# Forecasting epidemiological patterns using multi-scale semi-mechanistic models

## Privacy and Ethics in Pandemic Data Collection and Processing

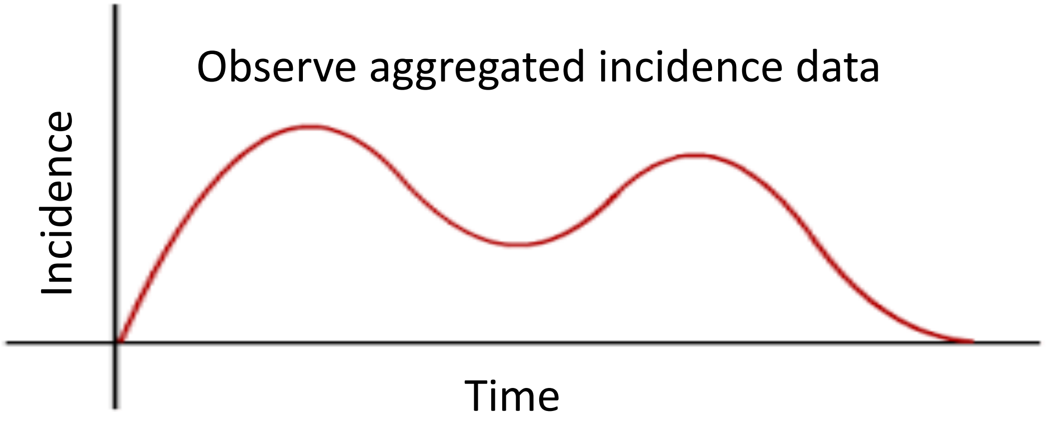
## ICERM @ Brown University

19 January 2023

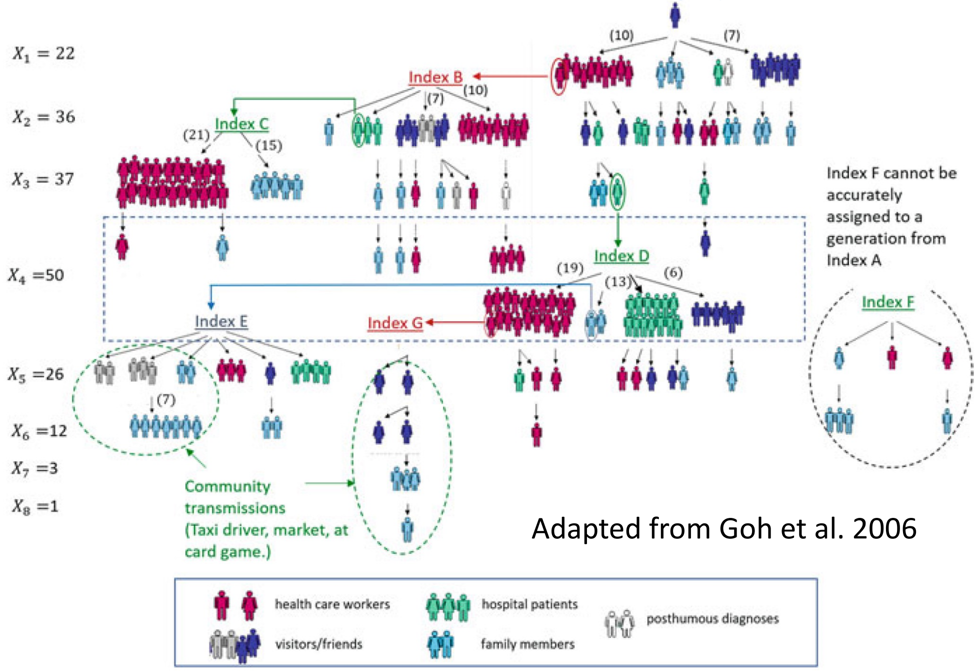
Gerardo Chowell, PhD

Georgia State University | SCHOOL OF PUBLIC HEALTH

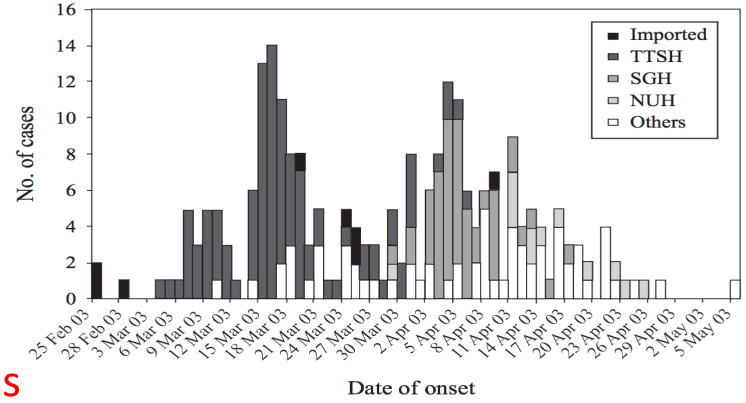
# Inferring the structure of aggregate epidemiological patterns



2003 SARS outbreak in Singapore

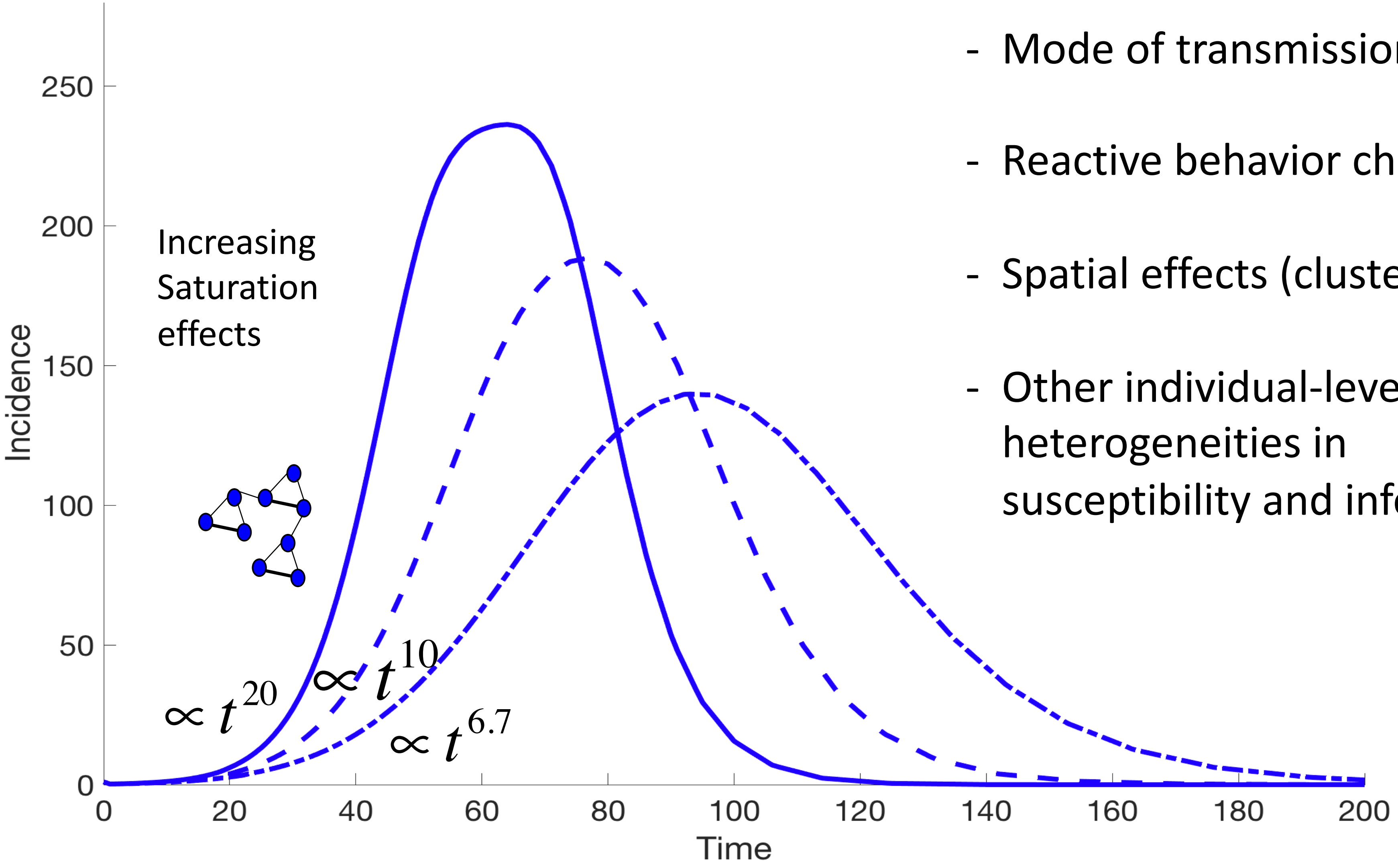


Adapted from Goh et al. 2006



More than one transmission tree could give rise to very similar epidemic curves

# Real epidemics exhibit variable epidemic growth scaling

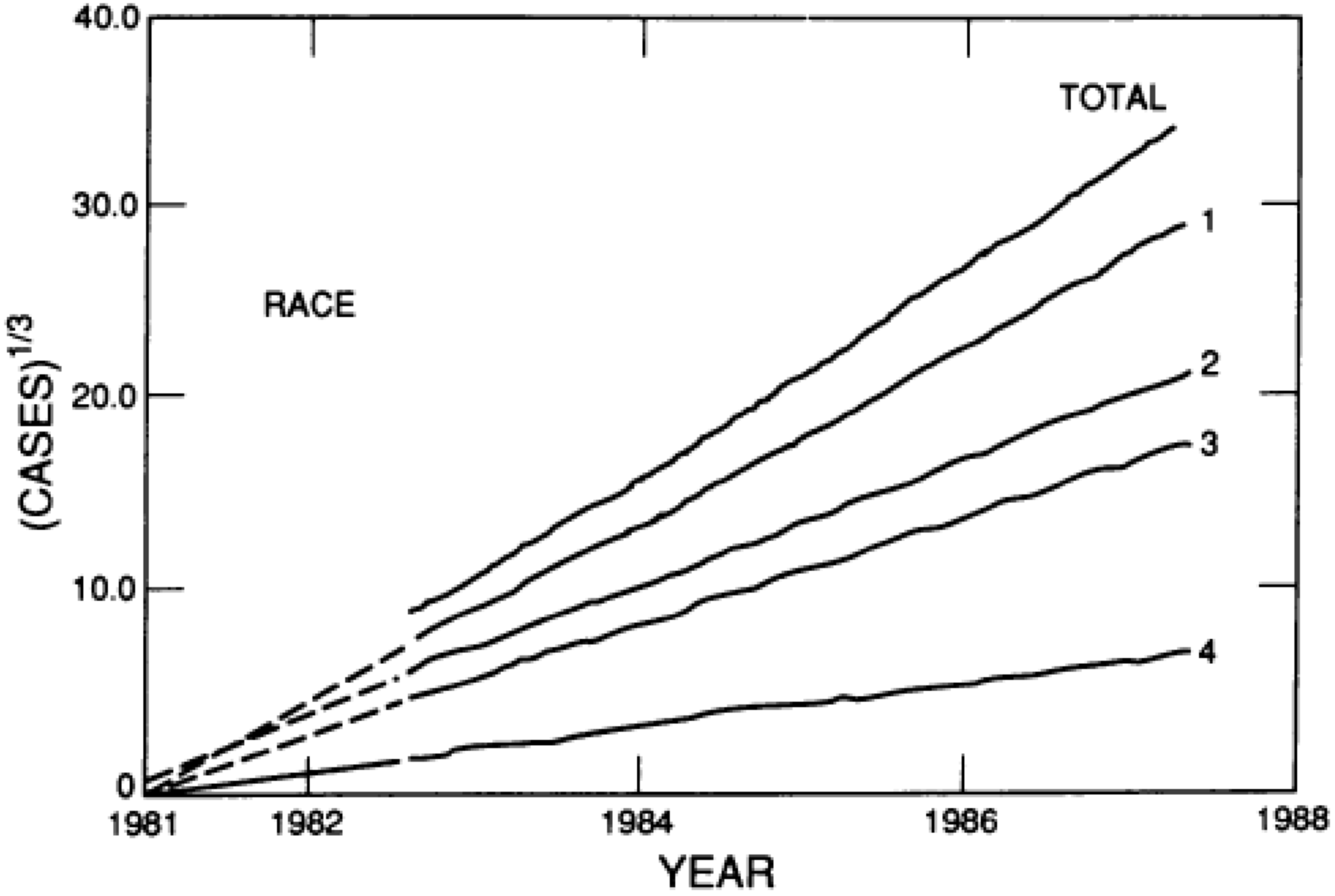


- Mode of transmission
- Reactive behavior changes
- Spatial effects (clustering)
- Other individual-level heterogeneities in susceptibility and infectivity

*Viboud et al. Epidemics 2016; Chowell et al. JR Interface 2016*

# The initial phase of the HIV/AIDS epidemic in the United States

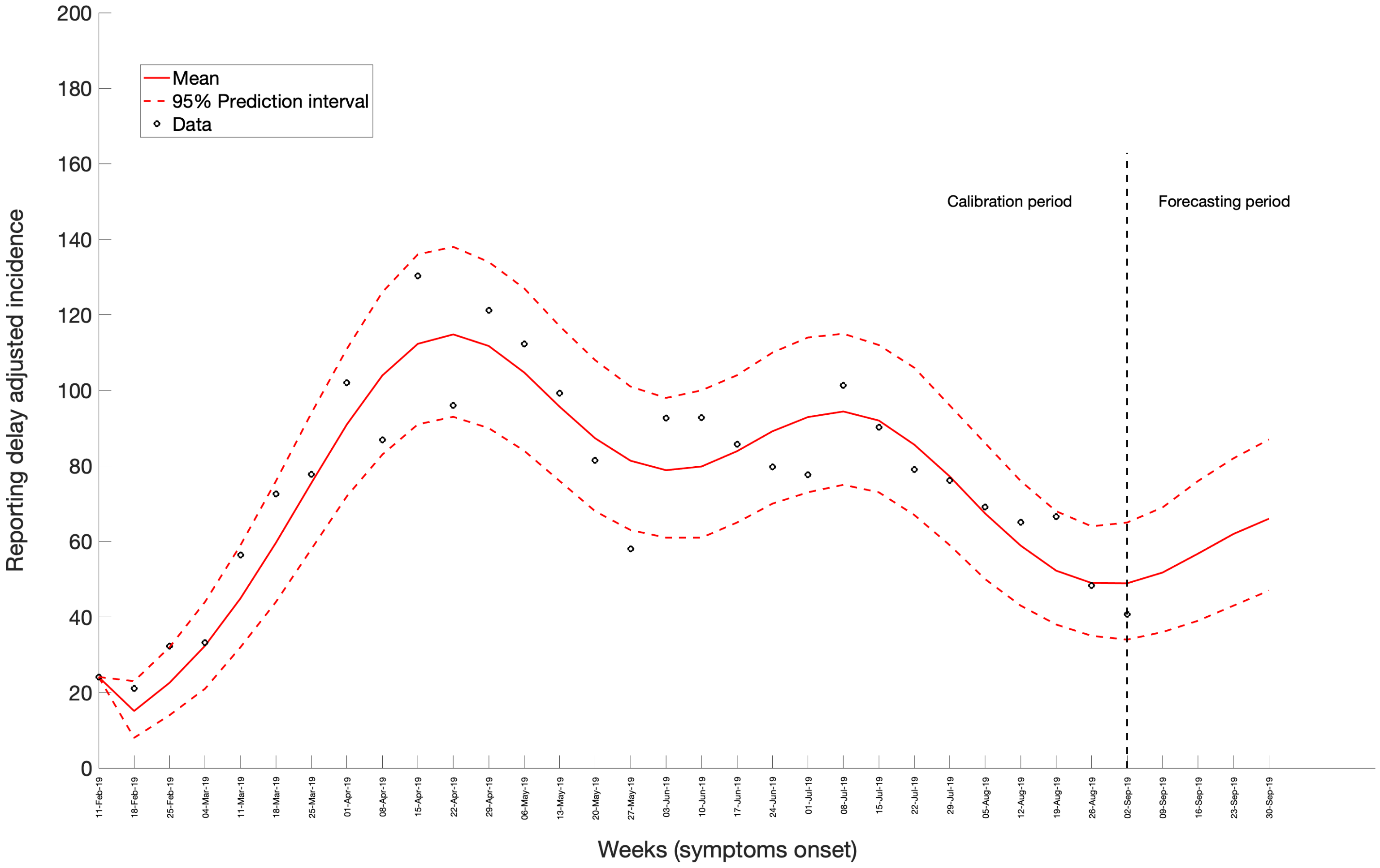
https://www.pbs.org/newshour/science/america-hiv-outbreak-origins-nyc-gaetan-dugas



**1=White; 2= Black; 3=Hispanic; 4=unknown**

*Colgate et al. PNAS 1989*

# Multimodal epidemics



- Aggregation of multiple underlying spreading mechanisms.
- Transmission occurs in high-risk groups first before affecting others.
- Transmission in different geographic areas occurs at different times.
- Emergence of new variants

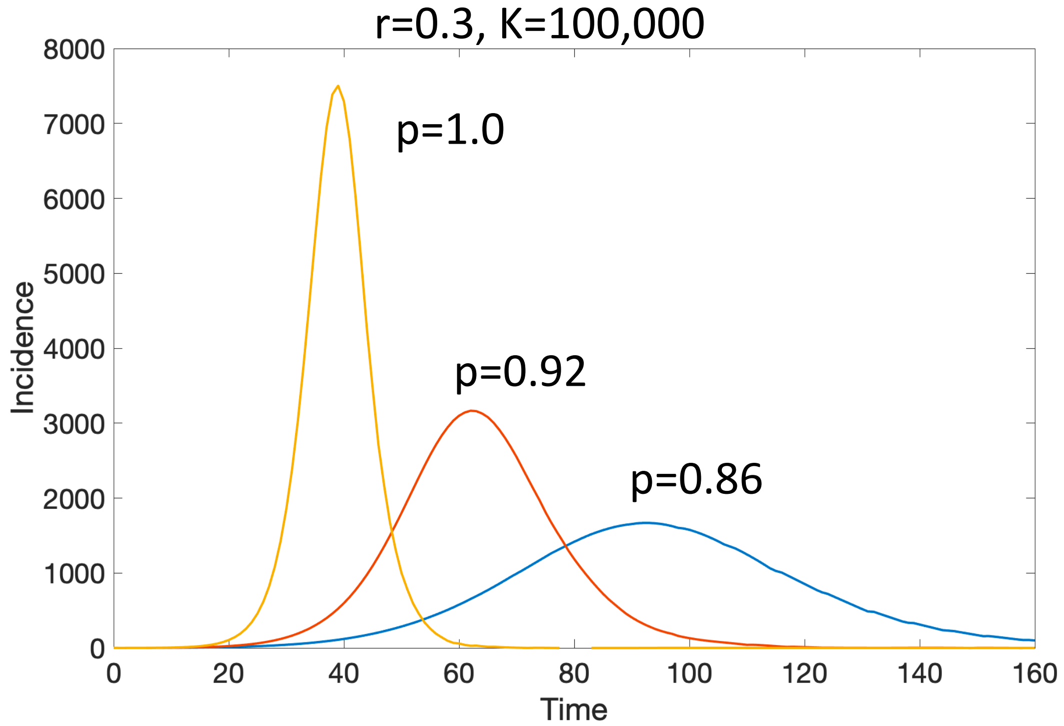
# Building block: generalized logistic growth model

**Growth rate**

**Deceleration of growth parameter**

$$C'(t) = rC^{p}(t)\left[1-\left(\frac{C(t)}{K}\right)\right]$$

**Epidemic size**



r=0.3, K=100,000

**Where:**

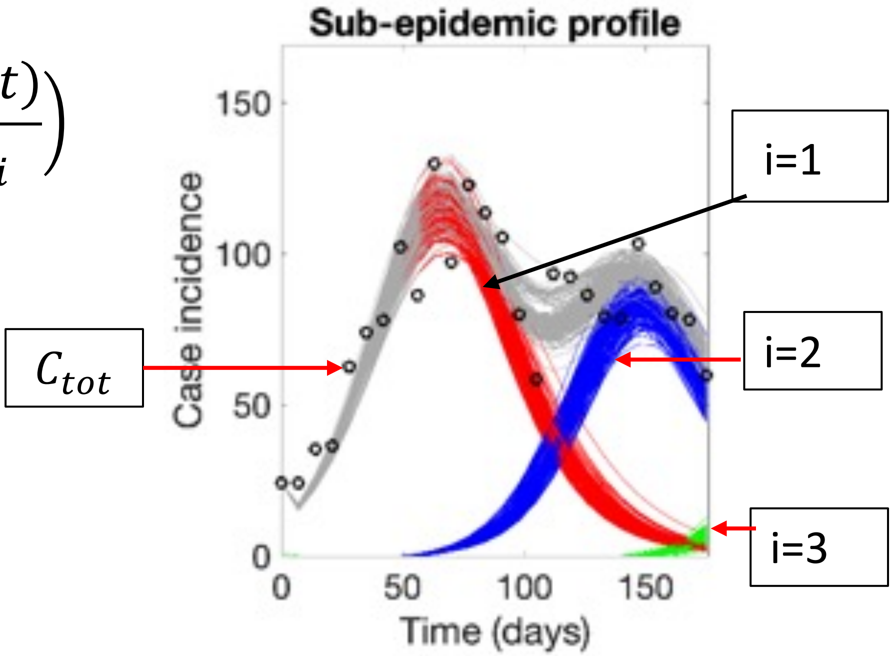
- C’(t) describes the incidence curve over time t
- r is a positive parameter denoting the growth rate
- p ∈ [0,1] is an “deceleration” growth parameter
- *K* is the final epidemic size parameter

Chowell et al. Plos Currents Outbreaks 2016.

