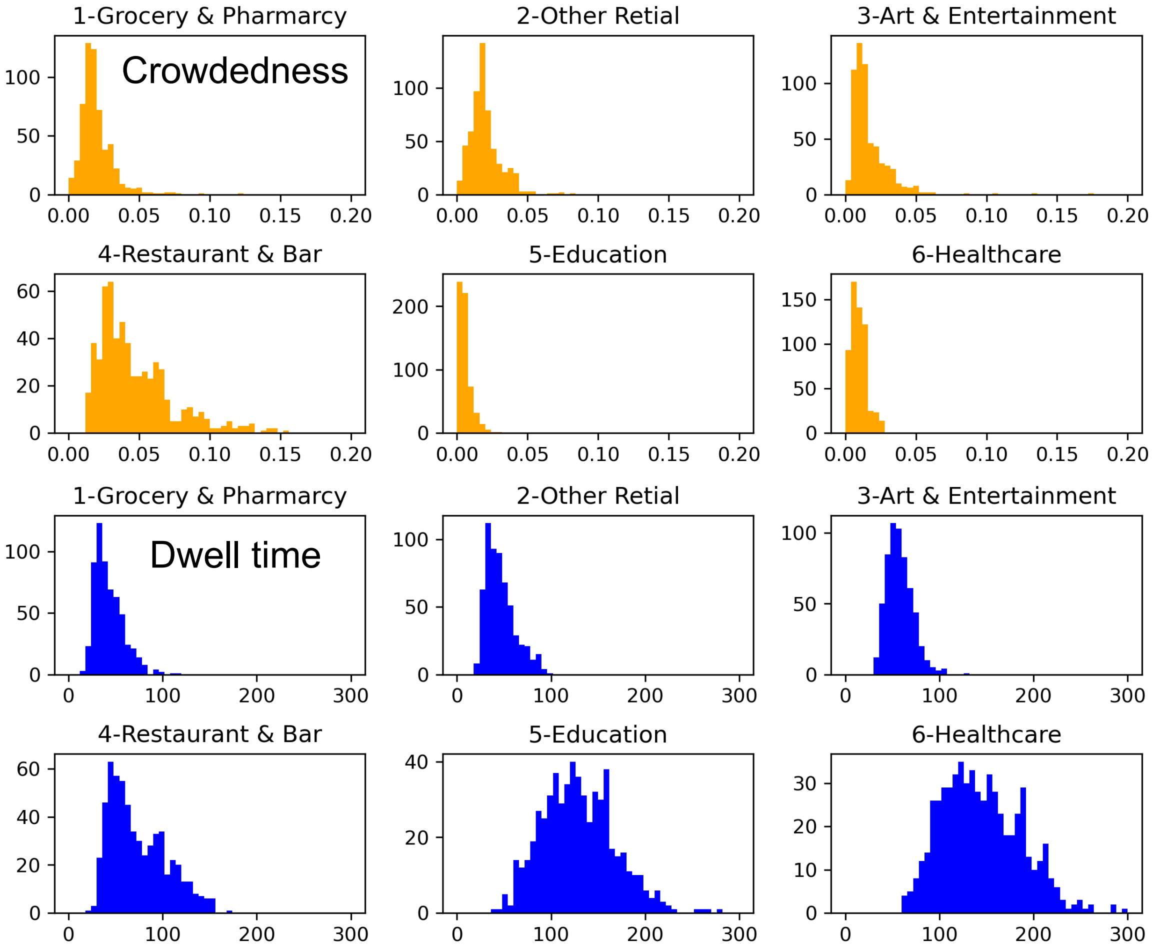
Restaurant & Bar Locations, New York City 2020





