Globally local: Hyper-local modeling for accurate forecast of COVID-19

Vishrawas Gopalakrishnan [a], Sayali Pethe [a], Sarah Kefayati [a], Raman Srinivasan [a], Paul Hake [a],
Ajay Deshpande [a], Xuan Liu [a], Etter Hoang [b], Marbelly Davila [b], Simone Bianco [c],
James H. Kaufman [c,*]

[a] IBM Watson Health, 75 Binney St, Cambridge, MA, 02142, USA
[b] Tampa General Hospital, 1 Tampa General Circle, Tampa FL 33605, USA
[c] IBM Research, Almaden Research Center, 650 Harry Rd, San Jose, CA, 95120, USA

A B S T R A C T

Importance: Assumption of a well-mixed population during modeling is often erroneously made without due analysis of its validity. Ignoring the importance of the geo-spatial granularity at which the data is collected could have significant implications on the quality of forecasts and the actionable clinical recommendations that are based on it.

Objective: This paper's primary objective is to test the hypothesis that the characteristic dynamics defining the trajectory of the pandemic in a region is lost when the data is aggregated and modeled at higher geo-spatial levels.

Design: We use publicly available confirmed SARS-CoV-2 cases and deaths from January 1st, 2020 to August 3rd, 2020 in the United States at different geo-spatial granularities to conduct our experiments. To understand the impact of this hypothesis, the output of this study was implemented in Tampa General Hospital (TGH) to provide resource demand forecast.

Results: The Mean Absolute Percentage Error (MAPE) in the forecast confirmed cases can be 30% higher for modeling at the state-level than aggregating model results at the scale of counties or clusters of counties. Similarly, modeling at a state-level and crafting policy decisions based on them may not be effective — county-level forecasts made by partitioning state-level forecasts are 3x worse for confirmed cases and 20x worse for deaths relative to the same model at the county level. By leveraging these results, TGH was able to accurately allocate clinical resources to tackle COVID-19 cases, continue elective surgical procedures largely uninterrupted and avoid costly construction of overflow capacity in the first two epidemic waves.

Conclusions and Relevance: Accurate forecasting at the county level requires hyper-local modeling with county resolution. State-level modeling does not accurately predict community spread in smaller sub-regions because state populations are not well mixed, resulting in large prediction errors. Actionable decisions such as deciding whether to cancel planned surgeries or construct overflow capacity require models with local specificity.

1. Introduction

The COVID-19 pandemic caused by the SARS-CoV-2 virus reveals the important contributions that can be made by accurate mathematical modeling of infectious disease. A variety of modeling approaches can be used to predict the course of an infection in a population. While the pandemic is global, the underlying disease dynamics (and the epidemiological parameters themselves) varies with specific local policies, non-pharmaceutical interventions (NPI), variations in demographics, healthcare infrastructure, etc. To realize the full potential of any modeling effort, forecasts must be accurate and meaningful to public health officials and governing authorities at a resolution relevant to their areas of responsibility (Haffajee and Mello, 2020). In most cases, this requires the forecasts be made at a hyper-local level. Models can only be used to guide closing or opening of business, forecast and allocate Intensive Care Units (ICU) resources, and establish social distancing guidance if their forecasts are meaningful at a scale relevant to these businesses, hospitals, and communities.

In practice, the ability to model at lower geo-spatial levels (e.g., county) is limited by the granularity and accuracy of ground truth data, as well as the underlying modeling assumptions (Chin et al.,

---

2020; Voss, 2009). Precision of public health SARS-CoV-2 case reports, including confirmed cases (reported incidence) and deaths (Woolf et al., 2020), is impeded by multiple noise sources (Bergman et al., 2020). Clinical encounters taking place adjacent to or outside a patient's residential district, testing capacity, public health reporting practice, and many other factors that can vary dramatically even between nearby districts affect the quality of the data. Some of the effects of these noise sources may be reduced by clustering administrative divisions or modeling at higher geo-spatial levels (e.g., state), but this approach may violate an assumption used by most compartmental models, that the population being modeled is well mixed (Tolles and Luong, 2020).

To illustrate the practical implication of this assumption, consider New York state. If we estimate the county forecast from the state-level model, we over-estimate the cases by 7x compared to the county model. It is straightforward to see that errors of such magnitude would have significant repercussions in the accurate planning of hospital resources and crafting the government policies. Our results reveal that, for the United States, even for reasonably small states, different counties make dominant contributions to daily confirmed cases at different times.

In this paper, we compare the performance of a SARS-CoV-2 model as a function of spatial scale, applying the model to county, county cluster, and state level divisions. For this purpose, we selected an epidemiological compartment model, as opposed to a regression model, to investigate changes in important disease parameters obtained at each geo-spatial level, and to avoid an approach where important parameters are under determined. Forecasting power and scaling behavior are expected to depend on model selection. An open source model and modeling framework were chosen so others can reproduce the results and test future extensions that may improve model performance. To evaluate the performance of the selected model we also compare the accuracy to other models with predictions publicly available through the US Center for Disease Control and Prevention (CDC) (Anon, 2020a; Ray et al., 2020). These other models include both epidemiological compartment models, regression models, and hybrid approaches. It should be noted that the objective of this paper is not to demonstrate the superiority of one model over the others, rather the performance comparison with the various CDC models is to establish confidence on the further downstream analysis on the role of geo-spatial granularity in forecast accuracy. Towards that end, one could do a similar exercise using standard SIR or SEIR models; however, those models do not perform well relative to the other models as the asymptomatic and unreported cases contribute a large proportion of the total COVID-19 case load. Therefore, we chose to use the model described here. Although, we show the comparison with CDC published results for only one time point, the current model was run for over 30 consecutive weeks for Tampa General Hospital and performed consistently over that time period.