# Overlapping sub-epidemic wave model

- Each sub-epidemic is modelled by a generalized logistic growth model
- An epidemic wave comprising of a set of *n* over lapping sub-epidemics is modelled using coupled differential equations

$$\frac{dC_i(t)}{dt} = rA_{i-1}(t)C_i(t)^p\left(1 - \frac{C_i(t)}{K_i}\right)$$



Where,

$C_i$= cumulative number of infections for sub-epidemic i,

$K_i$= size of the $\mathrm{i}_{\mathrm{th}}$ sub-epidemic

$A_i$= indicator variable

$$A_i(t)=f(x)=\begin{cases}1 & C_i(t) > C_{thr}\\ 0 & Otherwise\end{cases} \qquad \mathrm{i} = 1,2,3\ldots\ldots \mathrm{n}$$

**Number of model parameters is 5 (n>1)**

Chowell, G., Tariq, A., & Hyman, J. M. (2019). *BMC medicine*, *17*(1), 1-18.

# Modeling sub-epidemic sizes

Assuming the subsequent sub-epidemic sizes decline exponentially we have,

$$K_i = K_0 \, e^{-q(i-1)}$$

Where,

$K_0$ is the size of the initial sub-epidemic ($K_i = K_0$ )

The total size of the epidemic wave composed of n overlapping sub-epidemics is given by:

$$K_{tot} = \sum_{i=1}^{n} K_0 e^{-q(i-1)} = \frac{K_0(1 - e^{-qn})}{1 - e^{-q})}$$

Where,

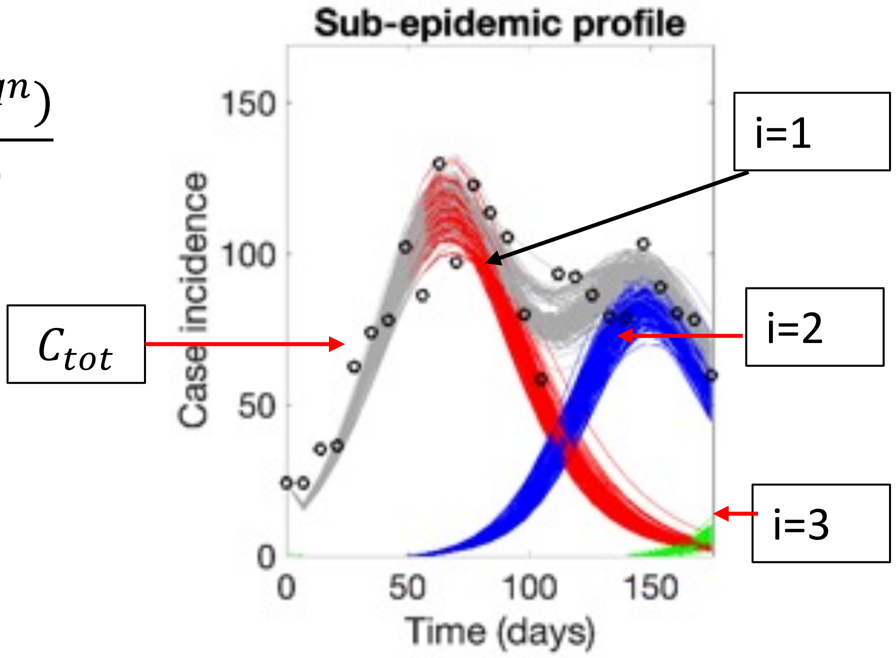
q= interventions or behavior change, q> 0 and n= number of sub-epidemics

$K_{tot} = nK_0$ , when $q$ =0

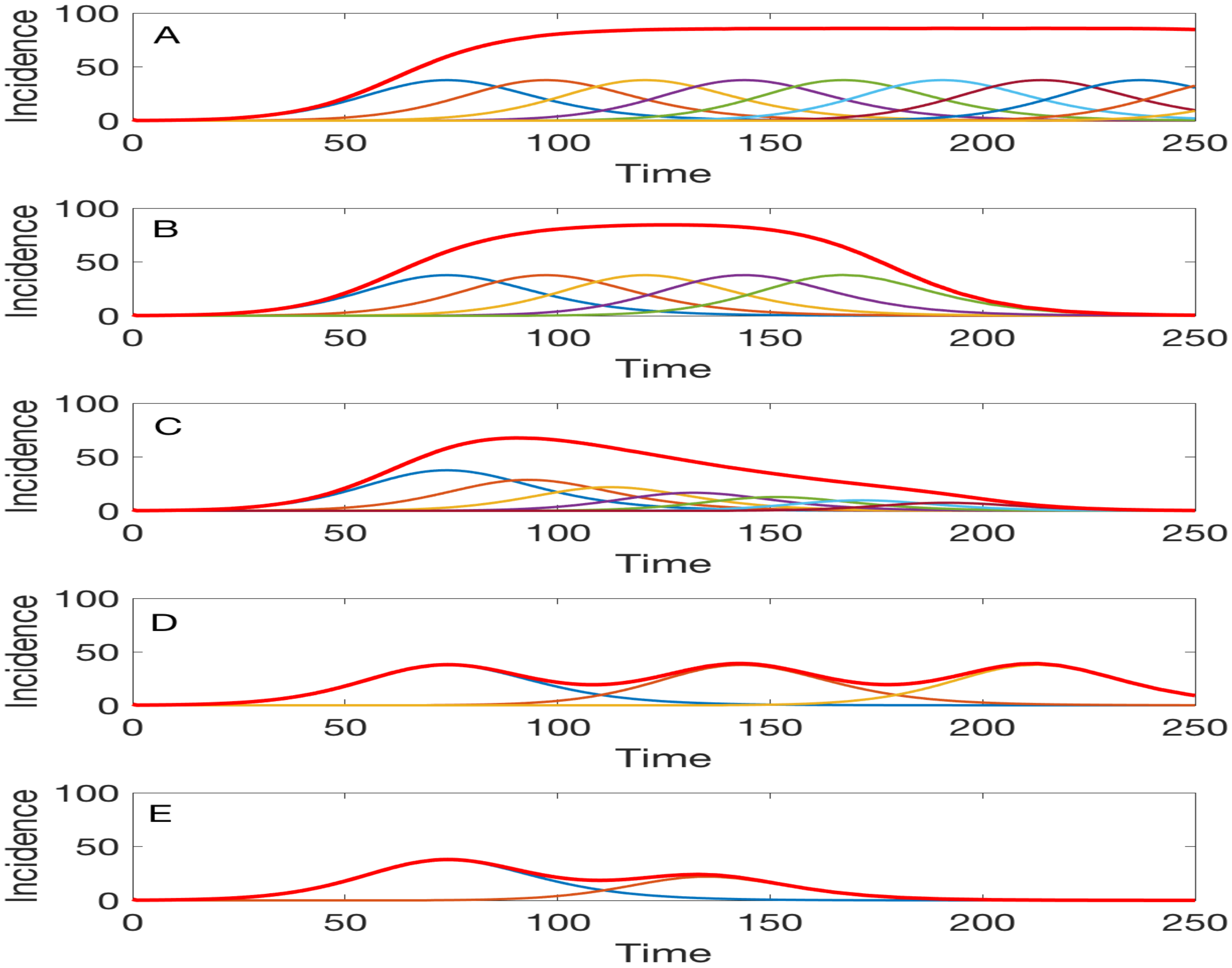
Where $K_{tot}$= total epidemic size

$$C_{tot}(t) = \sum_{i=1}^{n} C_i(t)$$

Where $C_{tot}(t)$ = cumulative curve of epidemic wave



# Representative epidemic waves composed of overlapping sub-epidemics



**Endemic state**

**Temporary endemic state**

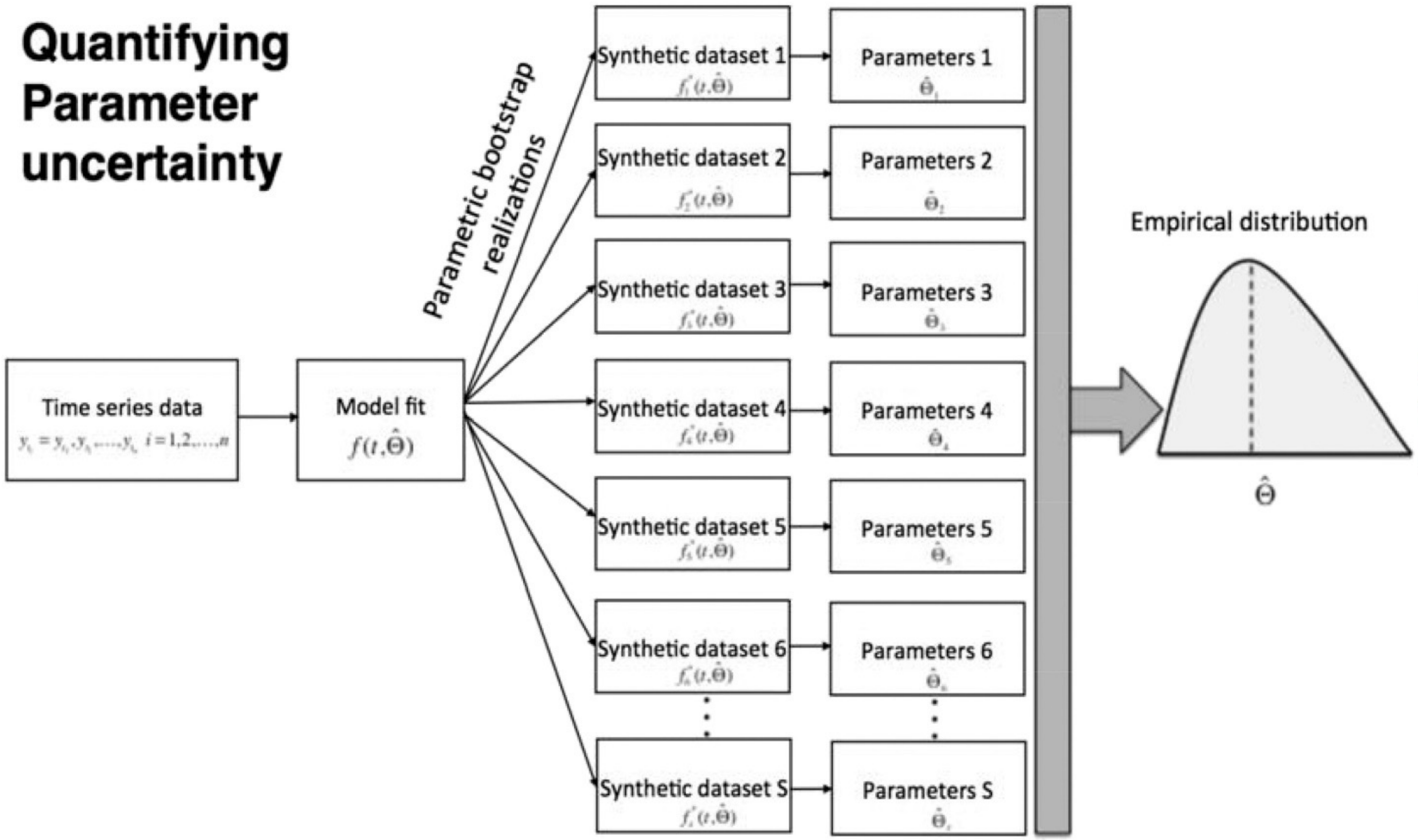
**Declining sub-epidemics**

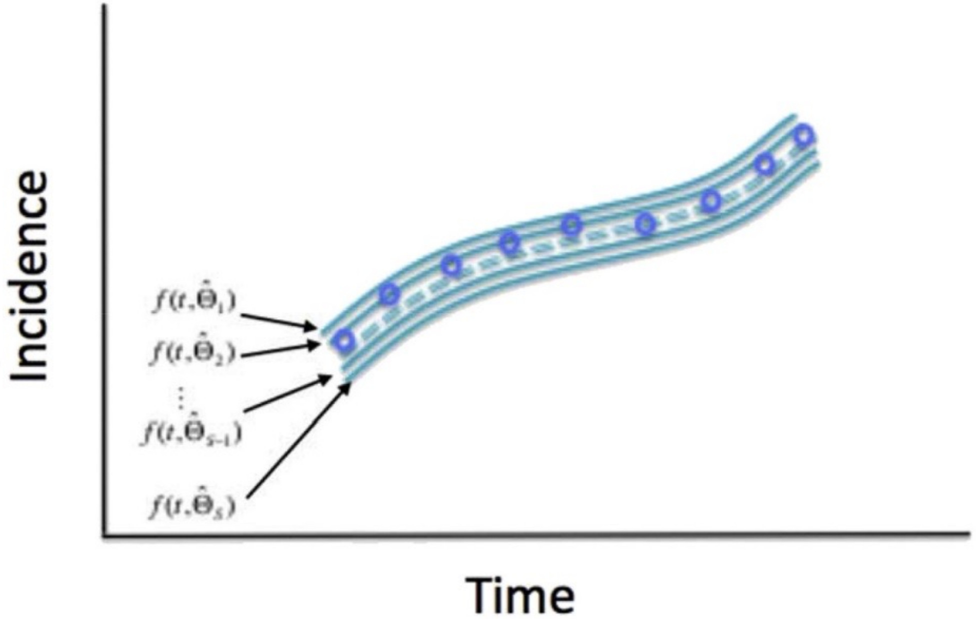
**Sustained oscillations**

**Damped oscillations**

Number of model parameters is 5 (n>1)

# Parameter estimation



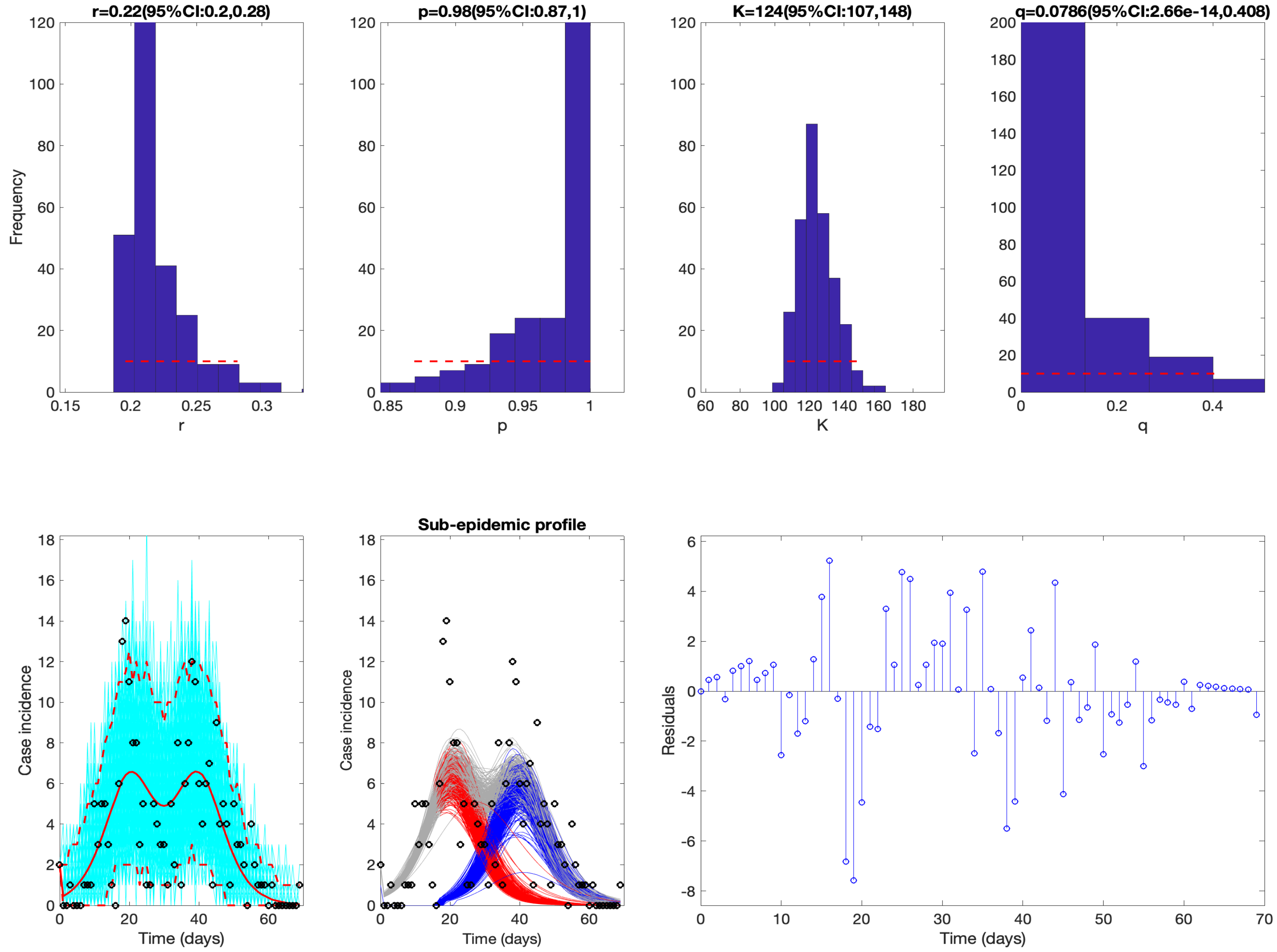


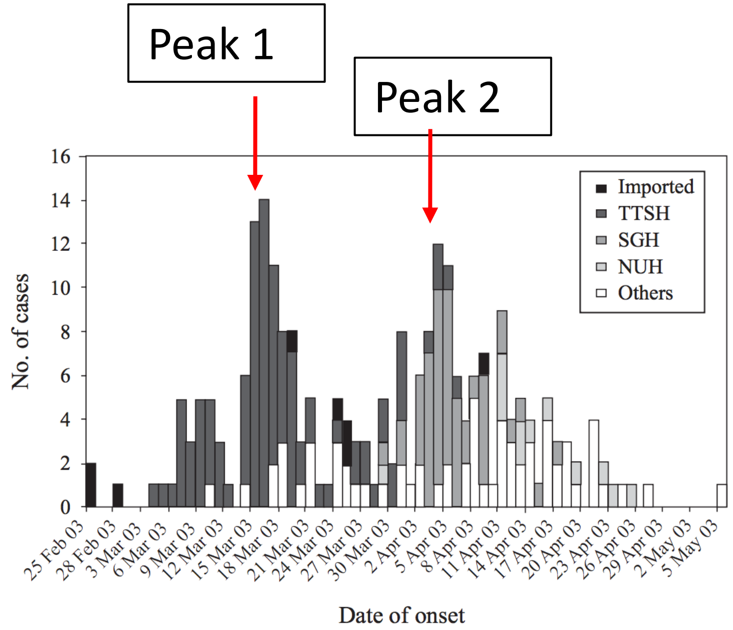
Model parameters are estimated by fitting the model to the aggregated incidence curve

Nonlinear least squares fitting or maximum likelihood estimation with parametric bootstrapping.

