

# Deploying COVID-19 Case Forecasting Models in the Developing World

Sansiddh Jain<sup>1</sup>, Avtansh Tiwari<sup>1</sup>, Nayana Bannur<sup>1</sup>, Ayush Deva<sup>1</sup>, Siddhant Shingi<sup>1</sup>, Vishwa Shah<sup>1</sup>, Mihir Kulkarni<sup>1</sup>, Namrata Dekka<sup>1</sup>, Keshav Ramaswami<sup>1</sup>, Vasudha Khare<sup>1</sup>, Harsh Maheshwari<sup>2</sup>, Soma Dhavala<sup>1</sup>, Jithin Sreedharan<sup>1</sup>, Jerome White<sup>1</sup>, Srujana Merugu<sup>1</sup>, and Alpan Raval<sup>1\*</sup>

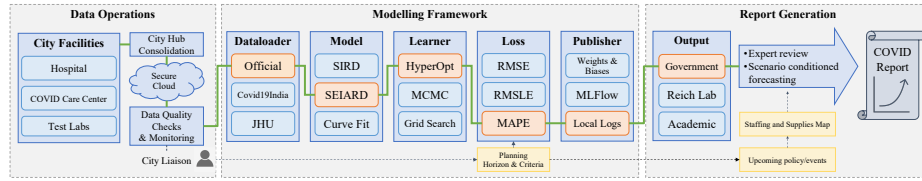
<sup>1</sup> Wadhvani Institute for Artificial Intelligence  
{sansiddh,avtansh}@wadhvaniai.org,  
{nayana.bannur99,ayushdeva97,sshingi19,vshah0350}@gmail.com,  
mihir@wadhvaniai.org, dnamrata@cs.ubc.ca,  
{ramasawami.keshav,khare.vas}@gmail.com,  
{soma,jithin,jerome}@wadhvaniai.org, smerugu@amazon.com,  
alpan@wadhvaniai.org

<sup>2</sup> Flipkart  
harshm121@gmail.com  
\*Corresponding author

**Abstract.** Forecasting infection case counts and estimating accurate epidemiological parameters are critical components of managing the response to a pandemic. This paper describes a modular, extensible framework for a COVID-19 forecasting system, primarily deployed during the first COVID wave in Mumbai and Jharkhand, India. We employ a variant of the SEIR compartmental model motivated by the nature of the available data and operational constraints. We estimate best-fit parameters using Sequential Model-Based Optimization (SMBO), and describe the use of a novel, fast and approximate Bayesian model averaging method (ABMA) for parameter uncertainty estimation that compares well with a more rigorous Markov Chain Monte Carlo (MCMC) approach in practice. We address on-the-ground deployment challenges such as spikes in the reported input data using a novel weighted smoothing method. We describe extensive empirical analyses to evaluate the accuracy of our method on ground truth as well as against other state-of-the-art approaches. Finally, we outline deployment lessons and describe how inferred model parameters were used by government partners to interpret the state of the epidemic and how model forecasts were used to estimate staffing and planning needs essential for addressing COVID-19 hospital burden.

## 1 Introduction

The COVID-19 pandemic has spurred intense interest in epidemiological forecasting models. The need for robust solutions has been especially pressing in



**Fig. 1: COVID forecasting framework.** The three-phase pipeline consists of a) *data operations*: working with government officials to consolidate and store data collected from various sources throughout the city, b) a modular *modelling framework* for fitting and forecasting, and c) *report generation* of customized forecasts for clients. The specific components for deployment with government partners are highlighted in orange and connected by a green line delineating the workflow.

dense populations across the developing world, with their limited health resource availability and long lead times for addressing shortfalls.

To reduce mortality, it is important to ensure that critical health care resources are available at the necessary capacity in a timely manner. There is therefore a need for forecasting reported infections at a local level to inform capacity planning, model the effects of policy changes (such as changes in lockdown levels, reduced mobility, and testing), and prepare for potential scenarios. In this chapter, we describe a deployed epidemiological forecasting framework, built with public health requirements in mind, that was used in Mumbai, India, one of the world’s most densely populated cities, as well as in other resource-constrained regions such as the state of Jharkhand, India, during the first COVID infection wave.

The framework, schematically displayed in Figure 1 (see Section 4 for a detailed explanation), consumes aggregate case count data culled by government officials from various types of health facilities, pipes it through a configurable modeling framework, and outputs burden estimates that were used by relevant authorities for subsequent planning of personnel and supplies. Partners for data and for usage of the solution were the Brihanmumbai Municipal Corporation and the Integrated Disease Surveillance Programme, at Mumbai and Jharkhand respectively.

Our main contributions, described in subsequent sections, are as follows.

*System and Process Design.* [Section 4]. We built a practical, modular, and extensible learning-based epidemic case forecasting system that is customizable to individual regions and application scenarios. The system consists of modules for data ingestion, preprocessing and exploratory analysis, model fitting, scenario-conditioned forecasting, and application-specific report generation.

*Modeling Methodology.* [Sections 5 and 6]. We developed and applied techniques for model and loss-agnostic estimation of parameters via sequential model based optimization (SMBO) [6]. Combining SMBO sampling with Bayesian model averaging enables fast approximate quantification of forecast uncertainty.

We found that this is empirically comparable to a more rigorous Markov Chain Monte Carlo (MCMC) approach but computationally faster. We developed smoothing methods to handle data issues arising out of reporting delays.

*Epidemiological Model Choice.* [Section 5]. We surmised that conditions of interpretability and identifiability warrant a simple variant of the Susceptible-Exposed-Infectious-Removed (SEIR) compartmental model especially when only confirmed, active, recovered and deceased case counts are observed, as was the case in our deployment scenario.

*Empirical Results.* [Section 6]. We carried out extensive empirical analyses detailing the optimization of relevant hyperparameters, field predictive performance for the city of Mumbai, and comparison with other state-of-the-art models hosted by Reich Lab [32].

Successful deployment of epidemiological models requires addressing various operational challenges inherent in the public health landscape. Some of these challenges are:

*Data limitations.* Data collection and management activities during a pandemic are highly impacted due to severe demands on public health authorities, leading to data discrepancies in the form of duplicate records, delayed and erroneous reporting, and incomplete and inconsistent data. In the face of these data challenges, it is important to choose model families whose complexity is commensurate with that of the available data.