To summarize, in this paper:

1. We experimentally verify our hypothesis that state-level modeling strongly violates the assumption of well-mixed populations and will lead to inaccurate forecasts at a county level even while performing well at the state level.
2. We show that even smaller regions have distinct characteristic dynamics that affect the forecasts of the entire state. Consequently, the consumer's requirements must guide modeling resolution, and one size does not fit all.
3. We applied the model with Tampa General Hospital (TGH) to forecast COVID-19 hospitalization and ICU demand at a regional level. Through this exercise, we validate the need for modeling at a hyperlocal level and its efficacy in a hospital setting wherein, based on the hyperlocal forecasts, the hospital took appropriate resource staffing decisions; thereby, ensuring an adequate number of beds in case of spikes and waves and at the same time restarting non-emergency surgeries when possible
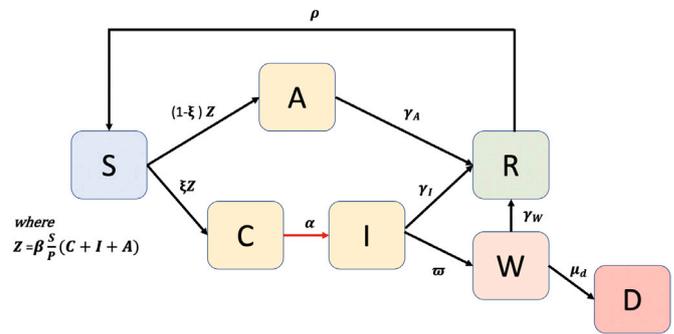

Fig. 1. Compartment model use to capture the disease dynamics of SARS-CoV-2 (see text).

2. Methods

2.1. Data description

This paper uses USAFacts COVID-19 datasets (Anon, 2020b), which in turn include daily COVID-19 confirmed cases and deaths obtained directly from CDC, state and local agencies. To ensure self-consistency of comparisons, all CDC model results and corresponding reference data were obtained close to the CDC publication date. This is necessary because reference data can be retrospectively updated, and we want to make a self-consistent comparison across models. This data is further curated; for instance, the post-processing ensures that the number of cumulative confirmed cases is a monotonically increasing curve. The data is refreshed every day around 9AM PST to reflect updates of the previous day. More information on the reference data (Adhikari et al., 2020; Lasry et al., 2020; Hsiang et al., 2020), and on the CDC challenge models (Anon, 2020a) used for model comparison is available in the supplement.

2.2. Base model

For this study we chose a modeling framework, the SpatioTemporal Epidemiological Modeler (STEM) (Douglas et al., 2019; Edlund et al., 2010), available through the Eclipse Foundation (Anon, 2006–2020a). The framework and model are open source and available under the Eclipse Public License (EPL2) (Alamoudi et al., 2020). A number of models for SARS-CoV-2 from multiple authors were evaluated (Schwartz et al., 2020; Gurbaxani et al., 2020; Anon, 2006–2020b). For this study we use the model shown in Fig. 1, which includes transmission from both asymptomatic/undetected (A), pre-symptomatic (C), and infectious individuals (I). Moreover, we capture infectious individuals who experience a worsening of symptoms (W) before either dying of the disease (D) or recovering (R). This model is a simple extension to the SACIR model available on the Eclipse site.

Model parameters and initial conditions were selected by a combination of literature references, grid search, and data-driven estimation through a Nelder–Mead AI simplex algorithm available in the STEM framework. The flow diagram of Fig. 1 translates into a system of ODEs reported in (Eq. 2) of the supplement. The definition of each compartment, all model parameters, literature values, and methods of parameter estimation are also discussed in the supplement. This basic epidemiological model was used to forecast at three spatial resolutions (counties, county clusters, and states). Definition of county clusters is described in the supplement.

The model was trained by minimizing the normalized root mean square error (NRMSE) of the forecasts of daily confirmed cases, daily rolling cumulative confirmed cases, and daily rolling cumulative deaths with respect to the corresponding ground truths. Consequently, we measure the model's forecasting performance by measuring the Mean

Absolute Percentage Error (MAPE) across these same three measures for forecasting time window ranging from one to four weeks into the future. The choice of including both cumulative and daily confirmed cases may be of interest to the readers. It was observed that without including the cumulative cases in the optimization, the model, in some cases, yields a total case count lesser than the ground truth even though the daily fit is extremely good. This was tracked to a 'parallel phase-shift' observed in the resultant cumulative curve resulting from the fitted daily model missing spike peaks but still largely tracking the rest of the daily cases. By including the cumulative confirmed cases in the objective function, the model adjusts the fitted daily curve such that the rolling cumulative cases as of that day is also correct. It should be noted that all three references — cumulative cases, daily cases, cumulative deaths, were given equal weights during the error minimization.

Since public health reporting usually takes place after a clinical encounter, the transition between the pre-symptomatic compartment (C) and the infectious compartment (I) was used to log daily confirmed cases (red arrow, Fig. 1). Here, we introduce a pre-symptomatic compartment C in place of the usual exposed compartment E based on documented evidence for replication-competent virus obtained from patients before the onset of symptoms (World Health Organization et al., 2020; Lauer et al., 2020; Liu et al., 2020; Wang et al., 2020; Wolfel et al., 2020). Thus, we describe the rate of individuals leaving the C compartment as a symptom appearance rate as opposed to an incubation rate. The processes are similar in that they both would contribute to a period of latency between exposure and clinical encounter. We chose the nomenclature to indicate that individuals in the C compartment contribute to the disease force of infection.

The steady state solution to the differential equations shown in the supplement (Eq. 2) provides the following expression for the basic reproductive number, R_0 as a function of the epidemiological parameters. Note that background birth and death rates are omitted, and individuals in the W compartment do not contribute to the force of infection.

$$R_0 = \frac{\beta}{\gamma_A} \left\{ 1 + \left(\frac{\gamma_A}{\gamma_I + \omega} \right) - \left[1 - \frac{\gamma_A}{\alpha} \right] \xi \right\} \tag{1}$$

To provide a scalable architecture for modeling, we built an automation pipeline that performs a variety of tasks ranging from data ingestion, smoothing, model invocation, and post-processing. The pipeline supports multi-processing, where each region is run by a separate instance of a docker image; allowing easy scaling of the compute cluster as required. This pipeline is discussed in the supplement. In response to changing policies and behaviors, disease parameters like transmission rate change through the pandemic. It is necessary to capture these variations, based on the epidemic curve, and to allow the model the flexibility to change these parameters in response to policies and behaviors. In this paper, we focus only on the transmission rate as an 'evolving' parameter. The objective of this paper is to highlight the importance of geo-spatial granularity. Hence, other parameters like case reporting rate are not allowed to vary dynamically. A model-centric discussion is not in the purview of this paper, and we refer interested audience to our paper (Gopalakrishnan et al., 2021). However, to summarize the methodology, we identify the discontinuities in the 'smoothed' daily confirmed cases curve by extracting the knees and elbows through the Kneedle algorithm (Satopaa et al., 2011). Each extracted discrete time point represents a change in the transmission rate that is passed to the differential equation solver to allow for the model to sufficiently adjust the parameters to fit the disease dynamics. Fig. 2 provides an illustration on how, based on the daily case curve in Hillsborough county, FL, the system detects the changes in the transmission rate. The figure further attempts correlate the detected changes in the transmission rate with the NPI on the ground. It is important to highlight here that the forecasting system does not explicitly look at the NPIs, i.e., it is purely data driven. The reason is because: (a) Our objective is to capture varying transmission rates and not predict effects of NPIs. (b) Related to (a), changes in transmission rate can happen even without NPIs, e.g., improved compliance of mask can decrease transmission rate or public holidays can result in increased population mixing and thus rise in transmission rate.