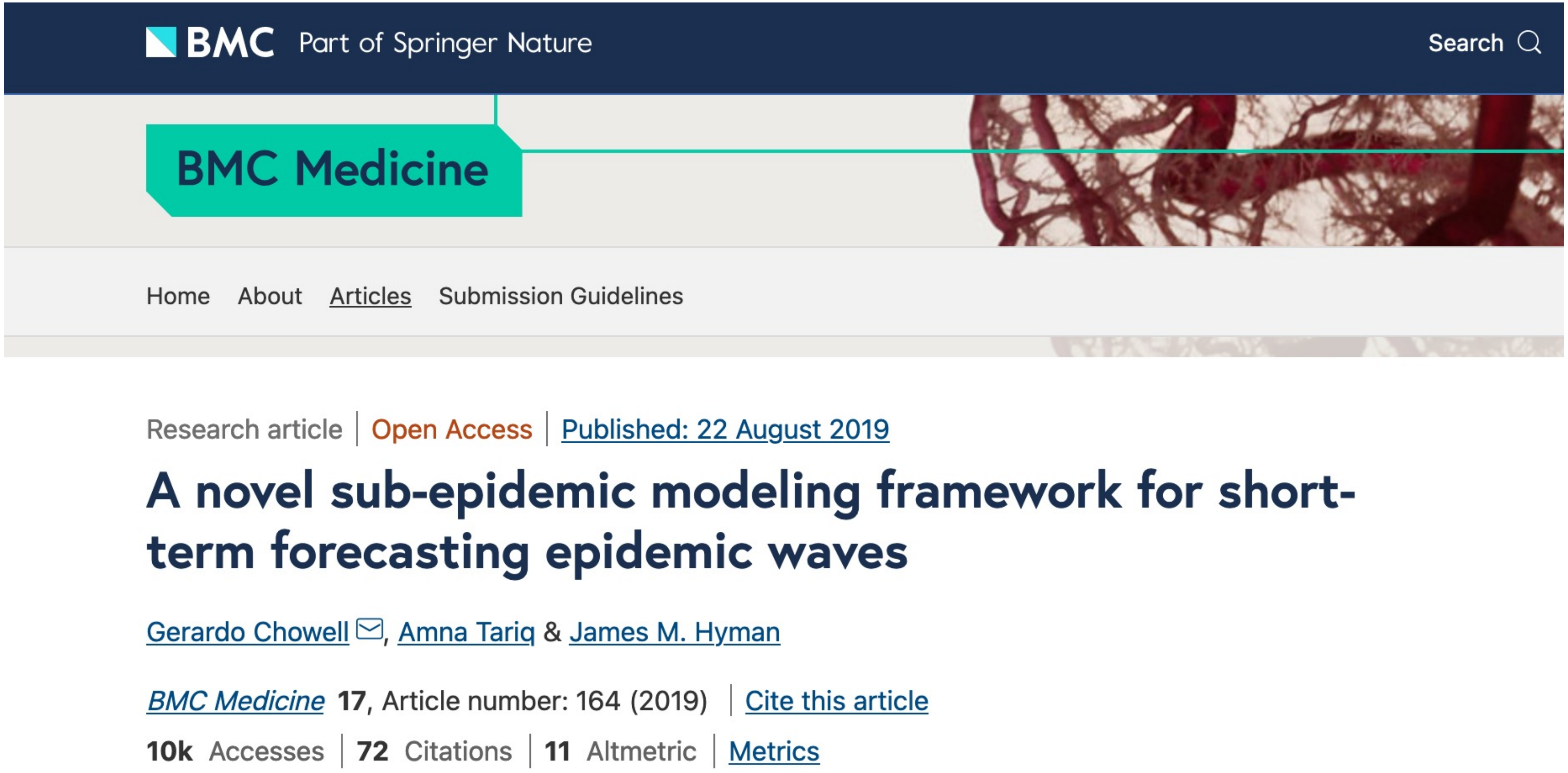
*Chowell et al. Infectious Disease Modeling. 2017*

# SARS outbreak in Singapore, 2003





Hence two sub-epidemics in the sub-epidemic profile



Research article | Open Access | Published: 22 August 2019

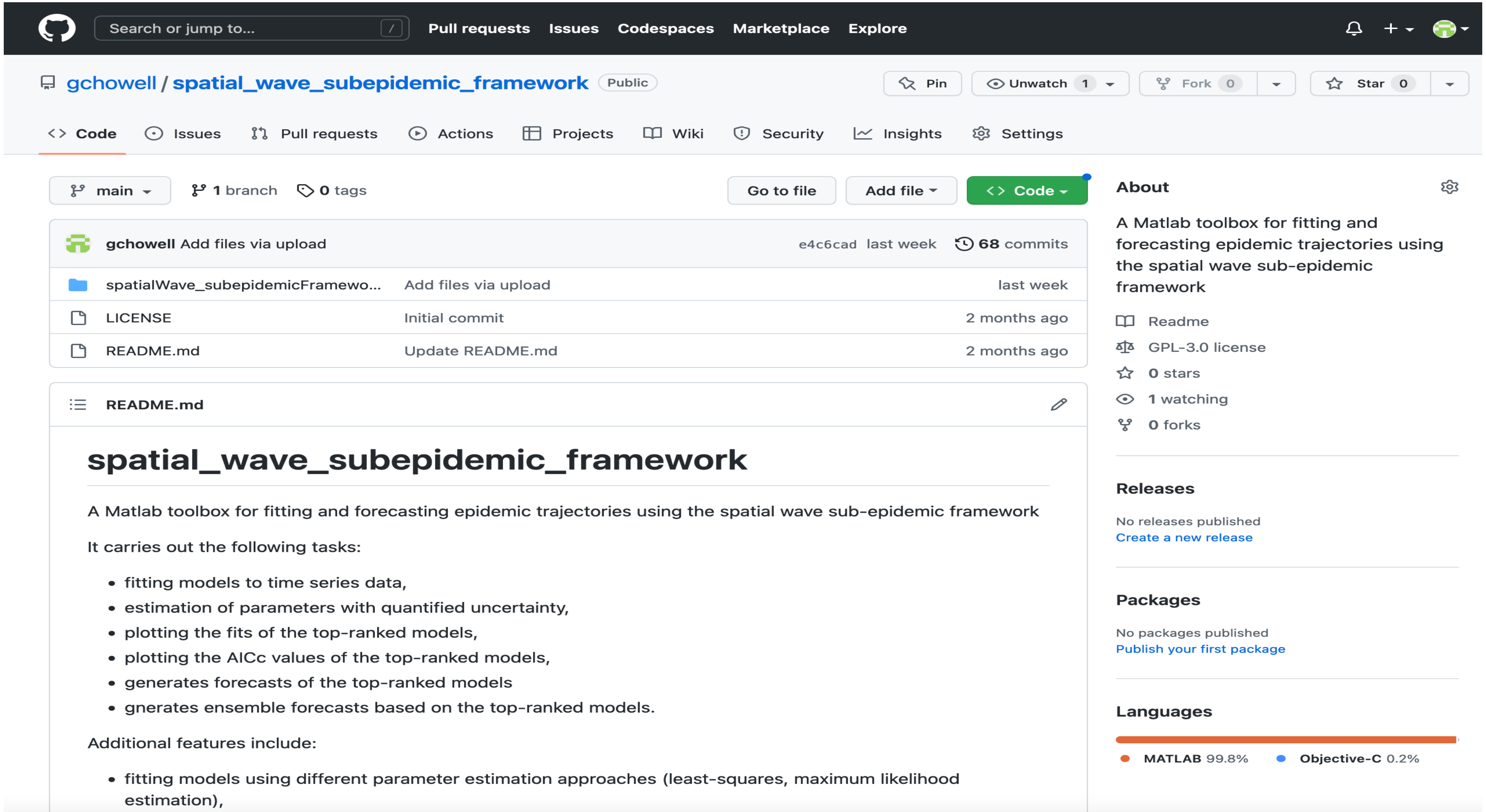
# A novel sub-epidemic modeling framework for short-term forecasting epidemic waves

Gerardo Chowell, Amna Tariq & James M. Hyman

*BMC Medicine* **17**, Article number: 164 (2019) | Cite this article

**10k** Accesses | **72** Citations | **11** Altmetric | Metrics

# Spatial wave sub-epidemic framework - MATLAB Toolbox



gchowell Add files via upload e4c6cad last week 68 commits

| | | |
|---|---|---|
| spatialWave_subepidemicFramewo... | Add files via upload | last week |
| LICENSE | Initial commit | 2 months ago |
| README.md | Update README.md | 2 months ago |

A Matlab toolbox for fitting and forecasting epidemic trajectories using the spatial wave sub-epidemic framework

README.md

## spatial_wave_subepidemic_framework

A Matlab toolbox for fitting and forecasting epidemic trajectories using the spatial wave sub-epidemic framework

It carries out the following tasks:

- fitting models to time series data,
- estimation of parameters with quantified uncertainty,
- plotting the fits of the top-ranked models,
- plotting the AICc values of the top-ranked models,
- generates forecasts of the top-ranked models
- gnerates ensemble forecasts based on the top-ranked models.

Additional features include:

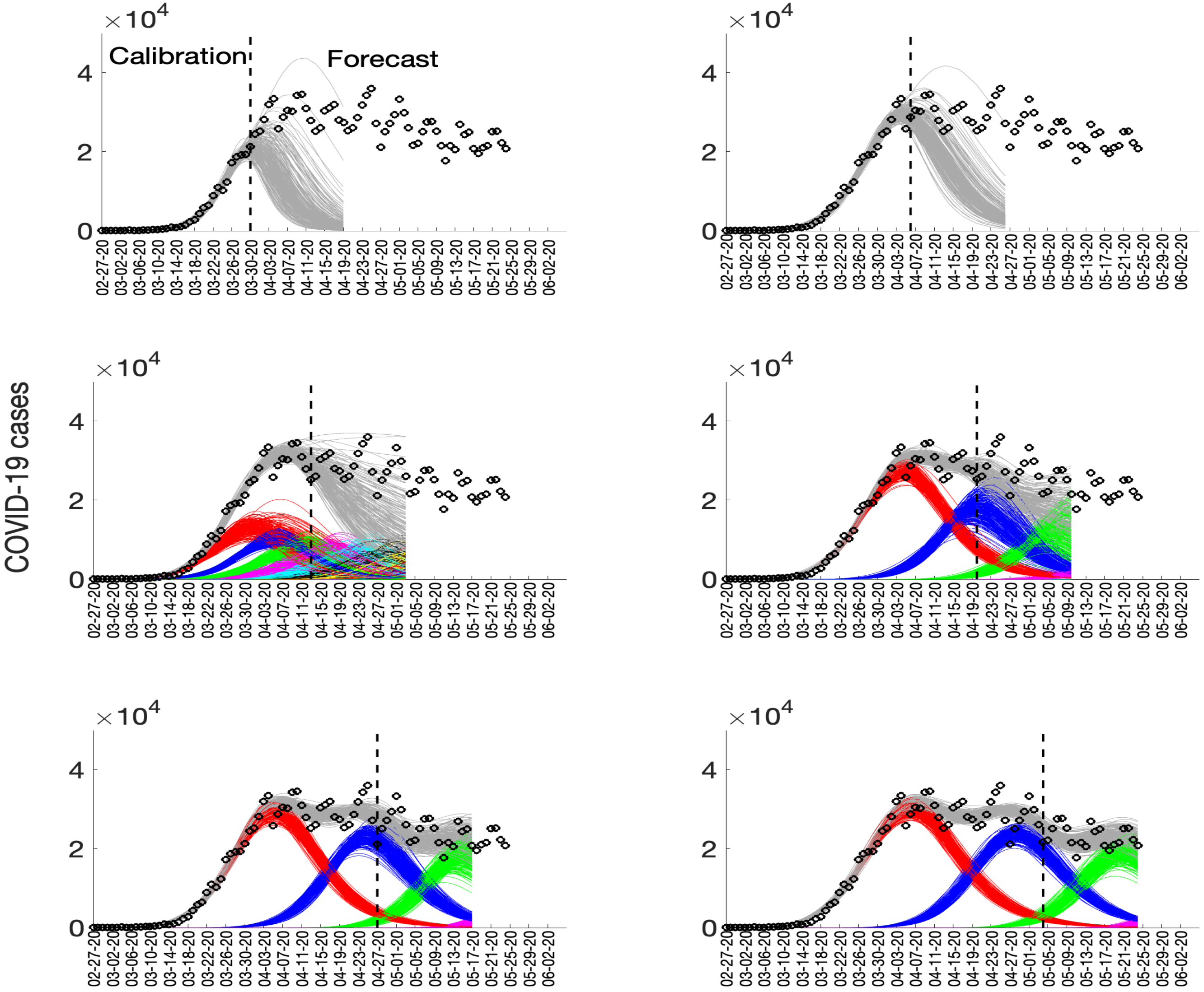
- fitting models using different parameter estimation approaches (least-squares, maximum likelihood estimation),

https://github.com/gchowell/spatial_wave_subepidemic_framework

# Real-time forecasts of the ongoing COVID-19 pandemic

https://publichealth.gsu.edu/research/coronavirus/

# Sub-epidemic wave model forecasts: USA



# Performance metrics

## Mean Absolute Error

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|f\left(t_i,\hat{\Theta}\right) - y_{t_i}\right|$$

## Mean Squared Error

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left(f\left(t_i,\hat{\Theta}\right) - y_{t_i}\right)^2$$

## Weighted Interval Score

$$WIS_{\alpha_{0:K}}(F,y) = \frac{1}{K+\frac{1}{2}}.\left(w_0.|y-m| + \sum_{k=1}^{K} w_k.IS_{\alpha_k}(F,y)\right)$$

$y_{t_i}$ = time series of incident cases describing epidemic wave

$t_i$= time points of time series data

$f(t_i,\widehat{\Theta})$= model fit

## Coverage of the 95% Prediction Interval

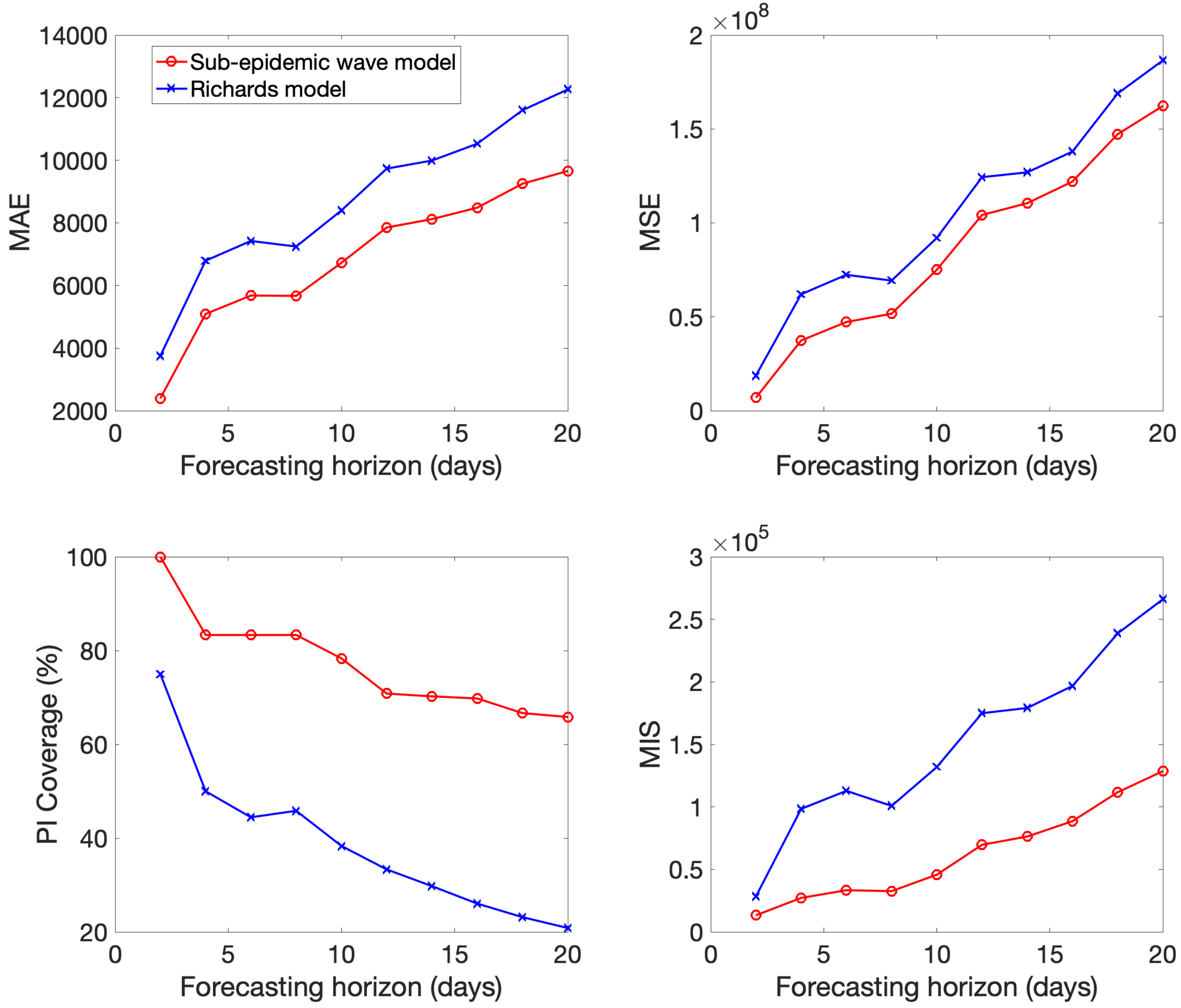
$$\frac{1}{n}\sum_{t=1}^{n} 1\{y_t > L_t \cap \; y_t < U_t\}$$

$L_t$= lower bound of 95% prediction interval

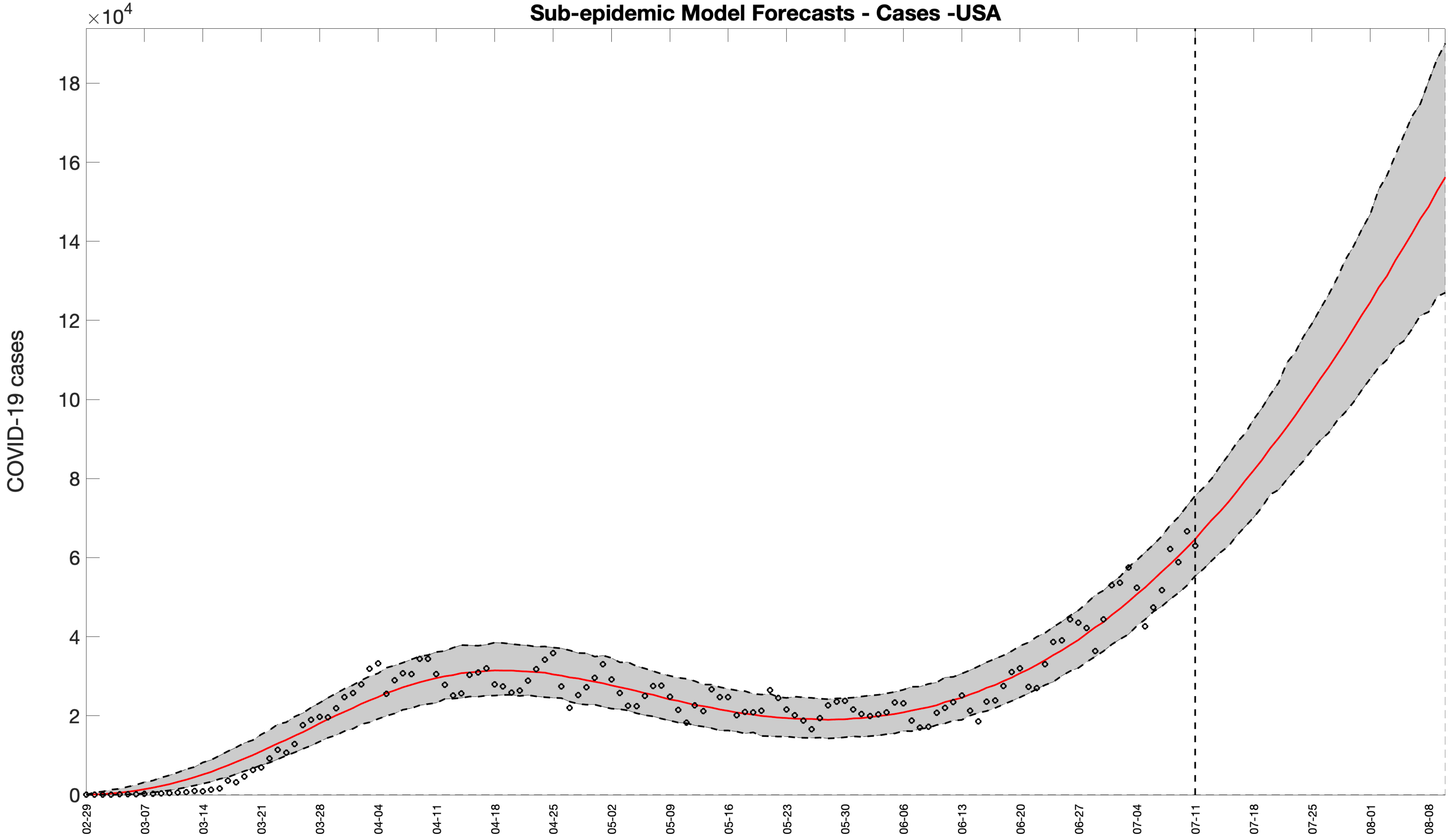
$U_t$= Upper bound of 95% prediction interval

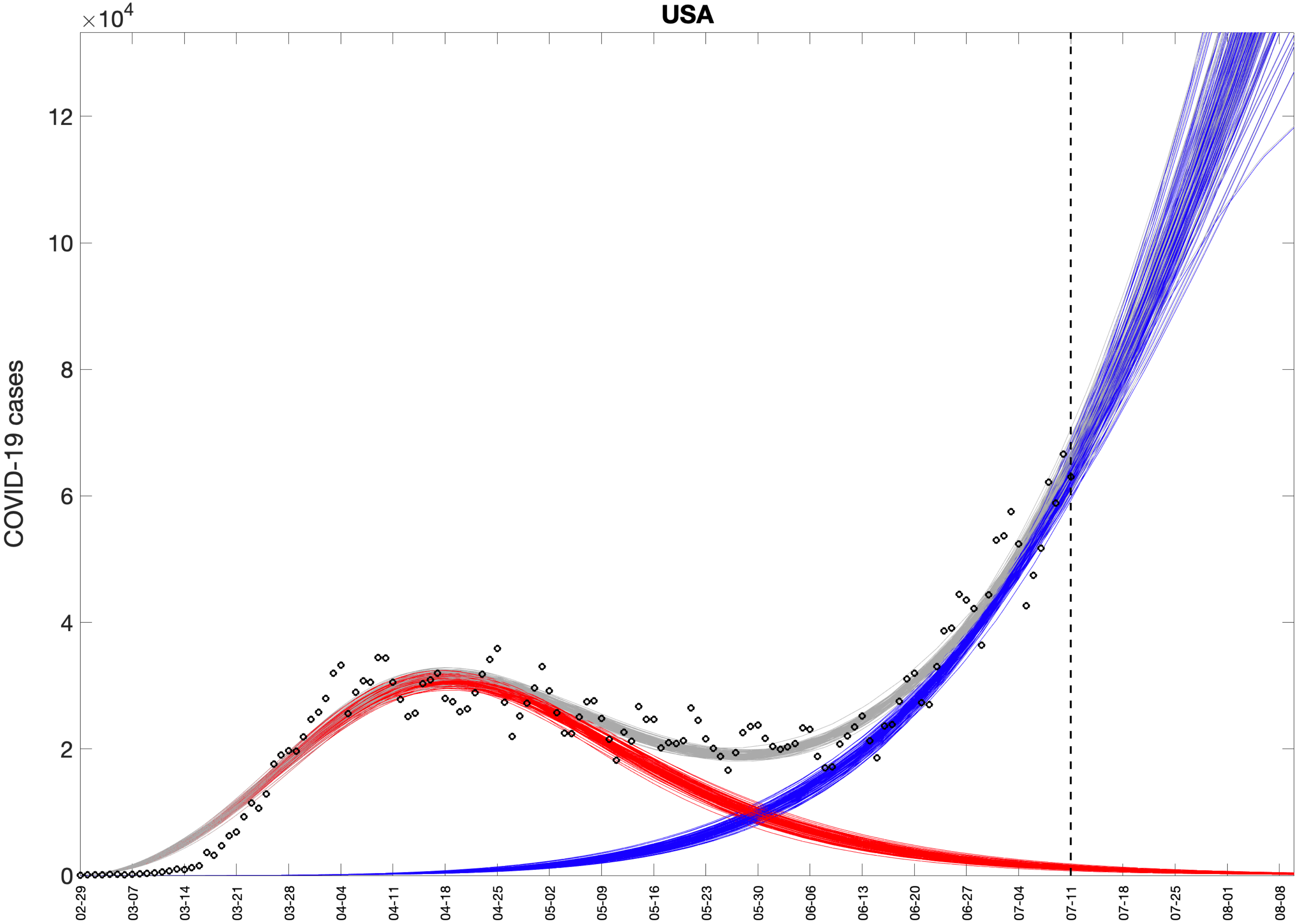
*Tilmann Gneiting & Adrian E Raftery (2007)*

