A semi-parametric, state-space compartmental model with time-dependent parameters for forecasting COVID-19 cases, hospitalizations, and deaths

Eamon B. O'Dea[1] and John M. Drake[1]

1 Odum School of Ecology and Center for the Ecology of Infectious Diseases, University of Georgia, Athens, Georgia, USA

* odea35@gmail.com

Abstract

Short-term forecasts of the dynamics of COVID-19 in the period up to its decline following mass vaccination was a task that received much attention but proved difficult to do with high accuracy. A major obstacle has been capturing variations in the underlying kinetics of transmission resulting from changes in public policy, individual behaviors, and evolution of the virus. However, the availability of standardized forecasts and versioned data sets from this period allows for continued work in this area. Here we introduce the Gaussian Infection State Space with Time-dependence (GISST) forecasting model. We evaluate its performance in 1-4 week ahead forecasts of COVID-19 cases, hospital admissions, and deaths in the state of California made with official reports of COVID-19, Google's mobility reports, and vaccination data available each week from June 29, 2020 to April, 26, 2021. Evaluation of these forecasts with a weighted interval score shows them to consistently outperform a naive baseline forecast and often score closer to or better than a high-performing ensemble forecaster. The GISST model also provides parameter estimates for a compartmental model of COVID-19 dynamics, includes a regression submodel for the transmission rate, and allows for parameters to vary over time according to a random walk. GISST provides a novel, balanced combination of computational efficiency, model interpretability, and applicability to large multivariate data sets that may prove useful in improving the accuracy of infectious disease forecasts.

Author summary

The COVID-19 pandemic has been unprecedented both in the volume of related data made available and the volume of infectious disease forecasts generated. Improvements in forecasting methods can allow us to more fully identify regular patterns within the available data, which may provide further insight into the epidemiology of COVID-19. Improved forecasting algorithms would also be of value during any future outbreaks of similar pathogens, and to develop models useful for scenario planning. In this paper, we introduce a new forecasting model developed for forecasting cases, deaths, and hospital admissions at the state level in the United States. We use versioned data sets to apply this to the same forecasting tasks that models performed during the pandemic. We find that this new model performs well relative to a leading ensemble forecasting method. We also show that this model can be useful for understanding how COVID-19 cases, hospital admissions, and deaths are related to each other and human mobility and vaccine administration, and how these relationships changed over time.

Introduction

Questions about the future trajectory of the COVID-19 pandemic were central to the personal [1] and public policy [2] decisions of most of the world in 2020. In principle, predictive models can be key tools for decision support during infectious disease outbreaks, but their value rests largely on the trustworthiness of their ability to capture the key relationships in complex systems subject to changing rules and conditions [3]. One way to evaluate the trustworthiness of such a model is to measures its predictive performance over short-term time horizons [4].

One of the major initiatives to develop models with short-term forecasting value is the COVID-19 Forecast Hub [5], which collected forecasts of COVID-19 cases, hospitalizations, and deaths over the course of the pandemic in a standard format [6]. The COVID-19 Forecast Hub not only makes it easier to compare the forecasts of different models made during the pandemic, but also creates a set of benchmark forecasting tasks which models developed later can use to demonstrate their predictive value. Of course, making good forecasts after the data are available is not as impressive as doing so beforehand, but this exercise still has value for developing models for the next pandemic, or resurgence of COVID-19. A model that can provide good prospective forecasts should also be able to provide good retrospective forecasts.

The non-stationary nature of COVID-19 pandemic [7] presents a challenge for forecasting by undermining the invariance assumptions that are the basis of most forecasting methods [3]. Ref [7] draws a contrast between simple statistical models, which may be readily adapted to the most recent observations in a time series, and epidemiological models, which–although invaluable for providing insight into the past—are not as readily adapted to the most recent observations and thus more prone to systematic forecasting errors when the data are nonstationary.

Here, we present a hybrid model, Gaussian Infection State Space with Time-dependence (GISST), which aims to combine the interpretability [8] of epidemiological models with the robustness to non-stationarity of statistical models by allowing key parameters of the epidemiological model to be fitted to the most recent observations. This model was developed for forecasting COVID-19 cases, hospitalizations, and deaths at the state level in the United States. The GISST model is interpretable in the sense that it incorporates our understanding of the high-level relationships among cases, hospitalizations, and deaths and produces outputs such as the effective reproduction number and estimates of the number of currently infected individuals. Further it includes a regression model for the effects human mobility and vaccination levels on transmission and susceptibility, and it has the potential to be used to identify new correlates of transmission from a large panel of candidate variables. Fig 1 provides a graphical summary of GISST. We describe the model in detail in Materials and Methods. In Results, we present parameter estimates and evaluations of the model's 1-day ahead forecasts and 1-4 week ahead forecasts for the state of California. We find that the GISST model's forecast performance approaches and sometimes exceeds that of the COVID-19 Forecast Hub Ensemble in 1-4 week ahead forecasts. It does particularly well for the hospital admissions indicator, which may be the indicator whose forecasts are of the greatest practical value. The GISST model does this while having some advantages in computational efficiency and interpretability compared with the COVID-19 Forecast Hub Ensemble model.

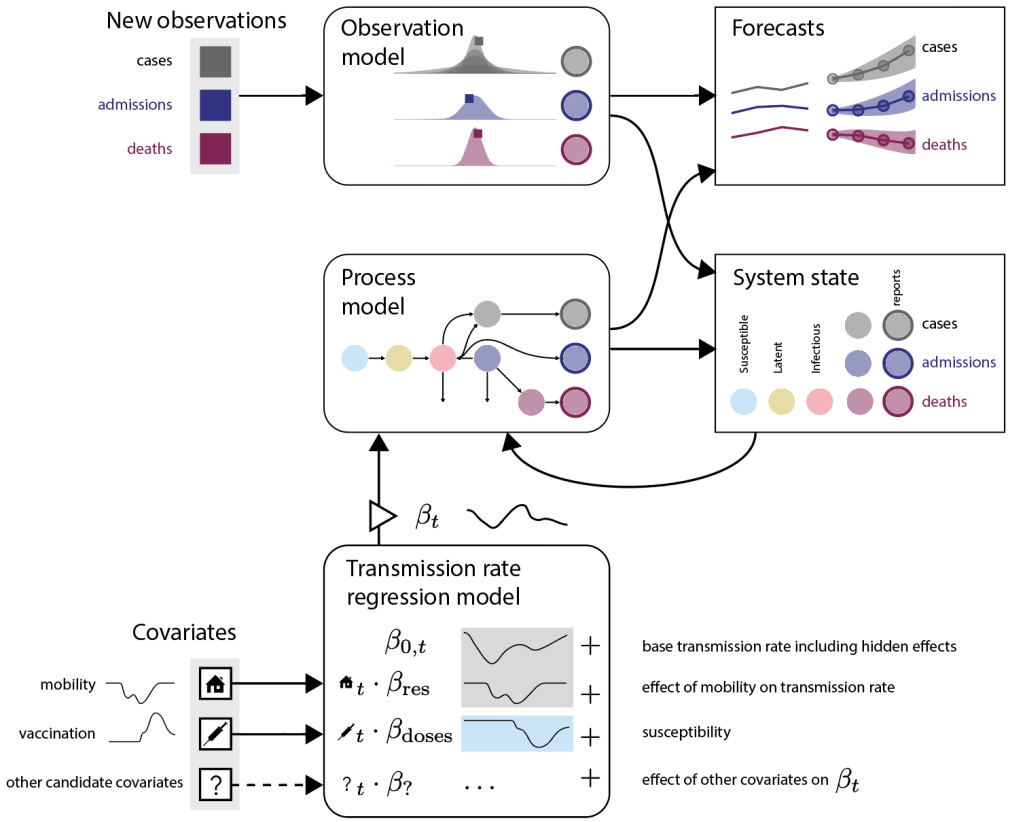

Fig 1. Graphical summary of our modeling and forecasting method, GISST. The main features are time-dependent parameters to accommodate the non-stationarity of COVID-19 dynamics, a regression model to incorporate data on covariates of transmission, and the integration of a compartmental epidemiological process model into a Gaussian state-space model for cases, hospital admissions, and deaths.

Materials and methods

Data

Forecasting targets

Many times series indicators of the status of the COVID-19 pandemic are available. To allow our model's forecasting skill to be compared with those of many other models, we use the forecasting targets established by the COVID-19 Forecast Hub [5,6]. These targets use data on cases and deaths distributed in the COVID-19 Data Repository `https://github.com/CSSEGISandData/COVID-19` created by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University [9].

They also make use of data on hospital admissions in the "COVID-19 Reported Patient Impact and Hospital Capacity by State Timeseries" and "COVID-19 Reported Patient Impact and Hospital Capacity by State" datasets provided by the U.S. Department of Health & Human Services (HHS) on `healthdata.gov`. We accessed these data through the mirror accessible through the Delphi Epidata API [10]. Specifically, this time series is the sum of the fields labeled `previous_day_admission_adult_covid_confirmed` and `previous_day_admission_pediatric_covid_confirmed` in the tables provided by HHS.

One challenging feature of these time series in regards to forecasting is that they are

subject to revision. To ensure a fair comparison between our model and those models that provided forecasts as the pandemic was progressing, for any forecast we used the version of the data available on the day for forecast submission to the forecasting hub. For evaluation of forecasts, we use a version of the data that is recent at the time of this writing.

The forecasting hub had a flexible submission format that allowed teams to submit for any subset of counties and states as well as the United States as a whole. We focus on targets at the state level in this work, in the largest state of California. One reason is that the county-level data are subject to more reporting anomalies. Another is that many important policy decisions are made at the state level. A third is that approximations used in our process model were better suited to modeling the dynamics of larger, state-level populations than those of smaller county populations.

The time series of cases, deaths, and hospitalizations are available at a daily resolution, but the targets are Sunday to Saturday weekly totals for cases and deaths. In keeping with the conventions of the COVID-19 Forecast Hub, we restrict our attention to 1-4 week ahead forecasts of cases and deaths and 1-28 day ahead forecasts of hospital admissions. The conventional wisdom is that forecasts at longer time horizons are generally not accurate enough to be useful [5]. Fig 2 shows the range of daily observation values of COVID-19 cases, hospital admissions, and deaths for California to illustrate the general trajectory of the time series used to fit our model as well as the extent to which it was revised over time.

The hospital admissions time series is unique in that it begins later than the cases and deaths time series, in the middle of summer, and in that it did not become available until November of 2020. At the beginning of the hospital admissions time series, there is also a misleading trend of growth that is simply due to the number of hospitals providing data increasing. To remove this artifact, we consider as missing the observations in which the field `previous_day_admission_adult_covid_confirmed_coverage` was below 90% of its last observed value. We apply this transformation both for the data sets used for fitting and those used for evaluation.

Regression model covariates

The GISST model includes a regression which expresses the risk of infection to susceptibles in terms of likely predictors of immunity and exposure. One such predictor is an aggregated measurement of the amount of time individuals spend in residential areas, as quantified in Google's Community Mobility Reports [11]. This metric is expressed as a change from a baseline. A value of 0% indicates no change relative to the median value of the indicator from January 3 through February 6, 2020. The baseline is calculated separately for each day of the week, which allows it to account for weekly cycles in mobility. One consequence of this form is that the percent change from baseline is typically lower on weekends because people already spent a lot of time in residential areas on weekends before the pandemic during the period when the baseline was calculated. In our model of COVID-19 transmission, we do not attempt to replicate weekly cycles of exposure, and rather model the average exposure of individuals over the course of a week. Using the raw metric in the reports would result in our model incorrectly predicting reduced exposure during the work week. Therefore, to obtain a more reasonable covariate for exposure, we use a rolling weekly average of the percent change from baseline in the original mobility reports. Fig 3 shows the resulting time series.

The details of the calculation of our mobility covariate are as follows. For a given forecast date, we obtained the latest snapshot of Google's `Global_Mobility_Report.csv` file on `http://web.archive.org` that was made prior

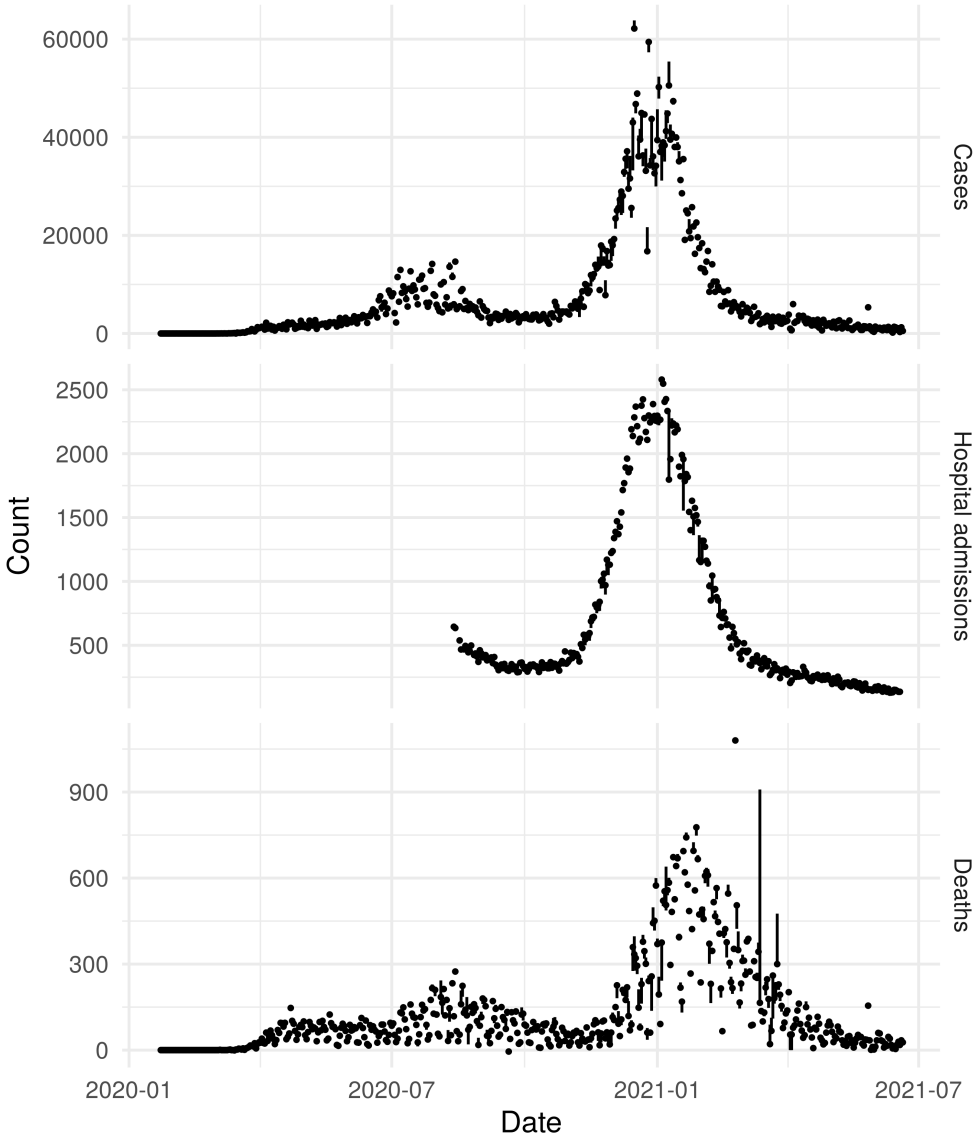

Fig 2. Time series of forecasting targets and range of revisions. Points are the California time series of COVID-19 indicators which our model is designed to forecast in the version of the data used to fit our model, which was the version available on June 21, 2021. The vertical line range gives the range of values that each observation had in all versions of the data used to fit our model, which were the versions of the data available on Mondays from June 29, 2020 to April 26, 2021.