2.3. Hospitalization and intensive care unit forecasting model

In order to deliver actionable projections for decision making by healthcare providers, as a post process we decomposed the confirmed case projections using rate equations to forecast both total hospitalization and ICU demand at the county level. The average rate (weighted by confirmed cases) of hospitalization per confirmed case was 0.35 ± 0.04 and the average rate of ICU admission per hospitalized patient as 0.10 ± 0.05.

Tampa General Hospital in Hillsborough County, Florida, USA, implemented this model to anticipate future resource demand from COVID-19 patients and to advise decisions regarding possible need for extra capacity and reduction in planned surgeries to free up bed space. It is the only Level 1 Trauma Center within the Tampa region. As COVID cases rose in early March 2020, TGH has treated an average of 20.4 percent of the Hillsborough COVID-19 cases with a max distribution of 45.0 percent of all COVID-19 cases in Hillsborough County. They established a regional collaboration for data and resource sharing facilitated by shared data collection and a reporting dashboard. More information on this model is available in the supplement. Projections based on both models were updated at least weekly since April 2020.

3. Results

3.1. Derived epidemiological parameters

The model optimization process automatically adds discrete changes to the transmission rate β, at discrete times, based on the dynamics observed in ground truth confirmed cases and deaths. In an attempt to distinguish between a "natural" range of transmission rates, β, reflecting human contact rates in the absence of interventions, and the larger range of β with interventions, the violin plots in Fig. 3(a) are split. The left hand (cyan) distribution includes only the maximum β observed by region. The right (purple) distributions include all β values (i.e. all changes in β from local policies and NPI, as well as the maximum values). Similarly the distribution of R_0 derived from Eq. (1)) is shown in Fig. 3(b). These distributions are also split with the left hand violin displaying R_0 values computed from maximum β values observed by region. The right hand violin plot (labeled R_{eff}) includes reproductive numbers computed independently from *all* β values observed in all regions. The distribution of transmission rates and reproduction numbers are multi-modal reflecting differences in disease dynamics between regions. Distributions for other parameters and a table listing L-statistics for all optimized parameters and reproduction numbers may be found in the supplementary Table 4 and 5 respectively. The Table also shows the average model convergence error (NRMSE) which was under 4% for simulation at the scale of U.S. States, systematically increasing to just over 10% for simulation of U.S. counties.

The optimized parameter values and derived values for the basic and effective reproduction numbers are within the range estimated from current clinical studies (World Health Organization et al., 2020; Lauer et al., 2020; Liu et al., 2020; Wang et al., 2020; Wolfel et al., 2020). Supplementary Table 5 provides statistics from the measured distribution of reproduction number values obtained using the epidemiological parameters in Table 4 at each spatial resolution.

3.2. Model performance and forecasts

Fig. 4 shows example output of the epidemiological model for two states. The model was run at county, county cluster, and state resolution for all states (and DC). Fig. 4 shows the actual and forecast (daily) confirmed cases and (cumulative) deaths vs. time for all county clusters within each of the two example states. The data is presented as a stack chart so the envelop over aggregated clusters reflects total confirmed cases and deaths at the state level. This is also indicated by the blue line above the stacked sub-regions for both the confirmed (reported)
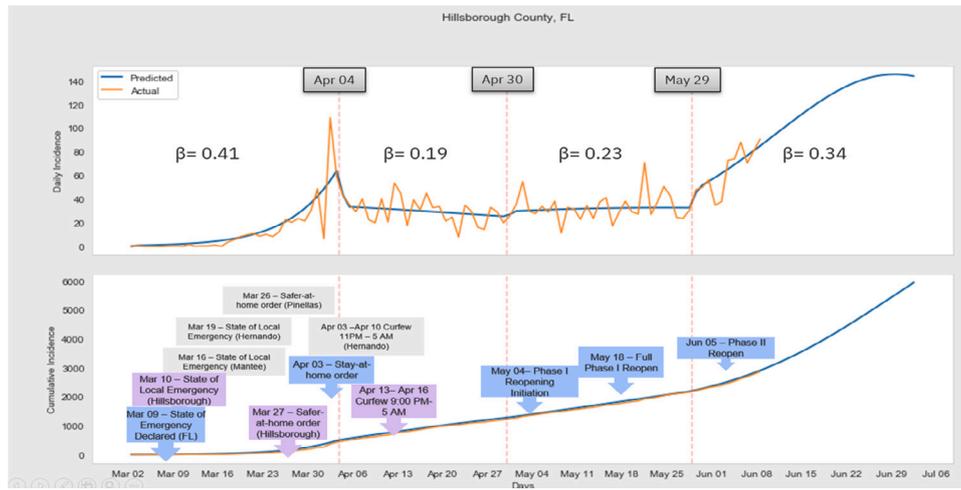
Fig. 2. Example of auto-detection of the changes in the transmission rate using the Kneedle algorithm on the daily case curves, superimposed with local and neighboring NPIs.
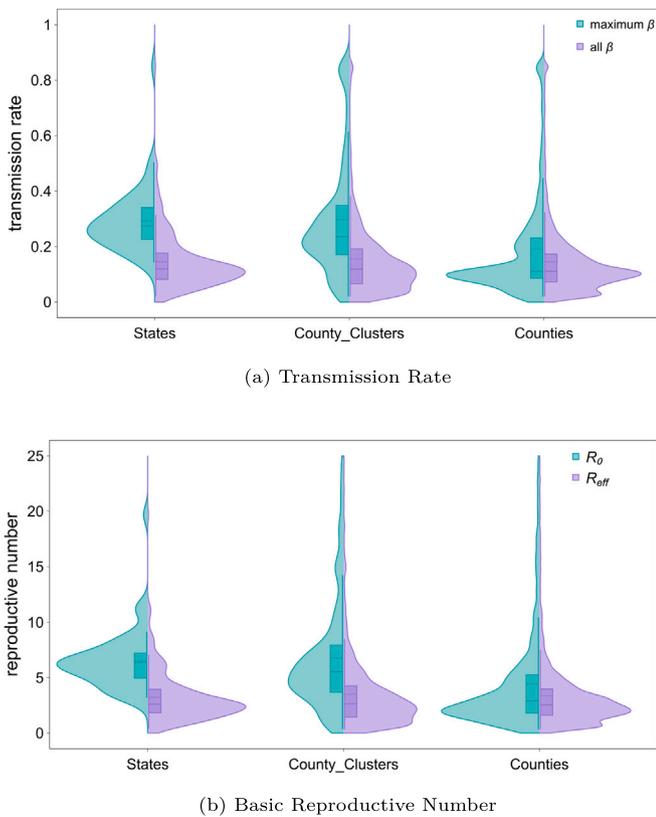