*Need for rich insights and what-if-scenarios.* The choice and implementation of correct policy is often predicated not just on expected case counts but also on changes in underlying epidemiological parameters. It is therefore important to construct models that are highly interpretable with independently verifiable parameters.

*Dynamic user requirements.* Forecasting needs are highly dynamic as the epidemic progresses. In the early stages, when containment is critical, accurate regional forecasts for short time horizons are valuable. In later stages, the emphasis shifts to estimation of hospital burden and to uncertainty around those projections. A robust yet flexible modeling framework that can support multiple application-specific needs is thus essential.

These challenges led to the development of a forecasting framework that can support data-driven parameter estimation and uncertainty quantification for interpretable models. Such models cannot be directly learned using gradient-based learning approaches. Further, variations in data semantics and user requirements necessitated a solution approach that is agnostic to the choice of model class and application-specific prediction quality criteria. These considerations motivated our work.

## 2 Problem Statement

The general epidemiological model parameter estimation problem can be stated as follows. For a given spatial region, let  $X[t]$  denote a multi-variate time series of case counts corresponding to different demographic groups and disease stages.

Let  $M_{\theta}$  be a black-box epidemiological forecasting model parameterized by  $\theta \in \Theta \subseteq \mathbb{R}^k$  that generates a future forecast from historical observations. Given a time series loss function  $L$ , the objective is to determine the optimal parameters  $\theta^*$  such that the loss between the predicted and observed data over a specified time horizon  $[t_i, t_j]$  is minimized:

$$\theta^* = \underset{\theta \in \Theta}{\operatorname{argmin}} L(\theta) = \underset{\theta \in \Theta}{\operatorname{argmin}} L\left(X[t_i : t_j], \hat{X}_{\theta}[t_i : t_j]\right), \quad (1)$$

where  $\hat{X}_{\theta}[t_i : t_j] = M_{\theta}(X[0 : t_i - 1])$ . To estimate parameter uncertainty, it is also necessary to generate the resulting forecast distribution over the time horizon.

In our deployment scenario, the primary data comprised counts of *confirmed*, *active*, *recovered*, and *deceased* cases. For certain regions, severity-stratified counts and testing data were also available. Reliable counts of in-flow and out-flow cases and within-region mobility indicators were rarely available. While public health partner requirements changed over time, they were largely focused on hospital burden estimation over a horizon of 30–45 days. The main need was to ensure adequate medical resources while minimizing excess capacity that would remain under-utilized. A readily interpretable relative accuracy criterion such as Mean Absolute Percentage Error (MAPE) in primary case count forecasts was well-aligned with our partner needs, but the relative importance of the four different case counts could vary.

### 3 Related Work

The work presented in this chapter builds on four areas of epidemiological modeling: a) forecasting, b) parameter estimation (with uncertainty), c) modeling with data limitations, and d) public health deployment.

#### 3.1 Epidemiological Forecasting

The COVID-19 pandemic and associated global forecasting challenges [25, 32] have spurred new research on modeling infectious disease spread. There are three broad classes of models.

(a) *Compartmental models* assume that individuals in a population at any given time are assigned to one of several states, called compartments, and move between these compartments. Examples include the SIR [26] and SEIR [22] models. Over the last year, variants of the SEIR model have been widely used to study the COVID-19 pandemic project hospital burden [37, 29, 38], incorporating aspects such as age-stratification [1, 21], asymptomatic transmission [35], and effects of social distancing measures [18, 7].

(b) *Agent-based models* [19, 39] simulate interactions and disease stage transitions of individual agents.

(c) *Curve-fitting models* fit parameterized curves to data. Examples include the exponential growth model and the IHME CurveFit model [23].

### 3.2 Model Parameters and Uncertainty Estimation

Multiple works address the problem of epidemiological parameter estimation with forecast uncertainty [17, 10, 12, 8], but these techniques are usually specific to the model class and involve assumptions on the generative process. Recently, there has been increased focus on the practical aspects of model fitting such as choice of training duration [30] and identifiability issues [33, 28, 15]. Model-agnostic evaluation of forecasts is another related area of interest [36, 32, 20]. While our deployment was based on compartmental models, the proposed techniques are model-agnostic and can be adapted to any application-specific loss function.

### 3.3 Modeling with Data Limitations

Epidemiological modeling in the developing world is plagued with data paucity and quality issues. Various studies have focused on understanding transmission dynamics in such limited data settings [34, 27, 3]. In our current work, we discuss some of these challenges and possible resolution via appropriate model choices and data preprocessing.

### 3.4 Public Health Deployment

Practical use of epidemiological models in public health response requires a holistic view of government priorities, policy levers, and processes. Work in economic epidemiology [31, 9] is targeted towards supporting decision-making related to interventions and policy choices. Multiple organizations [32, 13] currently share automated COVID case forecasts with relevant public health authorities, but the forecasts are not always customized for decision-making. Our deployment involved a two-way partnership with the government, providing precise capacity planning guidance and insights on the pandemic dynamics per requirements.

## 4 Resource Requirements and Solution Framework

This section provides an overview of our end-to-end epidemic forecasting system, including data inputs and reported outputs, as visualized in Figure 1.

### 4.1 Data Operations

Partnership with government and official data access was enabled in the early days of the pandemic by providing proof-of-concept forecasts that initially used daily time series of confirmed (C), active (A), recovered (R), and deceased (D) case counts. These were obtained from public sources such as COVID19India [14] and JHU [11] for regions in India and JHU data for 45 US states in order to benchmark our approach against other methods. We later obtained the same

type of time series data directly through official government sources in Mumbai city and in the state of Jharkhand, India.

Data schema for aggregate case counts was defined to facilitate automated forecasting. However, realisation of this schema across upstream providers (hospitals and health centers dedicated to COVID care) proved to be a challenge. Issues faced included those described as some of the challenges in Section 1: duplicate records, data entry errors, inconsistencies in schema across health facilities, and errors in reporting transitions between care levels and between disease stages. We therefore developed point-of-entry validation, along with partial automation of data consolidation and de-duplication. This semi-automated process was aided by sharing spreadsheets with partners on our cloud storage. Our pipeline performed quality checks and anonymized sensitive data before exposing it to the modelling team via an SQL-compliant database.<sup>3</sup> We also maintained descriptive data visualizations to monitor abnormal patterns. This pipeline was later replicated across partners.

