Spatiotemporal Dynamics, Nowcasting and Forecasting of COVID-19 in the US

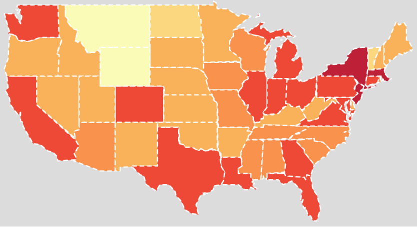
Lily Wang

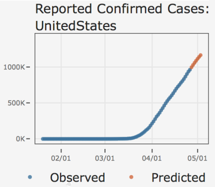
Department of Statistics
Iowa State University

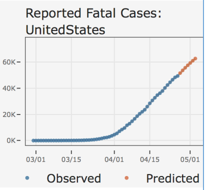
May 1, 2020

Background Introduction

Until April 29, 2020		
State	**Infection**	**Death**
United States	**1009258**	**52945**
New York	295137	17638
New Jersey	113856	6442
Massachusetts	58302	3153
Illinois	48102	2132
California	46570	1884
Pennsylvania	45323	2092
Michigan	39234	3566
Florida	32838	1170
Louisiana	27286	1758
Texas	26865	738
Connecticut	26312	2089
Georgia	23607	1022
Maryland	20113	929





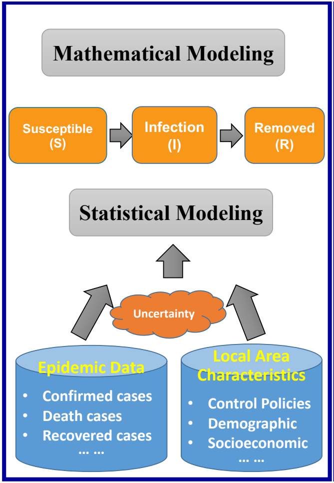


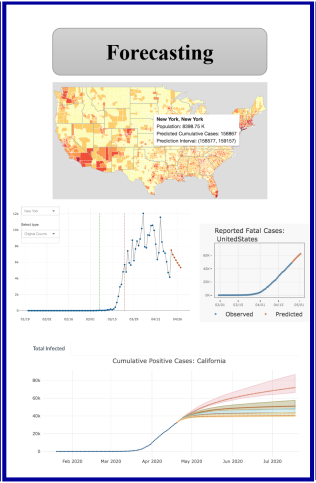
- 12-31-2019: WHO says mysterious pneumonia sickening dozens in China
- 01-11-2020: China reports 1st novel coronavirus death
- 01- 21,-2020: 1st confirmed case in the United States
- 01- 23-2020: China imposes strict lockdown in Wuhan
- 01-30-2020: WHO declares global health emergency
- 02-05-2020: Diamond Princess cruise ship quarantined
- 02-26-2020: 1st case of suspected local transmission in United States
- 03-03-2020: CDC lifts restrictions for virus testing
- 03-13-2020: Trump declares national emergency
- 03-15-2020: CDC warns against large gatherings
- 03-17-2020: Coronavirus now present in all 50 states
- 03-17-2020: Northern Californians ordered to “shelter in place”
- 03-20-2020: New York City declared US outbreak epicenter
- 03-26-2020: United States leads the world in COVID-19 cases
- 04-02-2020: Global cases hit 1 million

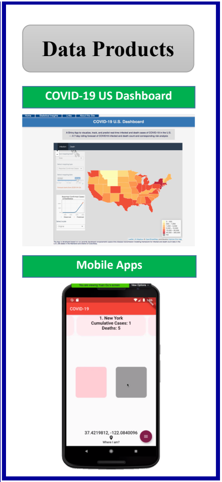
Research Goals

- **Goal 1.** Develop a dynamic epidemic modeling framework to study the spatial-temporal pattern of the spread of COVID-19.
- **Goal 2.** Investigate how factors contribute to the spread of COVID-19.
- **Goal 3.** Estimate and forecast the spatial-temporal pattern of the spread of the virus in the US up to the county level.
- **Goal 4.** Provide a user-friendly tool to visualize, track and predict the infected and death cases of COVID-19 in the US.

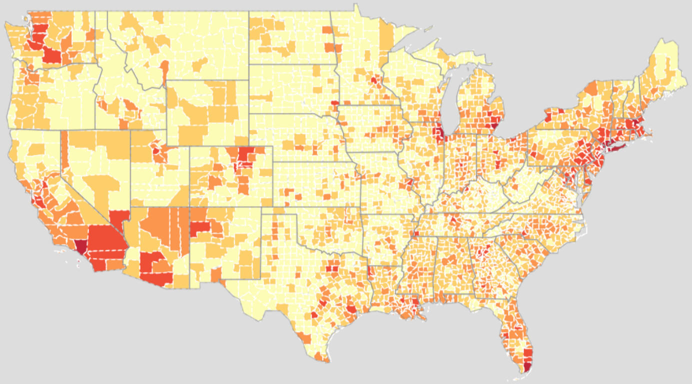
A Summary of Our Research and Products







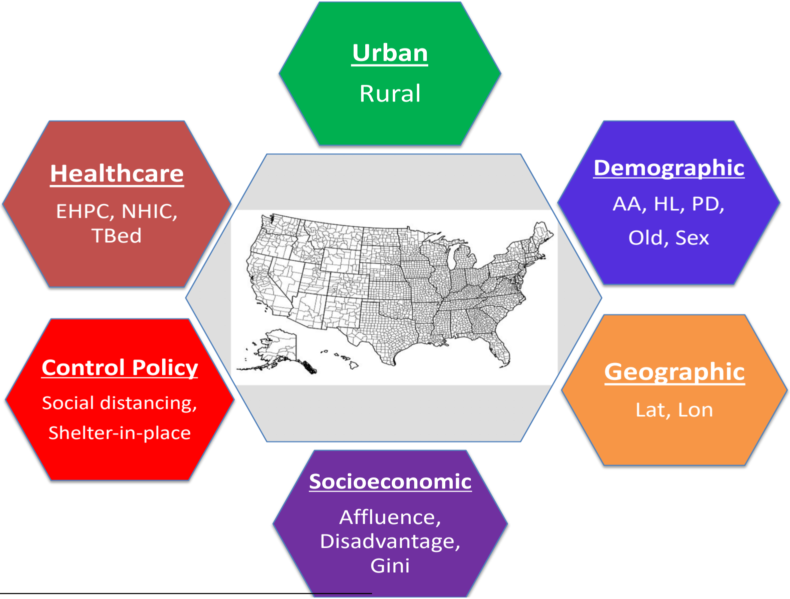
County-level Epidemic Data [1]



- 48 mainland U.S. states and the District of Columbia;
- 3,104 counties in total;
- Reported cases: infection, death, recovery.