(a) Transmission Rate

(b) Basic Reproductive Number

Fig. 3. Regional distributions of transmission rates and reproduction numbers at different modeling resolutions. The violin plots are split to visualize both the distribution of maximal regional values and all regional values to show how non-pharmaceutical interventions may have lowered transmission rates and reproductive numbers (see text). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

cases and deaths, and the corresponding forecasts. In each sub-figure the white background indicates the reference time period and data used for model optimization. The gray background shows actual and forecast confirmed cases and deaths for subsequent days. This latter region was used to compute the Mean Absolute Percentage Error (MAPE). The state level curve representing confirmed state level reports is shown in both the confirmed and forecast charts for comparison (black dotted line). Note that daily reported cases at the state level may not always match the sum of daily reported cases at the county level. This is a data reporting issue.

The data for California and Washington State (Figs. 4(a) and 4(c)) illustrate how the noise in daily confirmed cases can adversely impact forecasts made by an automated pipeline. This is evident as the model forecasts at the state level (blue curve) deviate from the ground truth data after a particularly large noise spike. Forecasts at individual and state aggregate county clusters do not show the same deviation. The stack charts also reveal that the relative contribution from different sub-regions to state level confirmed cases or deaths varies over time indicating the states themselves are not accurately described as well mixed.

Forecast accuracy as a function of spatial scale is expected to be model dependent. Therefore, before evaluating the effect of modeling resolution on statistical error, we sought to test the forecast accuracy of the SACIR model relative to other models at the same resolution to establish its baseline performance. For this purpose we selected models from the CDC challenge project (Anon, 2020a). The majority of these models provide forecasts at the state level. The pipeline forecasts daily and cumulative confirmed cases and cumulative deaths for all states at least four weeks into the future. Fig. 5(a), shows a self-consistent comparison of forecast accuracy, at state level resolution, for all models that forecast both confirmed cases and deaths at state level, for all states, and with data provided 1–4 weeks into the future. As discussed in the Section 2 section, some models were omitted from the comparison based on missing data.

Figs. 5(a) and 5(b) compare the statistical accuracy of several models, averaged over all states, at one to four weeks beyond the time of optimization. The order of each bar is based on the MAPE (low to high) after the first week. For each model the average MAPE increases with time after the optimization window. The baseline forecast performance of the SACIR model, when run at a state level resolution, compares favorably with the other CDC Challenge models run at the same resolution for the same ground truth data and time.

3.3. Effects of spatial scale

Having established the validity of the model with respect to other state level baseline models, Fig. 5(a), we next compare the MAPE for forecasts as a function of spatial scale. Figs. 6(a) and 6(c) test the assumption that spatial regions defined by the states represent well mixed populations. The figures show the statistical error in the forecasts of confirmed cases and deaths, respectively, as a function of time. The red curves correspond to the average MAPE for models run at the state level (i.e. treating each state as single well mixed
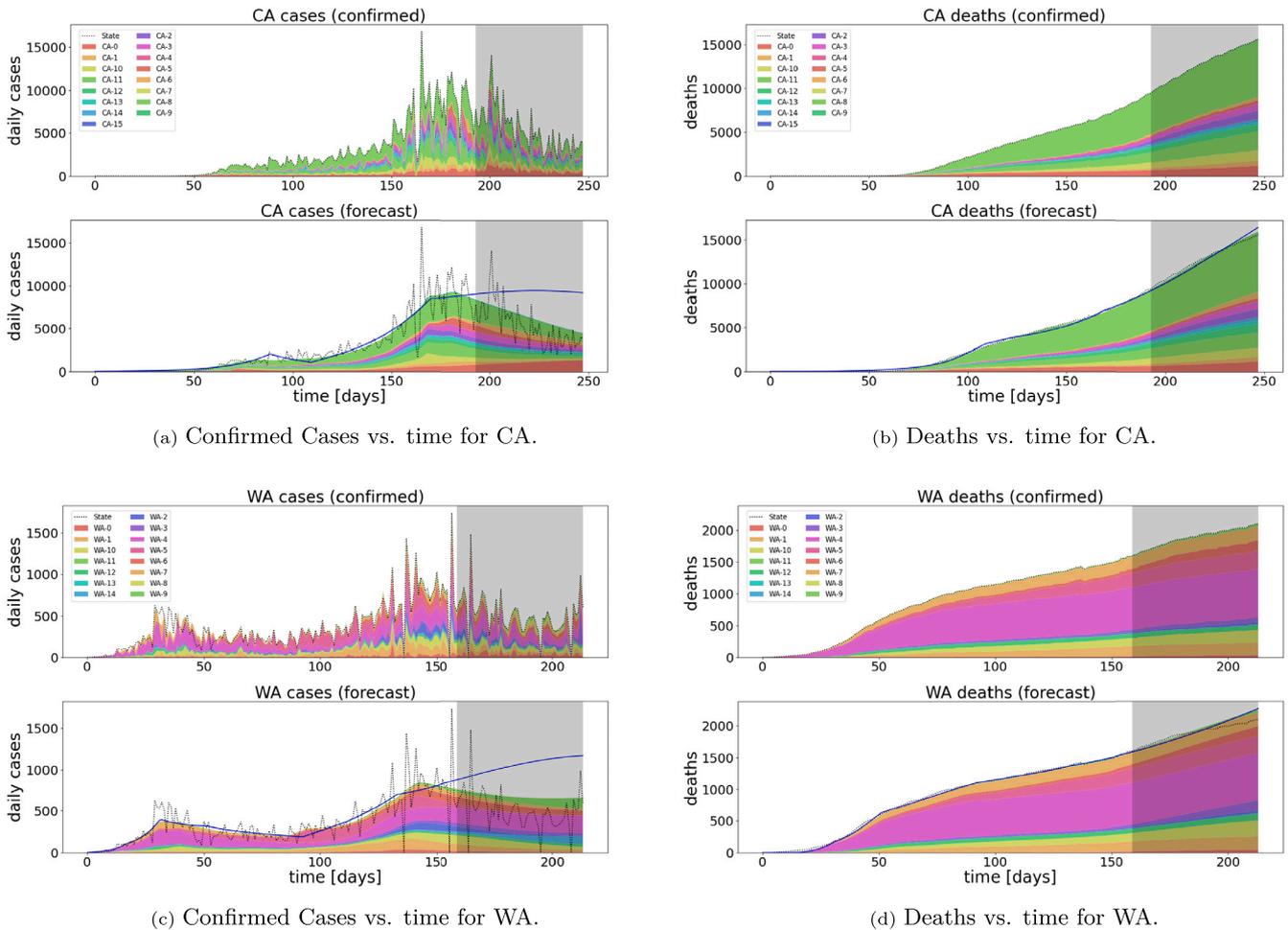
(a) Confirmed Cases vs. time for CA.