to the beginning forecast date in the UTC time zone. This was the the data available to all forecasting teams 22-23 hours before the submission deadline of 6PM ET. Although these data were not subject to revision, there was a gap of a few days between the last mobility data point and the forecast date which we are able to reproduce by using the archived data sets. To remove weekly seasonality, we calculated a moving 7-day average with a centered moving window. The last 3 observations of this 7-day average time series are missing because data for the full window is not yet available, and we also considered as missing output from any windows which included the observed date of the

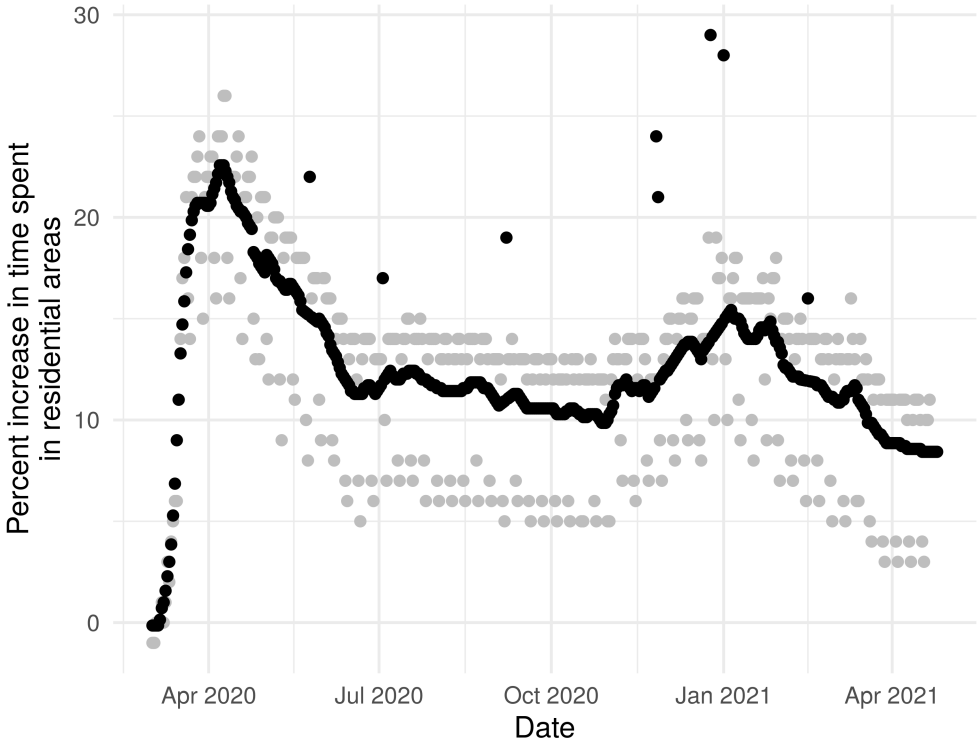

Fig 3. Indicators of the time spent at home. Grey points are the original data provided by Google of the amount of time people in California spent in residential areas relative to a pre-COVID-19 period. Black points are an example time series of the derived version which we use as a predictor of exposure in our model. Holiday effects in the original data are retained but weekly seasonality is removed.

holidays of President's Day, Memorial Day, Independence Day, Labor Day, Thanksgiving Thursday and Friday, Christmas, and New Year's Day. Missing values within the resulting 7-day average time series were imputed by linear approximation. The gap at the end of the time series was filled in by repeating the last non-missing value. Finally, values of the unsmoothed data on holidays were used to replace corresponding values in the smoothed time series. In this way, holidays do not affect the extrapolation of values in the gap between the end of the data and the forecast date, but they are still able to affect exposure in our model. Note, however, that we are only accounting for the reduction in exposure due to individuals being able to spend more time at home rather than any increased exposure due to social activity.

A second predictor of the risk of infection is the number of vaccine doses administered in a state. The modeling assumption is that average susceptibility declines with the number of doses administered. We model an effect on average susceptibility rather than number of susceptibles because a single dose of a COVID-19 vaccine does not generally fully immunize an individual. We sourced our time series of the number of doses administered from the Git repository at `www.github.com/govex/COVID-19`, which was created by Johns Hopkins University. These data came from either the states' public dashboards or the CDC Vaccine Tracker. Since the reported numbers are cumulative, the creators of this data set used the larger of the values from those 2 data sources each day as the most up-to-date value. To obtain a covariate free of missing values, we assumed all values before the first reported value were zero and that all values after the last reported value were equal to the last reported value. Missing values

within the time series were imputed with a linear approximation. Fig 4 shows the resulting time series.

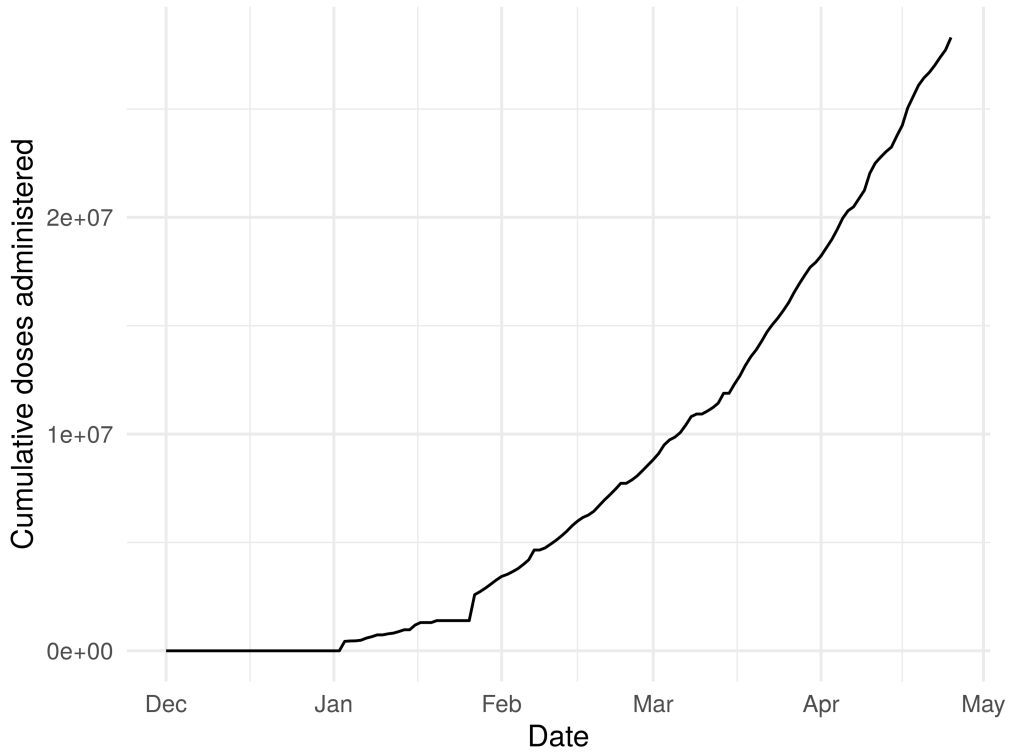

Fig 4. COVID-19 vaccination doses administered in California over time. This time series was used to estimate the effect of vaccination on reducing average susceptibility.

Model

The GISST model is a state space model, comprising a process model and an observation model. The process model and observation model both include stochasticity using Guassian random variables. For the process model, these Gaussian variables arise naturally by applying the system size expansion. We find the Gaussian assumptions justified by diagnostics of the fitted model, and it allows for the model's likelihood to be efficiently calculated using the extended Kalman filter [12]. This efficiency makes it feasible to estimate the effects of covariates on the process model's transmission rate in an embedded regression model, as well as allowing for parameters to vary over time according to a random walk to account for non-stationarity.

Process model

Our process models arises as the large-population limit of a continuous-time Markov chain defined for the following state variables:

X Uninfected and *susceptible* individuals. Susceptible individuals can become infected by individuals in the Y compartment.

L Individuals with *latent* infections who are not yet infectious. At the end of the L stage, these individual progress to the Y stage.

Y Individuals who are *infectious*. This compartment includes both symptomatic and asymptomatic individuals.

Z Individuals who have been diagnosed and will be reported as *cases*, but have not yet been reported. Those individuals are likely isolated and so are not infectious in our model. This compartment includes a count for the fraction of infectious individuals who recover without hospitalization, and it includes all infectious individuals who enter the hospitalized class.

Z_r This compartment keeps track of the number of cases reported each day, which we can compare with official reports.

H Individuals who have been *hospitalized*. Those individuals are modeled as not infectious due to isolation. A fraction of individuals in the H stage will recover, the remainder will pass.

A Individuals who are new hospital *admissions*. This compartment keeps track of the number of new hospitalized individuals each day, which we can compare with reported hospital admissions.

D Individuals who have *died* from the infection, but whose death has not yet been reported.

D_r The number of newly reported deaths each day.

The transition probabilities of the Markov chain are contained in Table 1. The definitions and values of model parameters appear in Table 2.

Table 1. Transition rules for the Markov chain.

Event	**Update**	**Propensity**
Transmission	$X \to L$	$\beta_t XY/N$
Incubation	$L \to Y$	ηL
Removal of infectious individual whose cases will be reported	$Y \to Z$	$(1-p_{h,t})\rho_t \gamma Y$
Removal of infectious individual whose cases will *not* be reported	$Y \to 0$	$(1-p_{h,t})(1-\rho_t)\gamma Y$
Reporting of cases	$Z \to Z_r$	$\gamma_{z,w} Z$
Hospitalization of an infection on a weekday	$Y \to H + A + Z$	$p_{h,t}\gamma Y$
Hospitalization of an infection on Saturday or Sunday	$Y \to H + A + Z$	$p_{h,t} p_{h,\text{weekend}} \gamma Y$
Death during hospitalization	$H \to D$	$p_d \gamma_h H$
Discharge from hospital	$H \to 0$	$\gamma_h (1-p_d) H$
Reporting of deaths	$D \to D_r$	$\gamma_{d,w} D$

Our model diverges from the typical SEIRD (Susceptible-Exposed-Infected-Removed-Deceased) structure for specific reasons. A state variable for the number of removed individuals—that is, individuals who completed a pass through the infectious state—is missing because it would be redundant. That quantity can easily be calculated as $N - X - L - Y$, and the counts of deceased individuals, which are tracked by our model, subtracted from that if needed. The hospitalization reaction involves 4 state variables due to the nature of the data we observe. Hospitalizations are counted as both cases and hospital admissions. However, our observations of cases are reports which lag the diagnosis by several days whereas we have a time series of actual date of hospital admission. Further, some hospitalizations conclude in deaths, which will on average be reported with a longer lag from the date of hospital admission than will cases. Thus the potential to lead to observable events with 3 different lag times makes it necessary for the hospital admission event to propagate to 3 pathways in our model.

Table 2. Parameters of Markov chain model.

Symbol	Definition	Value
β_t	Transmission rate	Estimated
N	Population size	Value given in JHU CSSE data set
η	Incubation rate	365.25 / 4 / year
γ	Removal rate	365.25 / 4 / year
$\gamma_{d,w}$	Rate of reporting deaths on day of week w	365.25 / 10 / year *
$\gamma_{z,w}$	Rate of reporting cases on day of week w	365.25 / 1 / year *
γ_h	Rate of exit from hospital	365.25 / 10 / year
$p_{h,t}$	Probability of an infection leading to hospitalization	Estimated*
$p_{h,\text{weekend}}$	Factor relating weekday hospital admissions to weekend admissions	Estimated*
p_d	Probability of a hospitalization leading to death	Estimated
ρ_t	Probability of a removal on day t being reported as a case	0.5*

* The parameter $p_{h,t}$ is fixed to 0.03 when hospitalization data are not available to avoid colinearity with p_d. Additionally, $p_{h,\text{weekend}}$ is assumed equal to 0.9 when hospitalization data are not available. The parameter ρ_t is estimated for dates up to June 2020 as described in *Estimation of* ρ_t, while $\gamma_{d,w}$ and $\gamma_{z,w}$ are fixed at this value for some days of the week and estimated for others as described in *Day of week effects on reporting*.

We could fit this Markov chain model using a simulation-based approach such as particle filtering. However, when the pandemic is in full strength and most state variables are large, it seems more advantageous to employ a central limit theorem to obtain a Gaussian distribution for the random fluctuations of the state variables. This choice allows for faster likelihood calculation via the Kalman filter. Therefore, computational resources may be redirected from calculating the minute details of our model's stochastic fluctuations to other tasks important for developing good forecasts such as experimenting with variations of the model's structure.