[1] Health Department Websites, NYT, COVID-19 Data Repository by JHU CSSE, COVID Tracking Project.

County-level Features [2]



[2] U.S. Census Bureau and U.S. Department of Homeland Security.

Challenges

MODELING

- The underlying disease transmission process is unobservable.
- There is a lot of uncertainty about what is observed.
- Contributions of the factors are unknown.
- The dynamics of the spread is highly nonlinear and complex.

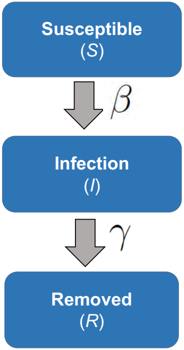
FORECAST

- Can we provide an accurate short-term forecast?
- How far the virus will spread and how many lives it will claim?
- Can we project the timing of the outbreak peak and the number of health resources required at a peak?
- What is the uncertainty associated with the forecast?

Spatio-Temporal Epidemic Modeling (STEM)

Kermack & Mckendrick (1927)

Bilinear Incidence Rates



$$\boxed{\frac{dS_t}{dt} = -\beta_t I_t S_t}$$

$$\boxed{\frac{dI_t}{dt} = -\beta_t I_t S_t - \gamma_t I_t}$$

$$\boxed{\frac{dR_t}{dt} = \gamma_t I_t}$$

SIR/SEIR related papers for COVID-19: Pan, et al. (2020), Sun, et al. (2020), Wang, et al. (2020), Zhang, et al. (2010), and others.

Kermack & Mckendrick (1927)

Bilinear Incidence Rates

Liu, *et al.* (1987)

Nonlinear Incidence Rates

Susceptible (S)

$$\boxed{\frac{dS_t}{dt} = -\beta_t I_t S_t} \xrightarrow[\alpha_2]{\alpha_1} \boxed{\frac{dS_t}{dt} = -\beta_t I_t^{\alpha_{1t}} S_t^{\alpha_{2t}}}$$

$\Downarrow \beta$

Infection (I)

$$\boxed{\frac{dI_t}{dt} = -\beta_t I_t S_t - \gamma_t I_t}$$

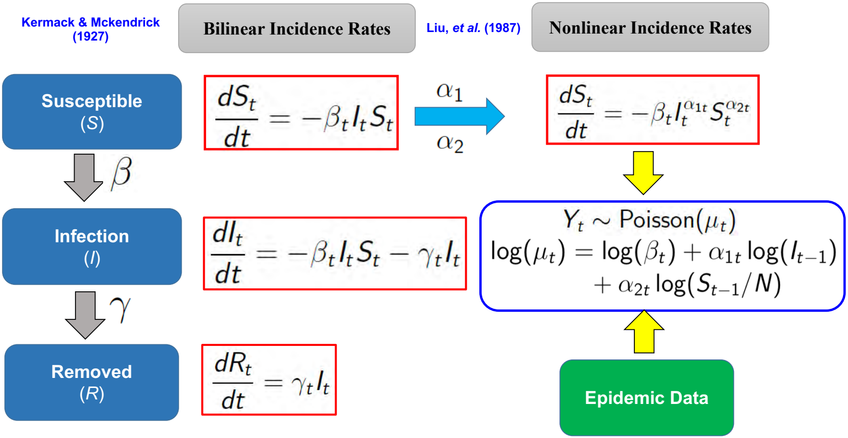
$\Downarrow \gamma$

Removed (R)

$$\boxed{\frac{dR_t}{dt} = \gamma_t I_t}$$

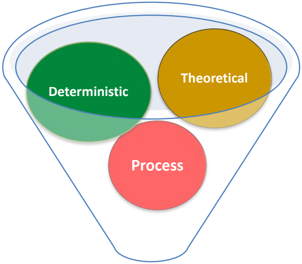
SIR/SEIR related papers for COVID-19: Pan, et al. (2020), Sun, et al. (2020), Wang, et al. (2020), Zhang, et al. (2010), and others.

SIR Models



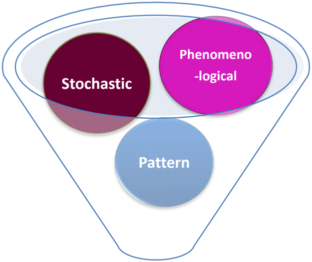
SIR/SEIR related papers for COVID-19: Pan, et al. (2020), Sun, et al. (2020), Wang, et al. (2020), Zhang, et al. (2010), and others.

Mathematical Models vs. Statistical Models



Mathematical Models

- Explicitly model nonlinearities in the process
- Represent the average behavior
- Focus is on model form, not parameter estimation for observed data



Statistical Models

- Data speak for themselves
- Provide a variety choices of errors
- Often describe the observed data
- Focus is on the pattern, little information about the mechanism

An Interface between Mathematical and Statistical Models

We investigate the disease dynamics by working at the interface of theoretical models and empirical data by combining the advantages of mathematical and statistical models.