(b) Deaths vs. time for CA.

(c) Confirmed Cases vs. time for WA.

(d) Deaths vs. time for WA.

Fig. 4. Confirmed and forecast (daily) cases and (cumulative) deaths vs. time for county clusters within states. The white background shows the time period used for model optimization. The gray background shows confirmed cases and deaths for subsequent days. The relative contribution to state level confirmed cases or deaths from different sub-regions varies over time. The dotted state level curve shows confirmed state level reports in the forecast charts for comparison.
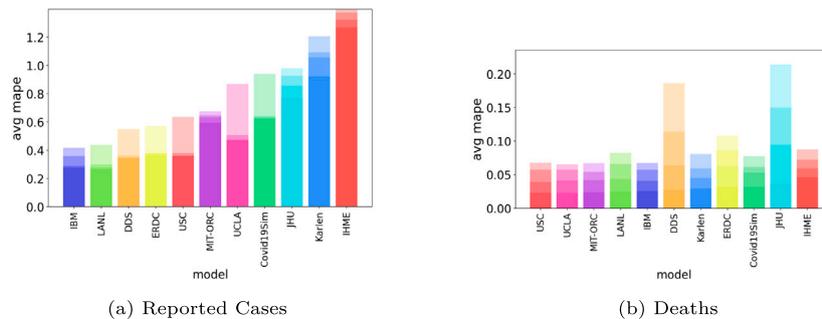

(a) Reported Cases

(b) Deaths

Fig. 5. MAPE by model after 1–4 weeks. Models are sorted by forecast accuracy after one week.

regions). The blue curves correspond to the average over states of the state level MAPE for models run at the county-cluster level. In this case the forecast for each state was obtained by summing all of the county cluster level forecasts within the state, and the state level MAPE computed by comparing the sum to the ground truth data for the state. The green curves were computed in the same way as the blue curves but running the model at county resolution. From Fig. 6(a) it is clear that modeling at either county or county-cluster resolution lowers the mean absolute percent forecast error by 30% for each of the four future weeks. County and county cluster resolutions yield comparable results. All three resolutions provide comparable forecast accuracy for cumulative deaths (Fig. 6(c)) at least in the first two weeks. State level

forecast accuracy of cumulative deaths degrades slightly faster after two week.

Figs. 6(b) and 6(d) test the complementary hypothesis. If larger regions are truly well mixed, then forecasts at higher resolution could be obtained by simply dividing or distributing forecasts made for larger regions by the fraction of population in each sub-region. Figs. 6(b) and 6(d) show the average MAPE for three methods to forecast confirmed cases or deaths at the county level. In each of the two figures, the red curve shows the average MAPE for county level forecasts obtained by dividing the state level model output to obtain county level forecasts. The blue curve represents county level forecasts obtained by dividing county cluster level model output. The green curve is simply the
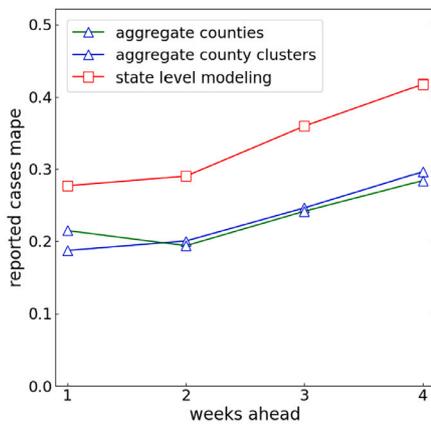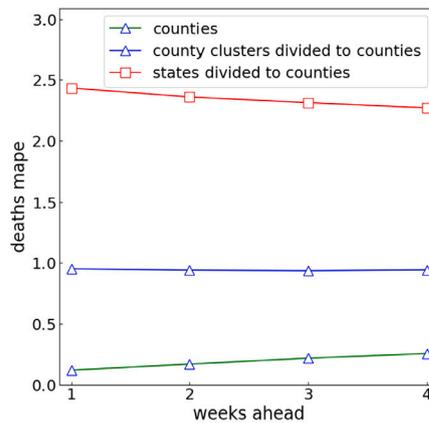
(a) Aggregating county cluster confirmed case forecasts provides more accurate results at the state level.
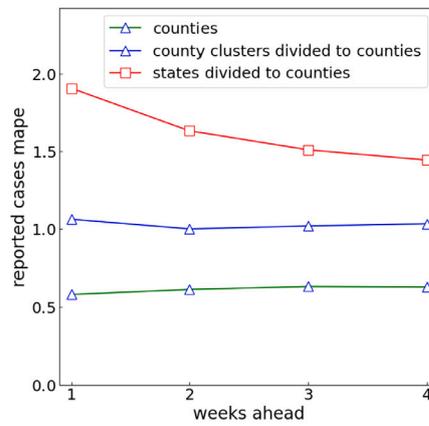
(b) Dividing state level confirmed case forecasts into smaller regions (by population) leads to much higher forecast error
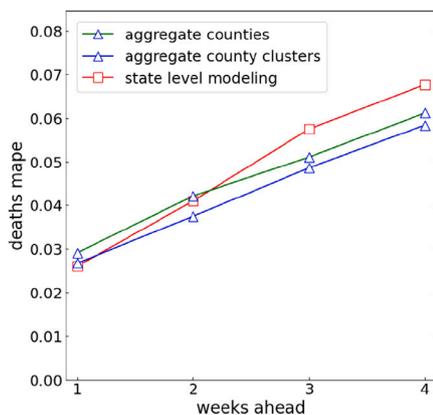
(c) Aggregating county cluster death forecasts provides more accurate results at the state level.