Such a distribution can be obtained by applying the system size expansion [13], which yields a system of ordinary differential equations for the expected values of the state variables and another system for the covariances of random fluctuations around the mean. The system of equations for the approximate expected values of the state variables is

$$\begin{aligned}
\langle \dot{X} \rangle &= -\beta_t \langle X \rangle \langle Y \rangle / N \\
\langle \dot{L} \rangle &= \beta_t \langle X \rangle \langle Y \rangle / N - \eta L \\
\langle \dot{Y} \rangle &= \eta \langle L \rangle - \gamma \langle Y \rangle \\
\langle \dot{Z} \rangle &= [p_{h,t} + \rho_t (1 - p_{h,t})] \gamma \langle Y \rangle - \gamma_{z,w} \langle Z \rangle \\
\langle \dot{Z}_r \rangle &= \gamma_{z,w} \langle Z \rangle \\
\langle \dot{H} \rangle &= p_{h,t} \gamma \langle Y \rangle - \gamma_h \langle H \rangle \\
\langle \dot{A} \rangle &= p_{h,t} \gamma \langle Y \rangle \\
\langle \dot{D} \rangle &= p_d \gamma_h \langle H \rangle - \gamma_{d,w} \langle D \rangle \\
\langle \dot{D}_r \rangle &= \gamma_{d,w} \langle D \rangle,
\end{aligned} \tag{1}$$

where the angular brackets indicate expected value and the overdots indicate time derivatives. Fig 5 provides a graphical representation of these equations. For the covariance matrix, $\mathbf{P}$, of fluctuations around the means, the system size expansion [13] yields

$$\dot{\mathbf{P}} = \mathbf{J}\mathbf{P} + \mathbf{P}\mathbf{J}^T + \mathbf{Q}, \tag{2}$$

where $\mathbf{J}$ is the Jacobian matrix of the system in Eq 1 and $\mathbf{Q}$ is the matrix defined by

$$\mathbf{Q} = \mathbf{S}^T \operatorname{diag}(f)\mathbf{S}, \tag{3}$$

which may be calculated from individual-level reactions in Table 1. The matrix $\mathbf{S}$ has elements S_{ij}, which are the stochiometric coefficients of species j in reaction i in the reactions in Table 1. The vector f collects the propensities column of Table 1. The solutions of Eqs 1 and 2 are the parameters of a Gaussian approximation to the distribution of the state variables in the original Markov chain. Thus, given an initial value of the means and $\mathbf{P}$, numerical integration of Eqs 1 and 2 allows values of the mean and variances at a future time to be calculated. We describe how such values are used to generate forecasts and fit the model in the subsection *Fitting methods*. We next explain some additional features of our model that proved necessary to obtain a good fit.

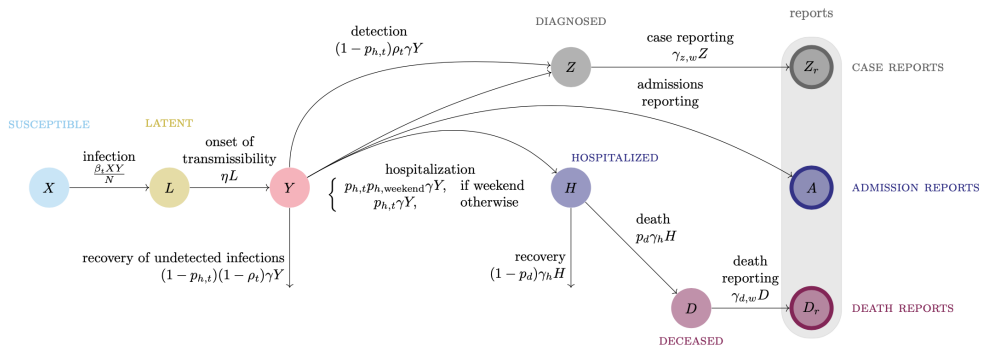

Fig 5. Flow diagram of differential equations for expected values of state variables

Day of week effects on reporting

Although we do not attempt to model variation in exposure due to the day of the week, the effect of the day of the week on reporting of cases and deaths cannot be ignored. We account for this by allowing the reporting rate of cases, $\gamma_{z,w}$, and the reporting rate of deaths, $\gamma_{d,w}$, to depend on the day of the week. Rather than introduce a parameter for each day of the week, we break the days of the week into 2 or 3 groups and keep parameter values the same within a group. For $\gamma_{z,w}$, there are two groups: (1) Saturday and Sunday and (2) Monday through Friday. For $\gamma_{d,w}$ there are three groups: (1) Sunday and Monday, (2) Tuesday and Wednesday, and (3) Thursday through Saturday. These groups were chosen to parsimoniously account for day of the week effects by inspecting 1 day ahead prediction errors on different days of the week. $\gamma_{z,w}$ is fixed to the value given in Table 2 on Monday through Friday and the weekend value is estimated. The parameter $\gamma_{d,w}$ is fixed to the value given in Table 2 on Thursday through Saturday and its values on the other two groups of days is estimated.

Hospital admissions were also affected by the day of the week, being noticeably lower on weekends most of the time. We model this depression by introducing the parameter $p_{h,\text{weekend}}$ with an allowed range of 0 to 1. As show in Table 1, the rate of hospital admission on weekends is calculated by multiplying the weekday rate by this reduction factor.

Regression model for the transmission rate

Non-pharmaceutical interventions (NPIs) and vaccination have been effective in reducing the growth rate of COVID-19 cases, and we employ a regression model to allow

for such effects in our model. Our covariates for quantifying these effects have been described in the *Data* subsection of *Materials and Methods*. We use $\texttt{doses}_t$ to denote the number of vaccine doses administered on day t and $\texttt{residential}_t$ to denote our indicator of the time spent in residential areas on day t. Our regression model for the transmission rate on day t, β_t, follows

$$\log \beta_t = \min(\beta_{0,t} + \texttt{doses}_t \cdot \beta_{\text{doses}} + \texttt{residential}_t \cdot \beta_{\text{res}}, \log(1000)). \tag{4}$$

The upper limit of 1000 on the value of β_t reflects our scientific understanding that the reproduction number is unlikely to exceed 10 even in the absence of control measures and it prevents numerical problems in accurately calculating a solution for our process model with very large transmission rates. Although for the sake of brevity we refer to β_t as a transmission rate, it serves to summarize effects on both transmissibility and susceptibility. That is $\beta_t =$ (transmission rate on day t) $\times$ (suseptibility on day t). The intended interpretation of our regression model in terms of this factorization of β_t is

$$\text{transmission rate on day } t = \exp(\beta_{0,t} + \texttt{residential}_t \cdot \beta_{\text{res}}) \tag{5}$$

$$\text{susceptibility on day } t = \exp(\texttt{doses}_t \cdot \beta_{\text{doses}}). \tag{6}$$

The coefficient $\beta_{0,t}$ allows for the effects of NPIs for which we lack covariates to affect the transmission rate, and also allows for the reproduction number to decline with the depletion of susceptibles at a rate consistent with heterogeneity in susceptibilty [14]. These effects are allowed to vary over time according to a random walk model, which we describe in the subsection *Time dependence of parameters*.

Observation model

Our process model accounts for variation around expected values of cases, hospital admissions, and deaths due to the intrinsic randomness of disease transmission. Another major source of variation comes from our imperfect observation of these variables. For the sake of simplicity and computational efficiency, we assume that these errors are zero-mean Gaussian random variables and are independent of each other. The variance of the observation error of hospital admissions, denoted τ_h, and that of deaths, denoted τ_d, are assumed to be constant over time. In contrast, allowing variation in the observation error of cases over time proved critical to obtaining a good fit. For this variable, we assume that variance of the observation error on day t is proportional to the model's estimate of the number of infected individuals at the beginning of day t. This allows for a mean-to-variance ratio similar to what would result from assuming each case was reported with a given probability. Additionally, we allow the proportionality constant, denoted $\tau_{c,t}$, to change over time according to a random walk model. This allowed observations to be given less weight in periods subject to large amounts of noise.

Time-dependence of parameters

The curve of cases in Fig 2, with its multiple peaks, looks nothing like the typical simulation of our process model with fixed parameters. Even after incorporating the covariates in our regression model for the transmission rate, we, like other modelers [15], find it necessary to allow the parameters of our model to change over time to obtain a good fit. Further, allowing variation in parameters is justifiable given the rapidly evolving epidemiology of COVID-19, in which human behavior played a large role. We allow some parameters in our model to vary over time by allowing the parameter values to vary according to a random walk. The use of random walks to non-parametrically model complex epidemiological time series has been successful in forecasting seasonal influenza at the state level [16]. Table 3 lists the parameters in our model which we

allow to take random walks, along with the frequency of steps in the walks, and the standard deviation of the normally distributed step size. Some of these parameters are transformed for the purpose of allowing normally distributed step sizes while staying within their natural domains. The frequency of steps and the standard deviations were determined through trial and error to achieve a good fit to the data. That is, they were treated as hyperparameters which were considered fixed when optimizing the likelihood with respect to our regular model parameters.

Table 3. Random walk models for time-dependent parameters

Parameter	Step frequency	Standard deviation of step size
$\log \beta_0$	every week	0.05
$\operatorname{logit} p_{h,t}$	every 4 weeks	0.5
τ_c	every 4 weeks	0.05

Estimation of ρ_t

Estimation of the time-dependence of ρ_t required special treatment. This parameter changed the most in the first few months of the epidemic as testing volume increased. We account for this by estimating ρ_t for days before June 22, 2020, from the ratio of cases to deaths. This time window corresponds to the first major wave of cases. The idea is that most COVID-19 deaths have been reported with a relatively high and constant probability. We assume that the ratio of the probability of cases being reported to deaths being reported is increasing up to June 21, 2020, and plateaus at 0.5 on that date. This ratio is estimated by fitting a MARS model with the earth package [17]. The model is fit to the log of the ratio of cases to deaths 20 days later. The exponential function of the predicted value of this model is used as an estimate of ρ_t after passing it through the `cummax` function in R [18] to ensure monotonicity and scaling it so that the final value is equal to 0.5. Fig 6 shows the resulting estimate of ρ_t for California.

Fitting methods

The GISST model has many parameters which must be estimated from the data before forecasts can be made. We do this using the method of maximum likelihood. The log likelihood of the model is calculated using an extended Kalman filter. The equations used in our extended Kalman filter implementation are nearly the same as those of the original Kalman filter introduced in Campagnoli *et al.* [12], the only difference being that the process model predictions are not generated by a linear function. We go through our calculations in detail in the next few paragraphs. This log likelihood was optimized using the L-BFGS algorithm [19].

Log likelihood calculation

The first step in calculating the log likelihood is to initialize the means and covariance of our process model's state variables to their values on the beginning of the day of the first observation, which corresponds to March, 2, 2020. The initial value of $\langle L \rangle$ is a model parameter, which we denote L_0. The initial values of $\langle Y \rangle$, $\langle Z \rangle$, $\langle H \rangle$, and $\langle D \rangle$ are calculated from L_0 assuming that the time derivatives in Eq 1 are zero, using the values of any time-dependent parameters when $t = 1$. This assumption is not intended to be highly accurate but rather to generate values that are plausible. The prediction-updates described later will serve to sharpen this initial estimate based on observations. The initial value of $\langle X \rangle$ is set to $N - \langle L \rangle - \langle Y \rangle - \langle H \rangle - \langle D \rangle$. The initial values of $\langle Z_r \rangle$, $\langle A \rangle$, and $\langle D_r \rangle$ are all set to zero, since these variables are meant to represent the

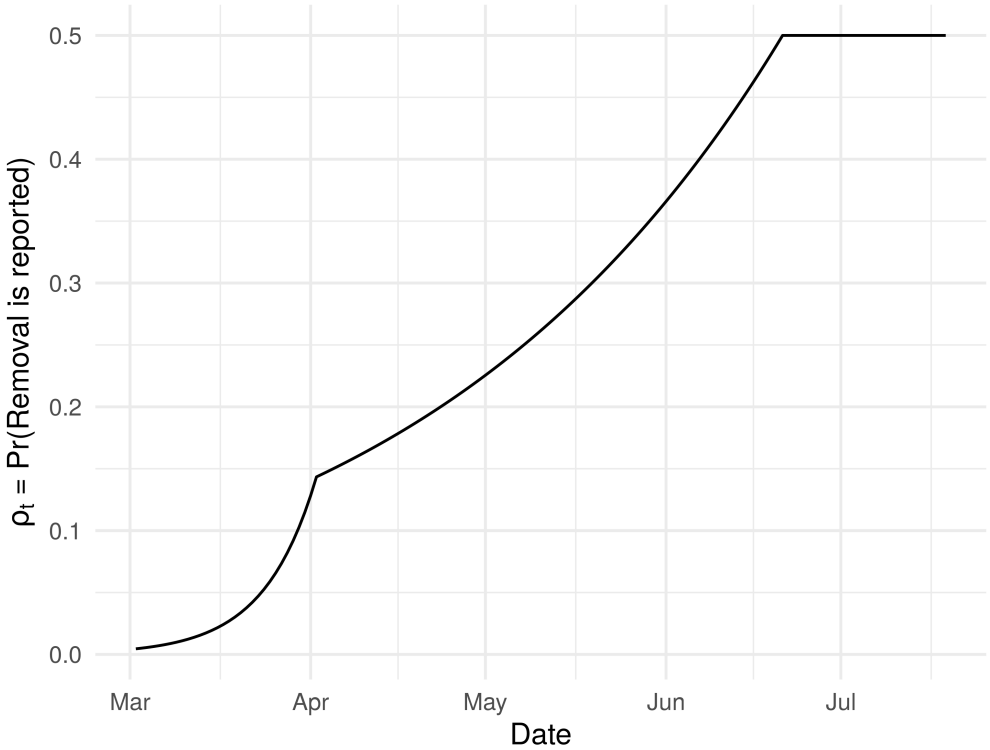

Fig 6. Estimate of the case reporting parameter ρ_t in California. This parameter is assumed to stay at 0.5 for dates beyond the last date plotted. See Tables 1 and 2 for the precise role of this parameter in our model.

accumulation reported cases, hospital admissions, and deaths over the coming day. Sometimes, during optimization, these initial values can become unreasonable and lead to numerical problems in calculating the likelihood. To avoid such problems, we do not allow any values to be greater than N, and we do not allow the intial value of $\langle X \rangle$ to be less than $N/10$. All of these initial values for means are collected in a vector denoted $\hat{\mathbf{x}}_{0|0}$. This is the initial vector of means for our process model. The initial covariance matrix for this model is denoted $\mathbf{P}_{0|0}$ and is set to a diagonal matrix in which all elements are 1 except for those corresponding to $\langle Z_r \rangle$, $\langle A \rangle$, and $\langle D_r \rangle$. These values are equal to zero at the beginning of the day by definition so their variances are set to zero. Similar to the means, the covariance matrix estimate is simply a rough approximation which the algorithm will sharpen. With the process model fully initialized, the next step is to generate a predictive distribution for the observed data based on the process model.