# Sub-epidemic wave model performance - USA



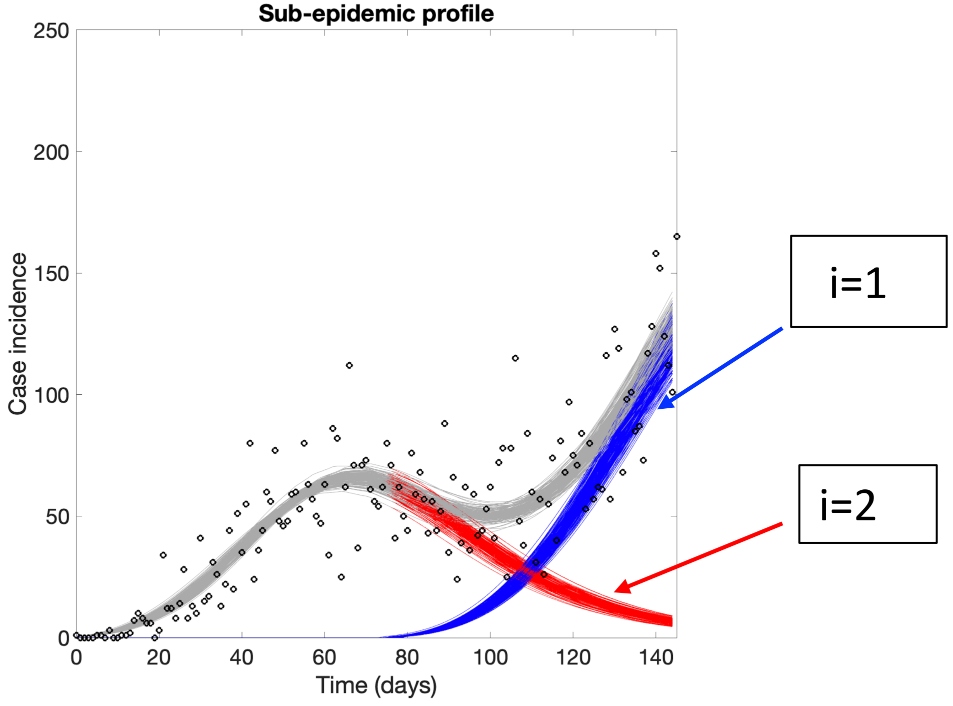
# COVID-19 Resurgence after relaxation of social distancing - USA




USA
×10^4
COVID-19 cases
0
2
4
6
8
10
12
02-29
03-07
03-14
03-21
03-28
04-04
04-11
04-18
04-25
05-02
05-09
05-16
05-23
05-30
06-06
06-13
06-20
06-27
07-04
07-11
07-18
07-25
08-01
08-08

# *n*-subepidemic modeling framework

- Each sub-epidemic is modelled by a generalized logistic growth model
- An epidemic wave comprising of a set of *n* overlapping sub-epidemics is modelled using coupled differential equations
- $$\frac{dC_i(t)}{dt} = {C_i}'(t) = A_i(t) r_i {C_i}^{p_i}(t)\left(1 - \frac{C_i(t)}{K_{0_i}}\right)$$



i=1

i=2

Where,

$C_i$= cumulative number of infections for sub-epidemic i,

$K_i$= size of the $\mathrm{i}_{\mathrm{th}}$ sub-epidemic

$A_i$= indicator variable

$$A_i(t) = f(x) = \begin{cases} 1 & C_i(t) > C_{thr} \\ 0 & Otherwise \end{cases} \qquad \mathrm{i} = 1,2,3 \ldots\ldots \mathrm{n}$$

**Number of model parameters is 3n+1**

# Top-ranking *K* sub-epidemic models via model selection

We considered a set of $n$-subepidemic models with $1 \leq n \leq 2$ and different values of $C_{thr}$ . Then we ranked them from best to worst according to the $AIC_c$ which is given by:

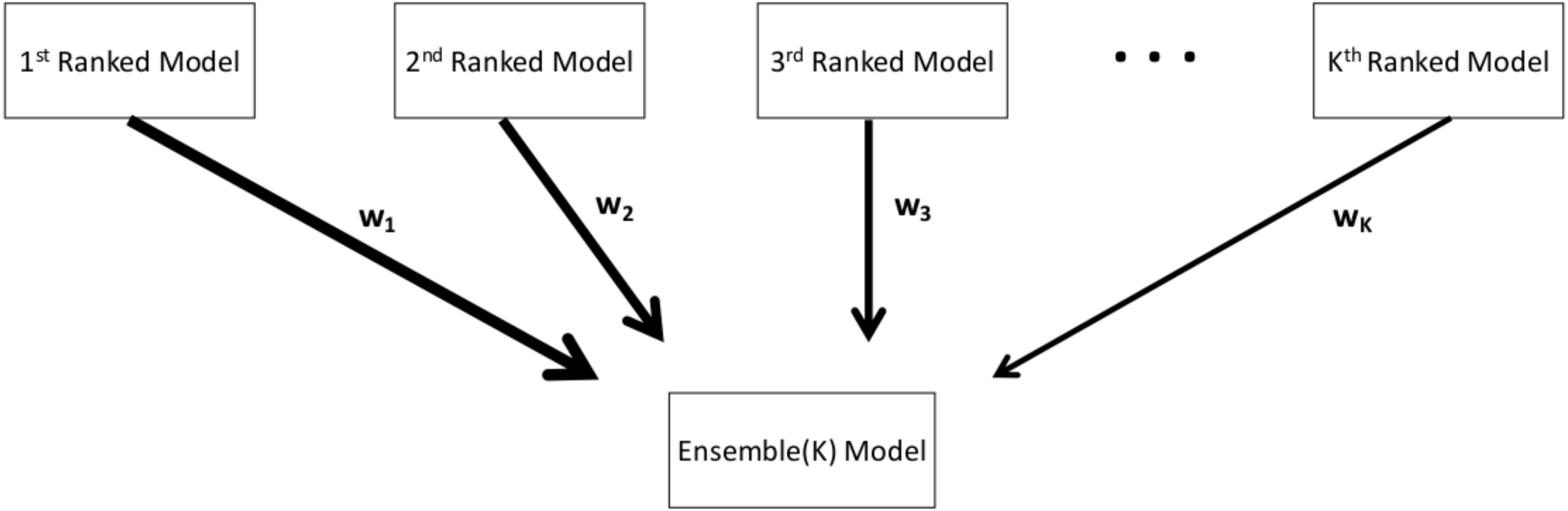
$$AIC_c = n_d log(SSE) + 2m + \frac{2m(m+1)}{n_d - m - 1}$$

where $SSE = \sum_{j=1}^{n_d}(f\left(t_j, \widehat{\Theta}\right) - y_{t_j})^2$, $m = 3n + 1$ is the number of model parameters and $n_d$ is the number of data points.

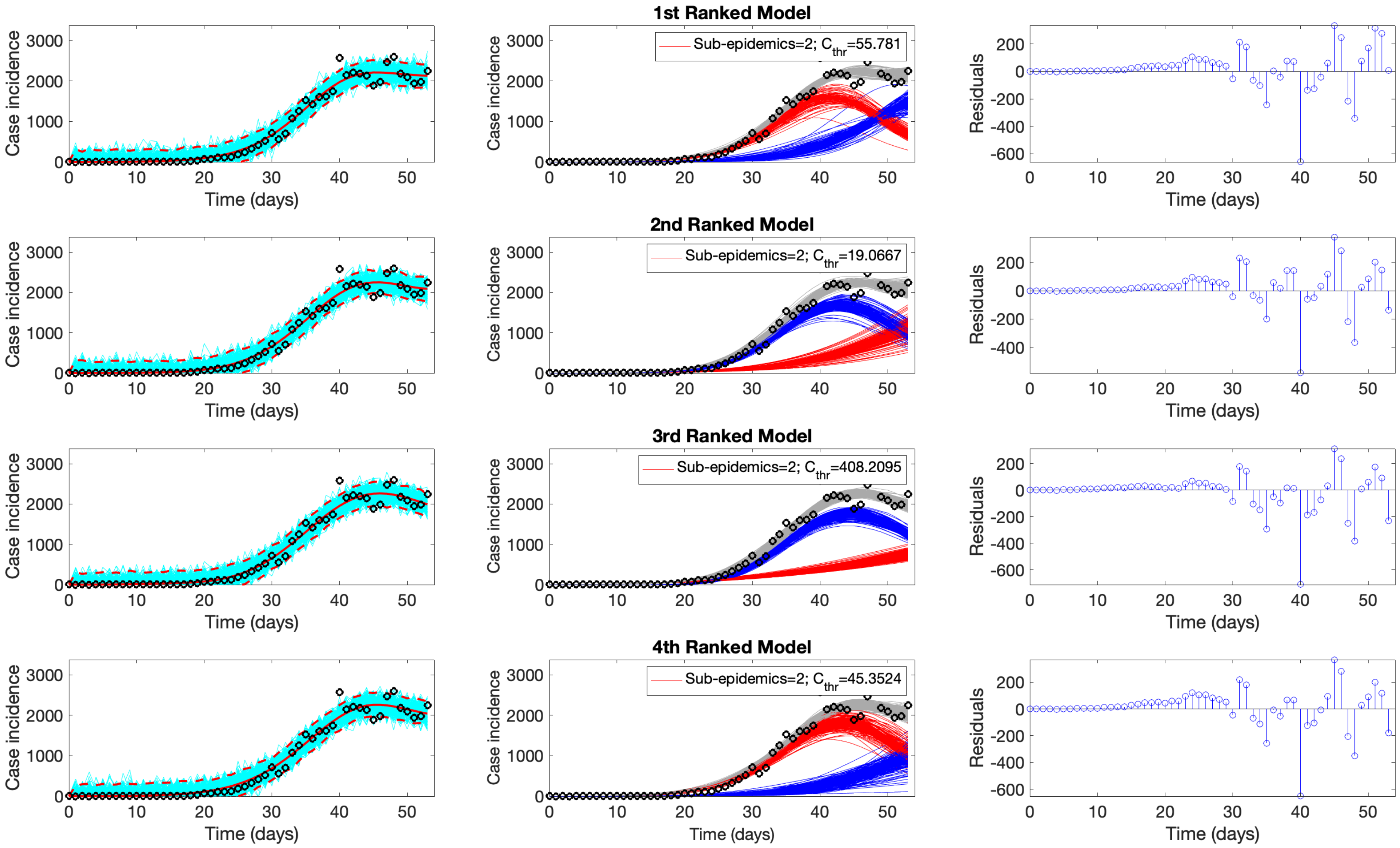
# Ensemble(K) models from top-ranking *K* sub-epidemic models

We generate ensemble models from the weighted combination of the highest-ranking sub-epidemic models as deemed by the AIC$_C$. An ensemble derived from the top-ranking ”K"  models is denoted by Ensemble(K).

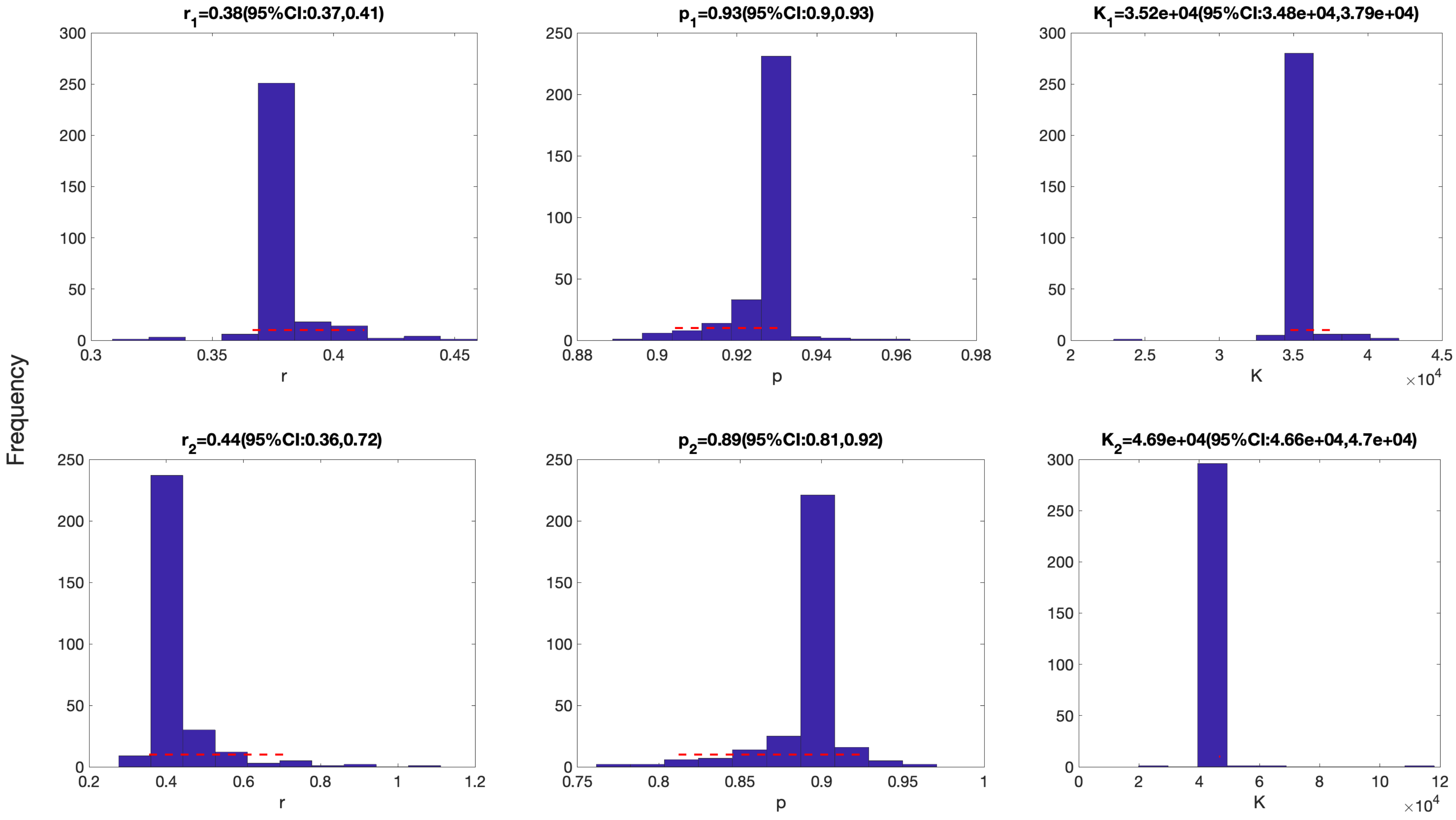
*Ensemble n-subepidemic modeling framework*



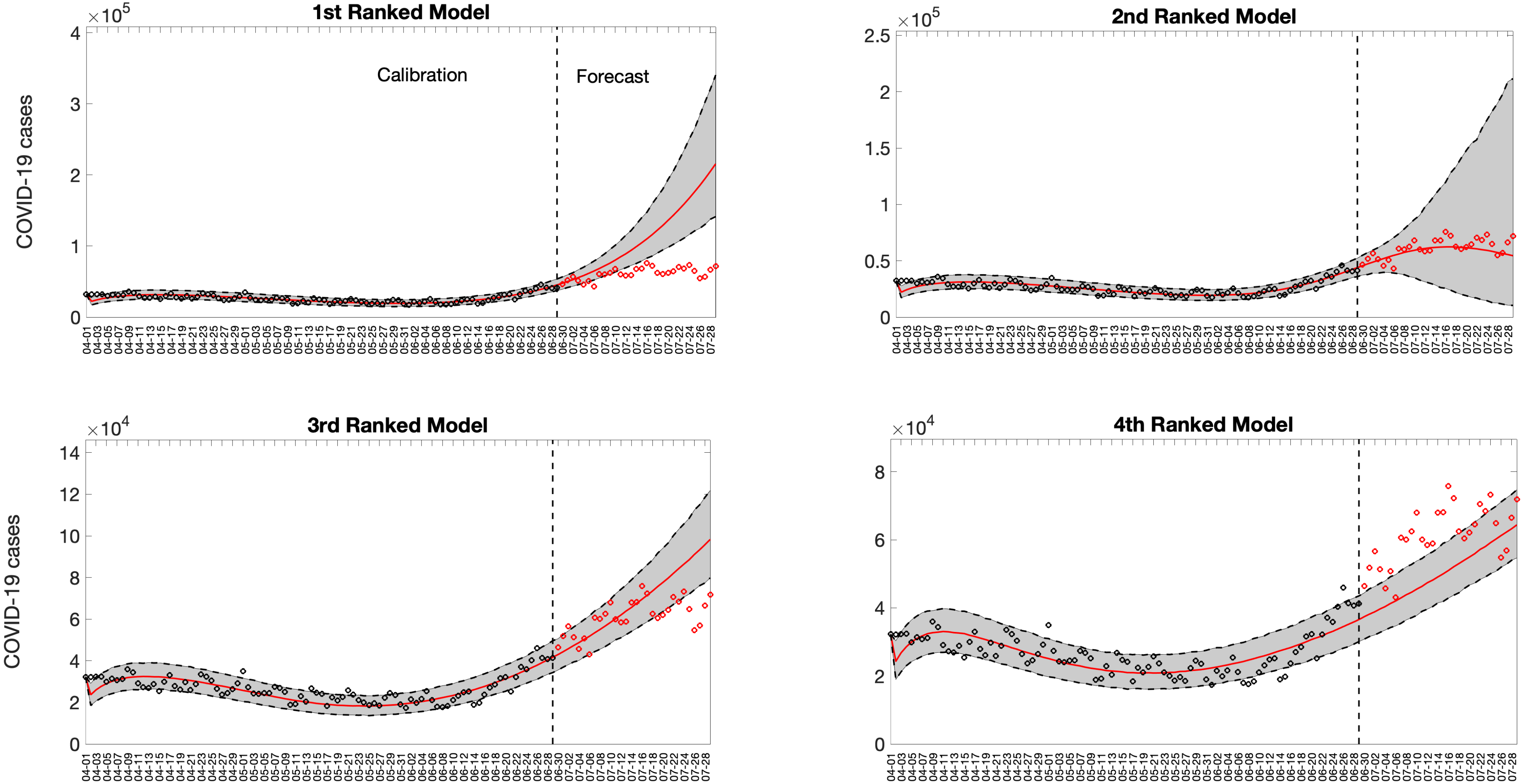
# Representative fits of the top-ranking sub-epidemic models