## 4.2 Modelling Framework

In order to flexibly support multiple forms of case count data, epidemiological models, and loss functions, our modelling framework has a modular design with five high-level components, each adapted to multivariate time-series forecasting: data loaders, models, learners, loss functions, and publishers.<sup>4</sup> This extensible and scalable framework allows rapid experimentation with different modelling techniques and data sources of varying complexity, (e.g., with or without severity-stratified case counts), as well as a platform for comparing results. Examples of these components are shown in Figure 1. Modelling details are discussed in Section 5.

## 4.3 Report Generation

The content and format of forecast presentations vary depending on the audience. Our solution framework allowed for model outputs to be consumed in a variety of ways: as time-series forecasts, planning reports (for public authorities) or data pushes (for submissions to Reich Lab (Section 6.5)).

Forecasts were either routine (for monthly planning) or *ad hoc* (to address urgent demands and rapidly evolving scenarios). Forecast requests, compiled by our team members physically located in COVID ‘war rooms’, included information on recent changes in mobility, expected changes in policy over the planning horizon, and notes on data. All of this information was used to appropriately define plausible parameter ranges as priors. Forecasts generated in response to the requests were organized into deciles of the forecast distribution (anchored on a future date determined by the planning horizon). Specific scenarios (mapped

<sup>3</sup> Within the Amazon Web Service framework: Lambda functions for ingesting and cleaning data, Glue for collating data, Athena for exposing data.

<sup>4</sup> Publishers log to ML tracking platforms such as MLFlow and Weights & Biases.

to deciles of the forecast distribution) were selected to represent an appropriate ‘planning scenario’, a ‘low case scenario’ and a ‘high case scenario’. These interactions not only allowed public authorities to understand the state of the pandemic through changes in model parameters, but also led to the consideration of information unavailable to the model (expected changes to testing strategy, treatment policy, and non-pharmaceutical interventions such as shelter-in-place orders).

In the capacity computation module, projected active case numbers and trends in utilization (occupancy rates and length of stay) were used to estimate facility-level demand for personnel, quarantine beds, oxygen-supported beds, and intensive care units. These estimates were distilled into final recommendations consisting of projected cases for a 30–45 day horizon and potential shortfalls. Based on health resource capacity proportions shared by government partners (see Figure 12), the estimates were also used to recommend increases in capacity.

## 5 Methodology

This section describes our approach to parameter estimation and uncertainty quantification. We also discuss smoothing strategies for handling data spikes and the choice of epidemiological model used in deployment.

### 5.1 Parameter Fitting with SMBO

Given the nonlinear nature of the optimisation problem of Equation 1, we employ a Sequential Model-Based Optimization (SMBO) method to explore the parameter searchspace. We use the Expected Improvement (EI) criterion [24] to return the best parameter set at each iteration and the Tree-structured Parzen Estimator (TPE) [5] to model the loss. We use the Hyperopt implementation of TPE [6]. Note that this black box optimization method is model-agnostic and applies to any zeroth order optimization problem.

### 5.2 Estimation of Forecast Uncertainty

We estimate parameter uncertainty by approximating the posterior distribution of the parameters given the data. Given a dataset  $X$  and loss function  $L(\theta)$  as defined in Equation 1, we assume that the likelihood function is tuned to the loss function, as given by

$$P(X|\theta) \propto e^{-\alpha L(\theta)}, \quad (2)$$

where  $\alpha$  is a concentration parameter to be estimated as described below. When the loss function  $L(\cdot)$  is a regular Bregman divergence, the distribution corresponds to a uniquely determined exponential family [4, 2]. Choosing the parameter prior  $P(\theta)$  to be from the conjugate prior family of the resulting exponential family yields a posterior distribution of the form

$$P(\theta|X) = e^{-\alpha L(\theta)} h(\theta) P(\theta), \quad (3)$$

where  $h(\boldsymbol{\theta})$  is uniquely determined by the loss function [16].

Expected values under the posterior distribution can be computed by first sampling from the prior  $P(\boldsymbol{\theta})$  with appropriate re-weighting, and then computing importance sampling estimates. Thus, given  $n$  parameter samples  $\{\boldsymbol{\theta}_i\}_{i=1}^n$  from the reweighted prior, and the corresponding loss function value  $L(\boldsymbol{\theta}_i)$  for each sample, the expected value of any function  $g(\boldsymbol{\theta})$  is estimated as

$$E[g(\boldsymbol{\theta})] \approx \frac{\sum_{i=1}^n e^{-\alpha L(\boldsymbol{\theta}_i)} \cdot g(\boldsymbol{\theta}_i)}{\sum_{i=1}^n e^{-\alpha L(\boldsymbol{\theta}_i)}}. \quad (4)$$

**Approximate Bayesian Model Averaging (ABMA).** A limitation of the above approach is that importance sampling with a broad prior as proposal distribution may require a prohibitively large number of samples. We therefore considered an approximation, hereafter referred to as Approximate Bayesian Model Averaging (ABMA), where we apply Equation 4 not to samples from the prior, but rather to the sequential samples obtained through SMBO. The intuition behind doing so is that these samples represent regions close to the optimal parameter values where the loss is likely to be small, and hence should capture the dominant behavior in sample averages like Equation 4. Additionally, the SMBO implementation is computationally efficient.

In Section 6, we show via comparison with an MCMC approach that this is a reasonable approximation. The term *ensemble mean*, used throughout this paper in the context of parameter and forecast averaging, refers to taking the sample means according to Equation 4.

In the application of ABMA, it remains to fix the values of  $\alpha$  and  $n$ . We estimate  $\alpha$  by minimizing the ensemble mean forecast and the ground truth on a validation time interval. To fix  $n$ , we just choose a value of  $n$  large enough such that the mean and variance of the parameter samples, computed through application of Equation 4, converges.