Temporal Structure

$$Y_t \sim \text{Poisson}(\mu_t)$$
$$\log(\mu_t) = \beta_{1t} + \beta_{2t}\log(I_{t-1}) + \alpha_t \log(S_{t-1}/N)$$

Epidemic Data

Spatiotemporal Structure

$$Y_{it} \sim \text{Poisson}(\mu_{it})$$
$$\log(\mu_{it}) = \beta_{1t}(\mathbf{U}_i) + \beta_{2t}(\mathbf{U}_i)\log(I_{i,t-1}) + \alpha_t \log(S_{i,t-1}/N_i)$$

Area-level Epidemic Data
(U_i: spatial coordinates)

Spatiotemporal + Covariates

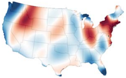
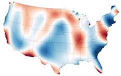
$$Y_{it} \sim \text{Poisson}(\mu_{it})$$
$$\log(\mu_{it}) = \beta_{1t}(\mathbf{U}_i) + \beta_{2t}(\mathbf{U}_i)\log(I_{i,t-1}) + \alpha_t \log(S_{i,t-1}/N_i) + \text{covariate effect}_i$$

- **Area-level Epidemic Data**
- **Control Measures**
- **Local Factors**

Spatio-Temporal Epidemic Modeling (STEM)

Suppose there are n counties. For the ith county on day t, we assume that the new increased infection cases:

$$Y_{it} | I_{i,t-1}, Z_{i,t-1}, \mathbf{A}_{i,t-r}, \mathbf{X}_i, \mathbf{U}_i \sim \mathrm{Poisson}(\mu_{it}),$$

$$\log(\mu_{it}) = \beta_{0t}(\mathbf{U}_i) + \beta_{1t}(\mathbf{U}_i)\log(I_{i,t-1})$$

$Z_{it} = \log(S_{it}/N_i)$

$$+\alpha_{0t} Z_{i,t-1} + \sum_{j=1}^{p} \alpha_{jt} A_{ij,t-r}$$

Time-varying Covariates (Control measures)

$$+\sum_{k=1}^{q} \gamma_{kt}(X_{ik})$$

County-level Covariates (Socioeconomic, demographic, … …)

with $\mathrm{E}(\gamma_{kt}) = 0$, $k = 1, \ldots, q$, for model identifiability.

STEM: Estimation

Moving Window Penalized Quasi-likelihood Method

For the current time t, and the estimation window $[t - t_0, t]$, we maximize the penalized quasi-likelihood:

$$\sum_{i=1}^{n} \sum_{s=t-t_0}^{t} L\left[g^{-1}\left\{\beta_0(\mathbf{U}_i) + \beta_1(\mathbf{U}_i)\log(I_{i,s-1}) + \alpha_0 Z_{i,s-1} \right.\right. \tag{1}$$

$$\left.\left. + \sum_{j=1}^{p} \alpha_j A_{ij,s-r} + \sum_{k=1}^{q} \gamma_k(X_{ik})\right\}, Y_{is}\right] - \frac{1}{2}\left\{\lambda_0 \mathcal{E}(\beta_0) + \lambda_1 \mathcal{E}(\beta_1)\right\},$$

where the energy functional is defined as:

$$\mathcal{E}(\beta) = \int_{\Omega} \{(\nabla_{u_1}^2 \beta)^2 + 2(\nabla_{u_1}\nabla_{u_2}\beta)^2 + (\nabla_{u_2}^2 \beta)^2\}\, du_1 du_2.$$

Moving Window Penalized Quasi-likelihood Method

Using spline basis expansion (Wang, et al., 2020 [3]) with smoothness constraints $\mathbf{H}\boldsymbol{\theta}_\ell = \mathbf{0}$, $\ell = 0, 1$, ($\boldsymbol{\theta}_\ell = \mathbf{Q}_2\boldsymbol{\theta}_\ell^*$), the penalized quasi-likelihood (1) can be changed to:

$$-\sum_{i=1}^{n}\sum_{s=t-t_0}^{t} L\left[g^{-1}\left\{\mathbf{B}(\mathbf{U}_i)^\top \mathbf{Q}_2(\boldsymbol{\theta}_0^* + \boldsymbol{\theta}_1^* \log(I_{i,s-1})) + \alpha_0 Z_{i,s-1} + \sum_{j=1}^{p}\alpha_j A_{ij,s-r} + \sum_{k=1}^{q}\boldsymbol{\Phi}_k^\top(X_{ik})\boldsymbol{\xi}_k\right\}, Y_{is}\right] + \frac{1}{2}\sum_{\ell=0}^{1}\left\{\lambda_\ell \boldsymbol{\theta}_\ell^{*\top}\mathbf{Q}_2^\top \mathbf{P}\mathbf{Q}_2\boldsymbol{\theta}_\ell^*\right\}.$$

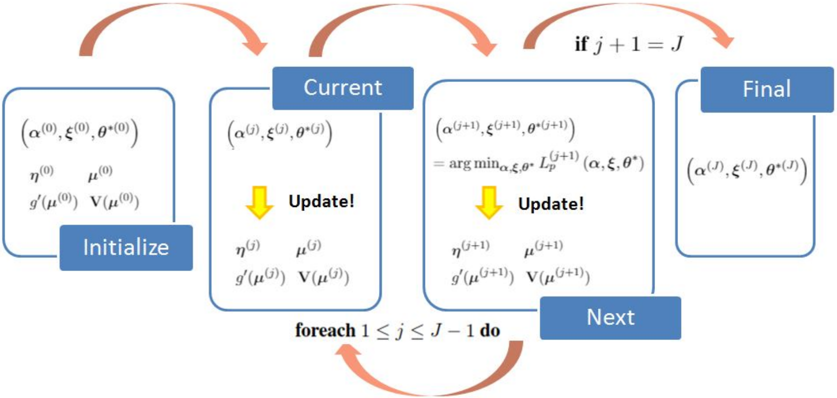
We obtain the estimators of α_j, $\beta_\ell(\cdot)$, and $\gamma_k(\cdot)$:

- $\widehat{\alpha}_{jt}$, $j = 0, \ldots, p$.
- $\widehat{\beta}_{\ell t}(\boldsymbol{u}) = \mathbf{B}(\boldsymbol{u})^\top \mathbf{Q}_2\widehat{\boldsymbol{\theta}}_{\ell t}^*$, $\ell = 0, 1$,
- $\widehat{\gamma}_{kt}(x_k) = \boldsymbol{\Phi}_k(x_k)^\top \widehat{\boldsymbol{\xi}}_{kt}$, $k = 1, \ldots, q$.