The predictive distribution for the next observations are generated by numerical integration of the system of equations in Eqs 1 and 2 over a time span of 1 day, where we used the convention of 365.25 days per year. For this step we used the `Tsit5` [20] integrator in the DifferentialEquations.jl package [21]. When solved with starting values from time t, the resulting vector of means is denoted $\hat{\mathbf{x}}_{t+1|t}$ and the resulting covariance matrix $\mathbf{P}_{t+1|t}$. If any elements of $\hat{\mathbf{x}}_{t+1|t}$ are negative, this must be due to numerical errors, so we set them to zero. Similarly, if any negative values appear on the diagonal of $\mathbf{P}_{t+1|t}$, the corresponding rows and columns are set to zero. The next step is to iteratively update these predictions for the mean and covariance based on the observed data.

The update of the prediction is done by using equations that minimize squared

prediction error when the process model is linear, and which remain relatively accurate when the model is non-linear. An intermediate step in this update is to calculate the Kalman gain matrix $\mathbf{K}_t$. This matrix determines the weight to give to two types of predictions: (1) predictions of the state variables on day t calculated as described in the previous paragraph by projecting past estimates of the state variables forward according to the process model, and (2) predictions of values of the state variables on day t based solely on observations on day t and the observation model. The equation for $\mathbf{K}_t$ is

$$\mathbf{K}_t = \mathbf{P}_{t|t-1}\mathbf{H}^T\mathbf{\Sigma}_t^{-1}, \tag{7}$$

where $\mathbf{H}$ is a matrix that maps the state variables to observed variables and $\mathbf{\Sigma}_t$ is the sum of the covariance matrix $\mathbf{R}_t$ of the observation model and the projection of the covariance matrix of the process model into the observed coordinates. The matrix $\mathbf{H}$ is 3×9 and contains zeros except for one column in each row, which corresponds to the state variables Z_r, A, and D_r respectively in rows 1, 2, and 3. The matrix $\mathbf{R}_t$ follows

$$\mathbf{R}_t = \begin{bmatrix} \tau_{c,t}\langle Y\rangle_{t-1} + 1 & 0 & 0 \\ 0 & \tau_h + 1 & 0 \\ 0 & 0 & \tau_d + 1 \end{bmatrix}, \tag{8}$$

where $\langle Y\rangle_{t-1}$ is the estimate of Y in $\mathbf{x}_{t-t|t-1}$. The ones on the diagonal of $\mathbf{R}$ prevent numerical problems that could occur if the τ parameters are all nearly zero. The matrix $\mathbf{\Sigma}_t$ follows

$$\mathbf{\Sigma}_t = \mathbf{H}\mathbf{P}_{t|t-1}\mathbf{H}^T + \mathbf{R}_t. \tag{9}$$

Next, a prediction error, denoted $\tilde{\mathbf{y}}_{t|t-1}$, for the process model is calculated as

$$\tilde{\mathbf{y}}_{t|t-1} = \mathbf{z}_t - \mathbf{H}\hat{\mathbf{x}}_{t|t-1}, \tag{10}$$

where $\mathbf{z}_t$ is a column vector containing the observed number of reported cases, hospital admissions, and reported deaths for day t respectively in rows 1, 2, and 3. The data-updated estimate of the state variables, $\hat{\mathbf{x}}_{t|t}$ now follows as

$$\hat{\mathbf{x}}_{t|t} = \hat{\mathbf{x}}_{t|t-1} + \mathbf{K}_t\tilde{\mathbf{y}}_{t|t-1}. \tag{11}$$

The data-updated covariance estimate, $\mathbf{P}_{t|t}$, satisfies

$$\mathbf{P}_{t|t} = (\mathbf{I} - \mathbf{K}_t\mathbf{H})\mathbf{P}_{t|t-1}, \tag{12}$$

where $\mathbf{I}$ denotes the identity matrix. After setting the means and covariances of the variables Z_r, A, and D_r in these updated estimates to zero, they can be used as initial values for projecting the process model forward to obtain $\hat{\mathbf{x}}_{t+1|t}$ and $\mathbf{P}_{t+1|t}$. In this way, process model predictions are obtained for all observations. Any missing observations, such as hospital admissions in data sets from early in the epidemic, are handled by setting the corresponding column in $\mathbf{K}_t$ to zero. Then, the results of these calculations may be used in an equation for the log likelihood.

The log likelihood of our model is calculated as the sum of the likelihood for the one-step ahead predictions of our process model and the step sizes in our random walk models for time-dependent parameters. The expression for the log likelihood of our process model predictions—*i.e.*, the marginal log likelihood of the Kalman filter—is

$$-0.5\sum_t[\tilde{\mathbf{y}}_{t|t-1,\mathrm{nm}}^T\mathbf{\Sigma}_{t,\mathrm{nm}}^{-1}\tilde{\mathbf{y}}_{t|t-1,\mathrm{nm}} + \log\det\mathbf{\Sigma}_{t,\mathrm{nm}} + d_{\mathrm{nm}}\log(2\pi)], \tag{13}$$

where the subscript "nm" indicates that rows and columns which correspond to missing observations in z_t have been omitted and d_{nm} is the number of non-missing observations in z_t. The log likelihoods for the random walk steps are simply mean-zero Gaussian log-likelihoods with the appropriate variances. Having completed the description of the calculation of the log likelihood of a set of parameters, we now turn to the methods we used to optimize the log likelihood.

Optimization

Our optimizer is a local optimizer which performs best when initialized in the neighborhood of the global optimum, so the first step in optimizing the likelihood was to choose good starting parameters for our optimizer. In most cases, we do this by using the parameters estimated from a version of the observed data issued one week earlier for a particular location, which we refer to as a warm start. Any new random walk step parameters introduced by extension of the number of observations are assumed to be zero. In the first data set containing hospital admissions, the parameter τ_h is initialized with the sample variance of all non-missing observations, and the parameter p_h is initialized as the quotient of the sum of all hospital admissions divided by the sum of all reported cases on days when hospital admissions were nonmissing.

In the first data set fitted, which in this study is the data set available on June 29, 2020, all parameter initializations had to be estimated from the data. We found that the following rough estimates lead to a satisfactory fit. All $\tau_{c,t}$ were intialized as the sample variance of cases divided by the mean number of cases. τ_d was initialized with the sample variance of observed deaths. If either of these values were less than 1, they were replaced with 1. An estimate of the unobserved time series of Y was calculated by dividing the observed cases time series by $\gamma\rho_t \mathrm{d}t$, in which $\mathrm{d}t$ is the daily time step in our process model. The mean of the first 7 values of this time series times γ/η was added to 1 to initialize L_0. The estimate of the time series of Y times $\gamma p_h/\gamma_h$ was used to generate an estimate of the unobserved hospitalization time series. The parameter p_d was initialized to the sum of observed deaths divided by the sum of these estimated hospitalizations. If this value was less than 0.01, the initialization was raised to 0.01. All $\gamma_{z,w}$ and $\gamma_{d,w}$ were initialized at the values given in Table 2.

An estimate of the time series of the effective reproduction number $\mathcal{R}_e = (\beta_t/\gamma)(X(t)/N)$ was calculated to generate initial values of β_{res} and $\beta_{0,t}$. Let cases on day t be denoted c_t. Our raw estimate of $\mathcal{R}_e$ was $(c_{t+7}/\rho_{t+7})/(c_t/\rho_t)$. These raw values were next clipped to be within 0.1 and 4 and then smoothed by applying a Savitzky-Golay filter of order 2 and length 21. To avoid bias from incomplete windows, the first and last 10 elements of output were considered missing. An estimate of a time series of X was calculated by first initializing it with the same value that would be used in the process model given the initial value of L_0. Then, successive values were calculated by subtracting c_{t+7}/ρ_{t+7} from the estimate of X on day $t-1$. Then, a time series of β_t estimates was calculated by multiplying the smoothed $\mathcal{R}_e$ estimate by γN and dividing by the estimate of X. Missing values on the ends of the time series were filled in by repetition of the first and last non-missing value. The slope from a linear regression of $\log \beta_t$ on $\texttt{residential}_t$ then provided the initial value of β_{res}. The residuals of this linear fit were grouped by week and averaged to provide the initialization of $\beta_{0,t}$. We have now described how all parameters were initialized.

Initial parameters were iteratively improved by optimizing with the L-BFGS algorithm in the R lbfgs package [22] to minimize the negative log likelihood. This algorithm was chosen for its ability to scale to a large numbers of parameters. Rate parameters were kept positive by fitting a log transform of them, and logits of probabilities were fitted to ensure they were valid probabilities. To keep all parameters on similar scales, the covariate $\texttt{doses}_t$ was divided by N and the covariate $\texttt{residential}_t$ was divided by 100. We provided a gradient function to the optimizer by using forward mode automatic differentiation via the ForwardDiff.jl [23] and DiffEqSensitivity.jl [24] packages. We ran the optimizer for 1000 iterations when fitting the first data set and when fitting the hospitalization data for the first time and 100 iterations when using a warm start. This number of iterations was typically sufficient to reduce the norm of the gradient to be 10 times smaller than the norm of the parameter vector. We stopped early if the norm of the gradient was 100 times smaller than the

norm of the parameter vector. The suitability of the returned parameters was evaluated by calculating the Hessian matrix with automatic differentiation and verifying that it was positive definite (*i.e.*, the negative log likelihood was convex). The inverse of the Hessian matrix was also used as a covariance matrix to generate Wald-type interval estimates for parameters [25].

Forecasts

Once a model is fitted, generation of forecasts is straightforward. In fact, the $\hat{\mathbf{x}}_{t|t-1}$ calculated as a part of the likelihood are already available as 1-step ahead point forecasts. We used these within-sample forecasts to evaluate the goodness of fit of a model. To generate out-of-sample forecasts, we simply simulated the process model in Eqs. 1 and 2 forward as far as needed beyond the day of the last observation and added the model-determined observation noise. In doing this, we used predicted values of β_t in the future from an AR-1 model fit to the daily times series of the deviations of β_t from γ in the past. The model has its mean fixed at zero such that only the AR coefficient for the deviations is estimated. We denote this coefficient ϕ and provide an example of an estimate in Table 4. The input estimates of β_t have a similar time series to the estimates of $\mathcal{R}_e$ plotted in Fig 7. In this figure, one can see that following an initial transient period, $\mathcal{R}_e$ had a tendency to fluctuate around and then revert to 1. This tendency may be a consequence of populations seeking to minimize the social and economic costs of non-pharmaceutical interventions (NPIs). Perhaps NPIs are strengthened only when $\mathcal{R}_e$ is clearly above one and COVID-19 indicators are on the rise and NPIs are relaxed soon after these indicators begin to drop. At any rate, we found that forecasts generated using predictions of β_t from an AR-1 model fitted to the estimates of β_t from days beyond May 1, 2020 up to the forecast date performed better than forecasts which assumed that β_t would remain at its most recently estimated value throughout the forecast period. The key benefit of the use of the AR-1 model seems to be that it limits the extent to which forecasts overshoot the peaks of pandemic waves.

In calculating forecasts, all parameters modeled as random walks were fixed at their final estimated value. To produce forecasts for the COVID-19 Forecast Hub, we recorded and zeroed the D_r and Z_r variables every 7 days to produce weekly forecasts of cases and deaths. Additionally, to calculate the observation model variance of a weekly observation, we summed the observation model variances that would have been applied to forecasts of individual days in the week. The Hub accepts probabilistic forecasts in the form of a set of 23 quantiles, {0.01, 0.025, 0.05, 0.1, 0.15, ..., 0.85, 0.90, 0.95, 0.975, 0.99}, for deaths and hospital admissions and a set of 7 quantiles, {0.025, 0.1, 0.25, 0.5, 0.75, 0.9, 0.975}, for cases. We obtained these quantiles using the `qnorm` function in R with the mean and variance of our model's forecast distributions. Any negative quantiles—resulting from predicted distributions with a large coefficient of variation, for example—were replaced with zeros.