**MCMC-based Estimation of Parameter Uncertainty.** A more rigorous method to sample parameter values is via Markov Chain Monte Carlo (MCMC) sampling from the posterior distribution of the parameters given the data, i.e., from  $\mathbf{P}(\boldsymbol{\theta}|X)$ . Assuming an appropriate prior over  $\boldsymbol{\theta}$  and a likelihood function  $P(X | \boldsymbol{\theta}, s)$ , where  $s$  denotes additional parameters in the likelihood function (analogous to  $\alpha$  in Equation 2), we may consider the larger problem of sampling from the extended posterior distribution

$$P(\boldsymbol{\theta}, s | X) \propto P(X | \boldsymbol{\theta}, s) P(\boldsymbol{\theta}) P(s), \quad (5)$$

where we assume  $\boldsymbol{\theta}$  and  $s$  have independent priors. This larger sampling problem is solved by a combination of MCMC and Gibbs sampling steps, using a truncated Normal distribution  $Q(\cdot)$  (truncated at the boundaries of the search space) as a proposal distribution for MCMC moves. We choose an initial  $s_0, \boldsymbol{\theta}_0$ , and then employ the following Metropolis-within-Gibbs procedure to alternately sample from the posterior distributions  $P(\boldsymbol{\theta} | X, s)$  and  $P(s | X, \boldsymbol{\theta})$ .



- 1) *MCMC sampling with target distribution  $P(\boldsymbol{\theta} \mid X, s)$ .* Given samples  $s_{k-1}$ ,  $\boldsymbol{\theta}_{k-1}$  from iteration  $k-1$ , sample a new  $\boldsymbol{\theta}_k$  according to the Metropolis criterion, using the proposal distribution  $Q(\boldsymbol{\theta}_k \mid \boldsymbol{\theta}_{k-1})$ .
- 2) *Sampling  $s$  from the posterior distribution  $P(s \mid X, \boldsymbol{\theta})$ .* Given a  $\boldsymbol{\theta}_k$  sampled in the previous step, sample  $s_k \sim P(s \mid X, \boldsymbol{\theta}_k)$ . This sampling can be made straightforward by choosing a prior for  $s$  that is conjugate to the likelihood distribution.

The steps above together constitute the Gibbs sampling updates, and are repeated alternately until convergence of the Markov chain. This eventually results in a sample from the extended distribution  $P(\boldsymbol{\theta}, s \mid X)$ .

**Estimating Quantiles for Forecasted Case Count.** Given an ensemble of ABMA parameter samples  $\{\boldsymbol{\theta}_i\}$ , one can compute forecast trajectories for each parameter set in the sample. For a forecast trajectory  $\hat{X}_{\boldsymbol{\theta}_i}(t)$ , the effective probability density function (pdf) is proportional to  $e^{-\alpha L(\boldsymbol{\theta}_i)}$ . The forecast quantiles on a given day  $t^*$  are obtained by sorting the forecasted case counts corresponding to different parameter sets on day  $t^*$  and using the cumulative distribution function (CDF) to find the quantile of interest.

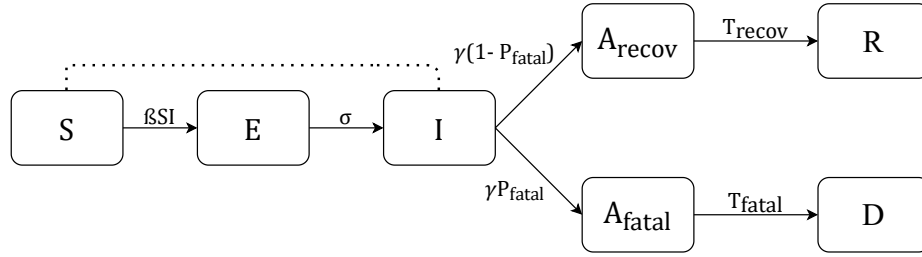
This procedure has the desirable property that a trajectory at a fixed quantile value corresponds to a single parameter set. However, forecast trajectories for different parameter sets can cross in time. Therefore quantiles computed at one time point are generally not preserved in time. A work-around is to re-estimate quantiles on each day and piece together a “fixed quantile” trajectory using the forecasted values for that quantile on each day. A trajectory so constructed, however, does not correspond to a single set of epidemiological parameters and is therefore difficult to interpret. We use the former method for the sake of interpretability in our deployment and specify the  $t^*$  at which quantiles are computed, while the latter method is used for “fixed quantile” forecast comparisons.

### 5.3 Data Spike Smoothing

One of the practical issues we faced was having to account for spikes in case count data. These spikes originate in delayed reporting of test results and delayed tracking of recoveries and deaths, thus leading to accumulated occurrences from the past being reported on a single day. Such spikes need to be smoothed by distributing them several days into the past, even though actual attributed dates were unavailable. We implemented a smoothing technique that carries out this redistribution under some assumptions.

Consider a spike in cumulative case counts  $X$  on day  $t_e$ , given by  $\Delta = X[t_e] - X[t_e-1]$  that can be attributed to delayed activity since day  $t_b$ . We construct a smoothed time series by redistributing  $\Delta$  as follows:

$$X_{\text{smooth}}[t] = \left( \sum_{j=t_b}^t w[j] \right) \Delta + X[t], \quad (6)$$



**Fig. 2: SEIARD Model Structure.** Individuals transition between states of disease progression: Susceptible (**S**), Exposed (**E**), Infectious (**I**), Recovered (**R**), and Deceased (**D**). Active cases are split between **A<sub>recov</sub>** and **A<sub>fatal</sub>**. Details in Section A.3.

where  $t \in \{t_b, \dots, t_{e-1}\}$  and  $\sum_{t=t_b}^{t_e} w[t] = 1$ . The weights  $w[t]$  can be computed in different ways:

- Uniform :  $w[t]$  constant.
- Proportional increments :  $w[t] \propto X[t] - X[t-1]$
- Proportional counts:  $w[t] \propto X[t]$

Note that a spike in one compartment could correspond to a compensatory dip or a spike in a different compartment such that the total case count stays constant. For example, a spike in recoveries must be compensated by a dip in active cases, since recoveries deplete active cases. Thus, over the duration of smoothing, the active cases are adjusted by additional recoveries, resulting in a smoothed version of active cases. The model is then fitted on the smoothed case time series rather than on the raw counts.

