Understanding the spread of COVID-19 in the U.S. using inter-state mobility data

Yu Chen
Supervisor: Dr. Oliver Ratmann

September 17, 2020

Current Covid-19 situation

- The severe Sars-Cov-2 situation in the United States.

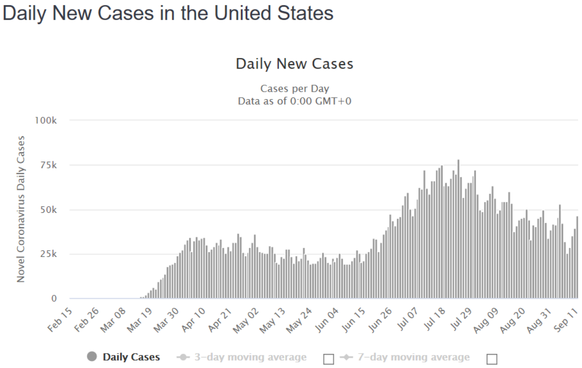

Figure: The daily infections in the U.S. up to 13th September 2020[1]

[1] Worldometers, *United States Coronavirus: 6,676,601 Cases and 198,128 Deaths - Worldometer*.

Basics of the semi-mechanistic COVID-19 model

We adapt the model based on the semi-mechanistic COVID-19 models[2].

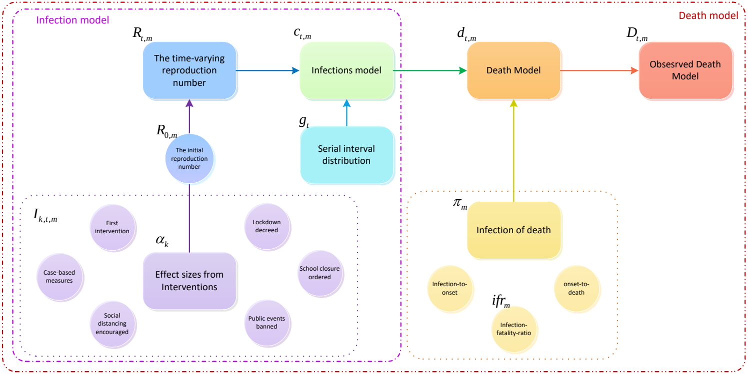

Figure: Conditional independence graph of semi-mechanistic model

[2] Flaxman et al., "Report 13: Estimating the number of infections and the impact of non-pharmaceutical interventions on COVID-19 in 11 European countries".

Infection model I

$$c_{t,m} = s_{t,m} R_{t,m} \sum_{\tau=0}^{t-1} c_{\tau,m} g_{t-\tau}, \tag{1a}$$

$$R_{t,m} = R_{0,m} \exp(-\sum_{k=1}^{6} \alpha_k I_{k,t,m}), \tag{1b}$$

t	day, $1, 2, \cdots$
m	locations (countries in Europe or states, counties in the U.S.)
$c_{t,m}$	the estimated number of infections on day t in location m
R_t	the effective reproduction number
R_0	the initial reproduction number
$I_{k,m,t}$	the kth intervention indicator on day t in location m
α	the effect size for the interventions

Infection model II

$$s_{t,m} = 1 - \frac{\sum_{i=1}^{t-1} c_{i,m}}{N_m}, \tag{2a}$$

$$g_s = \int_{\tau=s-0.5}^{s+0.5} g(\tau) d\tau, \quad s = 2, 3, \cdots, \tag{2b}$$

$$g_1 = \int_{\tau=0}^{1.5} g(\tau) d\tau. \tag{2c}$$

s	the proportion of susceptible individuals
N_m	the total population size in location m
$g(t)$	the serial interval distribution of time t
g_s	the discretised serial interval on day s