We conducted evaluations as follows. We used a version of data available on June 21, 2021. The extent of revision in our training data are summarized in Fig 2. These revisions were large enough to decrease the accuracy of forecasts but not so large that the decreases were catastrophic. Forecasts were scored using the weighted interval score [26]. We scored forecasts of each indicator and forecast horizon separately. For each of these, we use a set of forecast dates for which our model, the COVID-19 Forecast Hub Ensemble model, and the COVID-19 Forecast Hub Baseline models have produced forecasts. The Ensemble model and Baseline models are useful reference points because the Baseline model represents a level of forecast skill achievable with little effort whereas the Ensemble model represents a model with a high level of forecast skill. The Ensemble model is formulated as the median value of the quantiles of forecasts submitted to the Forecast Hub and has performed better than most individual

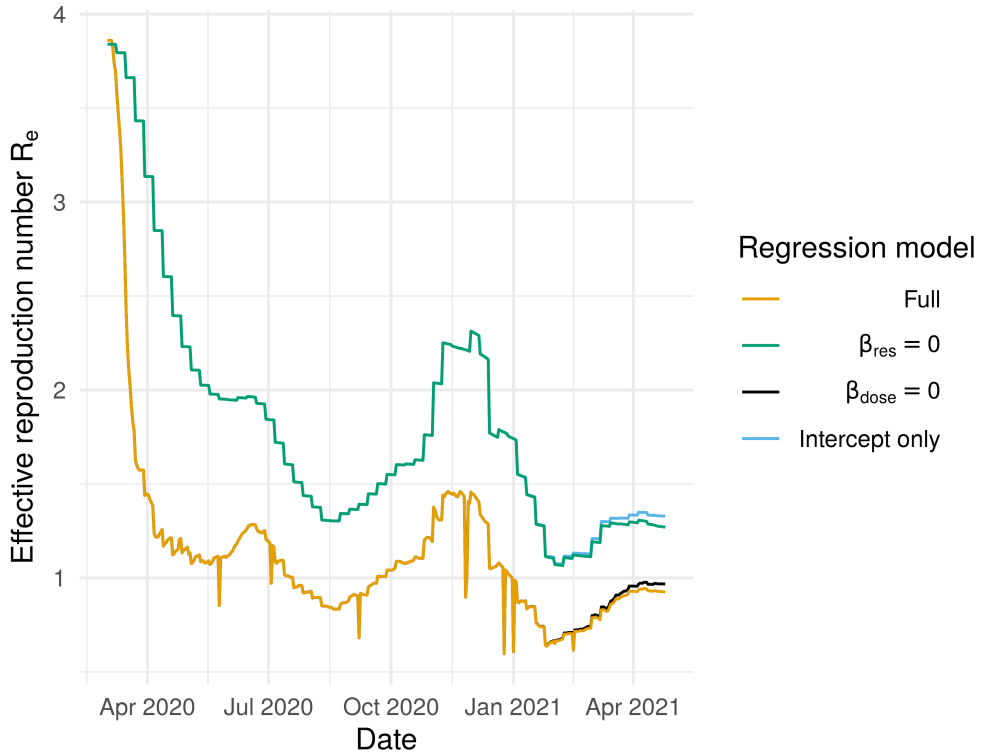

Fig 7. Effective reproduction number of fitted model for California. The model is fitted to data available on April 26, 2021. The full model corresponds to plugging β_t from Eq 4 into $\mathcal{R}_e = (\beta_t/\gamma)(X(t)/N)$. We also plot $\mathcal{R}_e$ with a β calculated by setting the effects of one or both of the covariates from their estimated values to zero to show their effects.

models. We analyze forecasts of data from dates ranging from June 29, 2020, to April 26, 2021, which includes the second and third major waves of spread. We do not evaluate forecasts of the first wave of spread because Ensemble forecasts are not available for most of it. Additionally, the interpretation of the indicators in the first wave is not straightforward as many surveillance systems were changing rapidly as the severity of the pandemic grew. We do not evaluate forecasts beyond April 2021 because the indicator counts were reduced and subject to more changes in the reporting schedule as the pandemic began to wane.

Reproducibility

All data used was publicly available. An archive of the code used to obtain the data and generate all results is available on Zenodo at `https://doi.org/10.5281/zenodo.5112578`.

Results

Goodness of fit

We experimented with model structure until two goodness of fit criteria were satisfied. The first was that errors in the 1-step ahead forecast had the expected normal

distribution. We used this diagnostic to evaluate whether that the model's normality assumptions were reasonable. Fig S1 indicates that they were, with the exception of a relatively small number of outliers.

The second criteria was that MASE (mean absoluate scaled error) [27] was below 1 for all model fits. Due to the strong weekly seasonality in some of the data, we calculated MASE with both a non-seasonal naive model and a 7-day seasonal naive model. Our criteria was that both versions be less than 1. This criterion provided a fast method to evaluate the forecasting skill of our model and its potential to outperform baseline models at the longer forecast horizons in our primary evaluation analysis. Fig S2 shows that our model for California satisfied these criteria.

Parameter estimates

An advantage of the mechanistic aspects of our model is that they provide some insight into why the a time series of indicators may be forecasted to go in a certain direction. To illustrate how this may be possible with our model, we next present an example of parameter estimates in our fitted models. The plausibility of these estimates in light of other sources of information about COVID-19 also provides a means to evaluate the assumptions of our model.

The effective reproduction number is perhaps the parameter of greatest interest in our model because it provides an easy-to-understand statement of whether the pandemic is growing or not [28]. The (instantaneous) effective reproduction number $\mathcal{R}_e$ in our model is a derived quantity equal to $\mathcal{R}_e = (\beta_t/\gamma)(X(t)/N)$. Fig 7 shows the value of this expression with our model's fixed values of γ and N, the predicted value of β_t resulting from our fitted regression model in Eq 4, and the estimates of $X(t)$ in $\mathbf{x}_{t|t}$ in Eq 11. The estimates seem reasonable, taking relatively large values early in the pandemic and then falling to and fluctuating around 1 after control measures were implemented. This figure also shows the contribution of individual regression model coefficients to the regression model's prediction of $\mathcal{R}_e$. All coefficients have effects in the expected direction, although the dose effect is small. Estimated values for these coefficients are in Table 4, along with estimates of other parameters that were not time-dependent.

Table 4. Maximum likelihood estimates of model parameters that are not time-dependent. Estimates are from the version of data available on April 26, 2021. ϕ is the coefficient for the AR-1 model of β_t used to make forecasts.

Parameter	Estimate	Standard error
β_{dose}	-0.06	0.34
β_{res}	-3.76	1.49
L_0	4.50	1.18
$\text{logit}\, p_d$	-1.17	0.02
$\log \gamma_{d,1} = \log \gamma_{d,2}$	2.85	0.06
$\log \gamma_{d,3} = \log \gamma_{d,4}$	3.61	0.03
$\log \gamma_{z,1} = \log \gamma_{z,7}$	5.79	0.04
$\log \tau_h$	6.61	0.27
$\text{logit}\, p_{h,\text{weekend}}$	2.28	0.07
$\log \tau_d$	8.41	0.07
ϕ	0.94	0.02

Because we repeatedly fit the model to longer and longer time series for the purpose of making forecasts, even parameters which are assumed to be constant in a single model fit can reveal how the model fit or the underlying epidemiology changes over

time. Fig S3 shows that the estimated effect of our indicator for the amount of time spent in residential areas was greatly reduced once hospital admissions data became available for fitting. We interpret this as a sign that our model was systematically incorrect in its assumed number of hospital admissions before being provided data, such that the availability of data triggered a step change in parameter estimates. Perhaps more epidemiologically interesting, the effect of this indicator trends toward zero from January onward. This behavior is consistent with NPIs becoming less important as vaccine-induced immunity builds up.

Estimates of time-dependent parameters are shown in Figs S4, S5, and S6. Overall, these parameters are estimated with a reasonable amount of precision and the trends seem reasonable. The probability of hospitalization is initially high and then quickly moves down to about 3 percent and fluctuates around that number. The rise in the random walk intercept $\beta_{0,t}$ in the fall of 2020 coincides with the onset of the third major wave of cases. The observation variance in cases $\tau_{c,t}$ per infected individual grows in proportion to the number of cases reported. This variance-to-mean relationship reflects overdispersion relative to a Poisson model, which is frequently the case for observation models of infectious disease time series.

Evaluation of short-term forecasts

Scores from our primary evaluation of forecast skill for cases and deaths are displayed in Fig 8. The GISST model performs best relative to the reference models at the 1- and 4-week ahead horizon for cases and the 3-4 week ahead horizons for deaths. For cases, if we had averaged scores on the log scale, the time windows leading to the greatest difference in scores might be the peak of the second wave and end of the third wave of the pandemic (Fig S7). However, we do not use the mean of the log of the weighted interval score as our summary, but rather the mean of the untransformed weighted interval score. Using the untransformed mean makes the summaray score proper, a desireable technical property for scores of probabilistic forecasts. A practical consequence of our use of the simple mean is that forecasts around the peak of the pandemic are given the most weight in the overall score. It is thus possible to see how the overall scores in Fig 8 are approximations of the average scores of observations in November through February in Fig S7. In turn, the correspondence between these scores and the difference between forecasted and observed trajectories can be seen by comparing Fig S7 with Figs 9A and 9C. One can see how the good interval scores for the GISST model's 4-week ahead deaths forecasts for observations in the final quarter of 2020 seen in Fig S7 do indeed correspond to forecasts that are sharper and more accurate than the COVID-19 Forecast Hub Ensemble in Fig 9C.

Scores from our primary evaluation of forecast skill for hospital admissions are displayed in Fig 10. The GISST model performs best relative to the COVID-19 Forecast Hub Ensemble model at all horizons except for 7 and 28 days. There is a clear periodicity in the weighted interval scores of the COVID-19 Forecast Hub Ensemble as a function of the horizon which likely originates from not modeling the reduction in hospital admissions on weekends (Fig 10). Examination of forecast trajectories in Fig 9B reveals that the overall better scores of GISST occurred in spite of poor forecasts early in the evaluation period, in which the 0.025 probability-level quantile of the the GISST forecast greatly exceeded the maximum value of hospital admissions in the December 14 forecasts. These poor forecasts were compensated for by consistent sharp and accurate forecasts from GISST for observations in the last half of January and throughout February. Performance at later dates was not as strong for GISST but these had less influence on the overall score due to the decreasing trend in hospitalizations.

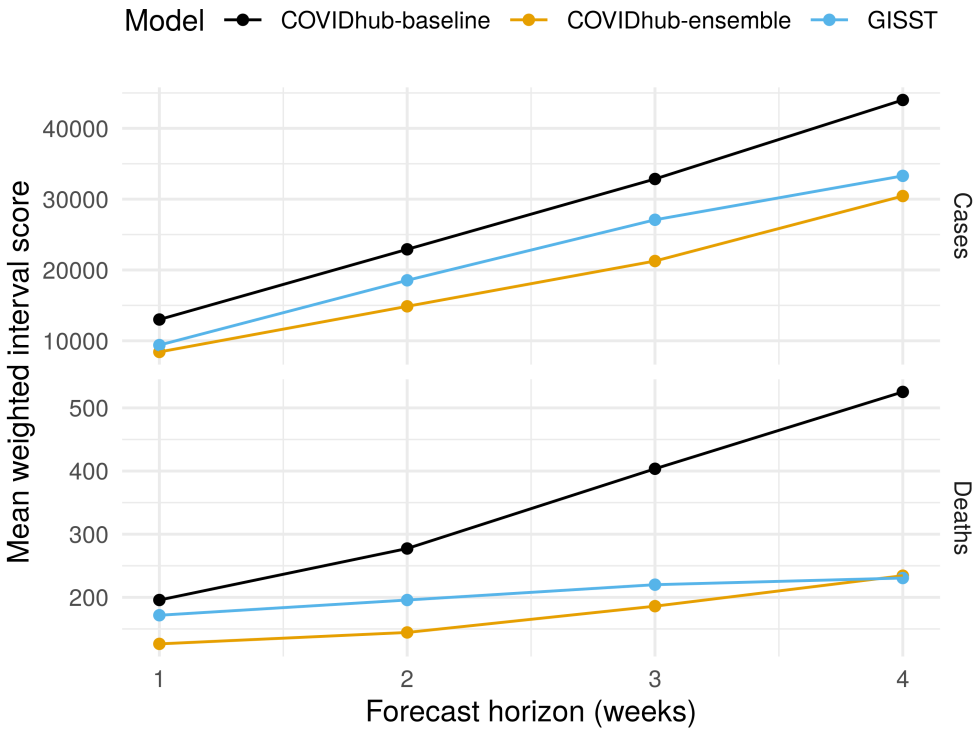

Fig 8. Overall performance of short-term forecasts of cases and deaths. Lower scores indicate better performance. The GISST forecast consistently outperforms the COVID-19 Forecast Hub Baseline forecast and performs relatively best for 4-week ahead death forecasts.

Discussion

We have introduced the GISST forecasting model and compared its performance in forecasting cases, hospital admissions and deaths with a simple baseline forecaster and a respected ensemble forecaster. Although some of its early December forecasts greatly overestimated future cases and hospital admissions, in summary evaluations of cases and deaths forecasts the GISST model dominated the baseline model and typically had a score closer to the ensemble model (Fig 8). In summary evaluations of forecasts of hospital admissions, the GISST model scored better than the ensemble in 28 out of 28 horizons (Fig 10). In addition to these positive performance attributes, GISST has the advantage of allowing trends in the forecasted indicators to be plausibly linked to each other and leading indicators such as metrics of human mobility. Additionally, it is computationally efficient. Even for the fits to April 2021 data sets, when our model had the largest number of parameters and observations, fitting an additional week of data required less than 4 hours of serial computation (Fig S9). In addition to permitting forecasts to be generated online for a large number of locations, this efficiency facilitates model development by shortening the time required to test changes in model structure.