(d) Dividing state level death forecasts into smaller regions (by population) leads to much higher forecast error

Fig. 6. Comparing aggregation of smaller counties or county clusters provides more accurate state level forecast of deaths and confirmed cases. Conversely, subdividing state level model results into counties or county clusters fails, demonstrating that most states are not approximated as well mixed regions. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

average MAPE from modeling at the county level itself (not aggregated by state as in Figs. 6(a) and 6(c)).

From the data in Figs. 6(b) and 6(d), we list in Table 1 the ratio of MAPE in forecasts of county confirmed cases or death based on state or county-cluster regions relative to modeling at the county level. The error ratios decrease with time (forecast week) as the MAPE increases for all resolutions over time. Looking at forecasts one week into the future, county level forecasts based on state level modeling have ∼3.3x higher error for forecast confirmed cases and over 20x higher higher error for forecast deaths. Modeling counties at the county-cluster level, compared to county level, results in 1.8x higher error for forecast confirmed cases and over 8x higher higher error for forecast deaths.

In terms of running time, we noticed that it takes longer for the algorithm to fit as we climb down the geo-spatial level ladder, i.e., the time taken for fitting at the county-level is greater than the county-cluster and the states. We believe this is due to the frequent changes in the case counts and data sparsity at lower geo-spatial levels as opposed to more gradual changes at the higher levels like states.

To demonstrate the practical implications of our analysis, we report the forecast hospital admissions, and ICU demand for the Tampa General Hospital (TGH) network. Forecasts of confirmed case data for Hillsborough County were performed at least weekly beginning in April 2020. Fig. 7 shows actual and forecast COVID-19 ICU patients (magenta and blue solid curves respectively). The forecasts were made weekly

Table 1
Table increasing MAPE (ratio) as spatial resolution is decreases from county to state (see Figs. 6(b) and 6(d)).

Region resolution	Week 1	Week 2	Week 3	Week 4
Incidence				
State/county	3.3	2.7	2.4	2.3
County-clusters/county	1.8	1.6	1.6	1.6
Deaths				
State/county	20.6	14.0	10.6	8.9
County-clusters/county	8.1	5.6	4.3	3.7

and given the hospital use-case, only the first week of the forecasts were used. The blue points indicate independent one-week future forecasts. All sequential forecasts are shown (solid blue) and the actual data are shown by the magenta curves. Total ICU patients, actual and forecast, are indicated by magenta and blue dashed curves respectively. Total ICU forecast in Fig. 7 was made by adding the forecast COVID-19 demand to a seven day moving average of non-COVID actual cases. The blue zone above and below the total ICU forecast represents the MAPE. The MAPE over the entire time range was 5.9%. The time-varying ICU bed capacity for the county is shown in red. Forecast and actual total ICU demand reached a maximum in June (before the peak COVID-19
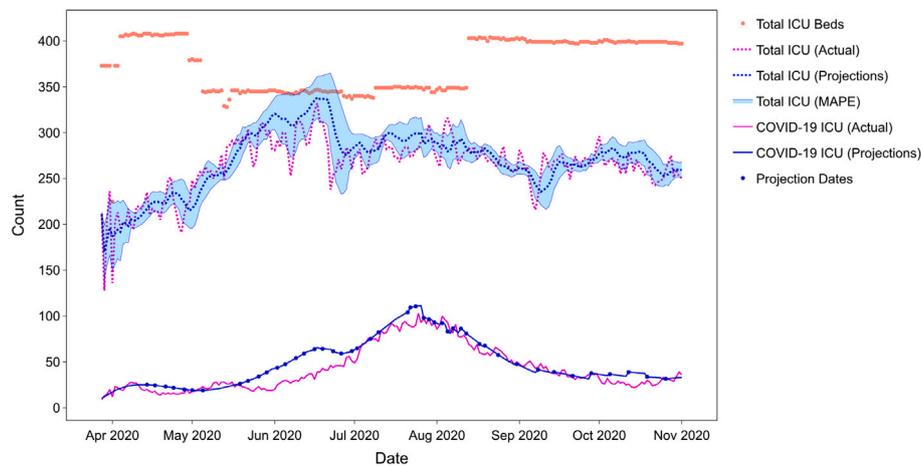
Fig. 7. ICU Status for Hillsborough County, FL. The red points indicated capacity (which varied over time). The figure shows COVID-19 ICU Patients, all ICU Patients, and the full series of COVID-19 ICU projections made between April and November 2020. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

demand) and never exceeded the ICU bed capacity of ∼350 beds at the time. Based on the total ICU demand projections, the hospital made the decision not to cancel other elective procedures and surgeries in the time period shown. Forecasts for county confirmed cases, total hospital admissions, and statistical error (MAPE) for individual projections are available in the supplement. Although the comparison to the other CDC models in Fig. 5 show relative performance for only one four week period, the data in Fig. 7 and in the supplement demonstrate that performance of the model is consistent and repeatable week-on-week for over 30 weeks.

4. Discussion

The data in Figs. 5(a) and 5(b) provides a quantitative comparison of the performance of several different models, including the current SACIR model, based on the MAPE calculated for one to four weeks into the future. Many of these models make different assumptions and consequently use markedly different approaches. Some are compartmental epidemiological models, some are based on regression, and some adopt a hybrid approach. The goal of this paper is not to identify a best model or paradigm, but to explore the question of modeling resolution itself.

Fig. 6 compares the predictive power of models at several spatial resolutions. Modeling at a spatial scale corresponding to administrative divisions of states may have advantages in terms of reduced computational complexity and cost. While public health reports by county, modeling at higher geo-spatial level allows aggregation of inherently noisy data which contributes to computational complexity and cost. However, if public health officials and policy makers require accurate insights at county level, modeling at the state level may not suffice.