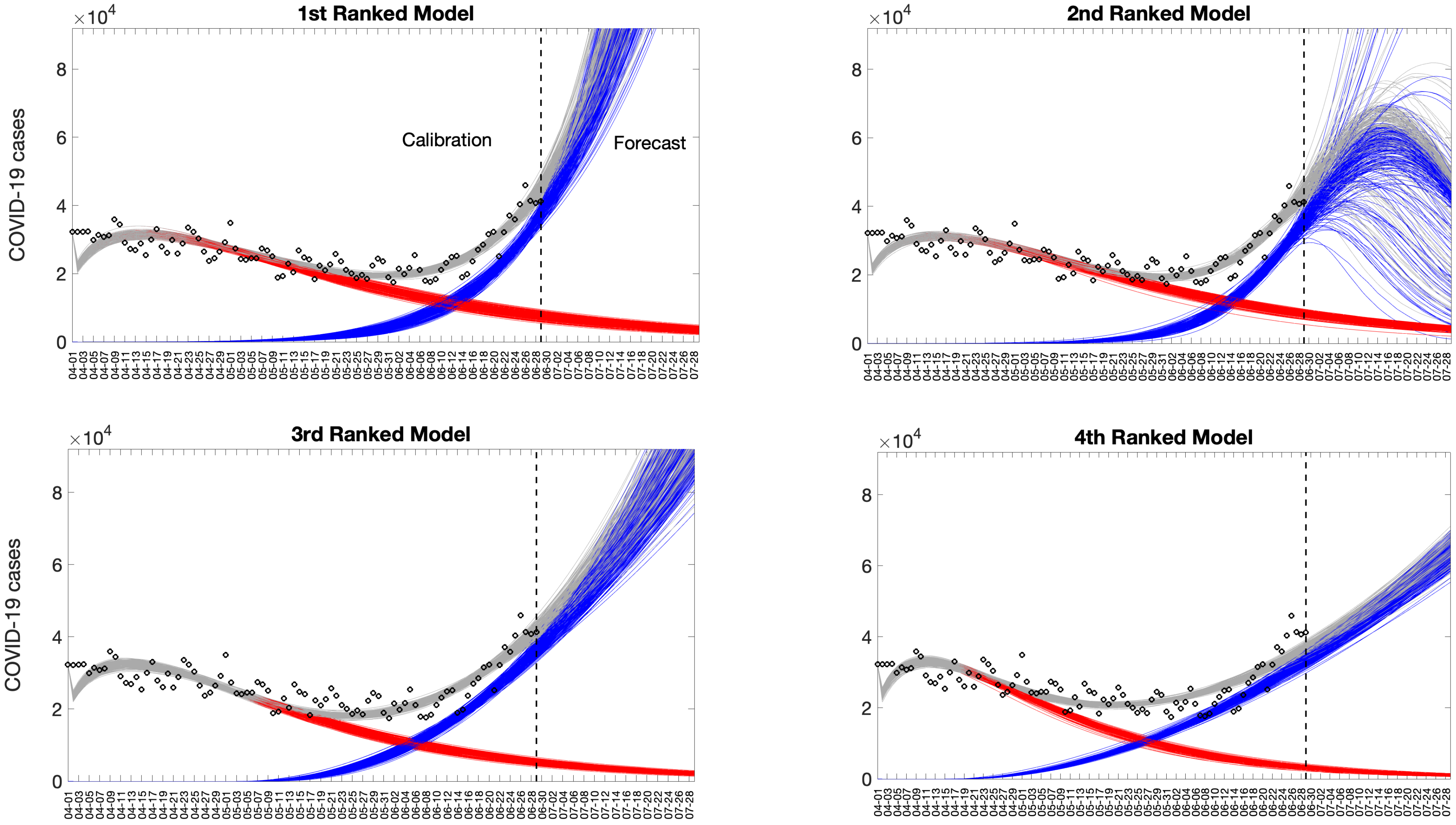
# Parameter estimates for the top-ranking sub-epidemic model



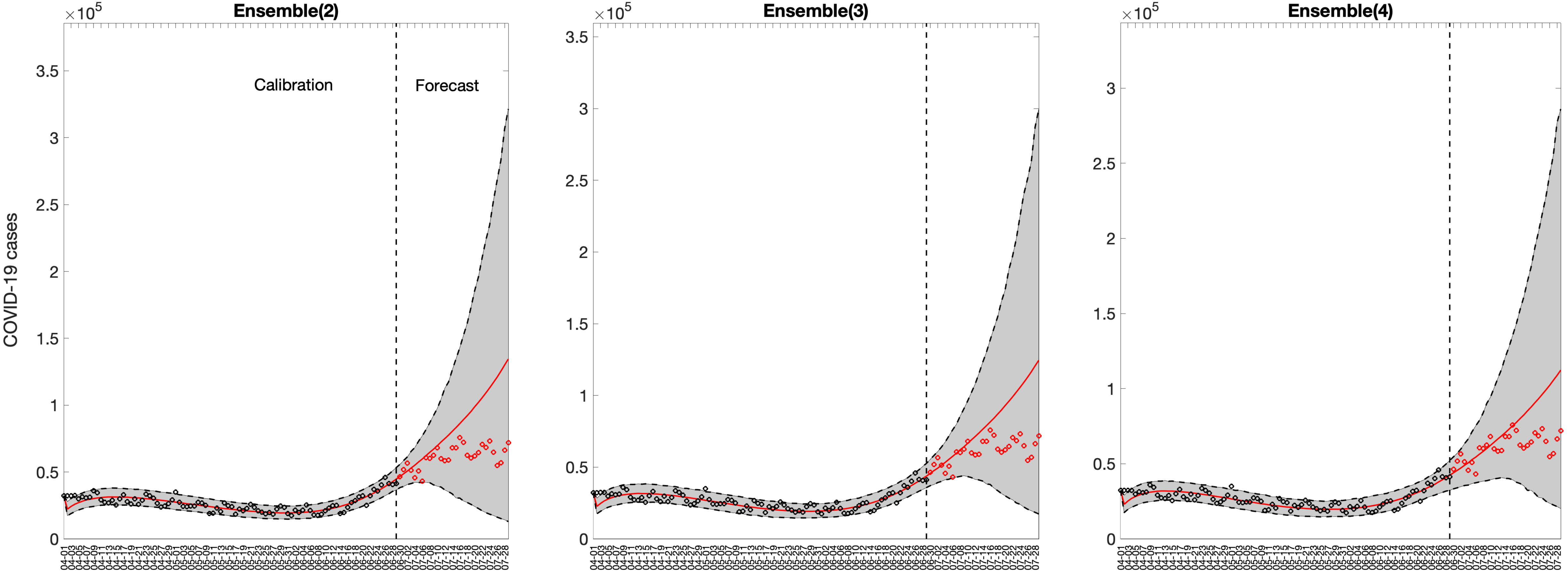
# Summer resurgence, USA



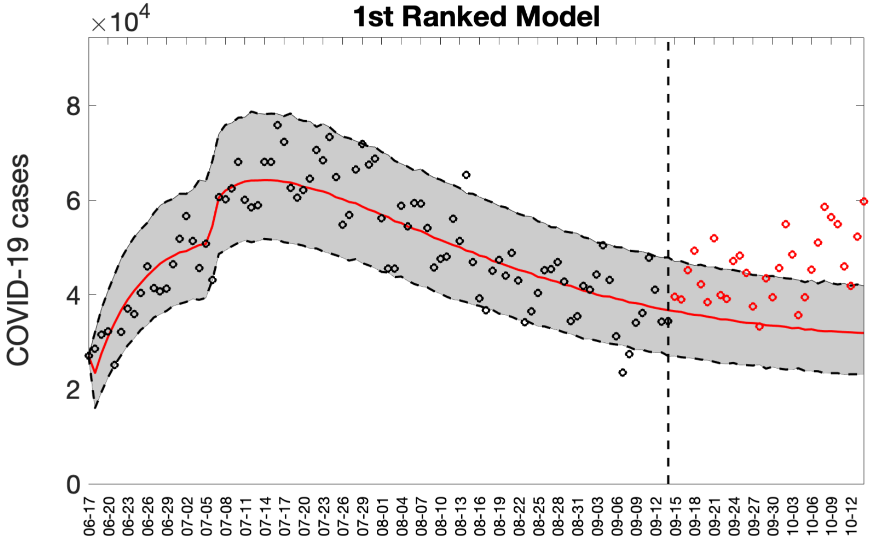
# Sub-epidemic profiles



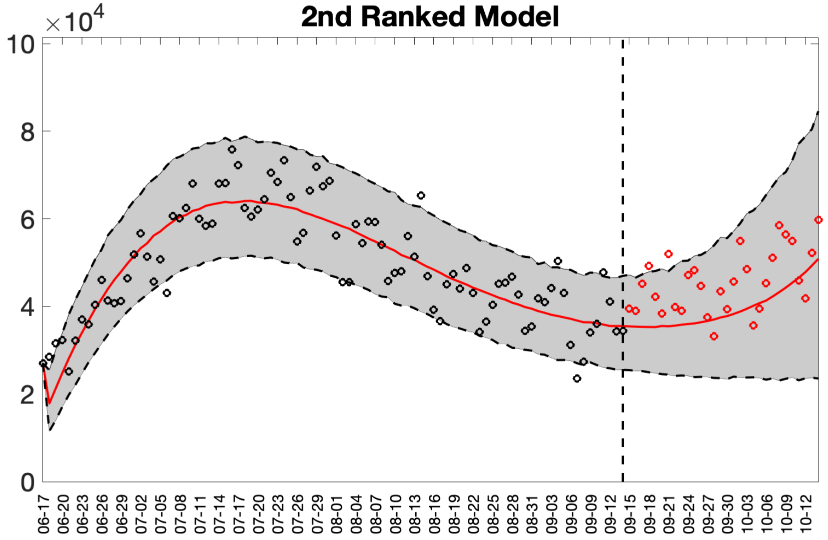
# Ensemble models – Summer resurgence, USA