[3] Check our R packages: `Triangulation` and `BPST` to generate splines over triangulation. https://github.com/funstatpackages

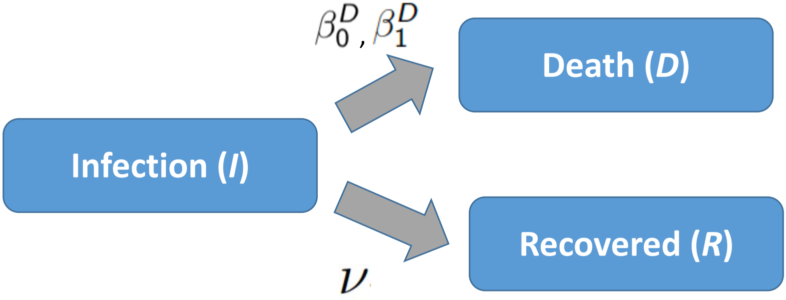
Spline Approximation and PIRLS Algorithm

- The optimization can be done via the penalized iteratively reweighted least squares (PIRLS):



- Wood (2015), Yu et al. (2019) and Kim and Wang (2020).

Modeling the Number of Fatal Cases



Modeling the Number of Fatal Cases

Temporal Structure

$$Y_t^D \sim \text{Poisson}(\mu_t^D)$$
$$\log(\mu_t^D) = \beta_{1t}^D + \beta_{2t}^D \log(I_{t-1})$$

Epidemic Data

Spatiotemporal Structure

$$Y_{it}^D \sim \text{Poisson}(\mu_{it}^D)$$
$$\log(\mu_{it}^D) = \beta_{1t}^D(\mathbf{U}_i) + \beta_{2t}^D(\mathbf{U}_i) \log(I_{i,t-1})$$

Area level Epidemic Data
(U_i : spatial coordinates)

Spatiotemporal + Covariates

$$Y_{it}^D \sim \text{Poisson}(\mu_{it}^D)$$
$$\log(\mu_{it}^D) = \beta_{1t}^D(\mathbf{U}_i) + \beta_{2t}^D(\mathbf{U}_i) \log(I_{i,t-1})$$
$$+ \text{ covariate effect}_i$$

- Area level Epidemic Data
- Control Measures
- Local Factors

- Suppose C_{it}, D_{it}, and R_{it} are the total confirmed cases, fatal cases and recovered cases, respectively.

- The number of active cases is: $I_{it} = C_{it} - D_{it} - R_{it}$.

- Let $Y_{it}^{D} = D_{it} - D_{i,t-1}$ be the new fatal cases on day t.

- **Death Model:**

$$Y_{it}^{D} | \mathbf{X}_i, \mathbf{U}_i, I_{i,t-1}, \mathbf{A}_{i,t-r} \sim \text{Poisson}(\mu_{it}^{D}),$$

$$\log(\mu_{it}^{D}) = \beta_{0t}^{D}(\mathbf{U}_i) + \beta_{1t}^{D} \log(I_{i,t-1}) + \sum_{j=1}^{p} \alpha_{jt}^{D} A_{ij,t-r} + \sum_{k=1}^{q} \gamma_{kt}^{D}(X_{ik}).$$

Modeling the Number of Recovered Cases

- Limit of recovered data during the disease spread.
- **Q**. How do you use the information on the recovered cases?

  A. Compartmental models in epidemilology (Anastassopoulou et al. 2020; Siettos and Russo 2013).
- Suppose that ν_t is the recovery rate (estimate or prior medical studies).
- **Recovery Model:**

$$\Delta R_{is} = \nu_t I_{i,s-1} + \varepsilon_{is}, \; s = t - t_0, \ldots, t.$$

Zero-Inflated Models at the Early Stage of Outbreak

- Early in an epidemic, there are many counties with zero daily new infections (Y_{it}) and new deaths (Y_{it}^D).

- Assume the observed counts $Y_{it}^* = Y_{it}$ or Y_{it}^D contributes to a Zero-Inflated Poisson (**ZIP**) distribution as follows:

$$P(Y_{it}^* = y^* | I_{i,t-1}, Z_{i,t-1}, \mathbf{A}_{i,t-r}, \mathbf{X}_i, \mathbf{U}_i) = \begin{cases} 1 - p_{it}^*, & y^* = 0, \\ p_{it}^* \frac{(\mu_{it}^*)^{y^*}}{\{\exp(\mu_{it}^*) - 1\} y^*!}, & y^* > 0. \end{cases}$$

- Also, $p_{it}^* = \mathrm{logit}(\eta_{it}^*)$ with $\eta_{it}^* = a_1 + \{b + \exp(a_2)\} \log(\mu_{it}^*)$ and a_1, a_2 are unknown parameters; see Wood et al. (2016).

COVID-19: Estimation and Inference

Estimation Models

STEM for infections:

$$\begin{aligned}\log(\mu_{it}) = {} & \beta_{0t}(\mathbf{U}_i) + \beta_{1t}(\mathbf{U}_i)\log(I_{i,t-1}) + \alpha_{0t}Z_{i,t-1} + \alpha_{2t}\mathrm{Control}_{i,2,t-7} \\ & + \gamma_{1t}(\mathrm{Gini}_i) + \gamma_{2t}(\mathrm{Urban}_i) + \gamma_{3t}(\mathrm{PD}_i) + \gamma_{4t}(\mathrm{Affluence}_i) \\ & + \gamma_{5t}(\mathrm{Disadvantage}_i) + \gamma_{6t}(\mathrm{Tbed}_i) + \gamma_{7t}(\mathrm{AA}_i) + \gamma_{8t}(\mathrm{HL}_i) \\ & + \gamma_{9t}(\mathrm{NHIC}_i) + \gamma_{10t}(\mathrm{EHPC}_i) + \gamma_{11t}(\mathrm{Sex}_i) + \gamma_{12t}(\mathrm{Old}_i)\end{aligned}$$