There are COVID-19 forecasts from many other forecasters besides the COVID-19 Forecast Hub Baseline and Ensemble models, and we next put our model in context with some of these other forecasters. The MechBayes model of Ref [15] provided accurate forecasts of COVID-19 deaths in real-time, and like GISST it does so by fitting a compartmental infectious disease model. But unlike GISST, the MechBayes model

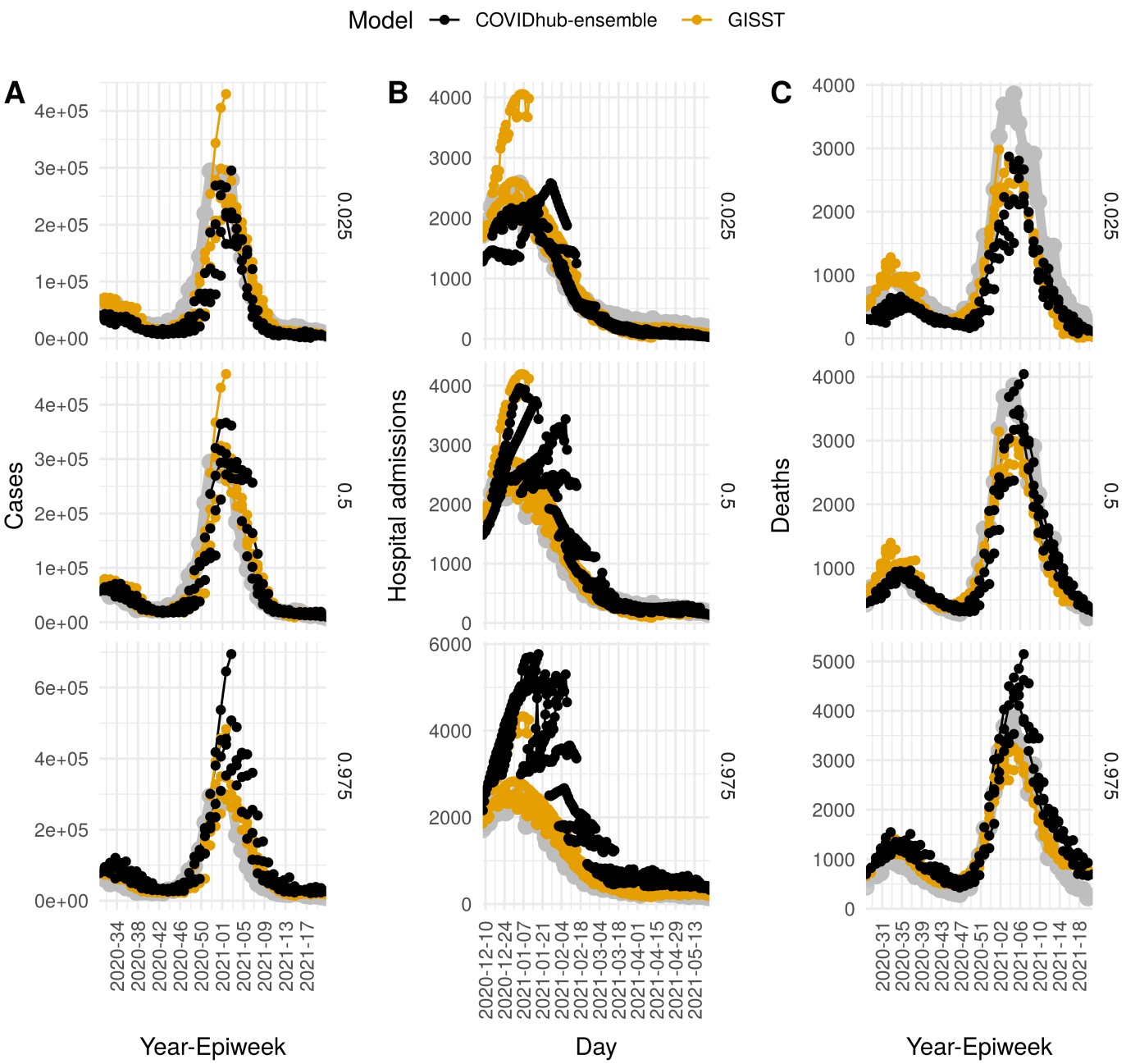

Fig 9. Forecast trajectories and target time series for cases (A), hospital admissions (B), and deaths (C). The panel labels at right are quantiles of the forecast. The target time series is plotted in grey. The performance penalty for underpredicting the true values is proportional to the quantile value, and the performance penalty for overprediction is proportional to 1 less the quantile value. Lines are drawn between forecasts of different horizons from a single forecast date.

does not include a regression model for the transmission rate and requires Markov Chain Monte Carlo (MCMC) sampling to fit, which likely makes it more computationally

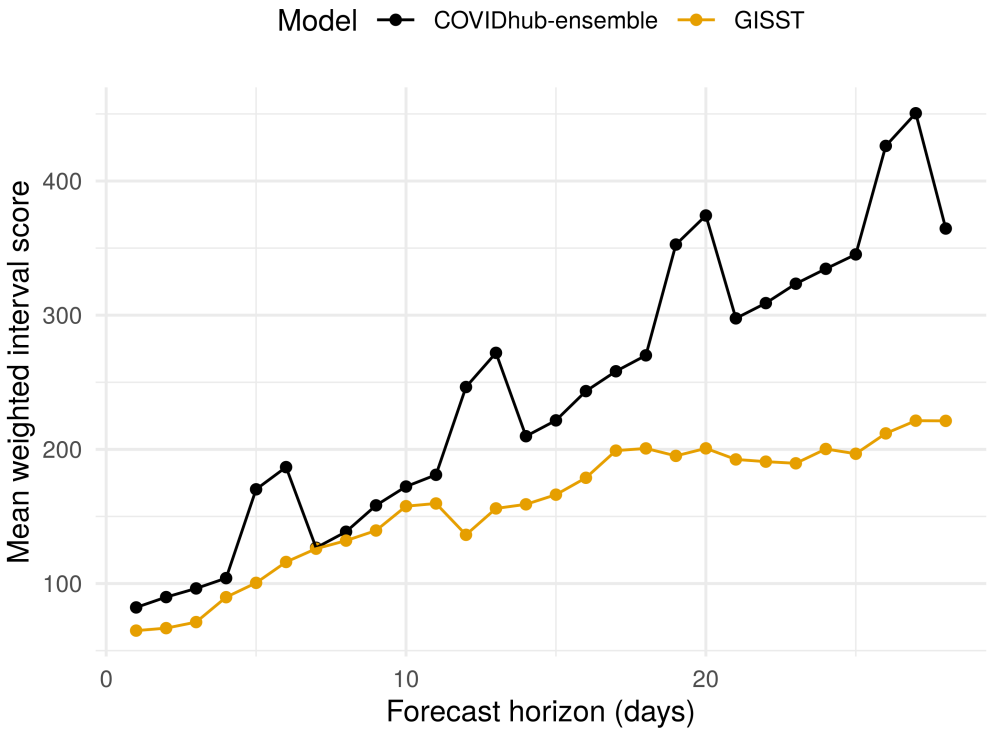

Fig 10. Overall performance of short-term forecasts of hospital admissions. Lower scores indicate better performance. The GISST forecasts score better than the COVIDhub Ensemble forecasts at all horizons. However, the difference in scores is slight at the 7 and 8 day horizons.

costly. Further, it does not include hospital admissions. The method introduced in Ref [29] is much more computationally efficient than GISST, but it does not forecast hospital admissions nor allow for transmission covariates. The Hub forecaster karlen-pypm introduced in Ref [30] has produced good forecasts of all 3 indicators via a mechanistic population model. More recent versions of this model explicitly account for the dynamics of multiple variants of SARS-CoV-2. However, it assumes parameters are fixed except for at changepoints, which must be estimated. Ref. [31] presents a fully Bayesian method for estimating such changepoints of transmission rates in a mechanistic population model. From the results of such estimation in time series of COVID cases in Germany, it seems that the data were only highly informative of about 1 out of 3 changepoint times. This result suggests to us that when prior information about the timing of changepoints is not readily available, a model that assumes a random walk in parameter values may be a better choice to avoid biases due to model misspecification.

Recently, Ref [32] presented results from an SEIR-type model of COVID in US states that included environmental covariates such as temperature and population density in addition to human mobility covariates. This was a spatially hierarchical Bayesian model fit with MCMC and likely more computationally demanding than GISST. Further, it forecasted deaths only and these forecasts were considered poor by the authors for horizons of more than 2 weeks. However, it is difficult to objectively compare accuracy with other models because the forecasts are not available in the standard Forecast Hub format.

The worst-scoring forecasts from GISST were made from dates in December (Figs S7

and S8) when the GISST model greatly over-predicted the number of cases and hospitalizations (Figs 9A and 9B). Although increasing the accuracy of such predictions may prove extremely difficult, future work should evaluate the benefits of incorporating uncertainty in future values of β_t, which seems to be one of the strengths of the COVID-19 Forecast Hub Ensemble forecaster.

Another limitation of the GISST model, as we have presented it here, may be that it is missing important covariates in the regression model for the transmission rate (Eq 4). For example, we have not tried using other mobility metrics provided in the Google Community mobility reports. On the one hand, predicting future values of such covariates could be just as challenging as predicting future values of COVID-19 indicators without them. One might address this problem to some extent by applying a penalty proportional to the absolute value of the regression coefficient of these covariates, which would shrink the effect of unimportant covariates to zero. This approach would be readily achievable with the optimizer we used by choosing the Orthant-Wise Limited-memory Quasi-Newton (OWL-QN) algorithm instead of the LBFGS algorithm. The use of this option is what we alluded to in the introduction as the ability of GISST to search for important variables from a large space of candidate variables. Variables identified in this manner may sufficiently lead the indicators we wish to forecast that they improve forecasts just by being available up to or close to the date of forecast. Knowing which variables are important for predicting the transmission rate, even if they are difficult to forecast, could also be valuable for the design of forecasts.

Supporting Information

Acknowledgments

We thank Éric Marty for creating the flow diagram of the ODE system and the graphical summary of our method.

Author contributions

Conceptualization: John M. Drake, Eamon B. O'Dea.
Funding acquisition: John M. Drake.
Supervision: John M. Drake.
Writing — review & editing: John M. Drake.
Data curation: Eamon B. O'Dea.
Formal analysis: Eamon B. O'Dea.
Investigation: Eamon B. O'Dea.
Methodology: Eamon B. O'Dea.
Software: Eamon B. O'Dea.
Validation: Eamon B. O'Dea.
Visualization: Eamon B. O'Dea.
Writing — original draft preparation: Eamon B. O'Dea.

References

1. Martin A, Markhvida M, Hallegatte S, Walsh B. Socio-Economic Impacts of COVID-19 on Household Consumption and Poverty. Econ Disaster Clim Chang. 2020; p. 1–27.

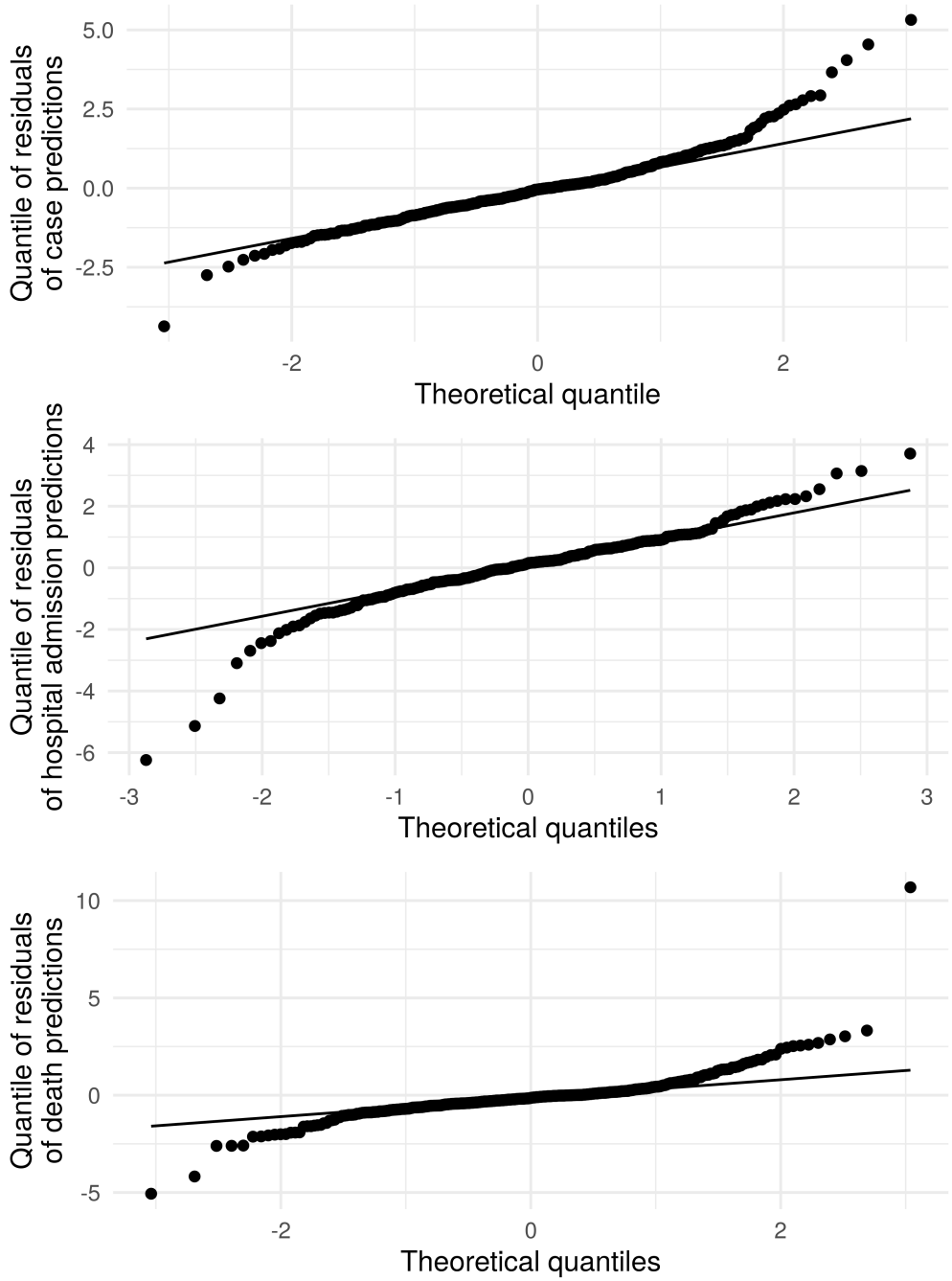

Fig S1. Quantile-quantile plot of distribution of 1-step ahead forecast errors in fit to California data. The specific residuals in the panels are the elements of $\tilde{\mathbf{y}}_{t|t-1}$ in Eq 10 divided by the square root of variances on the diagonal of $\mathbf{\Sigma}_t$ in Eq 9. These are quantiles from a fit to the data available on April 26, 2021, and thus represent a fit to data spanning the majority of the available data.