Death model

$$d_{t,m} = \sum_{\tau=0}^{t-1} c_{\tau,m} \pi_{t-\tau,m}, \tag{3a}$$

$$D_{t,m} \sim \text{Negative Binomial}\left(d_{t,m}, d_{t,m} + \frac{d_{t,m}^2}{\varphi}\right), \tag{3b}$$

$$\pi_m \sim ifr_m(\Gamma(5.1, 0.86) + \Gamma(18.8, 0.45)), \tag{3c}$$

$$\pi_{s,m} = \int_{\tau=s-0.5}^{s+0.5} \pi_m(\tau) d\tau, s = 2, 3, \cdots, \tag{3d}$$

$$\pi_{1,m} = \int_{\tau=0}^{1.5} \pi_m(\tau) d\tau.$$

$d_{t,m}$	the expected daily number of death on day t in location m
$D_{t,m}$	the reported death data on day t in location m
$\pi_m(t)$	the death distribution of time t
ifr_m	the infection-fatality-ratio in location m

Prior densities

$$\begin{aligned}
\varphi &\sim N_{[0,\infty)}(0, 5^2), && \text{(4a)}\\
c_{1,m}, c_{2,m} \cdots, c_{6,m} &\sim \exp(\frac{1}{\tau}), && \text{(4b)}\\
\tau &\sim \exp(0.03), && \text{(4c)}\\
\alpha_k &\sim \Gamma(1/6, 1) - \frac{\log(1.05)}{6}, && \text{(4d)}\\
R_{0,m} &\sim N(2.4, |\kappa|), && \text{(4e)}\\
\kappa &\sim N(0, 0.5^2), && \text{(4f)}\\
g &\sim \Gamma(6.5, 0.62). && \text{(4g)}
\end{aligned}$$

Incorporating transmission dynamics across population age groups

The renewable infection model Eq (1a) is modified by Oliver, Mélodie and other researchers in Imperial College COVID-19 response team,

$$c_{t,m,a} = \sum_{a' \in \mathcal{A}} s_{m,t,a} \rho_0 \gamma_{m,t,a',a} \sum_{s=1}^{t-1} c_{s,m,a'} g(t-s), \tag{5a}$$

$$\gamma_{m,t,a',a} = \eta_{m,t} \gamma_{m,a',a}, \tag{5b}$$

$$\eta_{m,t} = \exp\left(\sum_{k=1}^{3} X_{k,t,m} \alpha_k\right). \tag{5c}$$

a, a'	age groups
ρ_0	the probability of getting infected given contacts
$\gamma_{m,t,a',a}$	the contact intensities from age group a' to a
$\gamma_{m,a',a}$	the contact intensities prior to COVID-19 (the baseline)
$\eta_{m,t}$	the contact multiplier on day t in location m

Motivation

- COVID-19 is a respiratory disease, generally spreading across locations.
- Existing models fail to consider the human mobility **between** locations (states).

Data

- Inter-state mobility data set[3].
- Foursquare data set[4].
- Reported daily deaths data[5].

[3] United States Census Bureau, *American Community Survey (ACS)*.
[4] Waksman, *Phones, Lambdas and the Joy of Snap-to-Place Technology*.
[5] Johns Hopkins Coronavirus Resource Center, *COVID-19 Data in Motion - Johns Hopkins Coronavirus Resource Center*; The New York Times, *Coronavirus in the U.S.: Latest Map and Case Count - The New York Times*.

Inter-state mobility data set

ID	US state c	Estimated population N_c	State c' of residence one year ago ($N_{c' \to c}$)				
			Alabama	Alaska	Arizona	Arkansas	California
1	Alabama	4832358	4707940	419	1915	1051	2659
2	Alaska	727164	518	689936	910	722	3455
3	Arizona	7090137	2801	2743	6776152	1487	68516
4	Arkansas	2975961	438	413	602	2894068	3728
5	California	39114889	4591	2155	33670	2573	38325255
6	Colorado	5633029	3597	2102	7987	1442	28288
7	Connecticut	3538203	1054	8	1204	51	3870
8	Delaware	958770	39	0	115	0	551
9	District of Columbia	693798	174	0	204	0	3803
10	Florida	21092877	12870	1388	4542	3239	26888
11	Georgia	10410462	15589	917	2255	1900	16745
12	Hawaii	1403653	484	1509	2167	0	10456
13	Idaho	1734756	84	1179	3852	496	21018
14	Illinois	12599244	1280	1400	5818	1188	16518
15	Indiana	6612948	966	138	3009	1271	6811
16	Iowa	3119306	148	66	2021	410	4408
17	Kansas	2874208	476	378	1801	1507	5660
18	Kentucky	4414714	2055	115	1480	397	2052
19	Louisiana	4603081	1462	515	276	1653	5743
20	Maine	1324095	0	47	765	0	850
21	Maryland	5974521	647	485	2240	785	7924

Figure: Part of the inter-state mobility data

The mobility population from state c' to c is denoted by $N_{c' \to c}$.