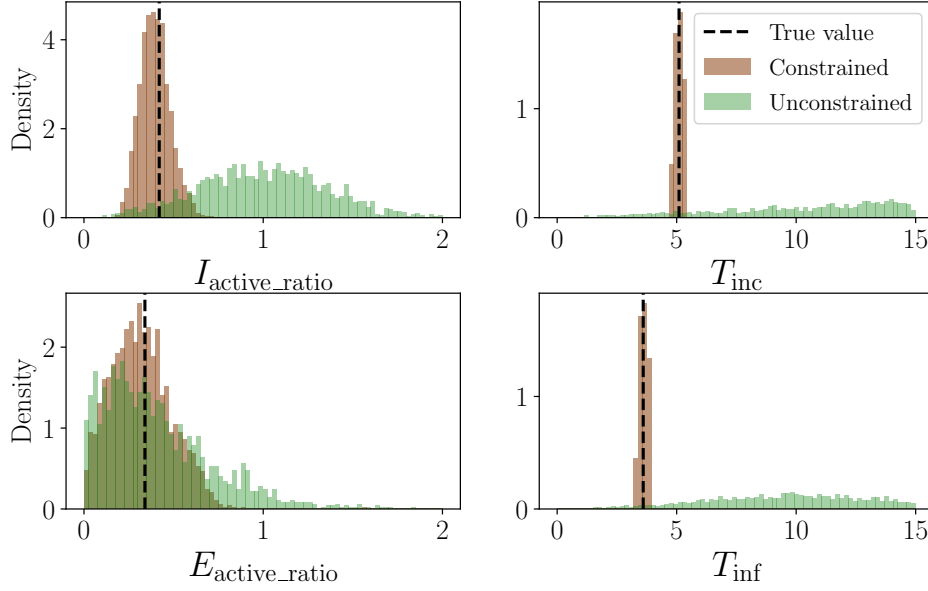
#### 5.4 Epidemiological Model Structure

In this section, we describe the motivation and details of the specific choice of epidemiological model used in deployment. As mentioned earlier, the modeling framework is agnostic to this choice.

**Model Choice Criteria.** The choice of epidemiological model was driven by four main criteria:

- *Expressivity*, or the ability to faithfully capture the disease dynamics,
- *Learnability* of parameters conditioned on the available data,
- *Interpretability* in order to understand the evolution of the pandemic,
- *Generalizability* to future scenarios by incorporating additional information.

We observed that simple exponential growth based models that do not account for decreasing susceptibility and finite populations became inaccurate as the



**Fig. 3: Identifiability of parameters in the SEIARD model.** Latent ratio posterior distributions when fitted on synthetic data generated by the true parameter values shown. Distributions were estimated using the MCMC method of Section 5.2 under a broad prior as well as under fixed choices of  $T_{\text{inc}}$ ,  $T_{\text{inf}}$ , and  $P_{\text{fatal}}$ .

pandemic progressed. Agent-based models [19, 39], on the other hand, are over-parameterized, with a low level of learnability. While curve-fit models and compartment models had similar levels of accuracy, the parameters of the latter were readily interpretable;<sup>5</sup> compartmental models, therefore, represented a natural choice.

**SEIARD Model.** We chose a simple compartmental model that only included compartments for primary case counts or those essential for expressivity. Accessible case count data typically comprised cumulative counts of confirmed cases ( $C$ ), active cases ( $A$ ), recovered cases ( $R$ ), and deceased cases ( $D$ ) with  $C = A + R + D$ . We assumed confirmed cases to be post-infectious owing to strict quarantining measures. We explicitly modeled the latent “incubation” and “infectious” stages that occur prior to detection so as to allow what-if scenario modeling with changes in isolation and testing policies. Figure 2 depicts the structure of this SEIARD model (dynamical equations are given in the Appendix). The population is split into compartments  $[\mathbf{S}, \mathbf{E}, \mathbf{I}, \mathbf{A}_{\text{recov}}, \mathbf{A}_{\text{fatal}}, \mathbf{R}, \mathbf{D}]$ . The states

<sup>5</sup> Application of compartmental models to a few Indian districts led to estimates of recovery period exceeding 50 days, leading to the discovery of a serious data reporting gap. Interpretability of these parameters, therefore, turned out to be important.

**S** (Susceptible), **E** (Exposed), and **I** (Infectious), and the associated parameters (transmission rate  $\beta$ , incubation period  $T_{\text{inc}} = \sigma^{-1}$ , and the infectious period  $T_{\text{inf}} = \gamma^{-1}$ ) are defined as in the classic SEIR model [22]. An individual who tests positive moves to a post-infectious but active phase, which is split into  $\mathbf{A}_{\text{fatal}}$  and  $\mathbf{A}_{\text{recov}}$  based on the eventual outcome: fatality (**D**) with probability  $P_{\text{fatal}}$  or recovery (**R**).  $T_{\text{fatal}}$  and  $T_{\text{recov}}$  are the mean durations from detection to eventual death or recovery respectively. Observed case count data is mapped to compartment populations as follows:  $R = |\mathbf{R}|$ ,  $D = |\mathbf{D}|$ ,  $A = |\mathbf{A}_{\text{recov}}| + |\mathbf{A}_{\text{fatal}}|$ .

Interpretable parameters such as the reproduction number  $\mathcal{R}_0 = \beta/\gamma$ ,  $T_{\text{recov}}$ , and  $T_{\text{fatal}}$  provide an understanding of the pandemic evolution and reporting gaps, and allow scenario-conditioned forecasting. Changes in the recovery policy and upcoming events such as festivals can be simulated by appropriate adjustment of  $T_{\text{recov}}$  and  $\beta$ .

**Parameter Fitting Considerations.** Estimating the parameters of this model from case counts posed several challenges:

- *Parameter drift due to changes in social distancing policy.* This drift necessitated refitting the model to a recent time window for every forecast round.
- *Unknown initial values of latent compartments<sup>6</sup>.* Initial values of **E** and **I** could not be set to 0 because the fitting time interval could occur at an intermediate stage of the pandemic. We treat these latent variables as additional parameters to be optimized during model fitting. More precisely, we consider the ratios of these variables with respect to the initial active counts (at the start of the fitting interval  $[t_i, t_j]$ ) as latent parameters:  $E_{\text{active\_ratio}} = E[t_i]/A[t_i]$  and  $I_{\text{active\_ratio}} = I[t_i]/A[t_i]$ .
- *Modeling undetected cases.* While undetected cases are not explicitly modeled, they are partially accounted for by letting  $E[t_i]$  and  $I[t_i]$  to be free parameters whose fitted values adjust to accommodate the effect of undetected cases.
- *Potential lack of identifiability of key parameters.* This issue was addressed by incorporating existing domain knowledge about  $T_{\text{inc}}$ ,  $T_{\text{inf}}$ , and  $P_{\text{fatal}}$  from raw (possibly incomplete) case line lists, whenever available (Figure 3; see also [15] for a deeper analysis of identifiability in this type of model).