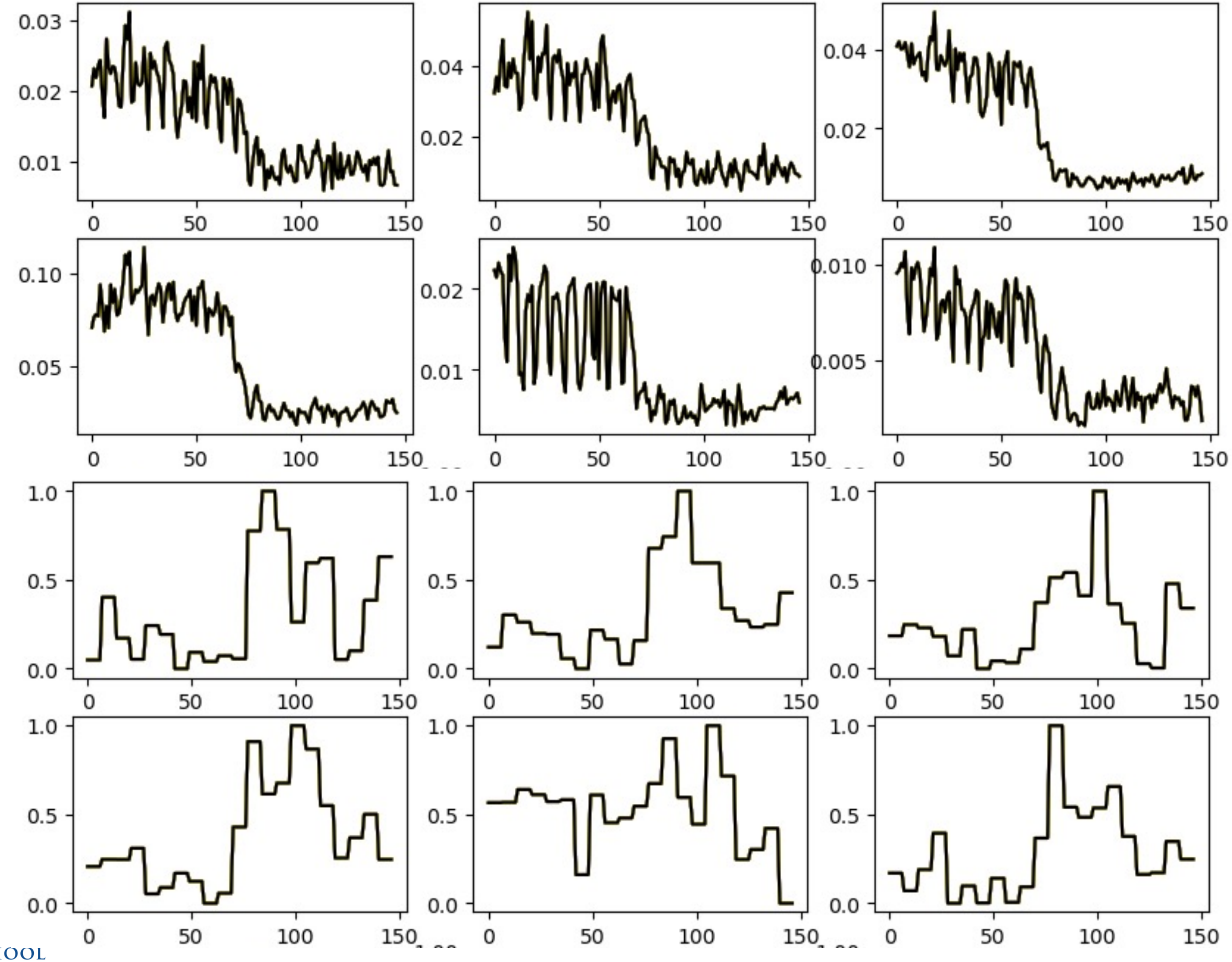


Grocery & Pharmacy, 1/31/2020, Friday


1-Grocery & Pharmarcy
Crowdedness
2-Other Retial
3-Art & Entertainment
4-Restaurant & Bar
5-Education
6-Healthcare
1-Grocery & Pharmarcy
Dwell time
2-Other Retial
3-Art & Entertainment
4-Restaurant & Bar
5-Education
6-Healthcare

Upper East Side

Crowdedness



Dwell time

Ongoing research

- Link behavior features with mobility change in different neighborhoods
  - Temporal discounting, loss aversion, agency, normative decisions
- Develop a parsimonious model for NYC informed by foot-traffic data
  - Represent population mixing in different settings
- Couple risk-driven behavior model with mobility-driven epidemic model
  - Feedback, retrospective forecasting

Reflections

- More complex models do not necessarily work better in real-world applications
  - Imperfect data, unrealistic assumptions, high computational cost
  - Complexity versus Parsimony

- Real-world data are imperfect
  - Underreporting, reporting delay, large observational noise, non-stationary
  - Develop methods to deal with imperfect data
  - Avoid perfectionism

- Communicate with end-users
  - Understand real needs, think about how model will be used
  - Push for better data collection

Acknowledgement

- Columbia: Jeffrey Shaman, Wan Yang, Kai Ruggeri, Sasikiran Kandula, Teresa Yamana, Marta Galanti, Charles Branas, Andrew Rundle, et al.
- NYC DOHMH, Test & Trace: Theodore Long, Jay Varma, Steffen Foerster, et al.
- China: Ruiyun Li, Zhanwei Du, Xiao-Ke Xu, Lin Wang, Renquan Zhang, et al.
- CDC FluSight challenge, COVID-19 forecasting hub, COVID-19 scenario hub

- Funding

NSF

CDC
CENTERS FOR DISEASE™
CONTROL AND PREVENTION

NIH
National Institutes of Health

DEFENSE THREAT REDUCTION AGENCY
UNITED STATES OF AMERICA

CSTE
COUNCIL OF STATE AND TERRITORIAL EPIDEMIOLOGISTS

Thank you!

sp3449@cumc.columbia.edu

www.columbia.edu/~sp3449/

@SenPei_CU

Columbia | Mailman School of Public Health