STEM for deaths:

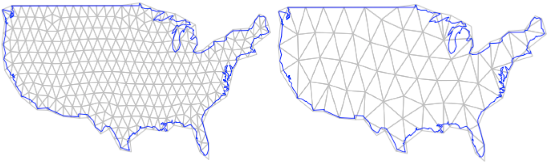
$$\begin{aligned}\log(\mu^{\mathrm{D}}_{it}) = {} & \beta^{\mathrm{D}}_{0t}(\mathbf{U}_i) + \beta^{\mathrm{D}}_{1t}\log(I_{i,t-1}) + \alpha^{\mathrm{D}}_{2t}\mathrm{Control}_{i,2,t-7} \\ & + \gamma^{\mathrm{D}}_{1t}\mathrm{Gini}_i + \gamma^{\mathrm{D}}_{2t}\mathrm{Urban}_i + \gamma^{\mathrm{D}}_{3t}\mathrm{PD}_i + \gamma^{\mathrm{D}}_{4t}\mathrm{Affluence}_i \\ & + \gamma^{\mathrm{D}}_{5t}\mathrm{Disadvantage}_i + \gamma^{\mathrm{D}}_{6t}\mathrm{Tbed}_i + \gamma^{\mathrm{D}}_{7t}\mathrm{AA}_i + \gamma^{\mathrm{D}}_{8t}\mathrm{HL}_i \\ & + \gamma^{\mathrm{D}}_{9t}\mathrm{NHIC}_i + \gamma^{\mathrm{D}}_{10t}\mathrm{EHPC}_i + \gamma^{\mathrm{D}}_{11t}\mathrm{Sex}_i + \gamma^{\mathrm{D}}_{12t}\mathrm{Old}_i.\end{aligned}$$

Estimation and Inference Settings

- Date: 03/23/20 – 04/25/20.
- Estimation window length: 9 days.

Estimation and Inference Settings

- Date: 03/23/20 – 04/25/20.
- Estimation window length: 9 days.
- Univariate splines: cubic splines, 2 interior knots.
- Bivariate splines:
  - 522 triangles, 306 vertices;
  - 119 triangles, 87 vertices.



Estimation and Inference Settings

- Date: 03/23/20 – 04/25/20.
- Estimation window length: 9 days.
- Univariate splines: cubic splines, 2 interior knots.
- Bivariate splines:
  - 522 triangles, 306 vertices;
  - 119 triangles, 87 vertices.
- Estimation: coefficients, coefficient maps, and curves of covariates.
- Inference: simultaneous confidence band (SCB).

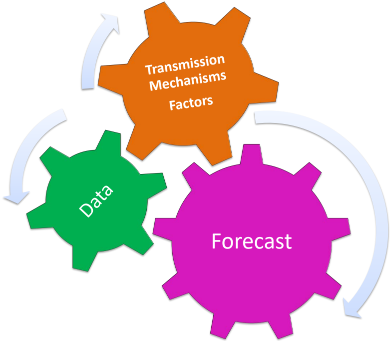
SCB: Population Density per Square Mile of Land Area

A Summary of County-level Factors

- Control Policy ("shelter-in-place") is highly significant;
- Infections increase with Population Density;
- Infections increase with African American Ratio;
- Infections increase with Hispanic Latino Ratio;
- Infections are higher in Urban areas;
- Infections are lower when there are more healthy care investments (Hospital Beds).

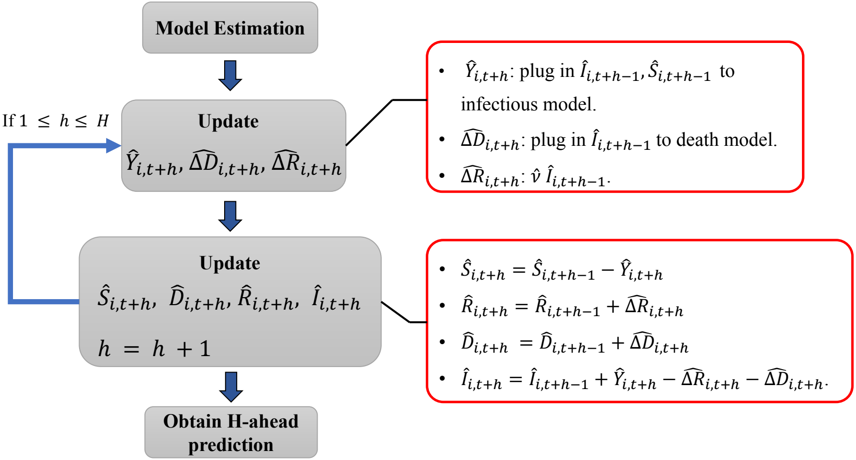
STEM: Forecasting

Forecasting COVID-19: How Difficult Is It?



1. **Short-term Forecast**
   - Clear time series trend;
   - Relatively easy, many existing methods are available;
   - Less uncertainty about what is observed.

2. **Long-term Forecast**
   - A lot of uncertainty;
   - Lack of good quality data;
   - Forecasts might affect what we are trying to forecast.


Model Estimation
If $1 \leq h \leq H$
Update
$\hat{Y}_{i,t+h}, \widehat{\Delta D}_{i,t+h}, \widehat{\Delta R}_{i,t+h}$
Update
$\hat{S}_{i,t+h}, \hat{D}_{i,t+h}, \hat{R}_{i,t+h}, \hat{I}_{i,t+h}$
$h = h + 1$
Obtain H-ahead prediction
• $\hat{Y}_{i,t+h}$: plug in $\hat{I}_{i,t+h-1}, \hat{S}_{i,t+h-1}$ to infectious model.
• $\widehat{\Delta D}_{i,t+h}$: plug in $\hat{I}_{i,t+h-1}$ to death model.
• $\widehat{\Delta R}_{i,t+h}$: $\hat{v}\, \hat{I}_{i,t+h-1}$.
• $\hat{S}_{i,t+h} = \hat{S}_{i,t+h-1} - \hat{Y}_{i,t+h}$
• $\hat{R}_{i,t+h} = \hat{R}_{i,t+h-1} + \widehat{\Delta R}_{i,t+h}$
• $\hat{D}_{i,t+h} = \hat{D}_{i,t+h-1} + \widehat{\Delta D}_{i,t+h}$
• $\hat{I}_{i,t+h} = \hat{I}_{i,t+h-1} + \hat{Y}_{i,t+h} - \widehat{\Delta R}_{i,t+h} - \widehat{\Delta D}_{i,t+h}$.