Inter-state mobility flow estimation

The outflow rate,

$$M(t) = \begin{pmatrix} m_{11}(t) & \cdots & m_{1c'}(t) & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ m_{c1}(t) & \cdots & m_{cc'}(t) & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{pmatrix} \tag{6a}$$

$$= \begin{pmatrix} \frac{N_{1\to1}(t)}{N_1(t)} & \cdots & \frac{N_{1\to c'}(t)}{N_1(t)} & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ \frac{N_{c\to1}(t)}{N_c(t)} & \cdots & \frac{N_{c\to c'}(t)}{N_c(t)} & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{pmatrix}. \tag{6b}$$

The estimated daily mobility flow patterns,

$$\hat{M} = [\hat{m}_{c,c'}]_{c,c'}, \tag{7a}$$

$$\hat{m}_{c,c'} = \frac{1}{365}\frac{N_{c\to c'}}{N_c}. \tag{7b}$$

Contact rate estimation

The total number of contacts,

$$C_c \propto \frac{N_c}{A_c}. \tag{8}$$

The number of contacts made by people in state c' and people in c,

$$C_{c' \to c} = \frac{N_{c'}\, m_{c',c}\, N_c}{\sum_d N_d\, m_{d,c}\, N_c}\, C_c = z \frac{N_{c'}\, m_{c',c}}{\sum_d N_d\, m_{d,c}} \frac{N_c}{A_c}. \tag{9}$$

The average number of contacts made by a person in state c' with people in state c prior to COVID-19 is

$$\gamma_{c',c} = \frac{C_{c' \to c}}{N_{c'}} = z \frac{N_c\, m_{c',c}}{A_c \sum_d N_d\, m_{d,c}}. \tag{10}$$

The mobility trends

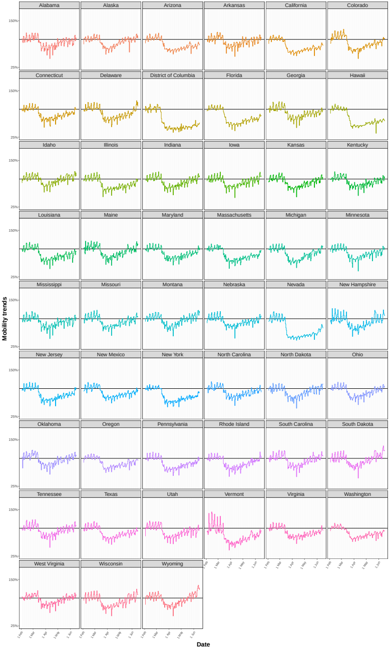

Figure: The mobility trends (change from the baseline) across 50 U.S. states and the District of Columbia

In our framework, the mobility trends X is then decomposed by the rebound point $t^{rebound}$.