2. Wu JT, Leung K, Leung GM. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. The Lancet. 2020;395(10225):689–697.

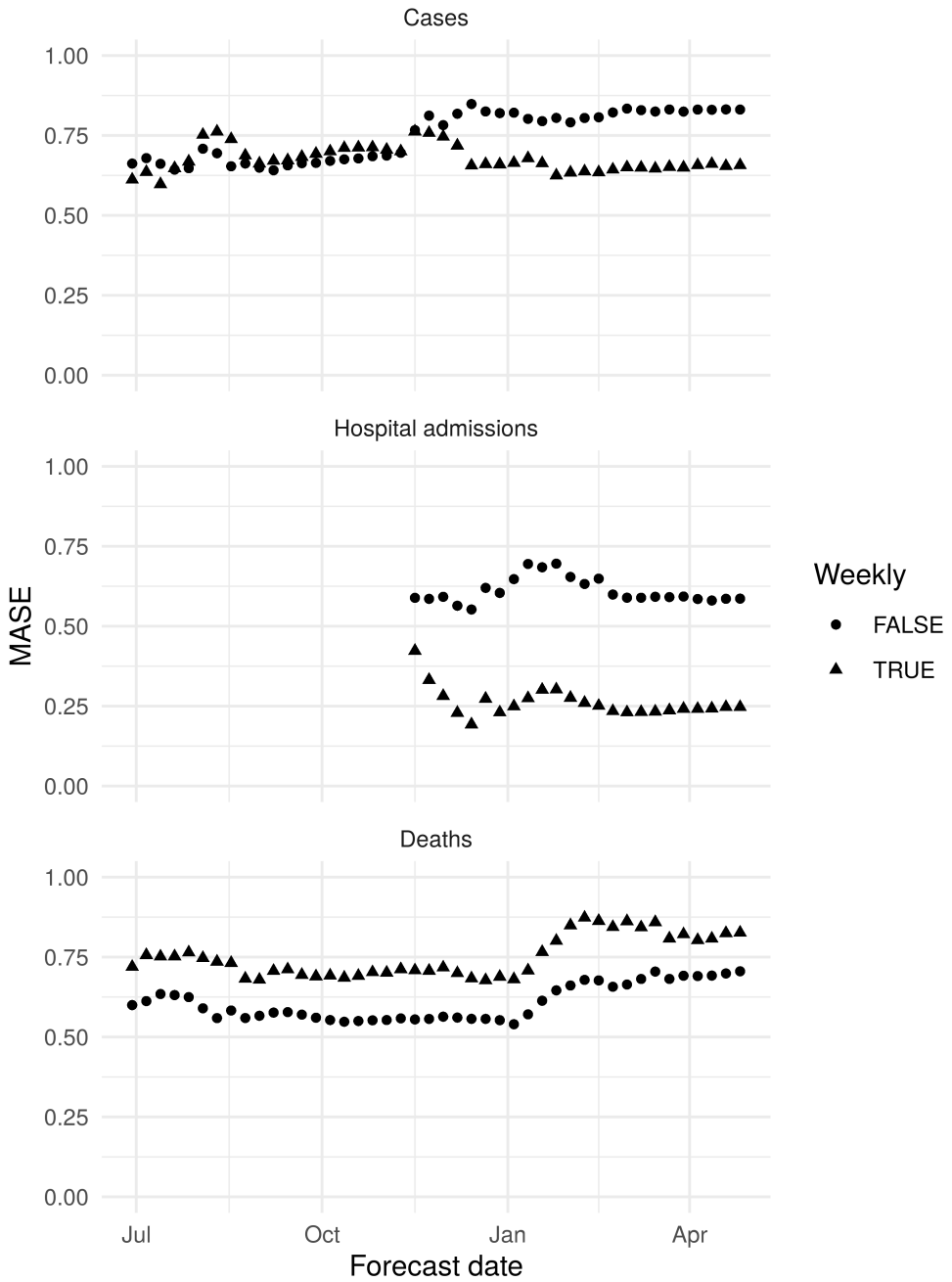

Fig S2. MASE of fits to California data. Weekly indicates whether MASE were calculated by using the previous observation from the same day of the week as the naive prediction (Weekly = TRUE) or whether MASE were calculated by using the previous day's observation as the naive prediction (Weekly = FALSE).

doi:10.1016/S0140-6736(20)30260-9.

3. Manheim D, Chamberlin M, Osoba O, Vardavas R, Moore M. Improving Decision Support for Infectious Disease Prevention and Control: Aligning Models and Other Tools with Policymakers' Needs. RAND Corporation; 2016. Available

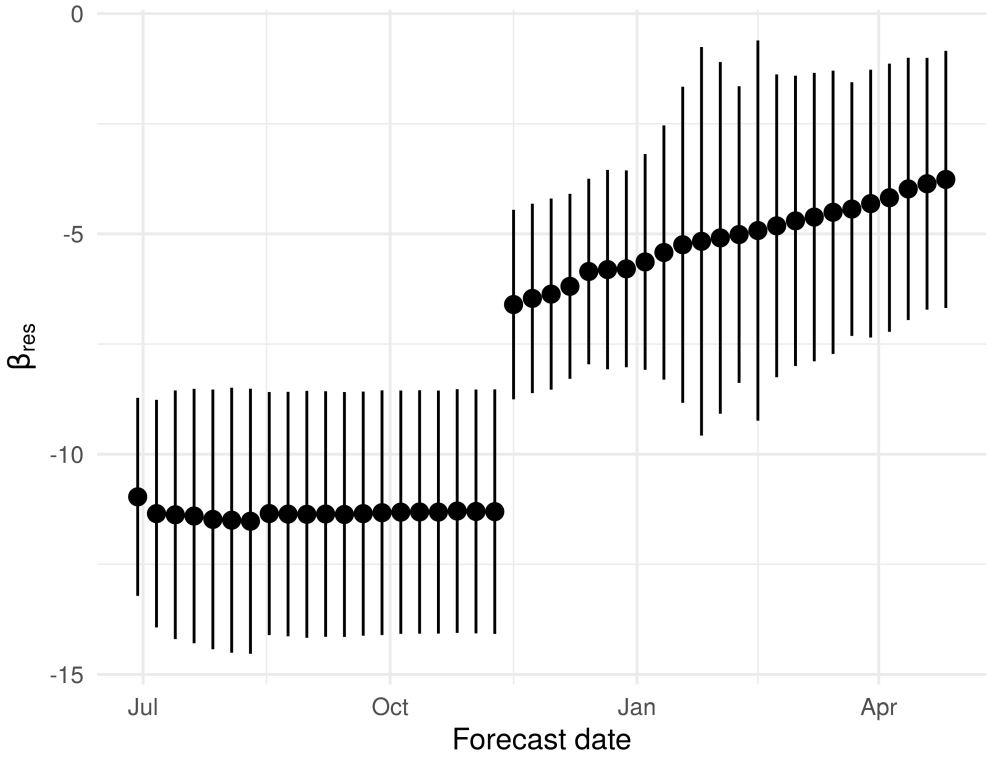

Fig S3. The effect of time spent in residential areas on the transmission rate $\beta_{\mathbf{res}}$ changes as more data become available. Error bars are 95% confidence intervals. Estimates come from modes fitted with data available up to the forecast date on the x axis.

from: `http://www.rand.org/pubs/research_reports/RR1576.html`.

4. Lauer SA, Brown AC, Reich NG. Infectious Disease Forecasting for Public Health. arXiv:200600073 [stat]. 2020;.

5. Ray EL, Wattanachit N, Niemi J, Kanji AH, House K, Cramer EY, et al. Ensemble Forecasts of Coronavirus Disease 2019 (COVID-19) in the U.S. medRxiv. 2020; p. 2020.08.19.20177493. doi:10.1101/2020.08.19.20177493.

6. Cramer E, Reich NG, Wang SY, Niemi J, Hannan A, House K, et al.. COVID-19 Forecast Hub: 4 December 2020 snapshot; 2020. Available from: `https://zenodo.org/record/4305938`.

7. Castle JL, Doornik JA, Hendry DF. The value of robust statistical forecasts in the COVID-19 pandemic. National Institute Economic Review. 2021;256:19–43. doi:10.1017/nie.2021.9.

8. Doshi-Velez F, Kim B. Towards A Rigorous Science of Interpretable Machine Learning. arXiv:170208608 [cs, stat]. 2017;.

9. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. The Lancet Infectious Diseases. 2020;20(5):533–534. doi:10.1016/S1473-3099(20)30120-1.

10. Farrow DC, Brooks LC, Tibshirani RJ, Rosenfeld R. Delphi Epidata API; 2020. Available from: `https://github.com/cmu-delphi/delphi-epidata`.

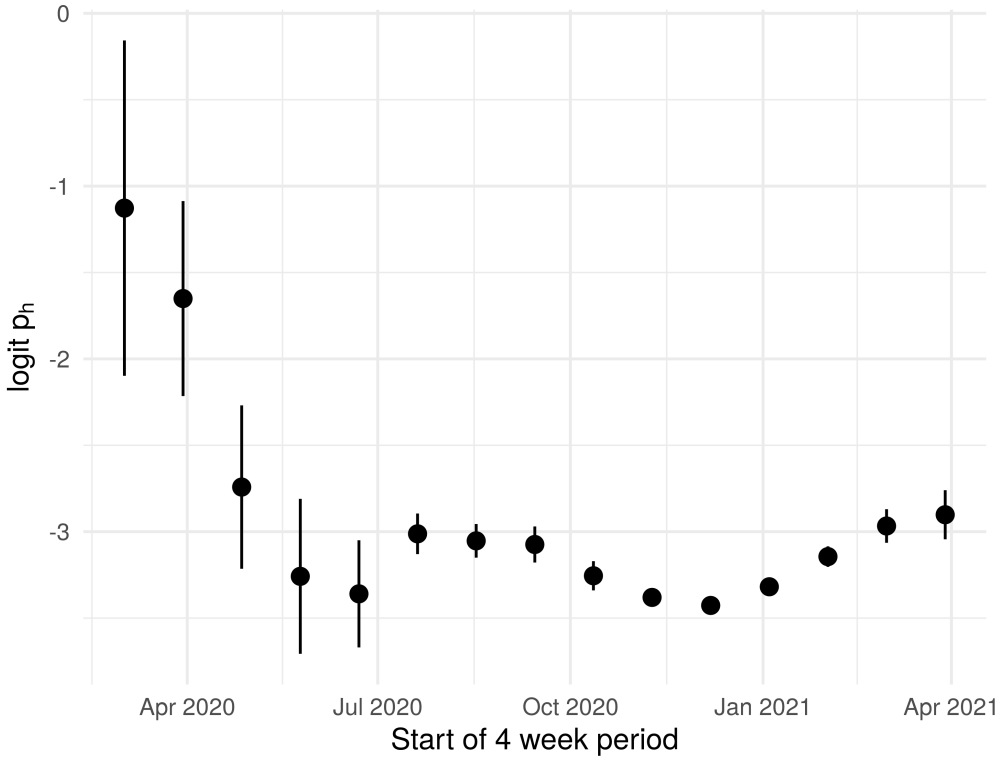

Fig S4. Estimated probability of hospitalization p_h over time in California model. Points are maximum likelihood estimates and error bars are 95% confidence intervals.

11. Google LLC . Google COVID-19 Community Mobility Reports; 2020. Available from: `https://www.google.com/covid19/mobility/`.

12. Campagnoli P, Petrone S, Petris G. Dynamic Linear Models with R. New York, NY: Springer New York; 2009.

13. van Kampen NG. Stochastic Processes in Physics and Chemistry, Third Edition. 3rd ed. Amsterdam ; Boston: North Holland; 2007.

14. Tkachenko AV, Maslov S, Elbanna A, Wong GN, Weiner ZJ, Goldenfeld N. Time-dependent heterogeneity leads to transient suppression of the COVID-19 epidemic, not herd immunity. PNAS. 2021;118(17). doi:10.1073/pnas.2015972118.

15. Gibson GC, Reich NG, Sheldon D. Real-time mechanistic Bayesian forecasts of COVID-19 mortality. medRxiv. 2020; p. 2020.12.22.20248736. doi:10.1101/2020.12.22.20248736.

16. Osthus D, Moran KR. Multiscale influenza forecasting. Nature Communications. 2021;12(1):2991. doi:10.1038/s41467-021-23234-5.

17. Milborrow S. earth: Multivariate Adaptive Regression Splines; 2011. Available from: `http://CRAN.R-project.org/package=earth`.

18. R Core Team. R: A Language and Environment for Statistical Computing; 2020. Available from: `https://www.R-project.org/`.

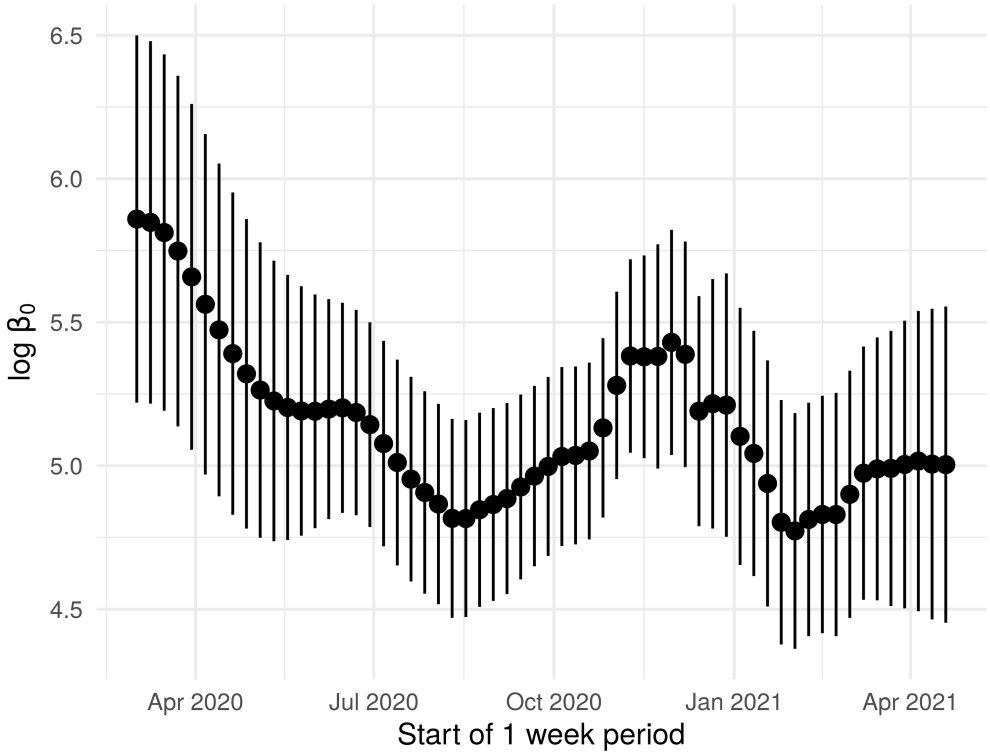

Fig S5. Estimated transmission rate intercept $\beta_{0,t}$ over time in California model. Points are maximum likelihood estimates and error bars are 95% confidence intervals.

19. Nocedal J, Wright SJ. Numerical optimization. 2nd ed. Springer series in operations research. New York: Springer; 2006.