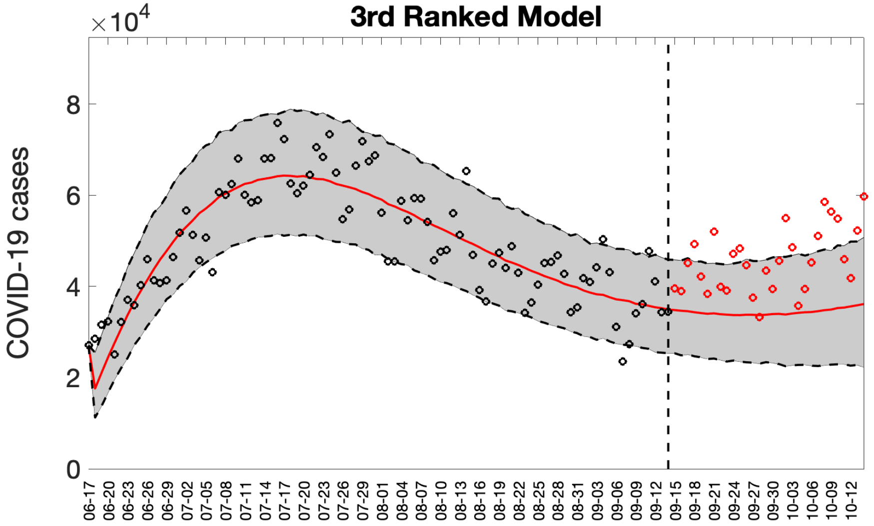
# Fall resurgence, USA



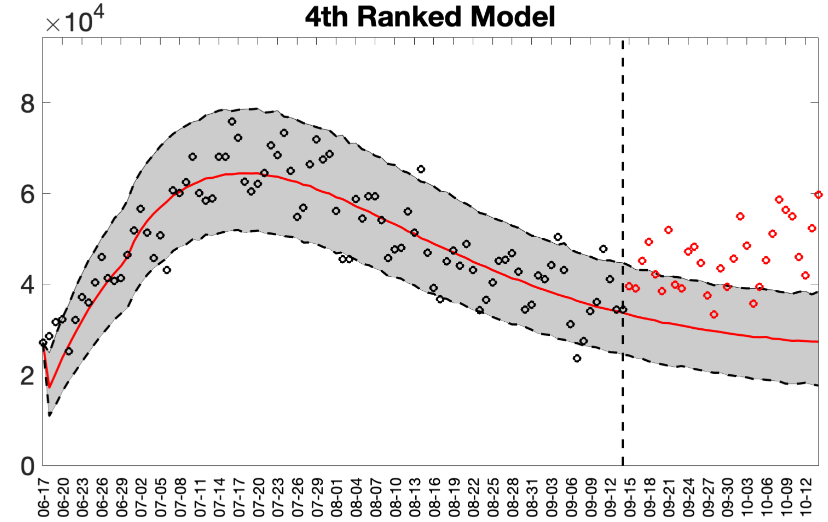
1st Ranked Model



2nd Ranked Model

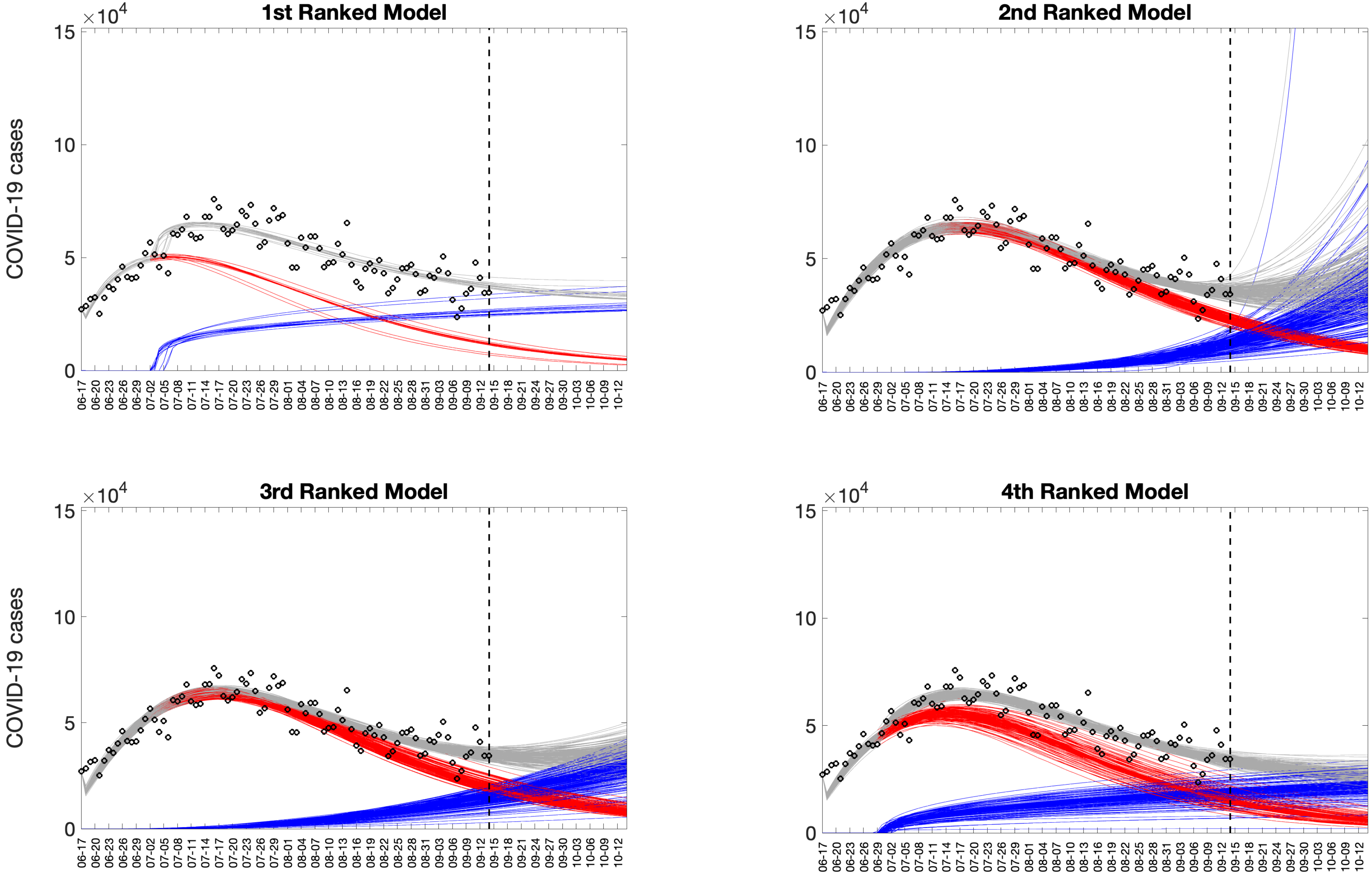


3rd Ranked Model

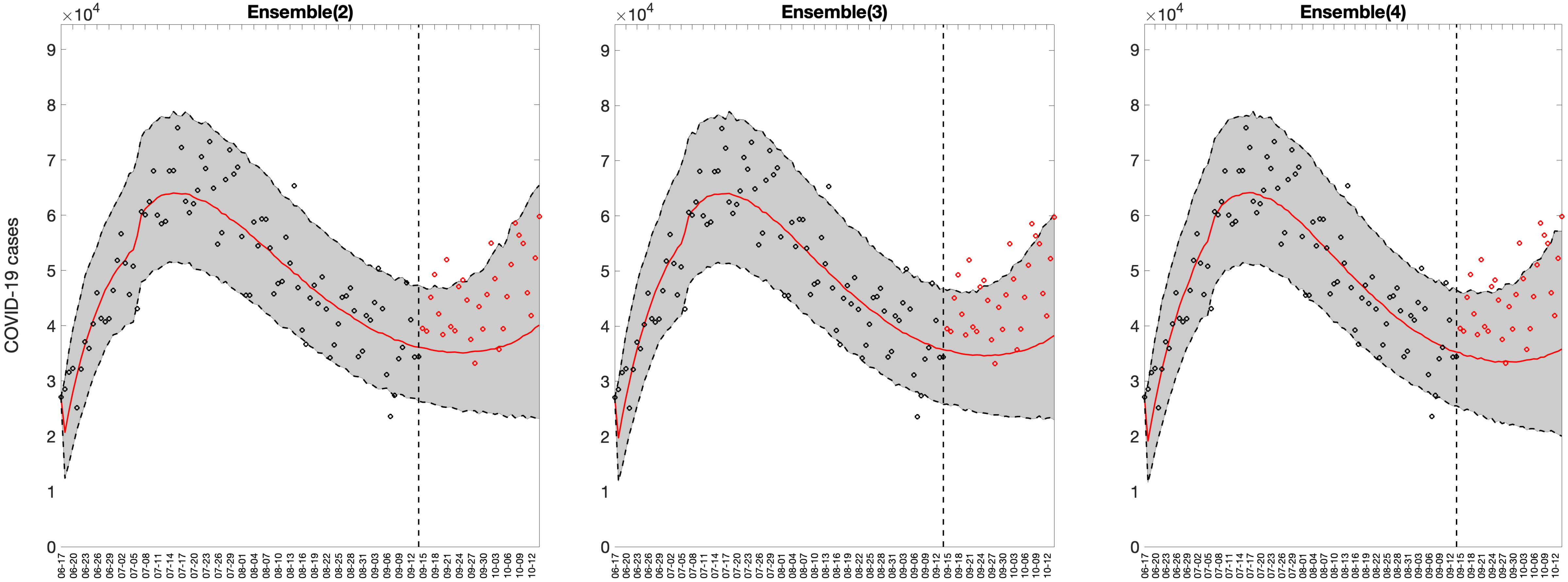


4th Ranked Model

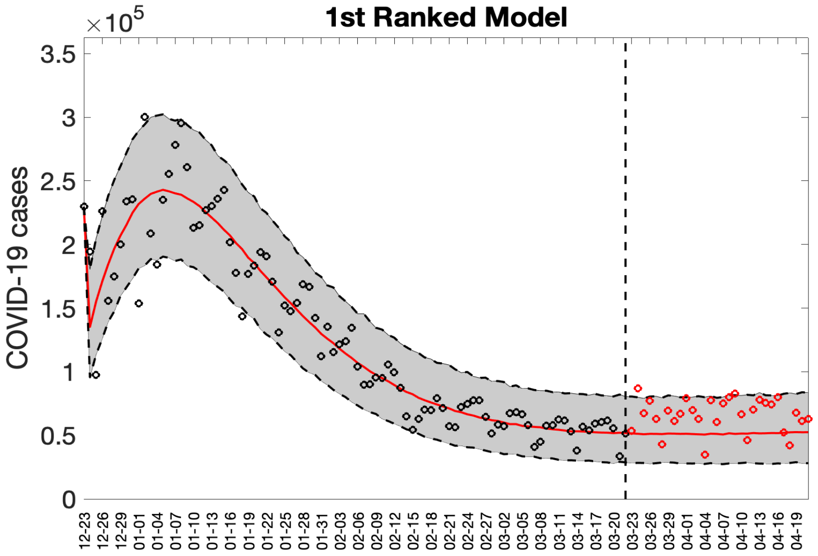
# Sub-epidemic profiles

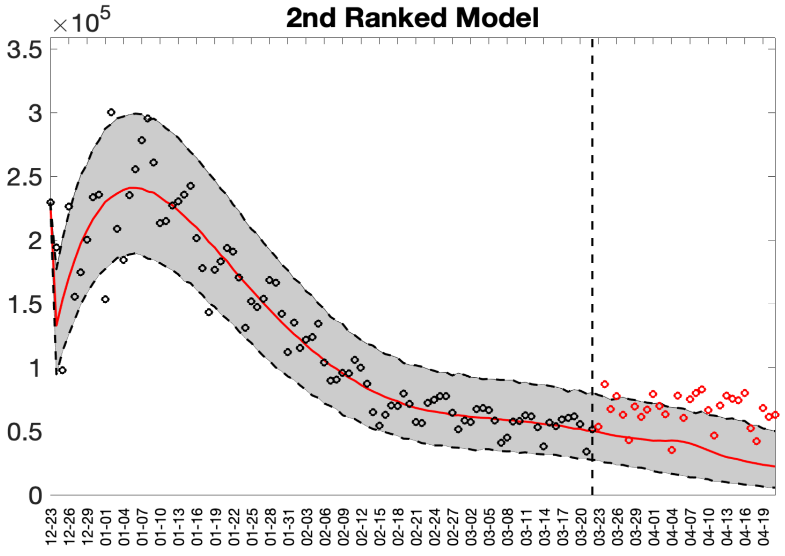


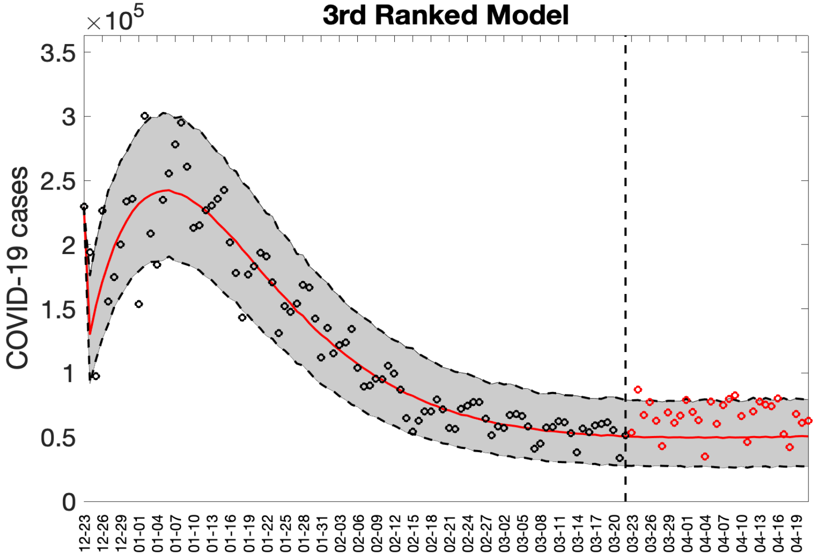
# Ensemble models – Fall resurgence

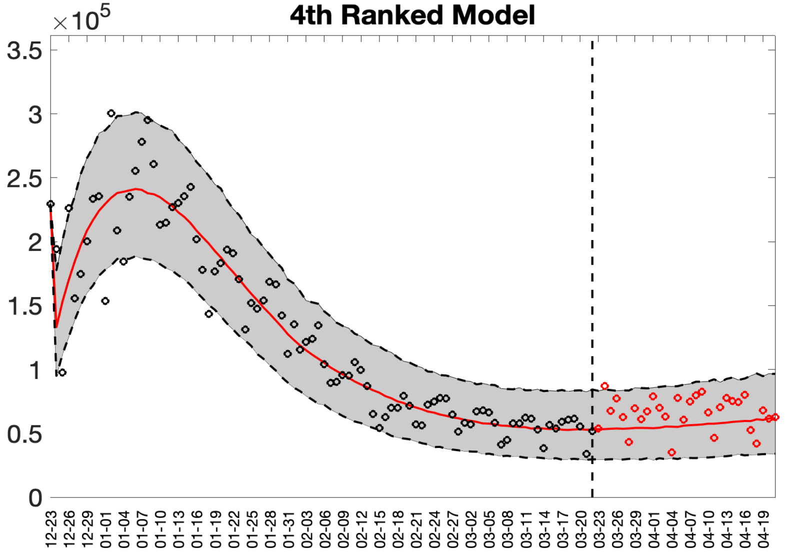


# Spring 2021

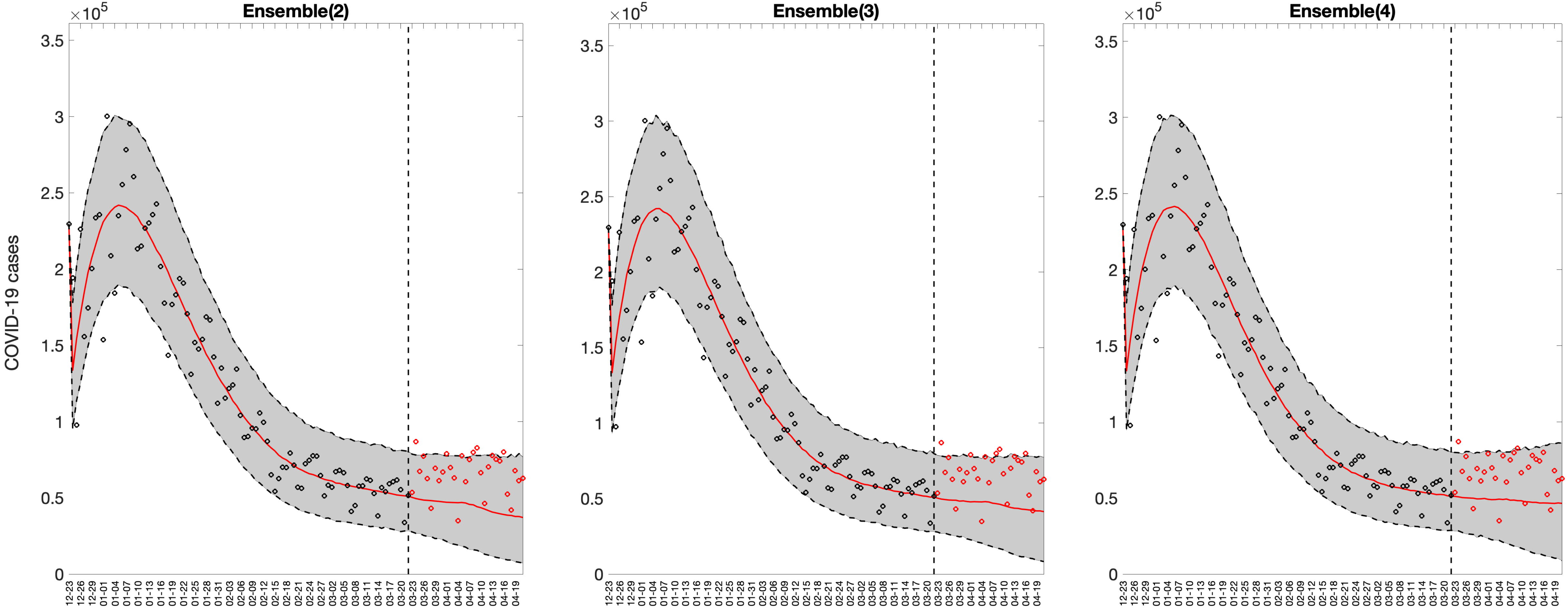

1st Ranked Model


2nd Ranked Model

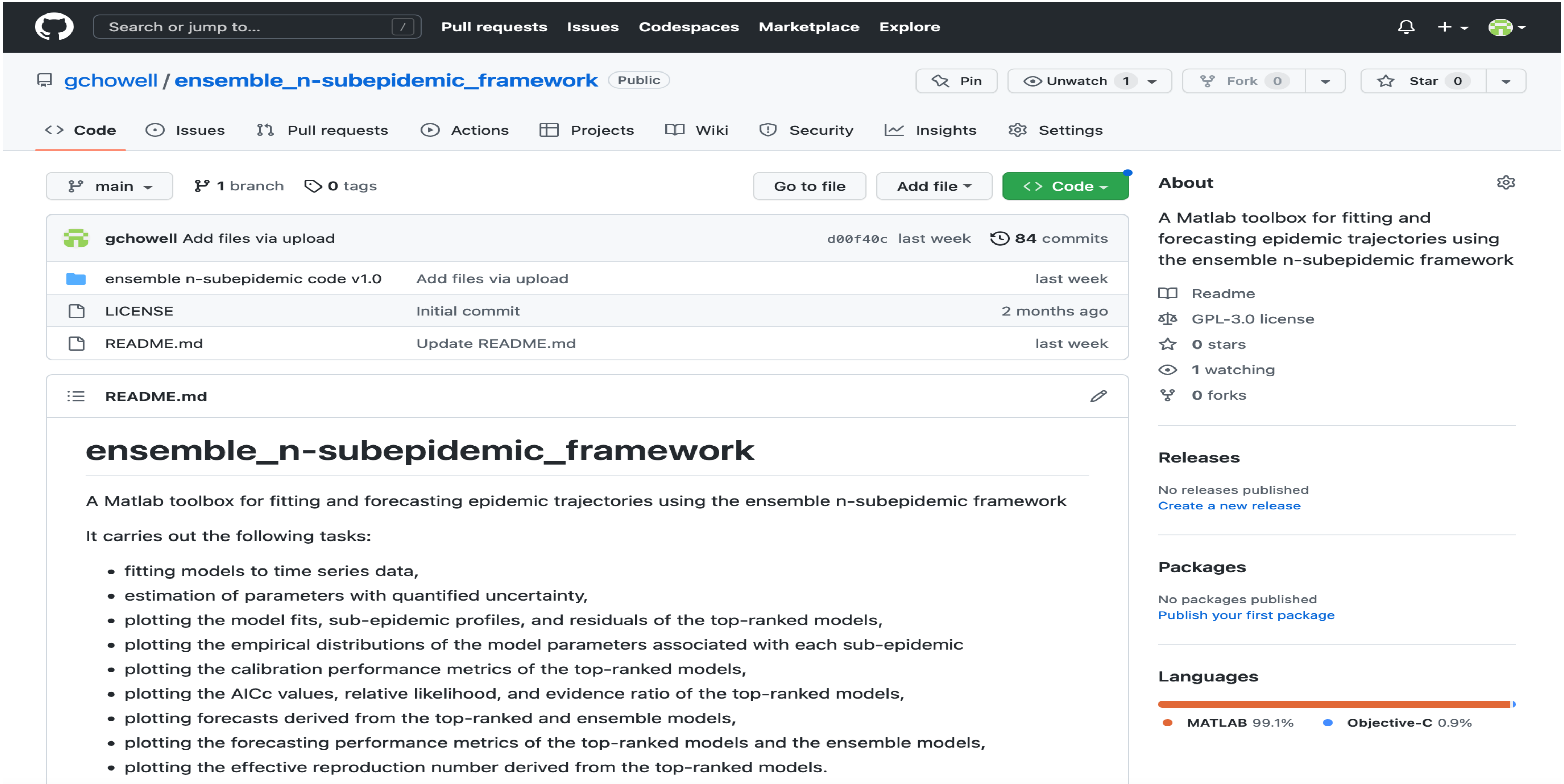

3rd Ranked Model


4th Ranked Model

# Ensemble models – Spring 2021



# Ensemble n-subepidemic framework - MATLAB Toolbox



gchowell Add files via upload d00f40c last week 84 commits

| | | |
|---|---|---|
| ensemble n-subepidemic code v1.0 | Add files via upload | last week |
| LICENSE | Initial commit | 2 months ago |
| README.md | Update README.md | last week |

README.md

## ensemble_n-subepidemic_framework

A Matlab toolbox for fitting and forecasting epidemic trajectories using the ensemble n-subepidemic framework

It carries out the following tasks:

- fitting models to time series data,
- estimation of parameters with quantified uncertainty,
- plotting the model fits, sub-epidemic profiles, and residuals of the top-ranked models,
- plotting the empirical distributions of the model parameters associated with each sub-epidemic
- plotting the calibration performance metrics of the top-ranked models,
- plotting the AICc values, relative likelihood, and evidence ratio of the top-ranked models,
- plotting forecasts derived from the top-ranked and ensemble models,
- plotting the forecasting performance metrics of the top-ranked models and the ensemble models,
- plotting the effective reproduction number derived from the top-ranked models.

**About**

A Matlab toolbox for fitting and forecasting epidemic trajectories using the ensemble n-subepidemic framework

Readme
GPL-3.0 license
0 stars
1 watching
0 forks

**Releases**

No releases published
Create a new release

**Packages**

No packages published
Publish your first package

**Languages**

MATLAB 99.1% Objective-C 0.9%

https://github.com/gchowell/ensemble_n-subepidemic_framework

# ARIMA models

>> Commonly used to forecast trends in finance and the weather.

The auto.arima function in the R package "forecast" is used to select orders and build the model. First, the degree of differencing $0 \leq d \leq 2$ is selected based on successive KPSS unit-root. Then given d, the orders p and q are selected based on the AICc for the d-times differenced data. For d=0 or d=1, a constant will be included if it improves the AICc value; for d>1, the constant $\mu$ is fixed as 0 to avoid the model having a quadratic or higher order trend.

**1) (log)ARIMA** using log-transformed data. Then we take the exponential of the forecasted values and the PI bounds to predict the incident death counts and get the PIs.

**2) ARIMA.** Any negative values are set as zero. Then, it is possible that the actual coverage probability of such PIs can be smaller than the nominal value (95%).

# Forecasting strategy

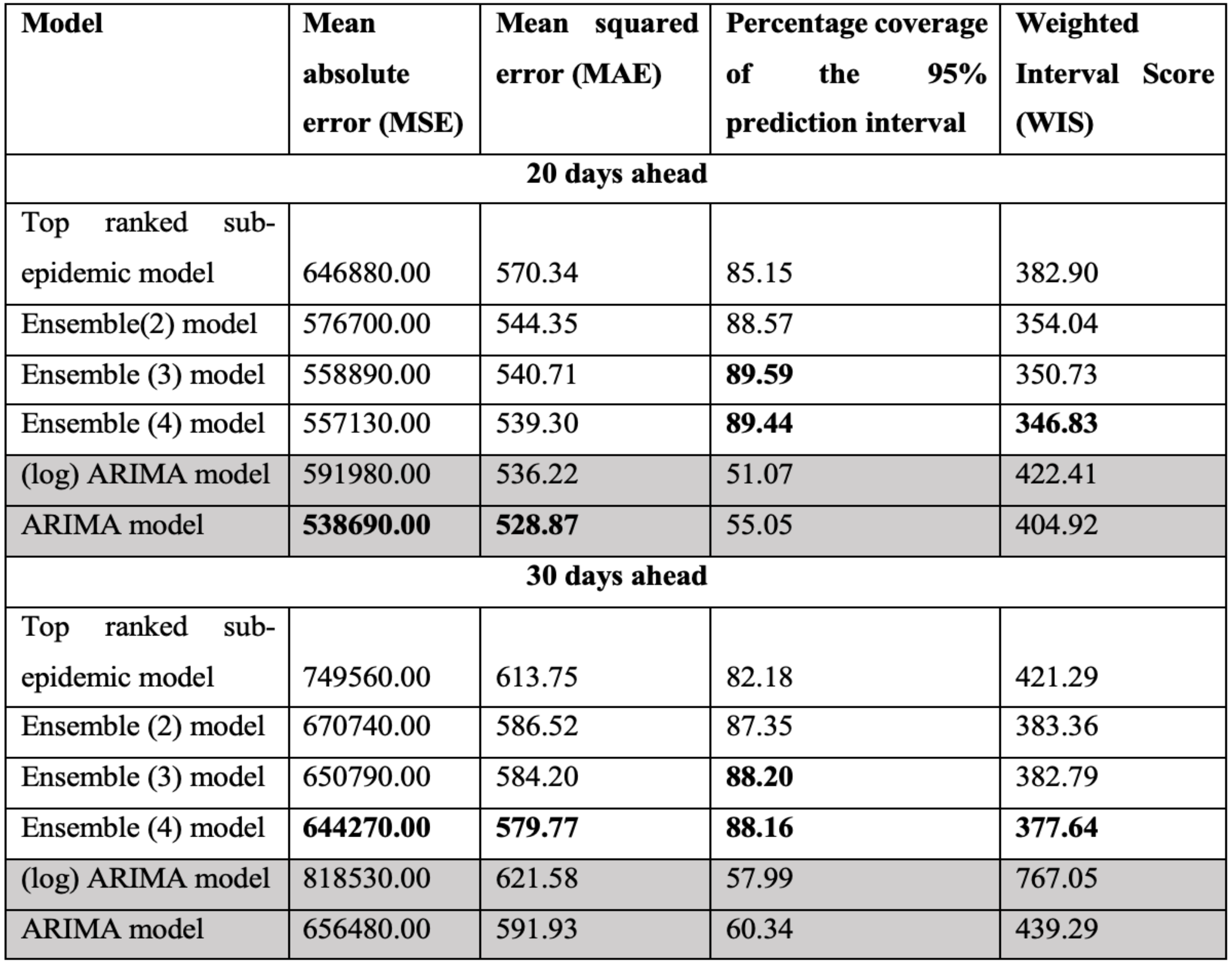
**Data.** We used daily COVID-19 deaths reported in the USA from the publicly available data tracking system of the Johns Hopkins Center for Systems Science and Engineering (CSSE) from 27 February 2020 to 30 March 2022.

**Models**. We conducted short-term forecasts using the top-ranking $n$-subepidemic model ($1 \leq n \leq 2$) and three ensemble models constructed with the top-ranking sub-epidemic models namely Ensemble(2), Ensemble(3), and Ensemble(4). For comparison, we also generated short-term forecasts using the previously described ARIMA models.

**Forecasting periods.** Using a 90-day calibration period for each model, we conducted a total of 98 weekly sequential 10-day, 20-day and 30-day ahead forecasts from 20 April 2020 to 28 February 2022, spanning five pandemic waves. The calibration period for each sequential forecast included seven additional days of data than the previous forecast.

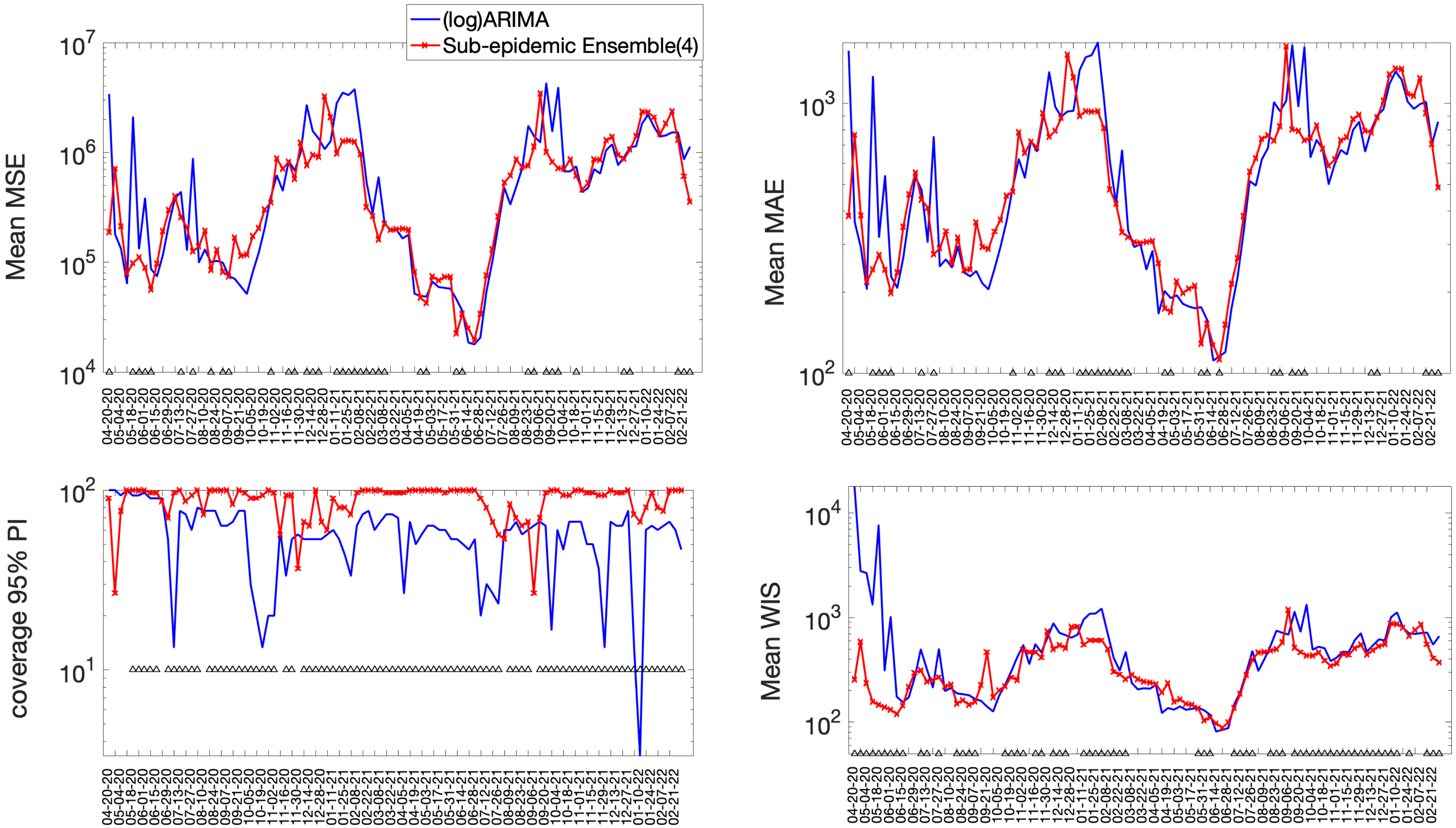
# Average forecasting performance metrics