$$t_c^{rebound} = \arg\min_t X_{t,c}^{avg}, \tag{11}$$

$$X_{t,c}^{avg} = \sum_{s=t-5}^{t+5} X_{s,c}, \tag{12}$$

$$X_{t,c} = X_{t,c}^{eased} \times X_{t,c}^{rebound}. \tag{13}$$

Two components of the mobility trends visualisation

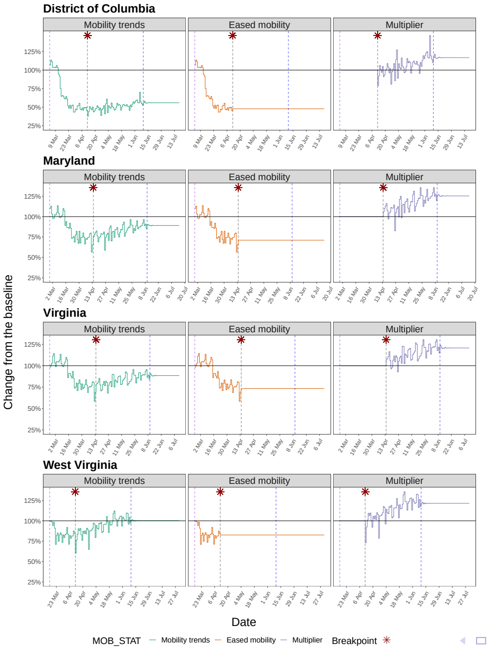

Figure: The plots of mobility trends and two covariates, decomposed by the rebound point (Breakpoint in the figure) with the average of the past five days as predictions.

Dynamic contact intensities construction

The contact multiplier is

$$\eta_{t,c} = \exp(\beta_1 \log(X_{t,c}^{eased}) + \beta_2 \log(X_{t,c}^{rebound})). \tag{14}$$

The contact intensities from c' to c over time are

$$\gamma_{t,c',c} = z \frac{(N_c \, m_{c',c}) \, \eta_{t,c} \, \eta_{t,c'}}{A_c \sum_d N_d \, m_{d,c} \, \eta_{t,d}}. \tag{15}$$

Infection process reconstruction

The renewable equation is modified as

$$c_{t,c} = \sum_{c' \in \mathcal{C}} s_{t,c}\ \rho_{0,c}\ \gamma_{t,c',c} \sum_{s=1}^{t-1} c_{s,c'}\ g(t-s), \tag{16}$$

where $\rho_{0,c} = \frac{R_{0,c}}{\gamma_c}$ is the state-specific effect and $\gamma_c = \sum_{c \in \mathcal{C}} \gamma_{c,c'}$.

The effective reproduction number is

$$R_{t,c} = \sum_{c' \in \mathcal{C}} s_{t,c'}\ \rho_{0,c}\ \gamma_{t,c,c'}. \tag{17}$$

Model fitting

- Fit model to death data
- Utilise Stan[6].
- Use an adaptive extension of the Hamiltonian Monte Carlo algorithm.
- Draw samples from two chains, each has 4000 iterations with 1500 warmups.
- Good performance and convergent chains.

[6] Carpenter et al., "Stan: A probabilistic programming language".