20. Tsitouras C. Runge–Kutta pairs of order 5(4) satisfying only the first column simplifying assumption. Computers & Mathematics with Applications. 2011;62(2):770–775. doi:10.1016/j.camwa.2011.06.002.

21. Rackauckas C, Nie Q. DifferentialEquations.jl – A Performant and Feature-Rich Ecosystem for Solving Differential Equations in Julia. Journal of Open Research Software. 2017;5(1):15. doi:10.5334/jors.151.

22. Coppola A, Stewart B, Okazaki N. lbfgs: Limited-memory BFGS Optimization; 2014.

23. Revels J, Lubin M, Papamarkou T. Forward-Mode Automatic Differentiation in Julia. arXiv:160707892 [cs]. 2016;.

24. Rackauckas C, Ma Y, Dixit V, Guo X, Innes M, Revels J, et al. A Comparison of Automatic Differentiation and Continuous Sensitivity Analysis for Derivatives of Differential Equation Solutions. arXiv:181201892 [cs]. 2018;.

25. Pawitan Y. In All Likelihood: Statistical Modelling and Inference Using Likelihood. 1st ed. Oxford University Press, USA; 2001.

26. Bracher J, Ray EL, Gneiting T, Reich NG. Evaluating epidemic forecasts in an interval format. arXiv:200512881 [q-bio, stat]. 2020;.

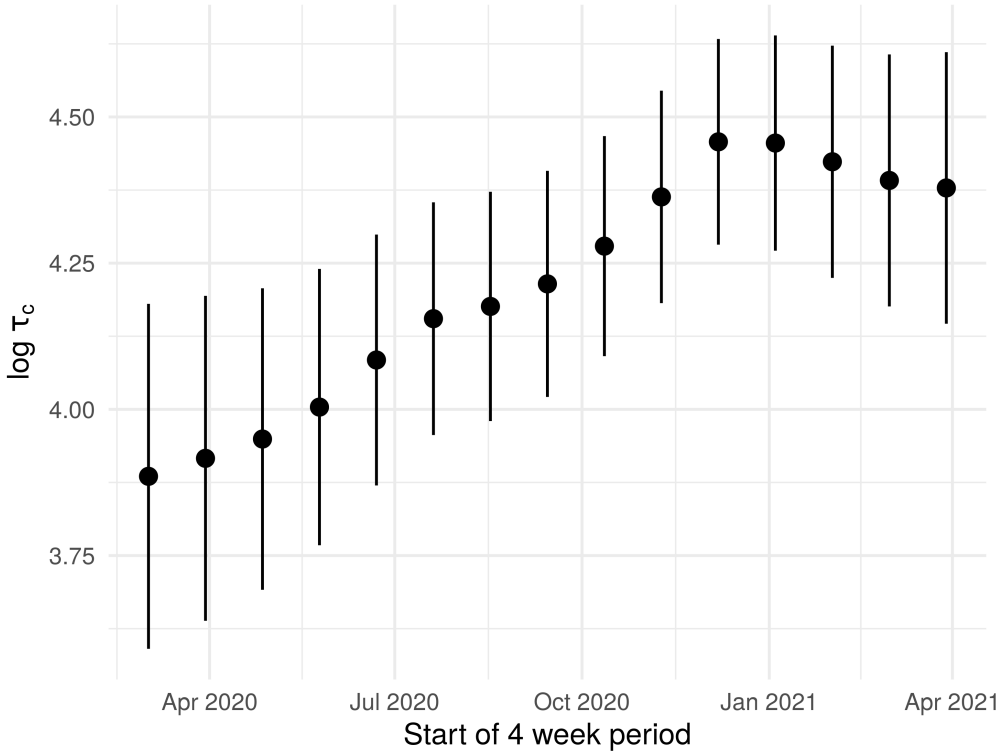

Fig S6. Variance in observation error τ_c over time in California model. Points are maximum likelihood estimates and error bars are 95% confidence intervals.

27. Hyndman RJ, Koehler AB. Another look at measures of forecast accuracy. International Journal of Forecasting. 2006;22(4):679–688. doi:10.1016/j.ijforecast.2006.03.001.

28. Gostic KM, McGough L, Baskerville EB, Abbott S, Joshi K, Tedijanto C, et al. Practical considerations for measuring the effective reproductive number, Rt. PLOS Computational Biology. 2020;16(12):e1008409. doi:10.1371/journal.pcbi.1008409.

29. Srivastava A, Xu T, Prasanna VK. Fast and Accurate Forecasting of COVID-19 Deaths Using the SIkJα Model. arXiv:200705180 [physics, q-bio]. 2020;.

30. Karlen D. Characterizing the spread of CoViD-19. arXiv:200707156 [physics, q-bio, stat]. 2020;.

31. Dehning J, Zierenberg J, Spitzner FP, Wibral M, Neto JP, Wilczek M, et al. Inferring change points in the spread of COVID-19 reveals the effectiveness of interventions. Science. 2020;369(6500). doi:10.1126/science.abb9789.

32. Smith TP, Flaxman S, Gallinat AS, Kinosian SP, Stemkovski M, Unwin HJT, et al. Temperature and population density influence SARS-CoV-2 transmission in the absence of nonpharmaceutical interventions. PNAS. 2021;118(25). doi:10.1073/pnas.2019284118.

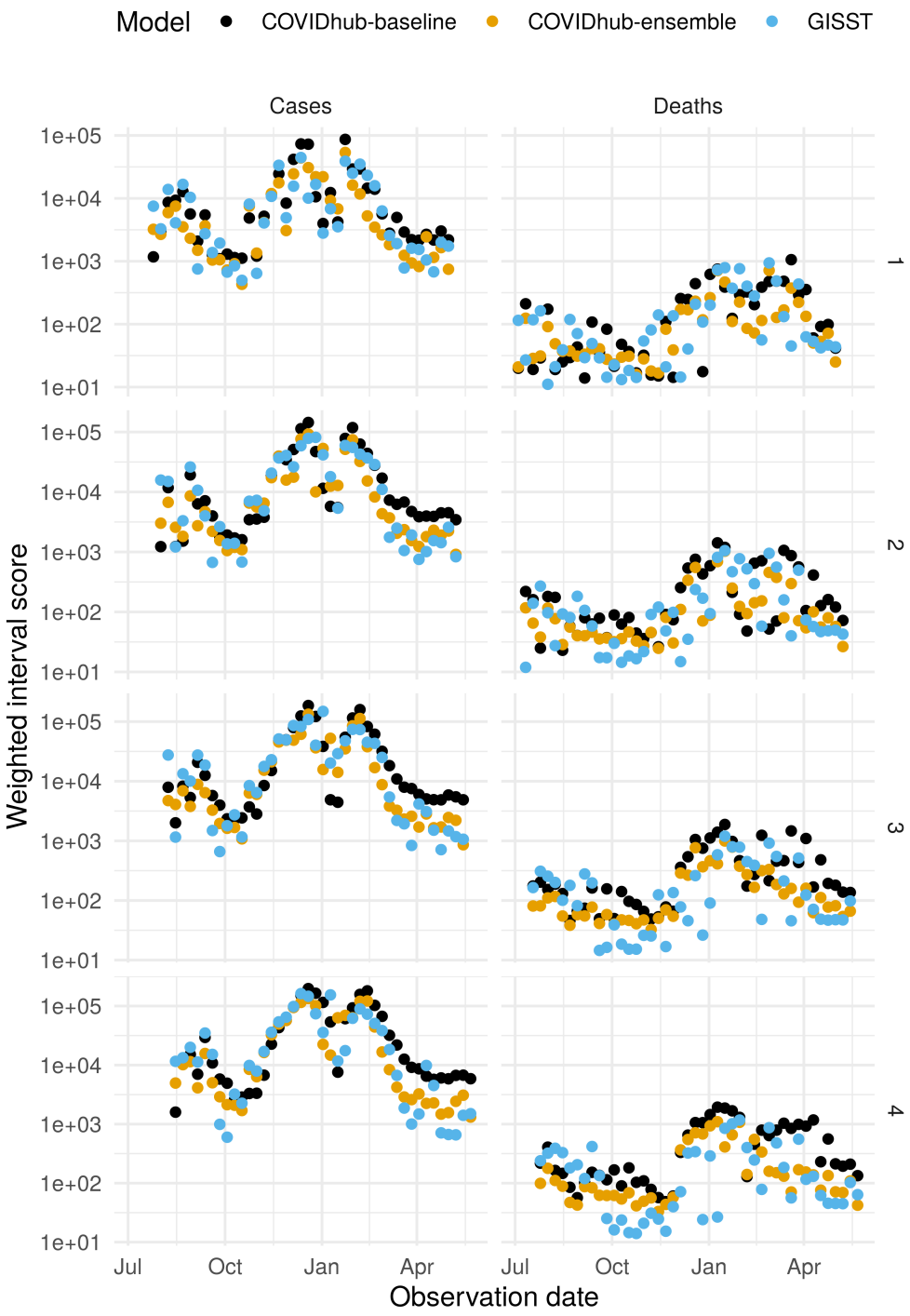

Fig S7. Performance of short-term forecasts of cases and deaths by date. The panel labels at right are the number of weeks in the forecast horizon. Lower scores indicate better performance.

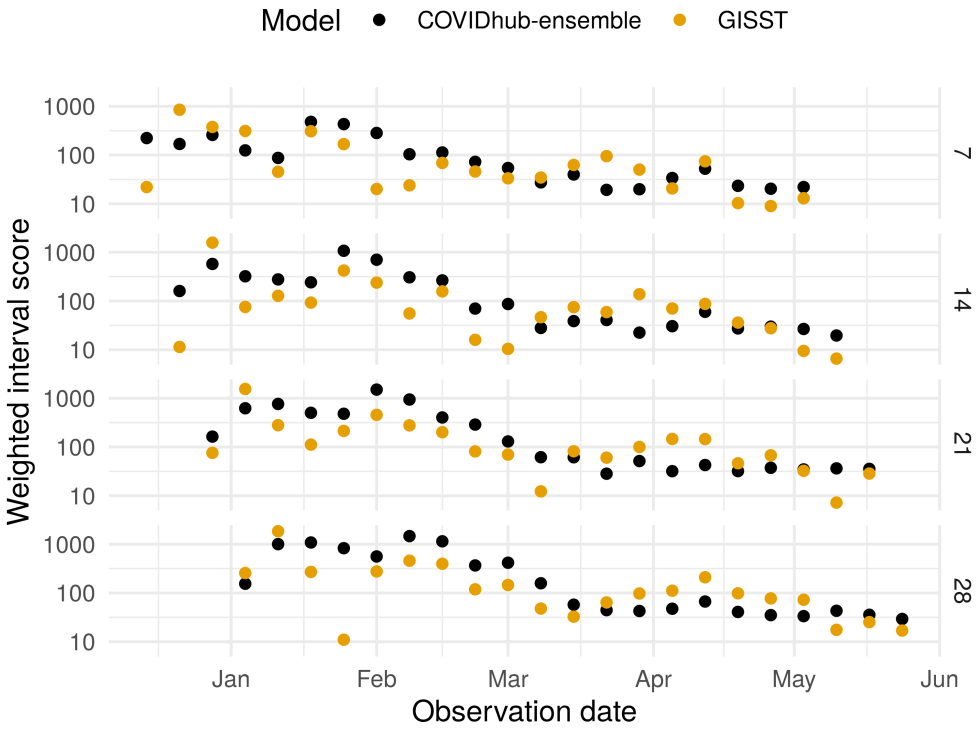

Fig S8. Performance of short-term forecasts of hospital admissions by date for selected horizons. The panel labels at right are the number of days in the forecast horizon. Lower scores indicate better performance.

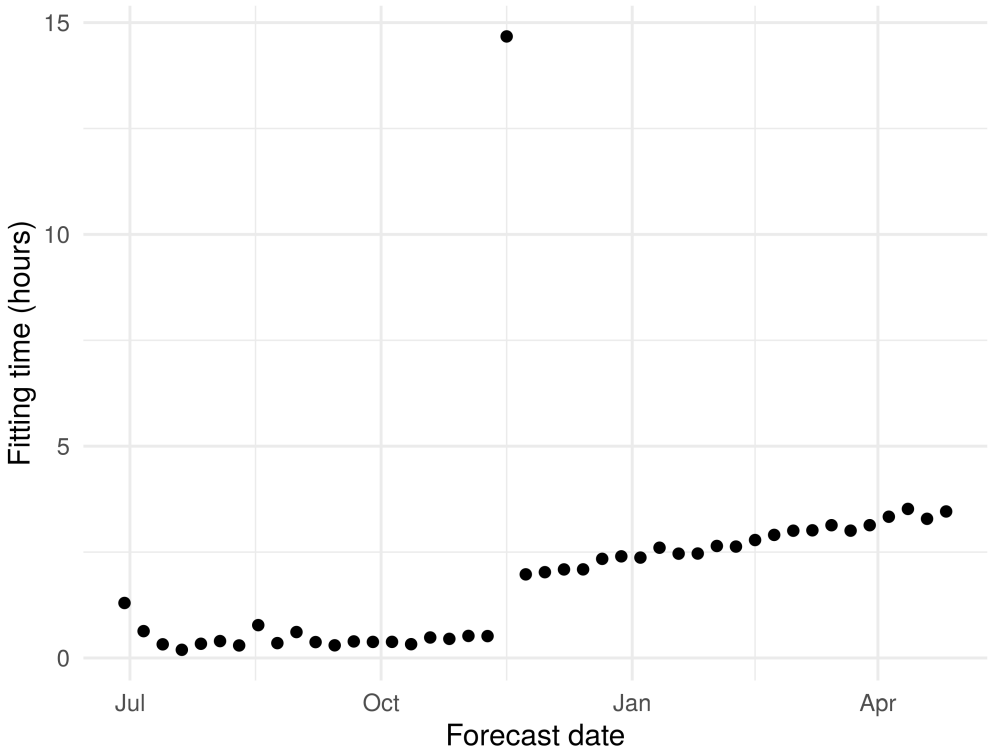

Fig S9. Time required to estimate GISST parameters. The spike in November was due to additional iterations due to the addition of hospital admissions data.