Fig. 4 and associated data in the supplement demonstrates – from raw data alone – that different sub-regions within states make dominant contributions to aggregate state level confirmed cases and deaths at different times. Fig. 6 and Table 1 show that the statistical error (MAPE) in the forecast of confirmed cases at the state level is 30% higher for modeling states compared to modeling at the level or counties or clusters of counties. Furthermore, testing the well mixed population assumption, the same figures reveal that county level predictions made by partitioning state level predictions (for a model performing well at the state level) are 3x worse for forecasting confirmed cases and 20x worse for forecasting deaths relative to the same model at the county level. Figs. 6(a) and 6(c) do indicate the possibility of diminishing returns as average MAPE in the forecast of confirmed cases and deaths has similar (or even slightly lower) error at county-cluster resolution than at county resolution. Aggregating data to county clusters does provide noise reduction. Some of this noise is associated with artifacts

of public health case reporting where confirmed cases may be reported based on clinical location as opposed to patient residence. In other cases reporting location may be skewed by local public health practices. This is evidenced by public health data made available early in the pandemic where reported case counts and deaths were zero in Bronx, NY, while the epidemic grew exponentially in all adjacent boroughs (Dong et al., 2020). Local reporting bias suggests that the optimal trade-off between noise reduction and spatial resolution may vary locally and even change over time as public health systems become more efficient over the course of an extended pandemic. Whatever decision one makes in optimizing modeling performance, the measure of success must be based on the accuracy achieved at the spatial resolution required by the consumers of the data. In the case of TGH Hillsborough, modeling at the county level the statistical error was < 6% (Fig. 7). Had the county prediction been based on state level modeling, the uncertainty would have increased by a factor of ∼3.3x to almost 20% (see Table 1), greatly reducing confidence in the forecast. It is worthwhile to note that while many of the NPIs like mask mandates and social distancing measures are generally decided at a state-level, individual counties are assigned different tiers (each of which is assigned a refined set of NPIs) to keep hospital and ICU demands at a safe level. However, while this addresses to some extent the issue of drill down from state to county, the issue of roll-up where the aggregations of counties yield a better estimate than the state itself is something that would be of interest to policy makers and epidemiologist.

One particular model parameter, the immunity loss rate (defined as ρ in Equations (2) in supplemental) deserves additional discussion. Not only does immunity loss rate in any ODE based model of infectious disease describe the waning of host antibody response but is also affected by evolution of the virus itself. If the protein epitopes of the virus mutate rapidly, the adaptive immune response of recovered patients may decrease independent of antibody counts. Literature values for the period of reinfection (inverse of immunity loss rate) for Corona Viridae has been reported in the range of 8 months to 2 years. In this paper we adopted a value of 10 months. Early in a pandemic the modeling is not particularly sensitive to the precise value of this parameter. Since the time the work reported here was completed, successive epidemic waves have occurred including the current wave based on the delta variant. Models for current data continue to perform well using this 10 month period of reinfection. One should expect that the epidemiological parameters, including immunity loss rate, may vary (stochastically) as the virus itself changes. For example estimates of R_0 for the delta variant indicate the transmission rate has risen to values similar to chicken pox (a 3–4x increase). This is evident in the recent (summer 2021) case reports in several southern states.

With relative accuracy from the model performance, in the first and the second waves, TGH was able to consume the COVID forecasting model in two primary ways. In mid-June, as COVID-19 cases began a rapid rise in Hillsborough County, TGH leadership faced a difficult decision on whether or not to begin canceling operations to make capacity for the surging COVID-19 cases. As demonstrated in Fig. 7, the model accuracy improved in June between actual vs forecast, and forecast a peak in confirmed cases on 7/20/2020 and peak ICU on 7/26/2020. TGH used these forecasts to chart a hospital capacity forecast and to inform leadership that TGH would have enough capacity to handle the surge. A contingency plan to address potential capacity shortfalls caused by the surge of COVID patients was drawn up and involved constructing an emergency overflow facility with 40 beds staffed with a ratio of 1 team member per bed. Based on the information the models presented, TGH leadership decided to continue surgical procedures and to not proceed with constructing the surge facility: allowing for continuity in patient care and financial stability of the health system. With the data monitored 2–3x per week, TGH strategically allocated resources, supplies, and staff to accommodate for the surge. For example, increases in lab capacity for pre-surgery PCR testing of patients and supplies of PPE and testers were reliably estimated using the model. The model was directionally reliable enough to generate and execute this strategic COVID capacity plan.

While Tampa General Hospital was able to meet the ICU demand in the first and second waves, the latest delta variant has placed enormous stress on ICU's throughout the state. At the time of this paper, during the third wave, TGH was able to dynamically adjust ICU capacity from a low of 125 to a high of 188. With this flexibility, the hospital was at capacity but did not exceed capacity for COVID-19 patients.

The outcome of this work demonstrates that epidemiological forecasting, when performed at the county level can have a significant impact on the strategic planning of a COVID response for a county and a hospital.

CRediT authorship contribution statement

Vishrawas Gopalakrishnan: Created the model, Created the automation pipeline, Analyzed the USAFacts COVID data, Performed the COVID model analysis, Wrote the manuscript. **Sayali Pethe:** Created the model, Created the automation pipeline, Analyzed the USAFacts COVID data, Performed the COVID model analysis, Wrote the manuscript. **Sarah Kefayati:** Performed the COVID model analysis, Analyzed the ICU data, Wrote the manuscript. **Raman Srinivasan:** Analyzed the USAFacts COVID data, Wrote the manuscript. **Paul Hake:** Analyzed the USAFacts COVID data, Analyzed the ICU data, Performed the ICU forecast, Wrote the manuscript. **Ajay Deshpande:** Conceived the work, Analyzed the USAFacts COVID data, Wrote the manuscript. **Xuan Liu:** Wrote the manuscript. **Etter Hoang:** Analyzed the ICU data, Performed the ICU forecast, Wrote the manuscript. **Marbelly Davila:** Analyzed the ICU data, Performed the ICU forecast, Wrote the manuscript. **Simone Bianco:** Analyzed the USAFacts COVID data, Performed the COVID model analysis, Wrote the manuscript. **James H. Kaufman:** Conceived the work, Created the model, Analyzed the USAFacts COVID data, Performed the COVID model analysis, Wrote the manuscript.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.epidem.2021.100510.

References

Adhikari, S., Pantaleo, N.P., Feldman, J.M., Ogedegbe, O., Thorpe, L., Troxel, A.B., 2020. Assessment of community-level disparities in coronavirus disease 2019 (COVID-19) infections and deaths in large US metropolitan areas. JAMA Netw. Open 3 (7), e2016938.

Alamoudi, E., Mehmood, R., Aljudaibi, W., Albeshri, A., Hasan, S.H., 2020. Open source and open data licenses in the smart infrastructure Era: Review and license selection frameworks. In: Smart Infrastructure and Applications. Springer, pp. 537–559.

Anon, 2006–2020a. The eclipse SpatioTemporal epidemiological modeler (STEM) website. http://www.eclipse.org/stem/.

Anon, 2006–2020b. Eclipse STEM wiki (full documentation). https://wiki.eclipse.org/STEM.