STEM: Projection Band

Estimation Variation + **Variation from Individual** → **Prediction Band**

- Generate bootstrap sample on time points $1, 2, \dots, t$.
- Estimate model based on bootstrap sample: $\hat{\boldsymbol{\beta}}^b$, $\hat{\boldsymbol{\alpha}}^b$, $\hat{\boldsymbol{\gamma}}^b$.
- Correct bias.

- Simulate a forecast path at time points $t+1, t+2, \dots, t+H$ based on $2\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}^b$, $2\hat{\boldsymbol{\alpha}} - \hat{\boldsymbol{\alpha}}^b$, $2\hat{\boldsymbol{\gamma}} - \hat{\boldsymbol{\gamma}}^b$.
- $\hat{Y}^b_{i,t+h}$, $\widehat{\Delta D}^b_{i,t+h}$, $\widehat{\Delta R}^b_{i,t+h}$ are generated from Poisson distribution.

- Repeat the bootstrap procedure for B times.
- Leave out αB the most extreme paths.
- Obtain $100(1-\alpha)\%$ prediction band.

- **Linear:** $\mathrm{E}(C_{it}|t) = \beta_{i,0} + \beta_{i1}t$, $\mathrm{Var}(C_{it}|t) = \sigma_i^2$, $i = 1, \ldots, n$;
- **Exponential, Poisson:**

$$\log\{\mathrm{E}(C_{it}|t)\} = \beta_{i0}+\beta_{i1}t,\ \mathrm{Var}(C_{it}|t) = \exp(\beta_{i0}+\beta_{i1}t),\ i = 1, \ldots, n;$$

- **Simple Epidemic Model (EM):**

$$\log(\mu_{it}) = \beta_0 + \beta_1 \log(I_{i,t-1}),\ \log(\mu_{it}^D) = \beta_0^D + \beta_1^D \log(I_{i,t-1})$$

Table: Average of root mean squared prediction errors (RMSPE_h) for the h-day ahead prediction, $h = 1, \ldots, 7$, based on 03/23–04/18, 2020.

	Method	RMSPE_1	RMSPE_2	RMSPE_3	RMSPE_4	RMSPE_5	RMSPE_6	RMSPE_7
Infection	Linear	40.332	56.581	74.074	94.038	117.661	143.440	167.763
	Exponential	>1000	>1000	>1000	>1000	>1000	>1000	>1000
	EM	41.323	69.217	97.766	130.247	166.116	199.284	236.642
	STEM	35.632	56.097	74.460	94.121	118.569	141.809	168.008
Death	Linear	6.899	9.917	13.297	16.944	21.272	25.393	29.586
	Exponential	>1000	>1000	>1000	>1000	>1000	>1000	>1000
	EM	3.799	7.322	10.405	13.617	17.221	20.304	23.282
	STEM	3.755	7.200	10.287	13.535	17.208	20.529	23.868

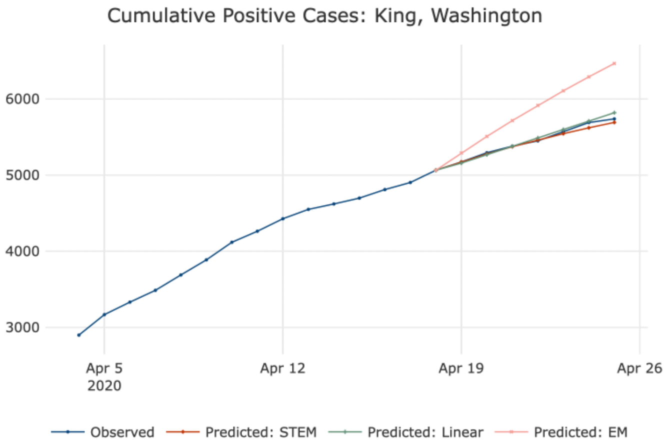


Figure: Comparison of 7-day ahead predictions using different methods.

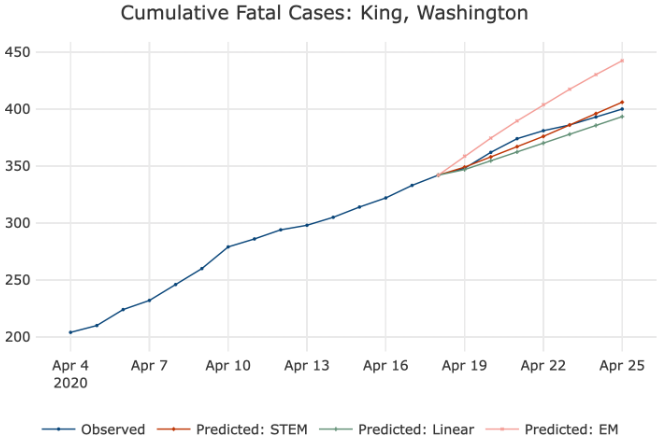
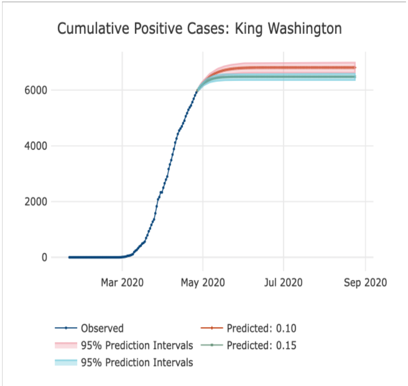
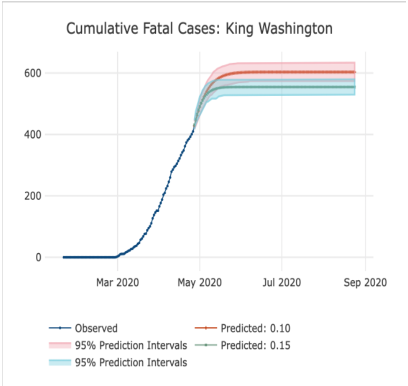


Figure: Comparison of 7-day ahead predictions using different methods.