20 April 2020 to 28 February 2022 (98 sequential forecasts)



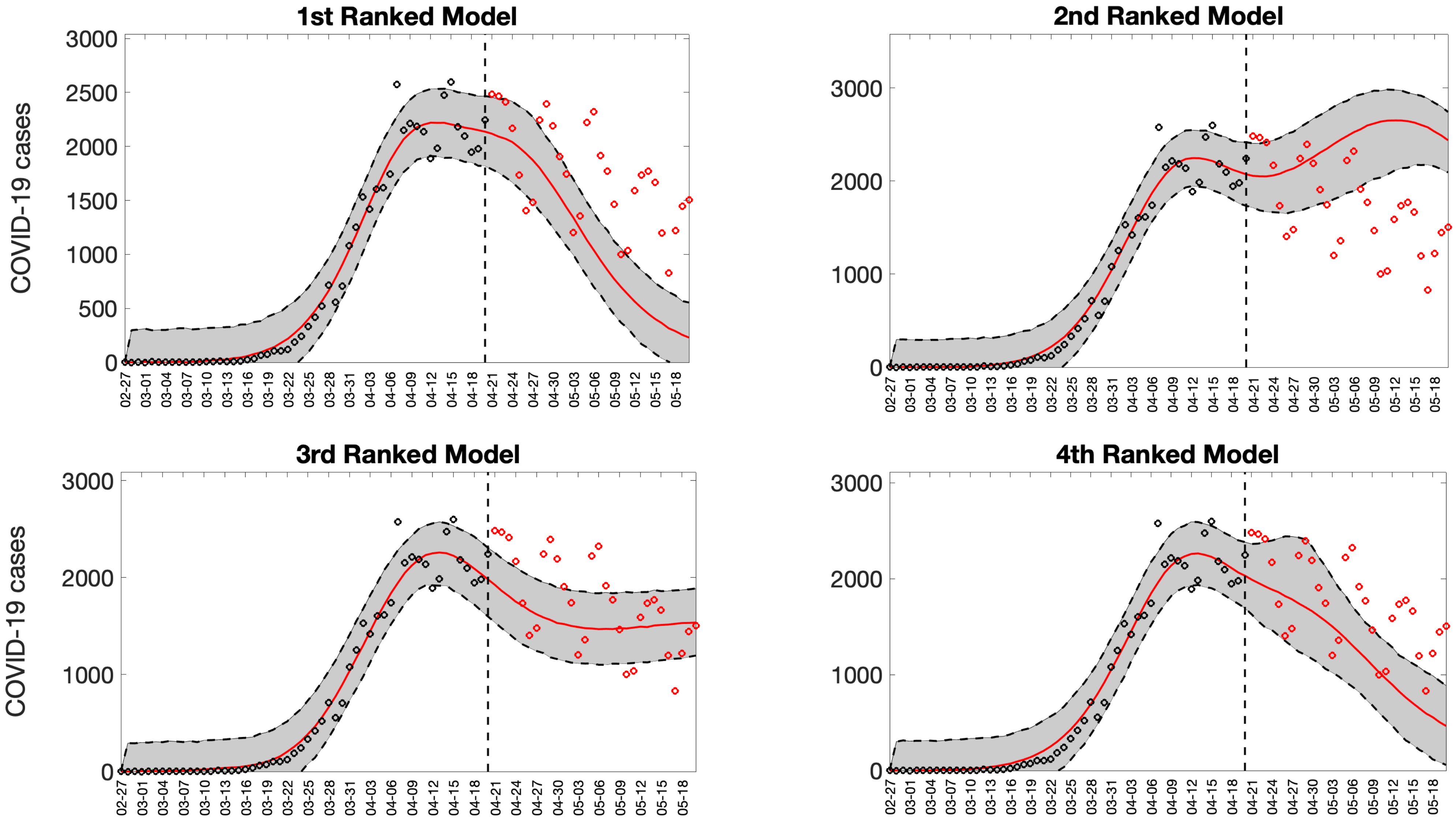
| Model | Mean absolute error (MSE) | Mean squared error (MAE) | Percentage coverage of the 95% prediction interval | Weighted Interval Score (WIS) |
|---|---|---|---|---|
| **20 days ahead** | | | | |
| Top ranked sub-epidemic model | 646880.00 | 570.34 | 85.15 | 382.90 |
| Ensemble(2) model | 576700.00 | 544.35 | 88.57 | 354.04 |
| Ensemble (3) model | 558890.00 | 540.71 | **89.59** | 350.73 |
| Ensemble (4) model | 557130.00 | 539.30 | **89.44** | **346.83** |
| (log) ARIMA model | 591980.00 | 536.22 | 51.07 | 422.41 |
| ARIMA model | **538690.00** | **528.87** | 55.05 | 404.92 |
| **30 days ahead** | | | | |
| Top ranked sub-epidemic model | 749560.00 | 613.75 | 82.18 | 421.29 |
| Ensemble (2) model | 670740.00 | 586.52 | 87.35 | 383.36 |
| Ensemble (3) model | 650790.00 | 584.20 | **88.20** | 382.79 |
| Ensemble (4) model | **644270.00** | **579.77** | **88.16** | **377.64** |
| (log) ARIMA model | 818530.00 | 621.58 | 57.99 | 767.05 |
| ARIMA model | 656480.00 | 591.93 | 60.34 | 439.29 |

# Performance metrics



20 April 2020 to 28 February 2022

# Representative top-ranking sub-epidemic forecasts



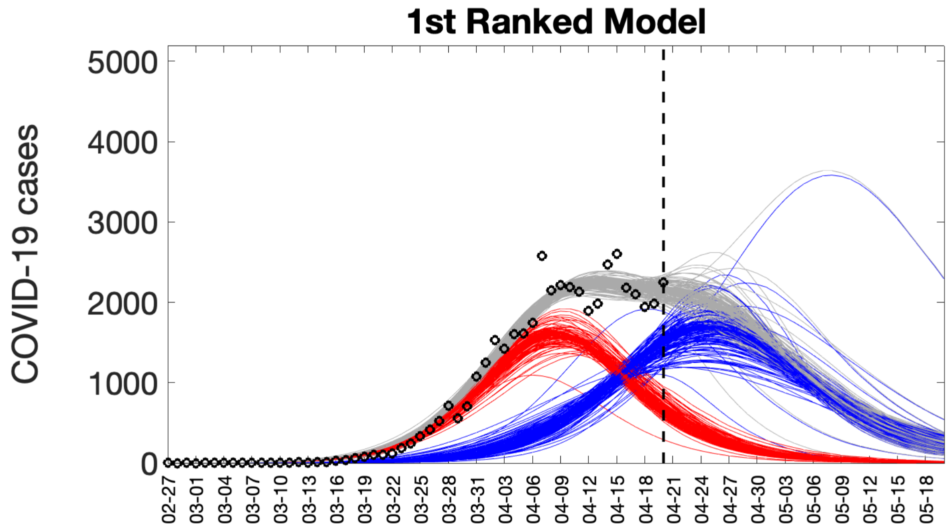
1st Ranked Model

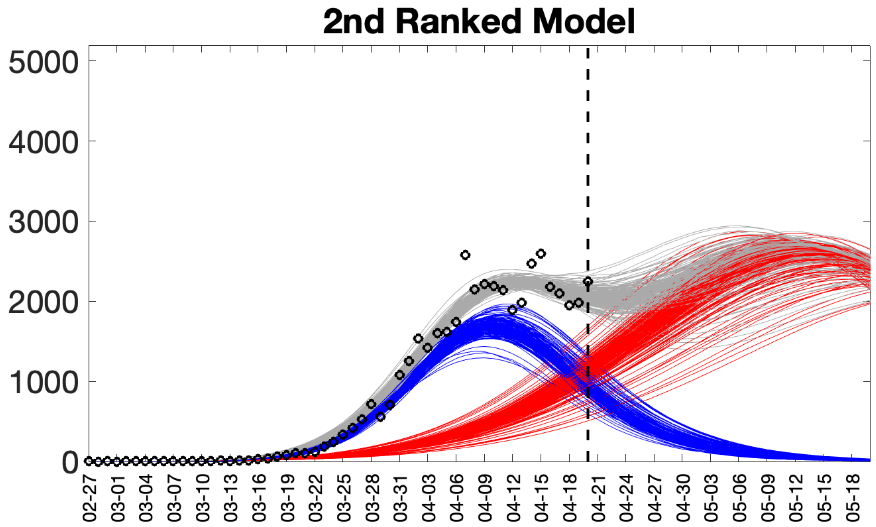
2nd Ranked Model

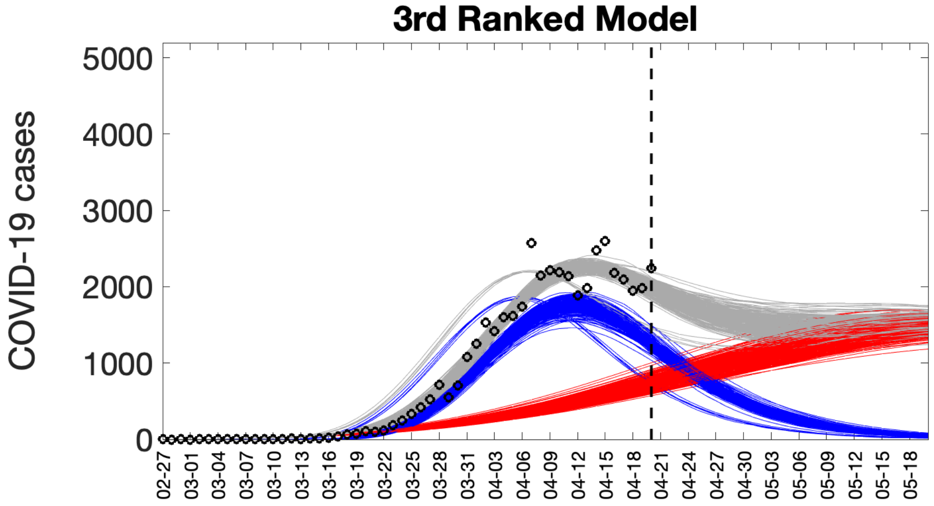
3rd Ranked Model

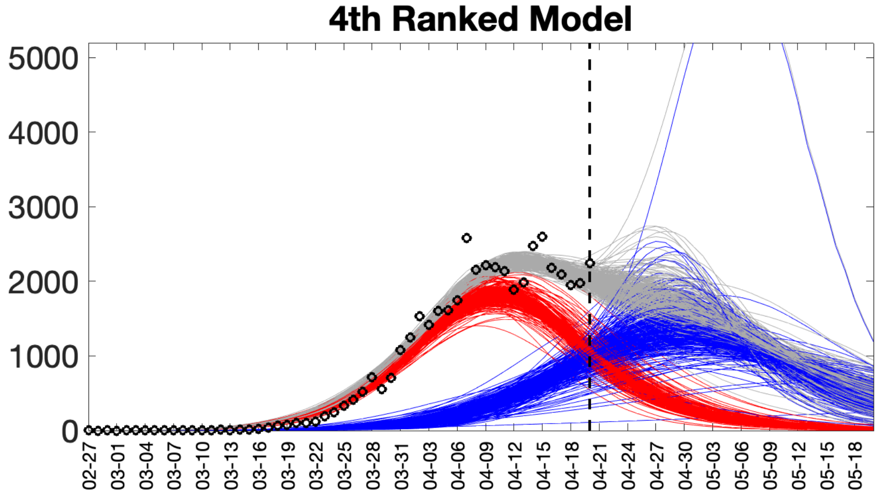
4th Ranked Model

# Representative top-ranking sub-epidemic profile forecasts

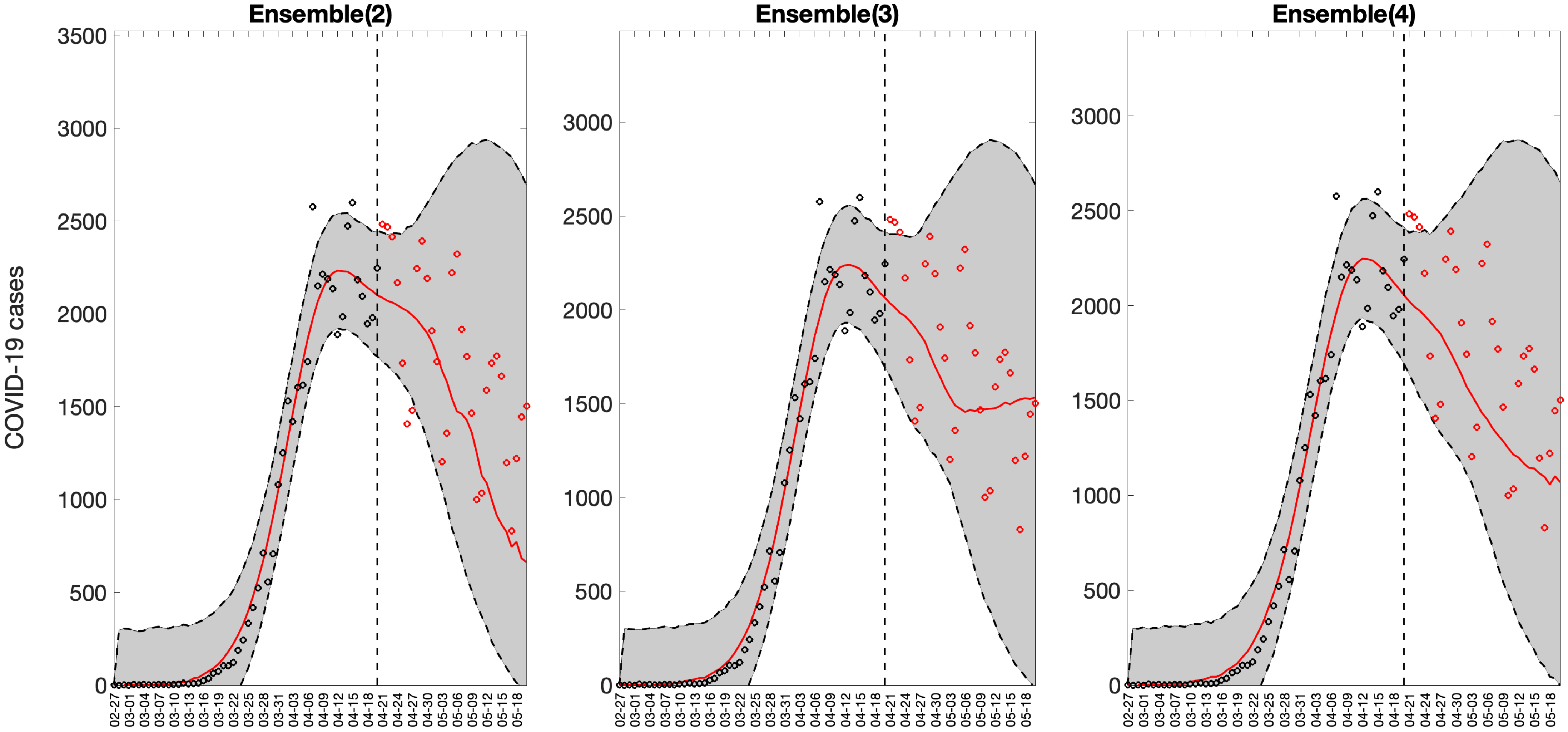

1st Ranked Model

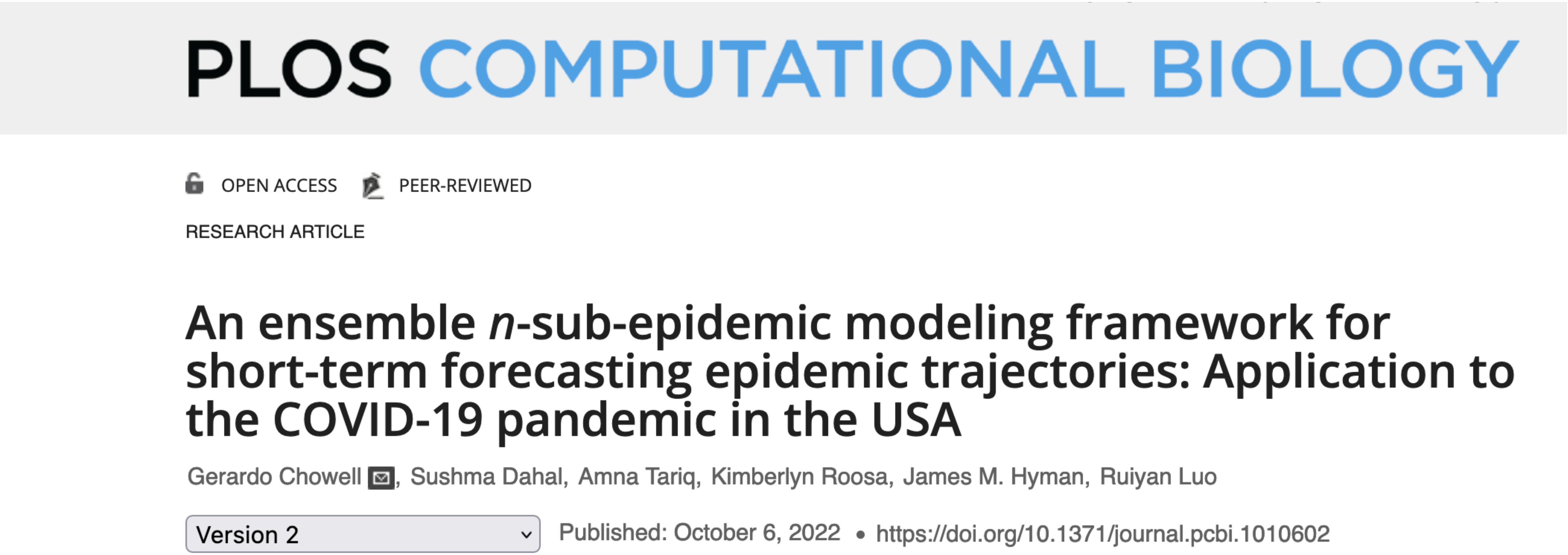

2nd Ranked Model


3rd Ranked Model


4th Ranked Model

# Representative ensemble sub-epidemic forecasts derived from top-ranking sub-epidemic models




PLOS COMPUTATIONAL BIOLOGY

RESEARCH ARTICLE

# An ensemble *n*-sub-epidemic modeling framework for short-term forecasting epidemic trajectories: Application to the COVID-19 pandemic in the USA

Gerardo Chowell, Sushma Dahal, Amna Tariq, Kimberlyn Roosa, James M. Hyman, Ruiyan Luo

Published: October 6, 2022 • https://doi.org/10.1371/journal.pcbi.1010602

# Summary

- Our ensemble sub-epidemic models outperformed top-ranking sub-epidemic models and a set of ARIMA models in weekly short-term forecasts covering the national trajectory of the COVID-19 pandemic in the USA from the early growth phase up until the Omicron-dominated wave.

- Forecasting performance consistently improved for the ensemble sub-epidemic models that incorporated a higher number of top-ranking sub-epidemic models.

- The sub-epidemic framework could also be used to forecast other biological and social growth processes, such as the epidemics of lung injury associated with e-cigarette use or vaping and the viral spread of information through social media platforms.

- Code and performance metrics are publicly available in a GitHub repository to facilitate comparison with other modeling approaches.

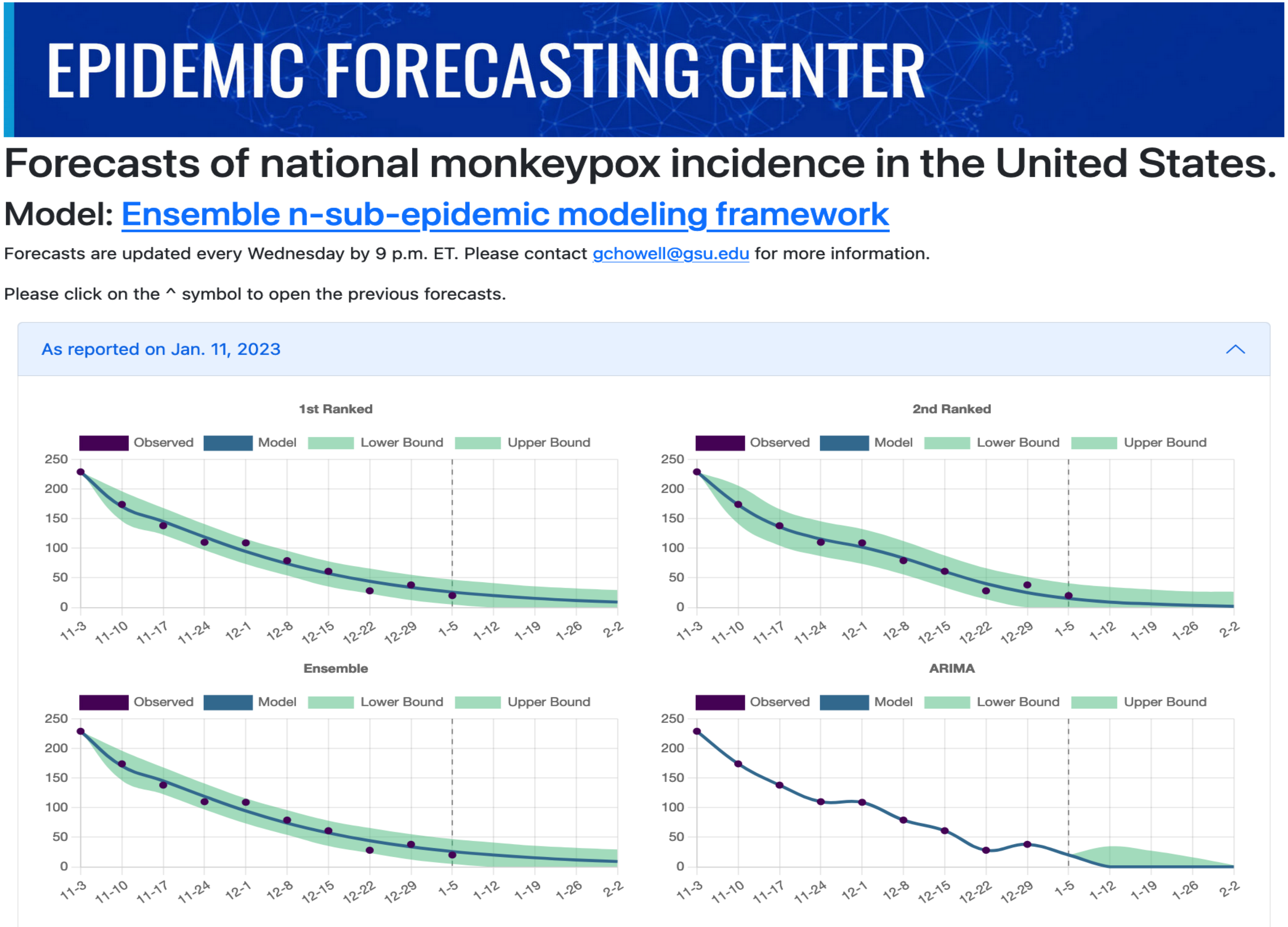
- Possibilities for further development of the framework.