## 6 Results and Field Evaluation

To demonstrate the value of the proposed framework, we report (a) empirical results validating choices of hyperparameters, uncertainty estimation, and data preprocessing, (b) field performance and impact of the deployed system in Mumbai, and (c) empirical comparison of our approach with other state-of-the-art (SOTA) models in Reich Lab on COVID-19 case data from the US. The main elements of our experiment setup are:

<sup>6</sup>  $S$  is not considered a latent variable since the net population can be assumed to be nearly constant.

- *Data*. We consider time series of confirmed ( $C$ ), active ( $A$ ), reported ( $R$ ), and deceased ( $D$ ) cases. Results shown are based on real case count data from the city of Mumbai except for the hyperparameter optimization discussion (COVID19India data) and the comparison with Reich Lab models (JHU data on 45 US regions).
- *Algorithms*. Various alternatives within our framework (ABMA vs. MCMC, smoothing variants) as well as 26 different models from Reich Lab were considered.
- *Evaluation Metric*. We use Mean Absolute Percentage Error (MAPE) for each of the four variables ( $C, A, R, D$ ), as well as aggregated over the compartments in a suitably weighted manner, as the primary metrics for evaluation and comparisons. Uncertainty estimation is discussed in terms of credible intervals.

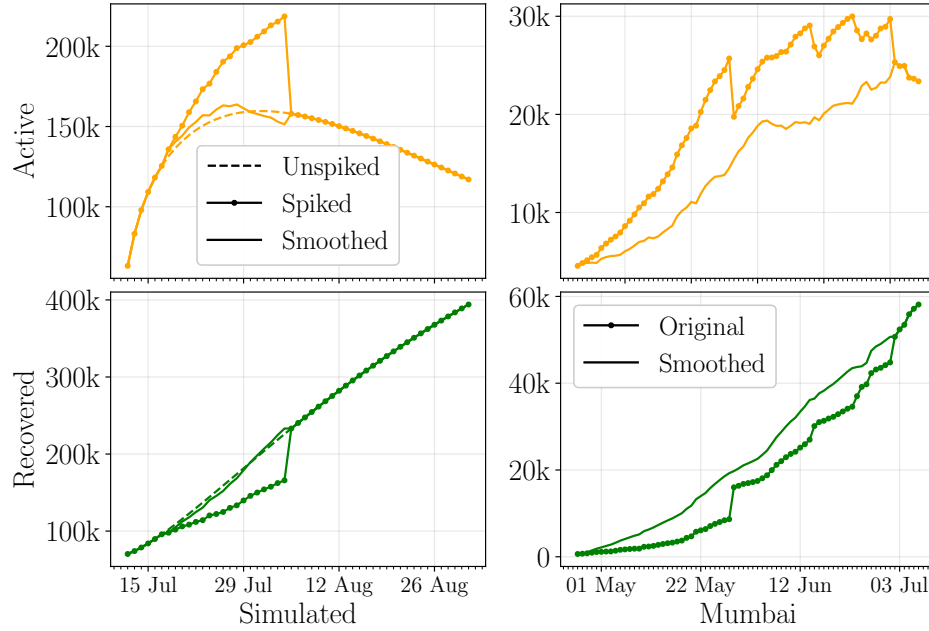
### 6.1 Data Spike Smoothing

Our experiments indicated that smoothing (in particular, the Proportional count variant) leads to better prediction of future cases. We evaluated the smoothing variants discussed in Section 5.3 on simulated incident (new) raw data  $X_{\text{inc}}[t]$ , with a simulated spike caused due to delays in data reporting between dates  $t_b$  and  $t_e$ . In order to mimic on-the-ground reporting discrepancies, we assume that only a fraction  $\delta_t$  of the true incident cases  $X_{\text{inc}}[t]$  on day  $t$  are actually reported on that day, i.e.,  $X[t] = X[t-1] + \delta_t X_{\text{inc}}[t]$ . The remaining  $(1 - \delta_t)X_{\text{inc}}[t]$  cases are reported on the end date  $t_e$ , resulting in a case spike on day  $t_e$ , i.e.,  $X[t_e] = X[t_e - 1] + X_{\text{inc}}[t_e] + \sum_{t=t_b}^{t_e-1} (1 - \delta_t) \cdot X_{\text{inc}}[t]$ . For the simulation, we assume that the  $\delta_t$  are i.i.d. samples from a Beta distribution (here, chosen as **Beta**(2, 2)). Other parameters are chosen according to Table 4 in the Appendix. Smoothing methods are then applied to the simulated data generated after the dependent case counts are adjusted to sum to the confirmed cases.

Table 1 and Figure 4 show the results of smoothing experiments. Overall, we found that the “Proportional counts” smoothing method works best at reconstructing the raw data with reasonable accuracy.

**Table 1: Comparison of smoothing methods on simulated spiky data.** Spikes were generated as described in the text on 5000 simulated time series, generated via uniform sampling of parameter values in Table 4 (a, b) (Supplement). MAPE loss on Active ( $A$ ) and Recovered ( $R$ ) compartments was computed between the smoothed and the true, unspiked series.

Smoothing Method	MAPE (A)	MAPE (R)
Proportional counts	$4.84 \pm 0.26$	$2.51 \pm 0.018$
Proportional increments	$5.15 \pm 0.18$	$3.17 \pm 0.022$
Uniform	$5.16 \pm 0.40$	$2.53 \pm 0.018$



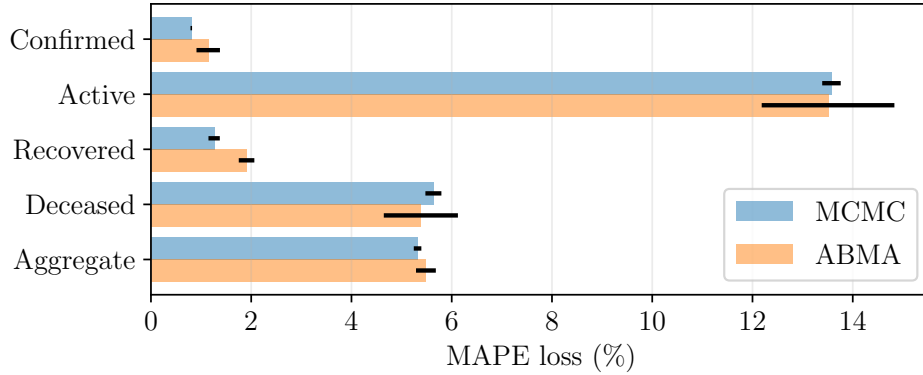
**Fig. 4: Performance of smoothing algorithm on simulated spike and Mumbai data.** Smoothing is done via the “Proportional counts” method.