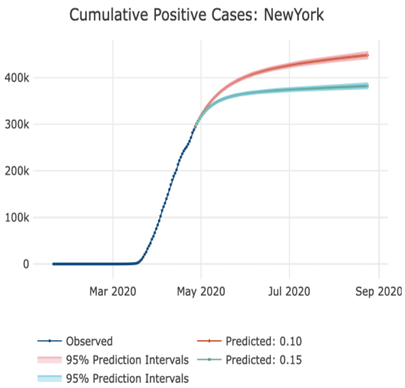
Long-term Projection at the County Level

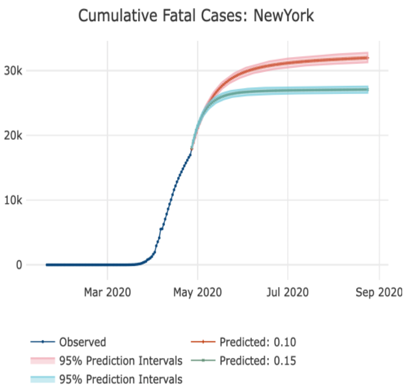




- We treat daily recovery rate as input parameters: 10% (red), 15% (green).
- As far as we know, we are the only one providing the county-level projection.

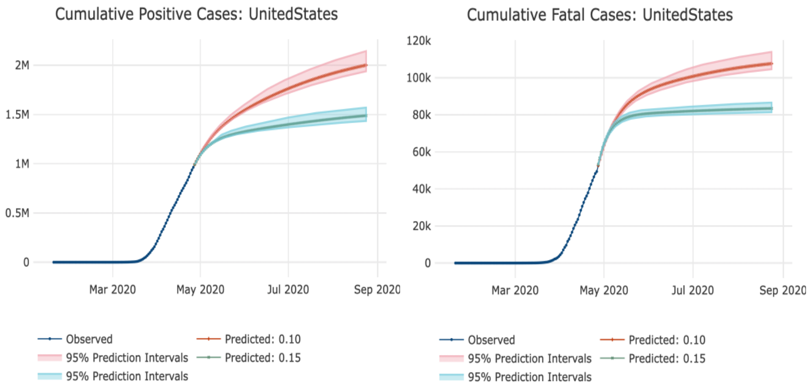
Long-term Projection at the State Level





- We treat daily recovery rate as input parameters: 10% (red), 15% (green).
- Our projection bands are much narrower than those provided by **IHME**.

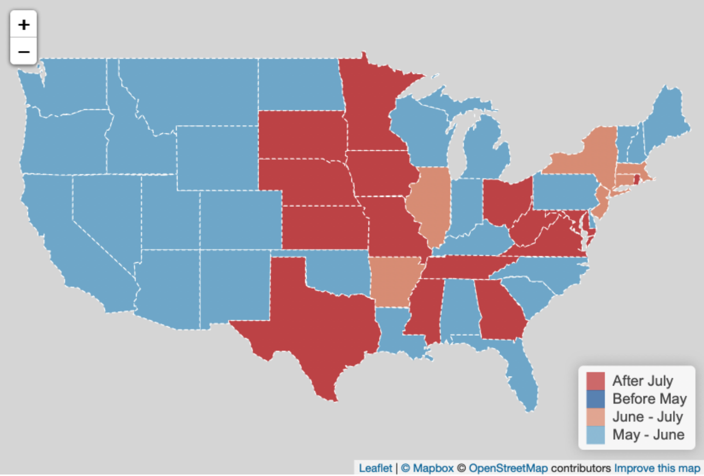
Long-term Projection for the U.S.



- We treat daily recovery rate as input parameters: 10% (red), 15% (green).
- Our projection bands are much narrower than those provided by **IHME**.

Long-term Projection

When will the COVID-19 end?

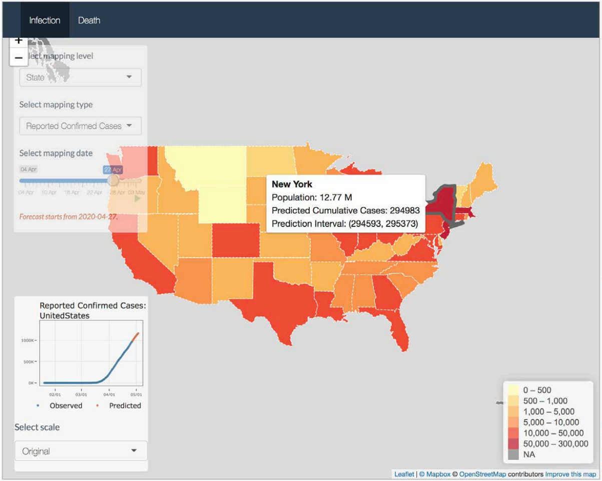


- Expected date when the fatal cases stop increasing at different states with recovery rate 10%. [Based on the data 04/18/20 – 04/26/20]

Data Products

COVID-19 Dashboard

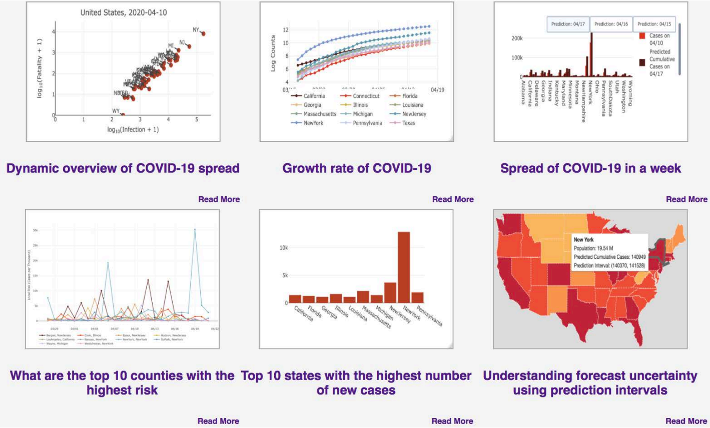
We provide a real-time 7-day forecast of infected and death counts at both the county level & state level, and the corresponding risk analysis. [https://covid19.stat.iastate.edu/]