Dynamic contact multipliers estimation

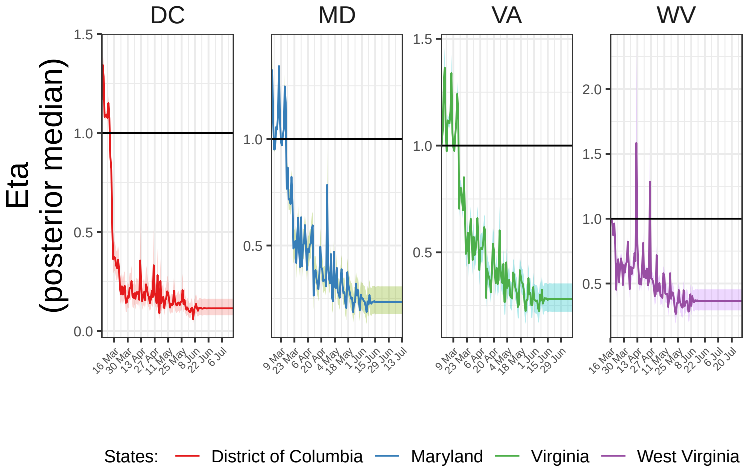

Figure: The posterior median of η with 95% credible interval

Sars-Cov-2 transmission dynamics estimation

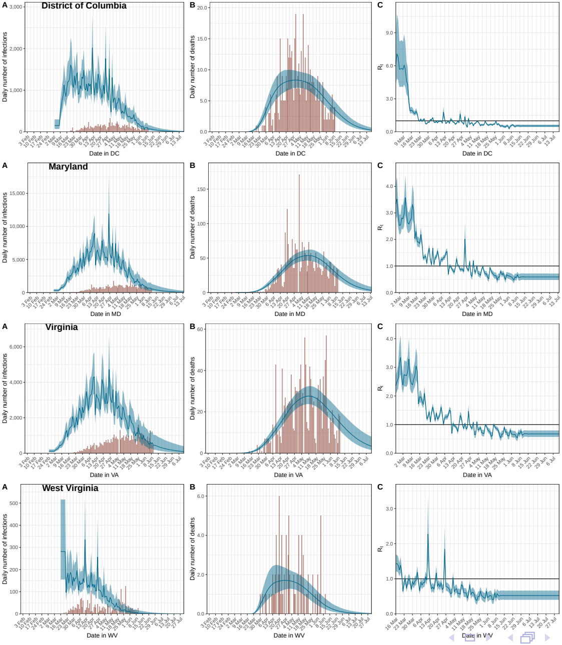

Figure: Three-panel plots: Estimates of daily infections, deaths and R_t for DC, MD, VA, and WV, respectively.

Comparison with previous work

The attack rate is

$$\text{attack rate}_c = \sum_t c_{t,c}/N_c. \tag{18}$$

state	our model		Unwin et al. (2020)[7] model	
	posterior mean	95% C.I.	posterior mean	95% C.I.
DC	10.76%	[7.68%, 14.80%]	11.6%	[8.4%, 16.2%]
MD	5.85%	[4.32%, 7.85%]	5.9%	[4.2%, 8.5%]
VA	2.20%	[1.63%, 2.94%]	2.5%	[1.7%, 3.8%]
WV	0.51%	[0.31%, 0.79%]	0.6%	[0.3%, 1.0%]

Table: Comparison the posterior mean with 95% credible interval of the attack rate for the selected four states as of 25th May 2020

[7] Unwin et al., "Report 23: State-level tracking of COVID-19 in the United States. 2020; 1–43.[cited 21 May 2020]".

Recommended future work

- Use the random effects for the effect size of the mobility trend covariates.
- Incorporate more states.
- Harness the updated data sets.

Conclusion

- The inter-state mobility flow patterns are proposed.
- The contact intensities model is reconstructed.
- The infection process is modified.
- The estimated contact multipliers decreased over time.
- The estimated R_t is less than 1 after June.
- As of 11th July 2020, the estimated attack rates for DC, MD, VA and WV are 11.12%, 6.15%, 2.39% and 0.50%, respectively.

References I

Carpenter, Bob et al. "Stan: A probabilistic programming language". In: *Journal of statistical software* 76.1 (2017).

Flaxman, Seth et al. "Report 13: Estimating the number of infections and the impact of non-pharmaceutical interventions on COVID-19 in 11 European countries". In: (2020).

Johns Hopkins Coronavirus Resource Center. *COVID-19 Data in Motion - Johns Hopkins Coronavirus Resource Center*. 2020. URL: https://coronavirus.jhu.edu/covid-19-daily-video (visited on 08/20/2020).

The New York Times. *Coronavirus in the U.S.: Latest Map and Case Count - The New York Times*. 2020. URL: https://www.nytimes.com/interactive/2020/us/coronavirus-us-cases.html (visited on 08/20/2020).

References II

United States Census Bureau. *American Community Survey (ACS)*. 2020. URL: https://www.census.gov/programs-surveys/acs/ (visited on 06/05/2020).

Unwin, H et al. “Report 23: State-level tracking of COVID-19 in the United States. 2020; 1–43.[cited 21 May 2020]”. In: (2020).

Waksman, Adam. *Phones, Lambdas and the Joy of Snap-to-Place Technology*. 2020. URL: https://enterprise.foursquare.com/intersections/article/phones-lambdas-and-the-joy-of-snap-to-place-techn/ (visited on 08/31/2020).

Worldometers. *United States Coronavirus: 6,676,601 Cases and 198,128 Deaths - Worldometer*. 2020. URL: https://www.worldometers.info/coronavirus/country/us/ (visited on 09/13/2020).