# Real-time forecasting monkeypox, July-October 2022

- **Data**. Weekly updates of the daily confirmed monkeypox cases by date of report from the CDC and the Global.health (G.h) GitHub repository. The CDC and G.h data sources define a confirmed case of monkeypox as a person with a laboratory-confirmed case of monkeypox.

- **Scope**. At the global level and for countries that have reported the great majority of the cases including Brazil, Canada, England, France, Germany, Spain, and the United States.

- **Forecasting periods.** Data updated on Wednesday evening from both the CDC and GitHub Global.health (G.h) repository.



# EPIDEMIC FORECASTING CENTER

## Forecasts of national monkeypox incidence in the United States.

### Model: Ensemble n-sub-epidemic modeling framework

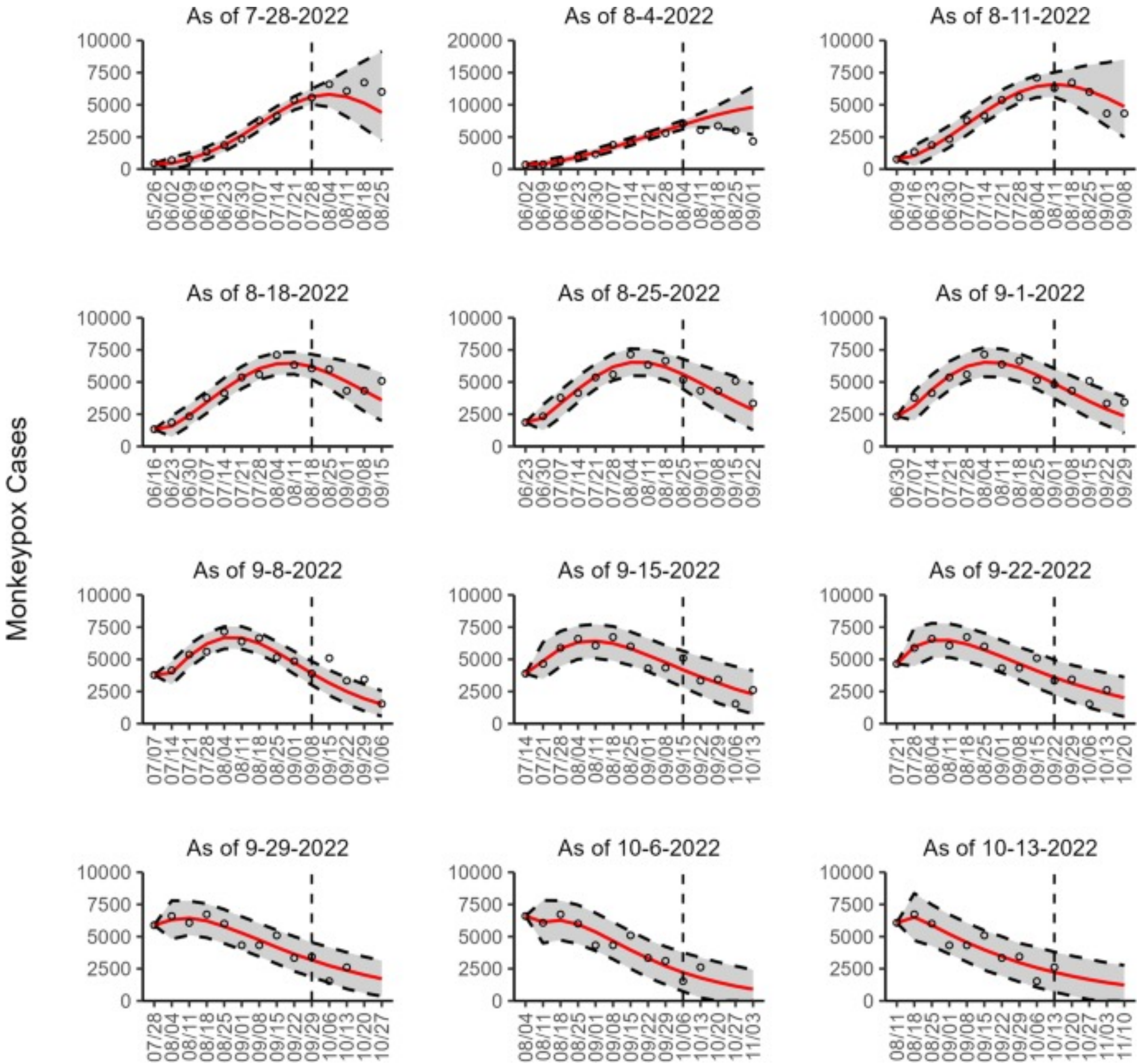
Forecasts are updated every Wednesday by 9 p.m. ET. Please contact gchowell@gsu.edu for more information.

Please click on the ^ symbol to open the previous forecasts.

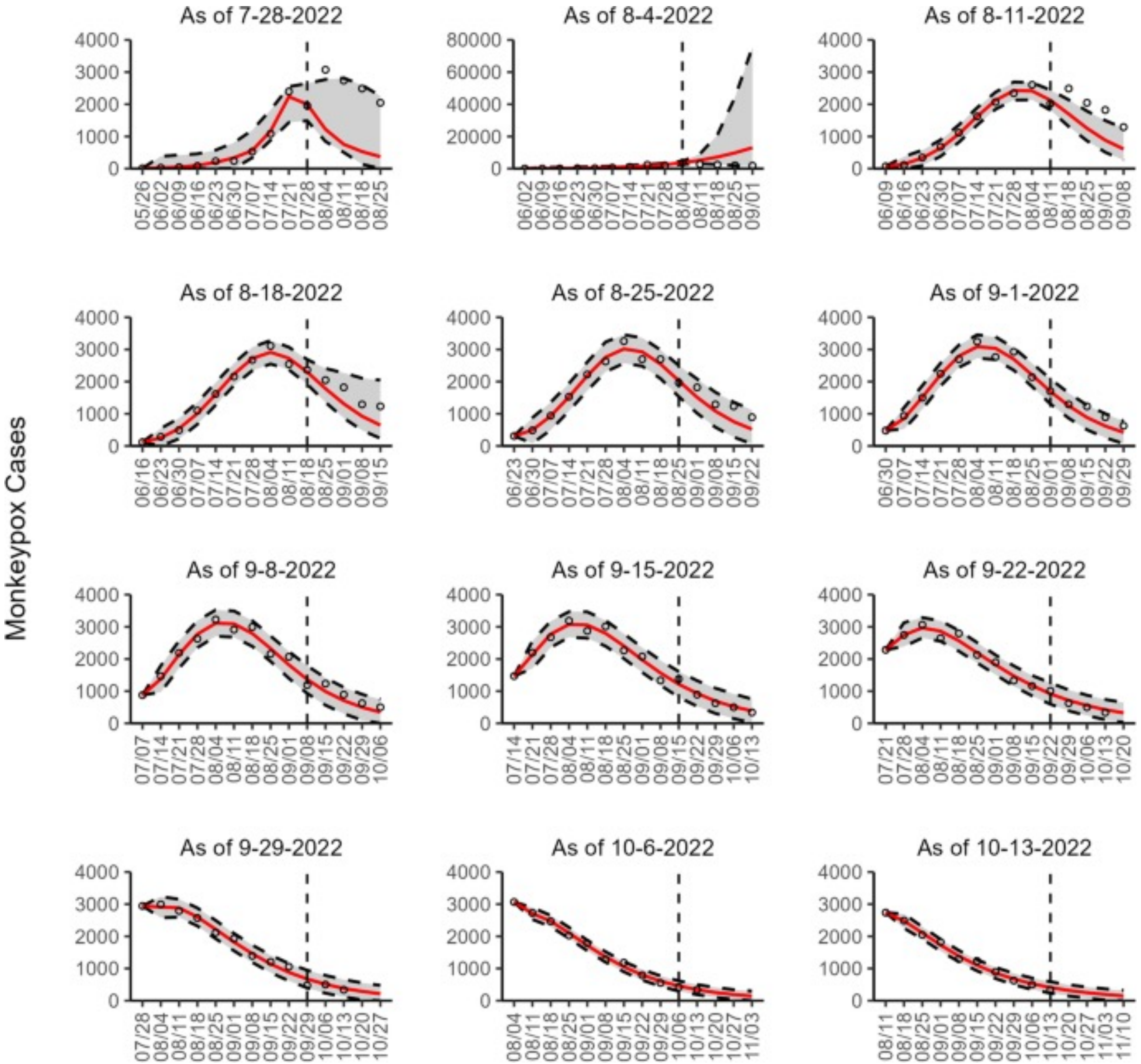
As reported on Jan. 11, 2023

https://publichealth.gsu.edu/research/monkeypox-forecasting-center/

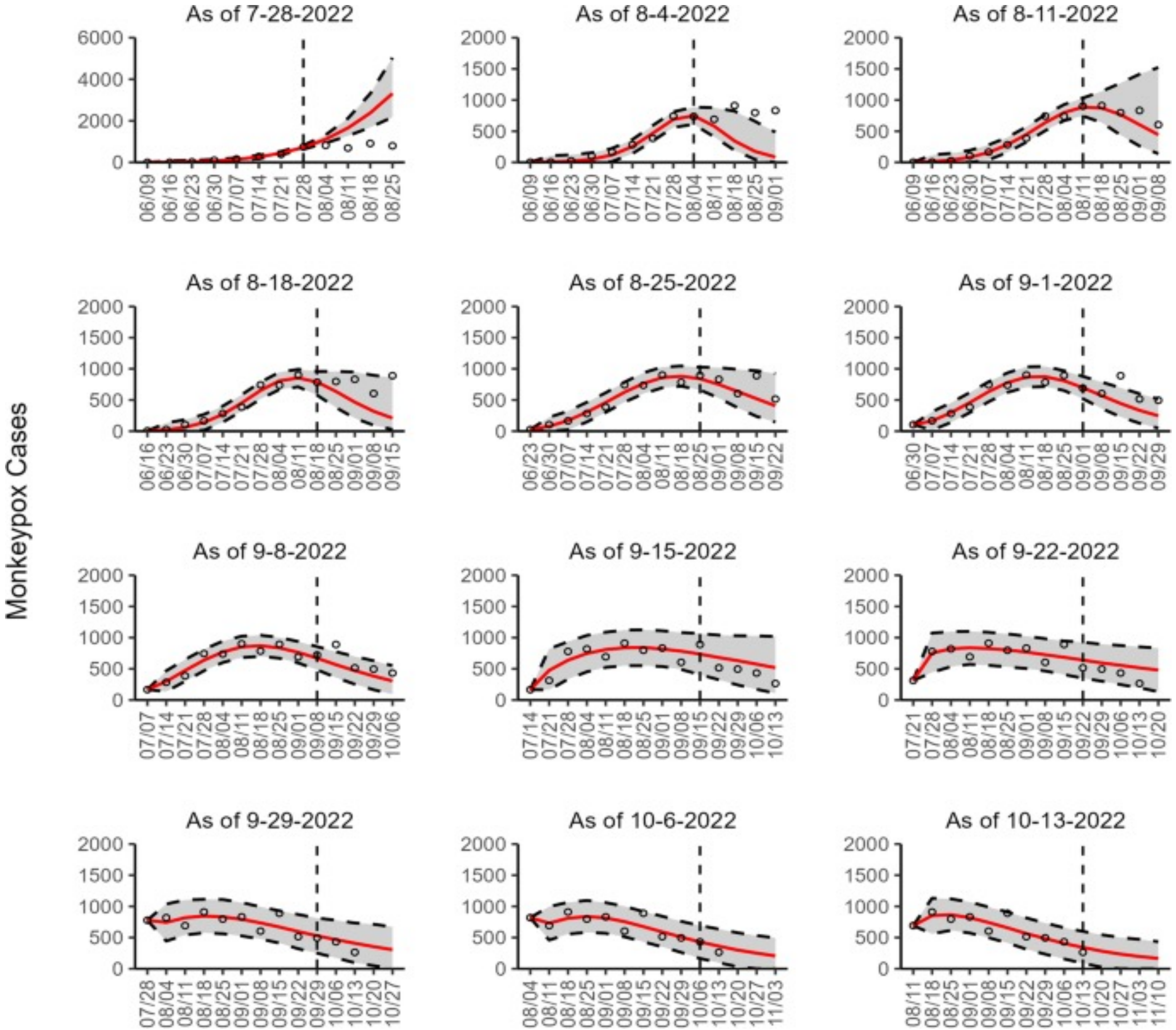
# Weighted ensemble model forecasts, Global



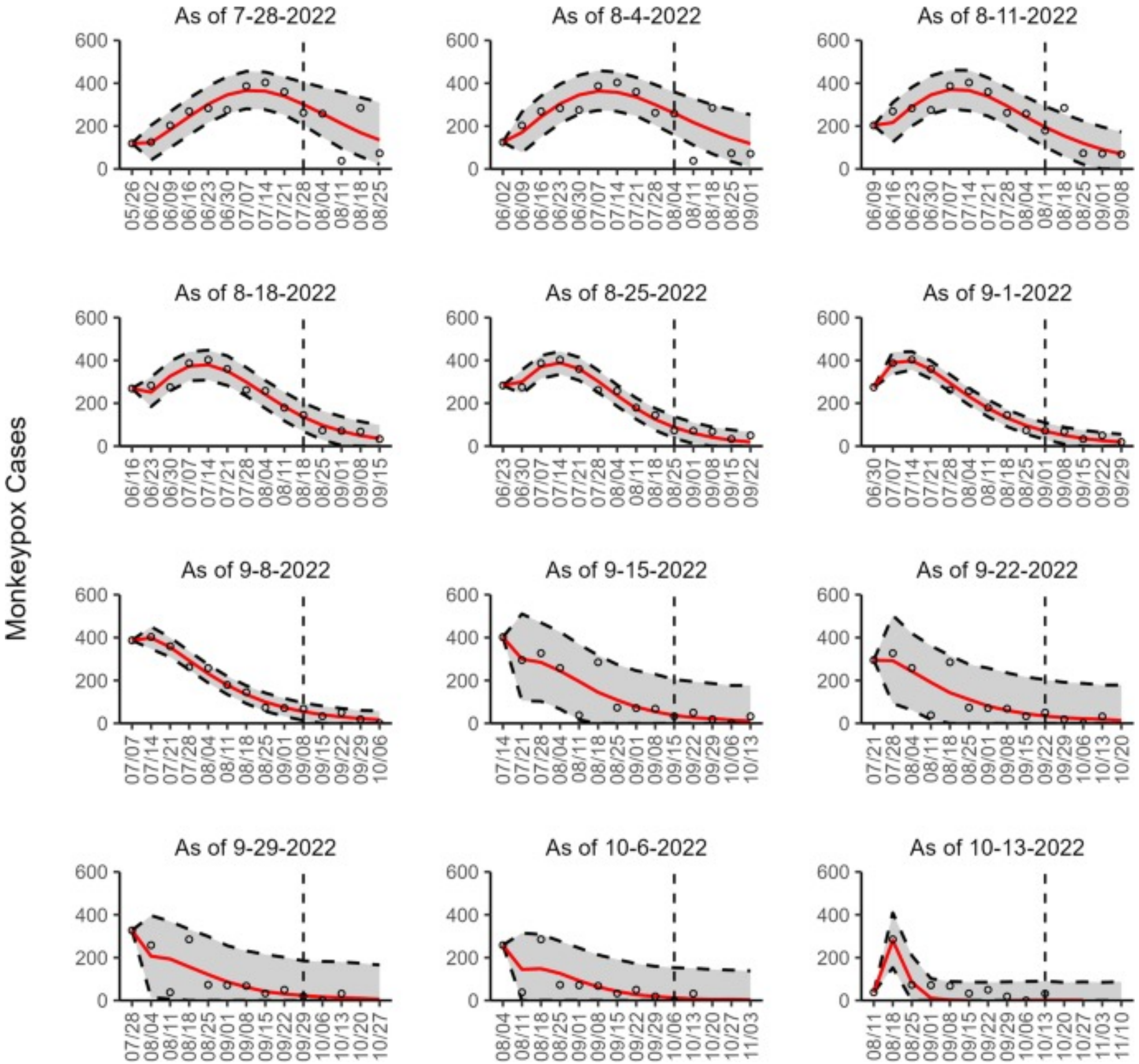
# Weighted ensemble model forecasts, USA

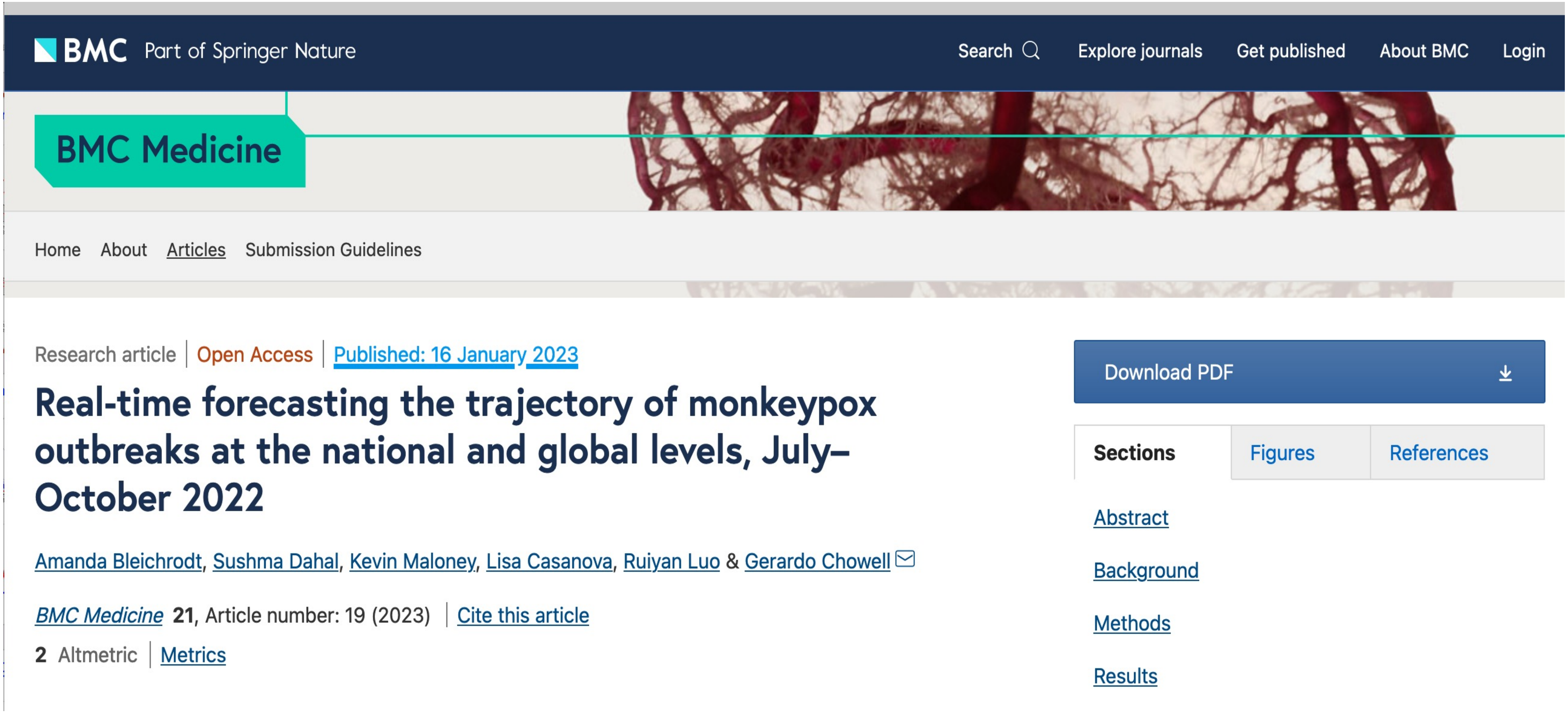


# Weighted ensemble model forecasts, Brazil



# Weighted ensemble model forecasts, England





Research article | Open Access | Published: 16 January 2023

# Real-time forecasting the trajectory of monkeypox outbreaks at the national and global levels, July–October 2022

Amanda Bleichrodt, Sushma Dahal, Kevin Maloney, Lisa Casanova, Ruiyan Luo & Gerardo Chowell

*BMC Medicine* **21**, Article number: 19 (2023) | Cite this article

**2** Altmetric | Metrics

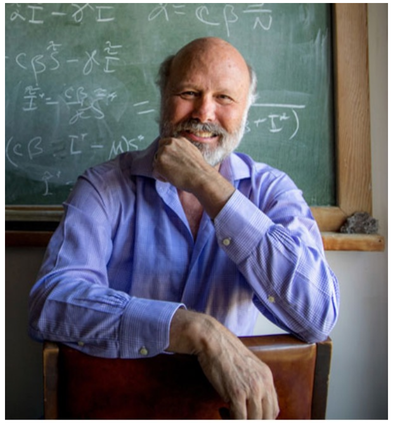
# Collaborators


Kimberlyn Roosa
Now postdoc


Amna Tariq
Now postdoc

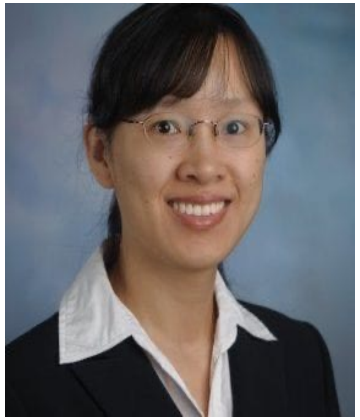

Sushma Dahal
PhD student


Amanda Bleichrodt
PhD student


Mac Hyman (Tulane)


Kevin Maloney (GSU)


Lisa Casanova (GSU)


Ruiyan Luo (GSU)

NSF