## 6.2 Hyperparameter Optimization

Within our modeling framework, the following hyperparameters need to be tuned to achieve good performance: a)  $T_{\theta}$ , the length of the fitting period in days to estimate model parameters  $\theta$ , b)  $T_{\alpha}$ , the number of days of data used to tune the concentration parameter  $\alpha$ , and c)  $n$ , the number of trials used to optimise  $\theta$ . In addition, it was also necessary to identify a suitable loss function and compare the performance of the TPE to other fitting techniques. Details of these tuning experiments are given in the Appendix.

## 6.3 Parameter and Uncertainty Estimation: ABMA vs. MCMC

We compared the ABMA method of Section 5.2 to the MCMC sampling approach described there, along three dimensions: a) comparison of the lowest loss values achieved, b) comparison of case count forecasts, and c) distributional comparison of the sampled parameters. For these comparison experiments we chose a duration of 28 days for fitting, from 1 October to 28 October 2020, and a duration of 21 days, from 1 November to 21 November 2020, for evaluation. The three days in between were used for fitting  $\alpha$  in the ABMA procedure. The time taken to generate MCMC samples was noticeably higher than that for ABMA: under a controlled compute setting, ABMA was found to be an order of magnitude faster. This was one of the principal motivations behind the choice of ABMA.



**Fig. 5: Comparison of test losses for MCMC and ABMA.** MAPE losses for the ensemble mean forecast on the test period are shown for each compartment under the two methods. Error bars denote the standard error across ten runs.

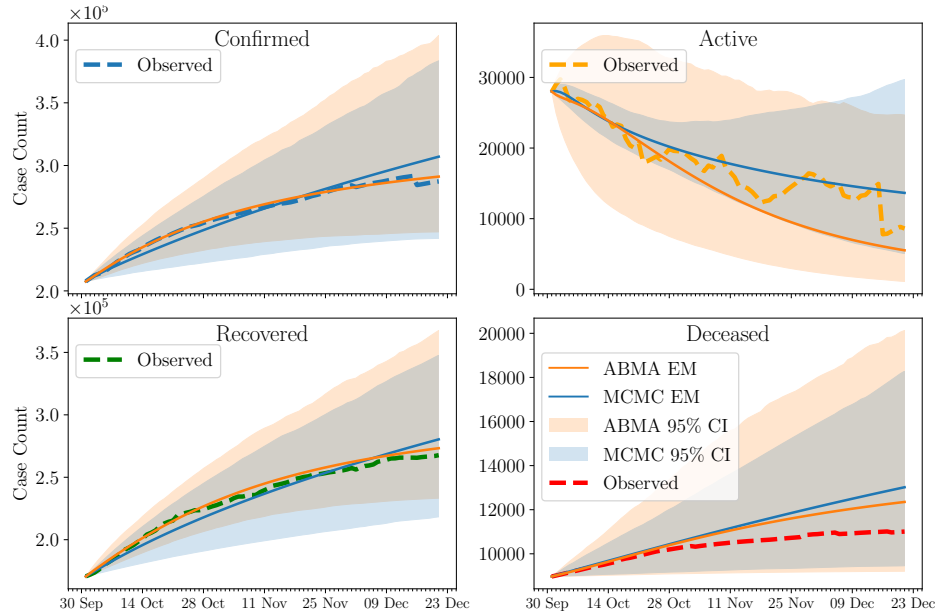
Details of MCMC sampling as well as comparison of sampled parameters are given in the Appendix.

**Comparison of the best loss values achieved.** We evaluated the MAPE losses across forecasted ensemble mean time series  $\hat{X}$  on the test duration, corresponding to parameters sampled by ABMA vs. MCMC algorithms (Figure 5). Note that for MCMC samples, the ensemble mean forecasts are generated by simply averaging the forecasts for the sampled parameter sets, while for ABMA samples, the ensemble mean forecasts are generated by weighting each forecast by its importance sampling factor. The figure illustrates that the differences between the MAPE loss values are minor when comparing both ensemble means on the test set.

**Comparison of case count forecasts.** We now turn to a comparison of the forecasts themselves. Figure 6 shows comparisons of ensemble mean forecasted values, and their 95% credible intervals for the four compartments studied which were generated using quantiles recomputed daily. We find that mean values and intervals agree quite well for all compartments except active cases. Also, for active cases, interestingly the true case count is observed to fluctuate between the MCMC and ABMA values.

#### 6.4 Field Deployment Performance

The ABMA framework was used to provide actionable insights for the city of Mumbai, India from May 2020 to October 2020. It was also used to forecast hospital burden in the state of Jharkhand, India. Figure 7 shows the ensemble mean



**Fig. 6: Comparison of 95% Credible Intervals and Ensemble Mean of MCMC and ABMA.** ‘EM’ denotes the Ensemble Mean forecast. Credible intervals (CI) are shown along with the point forecasts.

forecast against the ground truth and Table 2 shows the estimated parameters and loss values. These results confirm the high accuracy and interpretability of our methodology across different phases of the first wave of the pandemic.