Anon, 2020a. CDC activities and initiatives supporting the COVID-19 response and the President's Plan for opening America up again. https://www.cdc.gov/coronavirus/2019-ncov/downloads/php/CDC-Activities-Initiatives-for-COVID-19-Response.pdf. (Accessed 4 October 2020).

Anon, 2020b. US coronavirus cases and deaths. https://usafacts.org/visualizations/coronavirus-covid-19-spread-map. (Accessed 4 October 2020).

Bergman, A., Sella, Y., Agre, P., Casadevall, A., 2020. Oscillations in U.S. COVID-19 incidence and mortality data reflect diagnostic and reporting factors. MSystems 5 (4), 1–6. http://dx.doi.org/10.1128/msystems.00544-20.

Chin, V., Samia, N.I., Marchant, R., Rosen, O., Ioannidis, J.P., Tanner, M.A., Cripps, S., 2020. A case study in model failure? COVID-19 daily deaths and ICU bed utilisation predictions in New York state. Eur. J. Epidemiol. 35 (8), 733–742. http://dx.doi.org/10.1007/s10654-020-00669-6, arXiv:2006.15997.

Dong, E., Du, H., Gardner, L., 2020. An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect. Dis. 20 (5), 533–534.

Douglas, J.V., Bianco, S., Edlund, S., Engelhardt, T., Filter, M., Günther, T., Hu, K., Nixon, E.J., Sevilla, N.L., Swaid, A., et al., 2019. STEM: An open source tool for disease modeling. Health Secur. 17 (4), 291–306.

Edlund, S.B., Davis, M.A., Kaufman, J.H., 2010. The spatiotemporal epidemiological modeler. In: Proceedings of the 1st ACM International Health Informatics Symposium, pp. 817–820.

Gopalakrishnan, V., Navalekar, S., Ding, P., Hooley, R., Miller, J., Srinivasan, R., Deshpande, A., Liu, X., Bianco, S., Kaufman, J.H., 2021. Adaptive epidemic forecasting and community risk evaluation of COVID-19. arXiv preprint arXiv:2106.02094.

Gurbaxani, B.M., Schwartz, I.B., et al., 2020. SARS-CoV-2 open source models. https://wiki.eclipse.org/STEM.

Haffajee, R.L., Mello, M.M., 2020. Thinking globally, acting locally — The U.S. response to Covid-19. N. Engl. J. Med. 382 (22), e75. http://dx.doi.org/10.1056/NEJMp2006740, nejm.org URL: http://www.nejm.org/doi/10.1056/NEJMp2006740.

Hsiang, S., Allen, D., Annan-Phan, S., Bell, K., Bolliger, I., Chong, T., Druckenmiller, H., Huang, L.Y., Hultgren, A., Krasovich, E., et al., 2020. The effect of large-scale anti-contagion policies on the COVID-19 pandemic. Nature 584 (7820), 262–267.

Lasry, A., Kidder, D., Hast, M., Poovey, J., Sunshine, G., Zviedrite, N., Ahmed, F., Ethier, K.A., 2020. Timing of community mitigation and changes in reported COVID-19 and community mobility–four US metropolitan areas, February 26–April 1, 2020. CDC Stacks.

Lauer, S.A., Grantz, K.H., Bi, Q., Jones, F.K., Zheng, Q., Meredith, H.R., Azman, A.S., Reich, N.G., Lessler, J., 2020. The incubation period of coronavirus disease 2019 (CoVid-19) from publicly reported confirmed cases: Estimation and application. Ann. Intern. Med. 172 (9), 577–582. http://dx.doi.org/10.7326/M20-0504.

Liu, Y., Yan, L.M., Wan, L., Xiang, T.X., Le, A., Liu, J.M., Peiris, M., Poon, L.L., Zhang, W., 2020. Viral dynamics in mild and severe cases of COVID-19. Lancet Infect. Dis.

Ray, E.L., Wattanachit, N., Niemi, J., Kanji, A.H., House, K., Cramer, E.Y., Bracher, J., Zheng, A., Yamana, T.K., Xiong, X., Woody, S., Wang, Y., Wang, L., Walraven, R.L., Tomar, V., Sherratt, K., Sheldon, D., Reiner, R.C., Prakash, B.A., Osthus, D., Li, M.L., Lee, E.C., Koyluoglu, U., Keskinocak, P., Gu, Y., Gu, Q., George, G.E., España, G., Corsetti, S., Chhatwal, J., Cavany, S., Biegel, H., Ben-Nun, M., Walker, J., Slayton, R., Lopez, V., Biggerstaff, M., Johansson, M.A., Reich, N.G., 2020. Ensemble forecasts of coronavirus disease 2019 (COVID-19) in the U.S. http://dx.doi.org/10.1101/2020.08.19.20177493, https://www.medrxiv.org/content/early/2020/08/22/2020.08.19.20177493.

Satopaa, V., Albrecht, J., Irwin, D., Raghavan, B., 2011. Finding a "kneedle" in a haystack: Detecting knee points in system behavior. In: 2011 31st International Conference on Distributed Computing Systems Workshops. IEEE, pp. 166–171.

Schwartz, I.B., Kaufman, J.H., Hu, K., Bianco, S., 2020. Predicting the impact of asymptomatic transmission, non-pharmaceutical intervention and testing on the spread of COVID19 COVID19. medRxiv.

Tolles, J., Luong, T., 2020. Modeling epidemics with compartmental models. JAMA 323 (24), 2515–2516. http://dx.doi.org/10.1001/jama.2020.8420, URL: https://jamanetwork.com/journals/jama/fullarticle/2766672.

Voss, J., 2009. Encoding changing country codes for the Semantic Web with ISO 3166 and SKOS. In: Metadata and Semantics. Springer, pp. 211–221.

Wang, W., Xu, Y., Gao, R., Lu, R., Han, K., Wu, G., Tan, W., 2020. Detection of SARS-CoV-2 in different types of clinical specimens. JAMA 323 (18), 1843–1844.

Wolfel, R., Corman, V., Guggemos, W., et al., 2020. Virological assessment of hospitalized cases of coronavirus disease. Nature 581, 465–469.

Woolf, S.H., Chapman, D.A., Sabo, R.T., Weinberger, D.M., Hill, L., Taylor, D.D., 2020. Excess deaths from COVID-19 and other causes, March–July 2020. JAMA 324 (15), 1562–1564. http://dx.doi.org/10.1001/jama.2020.19545.

World Health Organization, et al., 2020. Modes of Transmission of Virus Causing COVID-19: Implications for IPC Precaution Recommendations: Scientific Brief, 27 March 2020. Tech. Rep., World Health Organization.