SCAN ME

COVID-19 Dashboard – Insights

We provide some indepth statistical insights based on our analysis of COVID-19 infected/death count.
[https://covid19.stat.iastate.edu/]



SCAN ME

Conclusions

- Bridge the gap between mathematical models and statistical analysis in the infectious disease study.
- Enhance the dynamics of the SIR mechanism by means of nonparametric spatiotemporal analysis.
- Investigate the spatial associations between the infection/death count, and area-level factors/characteristics across the US.
- Can be used as an important tool for understanding the dynamic of the disease spread, as well as to assess how this outbreak may unfold through time and space.
- Provide a very accurate short-term forecast, and can also be used for long-term prediction.

Future Works

1. Methodology
   - Disease mapping: to illustrate high-risk areas, and help policy making and resource allocation.
   - Extensions and applications:
     - epidemic models in which there are several types of areas with potentially different characteristics;
     - more complex models that include features such as latent periods or more realistic population structure.
2. Data products
   - Mobile App is underdevelopment: real-time forecast up to county level;
   - Risk Analysis Apps for communities, schools, businesses and companies will be developed.

Our Products and Contact Information

- Details of our research can be found in the **arXiv paper**

  http://arxiv.org/abs/2004.14103

- The **R package** of the proposed method can be downloaded from the Github Repository:

  https://github.com/covid19-dashboard-us/STEM

- The **R shiny apps** demonstrating the proposed methods can be found from

  https://covid19.stat.iastate.edu/

- Questions, comments, suggestions: please **email** me at

  Email: lilywang@iastate.edu

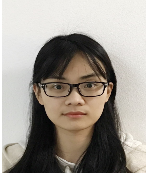
Contributors

Lily Wang
Iowa State

Guannan Wang
William&Mary

Lei Gao
Iowa State

Xinyi Li
SAMSI/UNC-CH



Shan Yu
Iowa State

Myungjin Kim
Iowa State

Yueying Wang
Iowa State

Zhiling Gu
Iowa State

Acknowledgement

- Dr. Sarah Nusser, Vice President for Research, Iowa State University (ISU);
- Dr. Dan Nettleton, Chair of Department of Statistics, ISU;
- Mr. Levi Baber and IT support at ISU;
- Colleagues from ISU, WM, SAMSI, UNC-CH, Tsinghua, Clemson, MSU, UVA, GWU, UGA, UCI, UCR, Columbia, Rutgers, TAMU, and many others;
- Users of COVID-19 US Dashboard;
- NSF 1934884 [HDR:TRIPODS: D4 (Dependable Data Driven Discovery) Institute].

To Our Healthcare Workers:

Thank you for putting yourself in the way of danger to save others and save the public!
Thanks for being heroes of this country in the pandemic!
We are with you!

References I

- Anastassopoulou, C., et al. (2020). Data-based analysis, modelling and forecasting of the COVID-19 outbreak. PLOS ONE, 15, 1–21.
- Finkenstädt, B. F. & Grenfell, B. T. (2000). Time series modelling of childhood diseases: a dynamical systems approach. JRSSC, 49, 187–205.
- Kermack, W.O. & McKendrick, A.G. (1927). A contribution to the mathematical theory of epidemics. Proceedings of the royal society of london,115, 700-721.
- Kim, M. & Wang (2020). Generalized spatillay varying coefficient models. JCGS. Accepted.
- Lai, M. J. & Schumaker, L. L. (2007). Spline Functions on Triangulations. Cambridge University Press.
- Lai, M. J. & Wang, L. (2013). Bivariate penalized splines for regression. Statistica Sinica, 23, 1399–1417.
- Lawson, A. B., et al. (2016). Handbook of spatial epidemiology. CRC Press.
- Liu, W. M., et al. (1987), Dynamical behavior of epidemiological models with nonlinear incidence rates. Journal of mathematical biology, 25, 359–380.

References II

- Mu, J., Wang, G. & Wang, L. (2018). Estimation and inference in spatially varying coefficient models. Environmetrics, 29:e2485.
- Pan, A., et al (2020). Association of public health interventions with the epidemiology of the COVID-19 outbreak in Wuhan, China. JAMA, accepted.
- Sun, H., et al. (2020). Tracking Reproductivity of COVID-19 Epidemic in China with Varying Coefficient SIR Model. Journal of Data Science, Accepted.
- Siettos, C. I. & Russo, L. (2013). Mathematical modeling of infectious disease dynamics. Virulence, 4, 295–306.
- Staszewska–Bystrova, A. (2009). Bootstrap confidence bands for forecast paths. Available at SSRN.
- Wakefield, J., Dong, T. Q., & Minin, V. N. (2019). Spatio–temporal analysis of surveillance data. Handbook of Infectious Disease Data Analysis, 455-476.
- Wood, S. N., Pya, N., and Säfken, B. (2016). Smoothing parameter and model selection for general smooth models, JASA, 111, 1548-1563.
- Wood, S. (2017). Generalized additive models: an introduction with R. CRC press.

References III

- Wang, L., et al. (2020). Efficient estimation of partially linear models for data on complicated domains by bivariate penalized splines over triangulations. Statistica Sinica. In Press.
- Wang, L., et al. (2020). An Epidemiological Forecast Model and Software Assessing Interventions on COVID-19 Epidemic in China. Journal of Data Science, accepted.
- Yu, S., Wang, G., Wang, L., Liu, C. & Yang, L. (2019). Estimation and inference for generalized geoadditive models. JASA, preprint.
- Zhang, Y., et al. (2020). Prediction of the COVID-19 outbreak based on a realistic stochastic model. medRxiv, preprint.