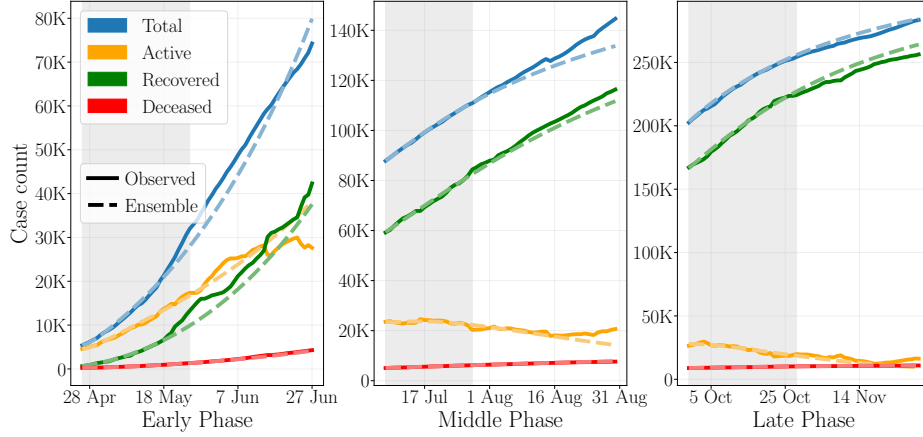
Recommendations made using this forecasting framework helped increase Mumbai’s ICU bed capacity by over 1200 units. About 95% of ICU capacity across the city, created in response to forecasting, was utilized when hospitalizations peaked. Moreover, over the deployment period, no absolute shortfall of critical health care resources became apparent.

### 6.5 Comparison with SOTA Models

We used the Reich Lab forecasting hub [32] for US states as a basis for comparing the performance of our forecasting framework with other methods (Table 3). The source of data for the forecasts is the JHU CSSE data [11]. We evaluated our method only on regions for which all four primary case counts were available (44 states plus Washington D.C.). Parameter and hyperparameter choices for the evaluation are in the Appendix (Table 4). The basis for comparison with other models is the MAPE value on deceased case counts.

We found that 26 models submitted to Reich Lab had submissions for at least 45 regions over the duration studied, with a range of median MAPE values between 1.21% and 4.26%. The top 10 models by median MAPE are shown





**Fig. 7: Performance of ABMA on Predicting Mumbai Caseload.** Predicted and ground truth case counts for Mumbai city across compartments and phases of the pandemic. Here, the parameter fitting period is  $T_\theta = 30$  days and the fitting period for the hyperparameter  $\alpha$  is  $T_\alpha = 3$  days (see Supplement). Forecasts are shown 30 days beyond the end of the  $T_\alpha$  fitting period. All dates refer to the year 2020.

in Table 3. Our forecasts have low error and compare well to other models without the need for any region-specific customisation of hyperparameters or fitting method.

We further analyzed the distribution of errors across regions (Figure 8) and assessed performance relative to other modeling approaches for each US region using the normalized MAPE metric, defined as

$$\text{NMAPE}(m) = \frac{L(m) - \text{MED}_{m' \in \mathcal{M}}\{L(m')\}}{\text{MAD}_{m' \in \mathcal{M}}\{L(m')\}}, \quad (7)$$

where  $L(m)$  is the MAPE loss for model  $m$ ,  $\mathcal{M}$  is the set of all models, MED and MAD denote the Median and the Median Absolute Deviation operators respectively. The variation in relative model performance across states could be attributed to changes in social distancing policies and the consequent variations in disease spread dynamics.

## 7 Lessons Learned

The lessons learned from our deployment are summarized below.

- *Dynamic data entails human in the loop.* The data-related challenges mentioned in Section 1 led to us relying heavily on humans in the loop for tracking evolving data definitions and carrying out semi-automated quality checks. Data versioning and pre-processing prior to modeling (Section 6.1) were also essential.

**Table 2: Quantitative Performance of ABMA on Mumbai city data.** Ensemble mean (ABMA) *parameters*, and test MAPE loss (%) for each *compartment* of the ABMA forecasts. Note that the ABMA forecast is the mean forecast, not the forecast of the mean parameters. Also, the subscript ‘ar’ refers to ‘active ratio’.

Phase	Parameter values								Compartment MAPE loss (%)			
	$\mathcal{R}_0$	$T_{\text{inf}}$	$T_{\text{inc}}$	$T_{\text{recov}}$	$T_{\text{fatal}}$	$P_{\text{fatal}}$	$E_{\text{ar}}$	$I_{\text{ar}}$	C	A	R	D
Early	1.18	3.59	4.29	30.14	14.70	0.07	0.95	0.12	6.91	11.33	12.91	3.78
Middle	0.83	3.86	4.17	19.32	24.32	0.05	0.30	0.18	3.10	9.04	2.39	1.05
Late	0.81	3.72	4.42	11.75	11.87	0.02	0.35	0.37	1.20	17.54	2.39	1.43

- *Model interpretability and identifiability is paramount.* Our choice of the SEIARD model (Section 5.4) over other model families was motivated by planning needs and policy choices. Therefore, the parameters had to be interpretable, independently verifiable where possible, and robustly estimable from the available data.
- *Application needs should dictate modeling choices.* Focus on capacity planning had a tangible meaning for the horizon of interest: policies around capacity took about a month to implement. We customized our model fitting with loss computed over this time horizon. When provided with information on upcoming policy changes and events—festivals, for example—we adapted the models accordingly to generate accurate forecasts.
- *Insights must be actionable.* Model insights had to be translated to concrete action guidance to enable smooth planning. Uncertainty estimation (Section 5.2) allowed us to provide three relevant scenarios: a) a planning b) a high case, c) a low case scenario. Region-specific testing levels, evolving severity of cases and sero-surveillance information to understand the state of the pandemic were important factors informing the selection of planning scenarios.
- *Capacity gaps at the last mile are hard to anticipate.* Although there were no apparent capacity shortfalls at a city level, the ability of a critical patient to access these resources is mediated by granular factors such as access to information on the availability of beds, local emergency transportation, ability to pay, and other equity considerations, which need to be factored in.

## 8 Conclusions and Future Directions

We have presented here a flexible, data-driven modeling framework and demonstrated its value for epidemic forecasting using the kind of case count aggregate data that is typically available in a resource-constrained public health setting. The deployed system was used to drive decision-making and planning with good accuracy (worst case MAPE < 20%) during the COVID-19 pandemic in Mumbai

**Table 3: Performance of top 10 Reich Lab models submitting forecasts for at least 45 regions on cumulative death counts.** All models were fitted on data from 18 Aug to 19 Sept 2020. Based on hyperparameter optimisation for Mumbai, we chose  $T_\theta = 30$ ,  $T_\alpha = 3$ . Model forecasts four weeks into the future starting 20 Sept 2020, aggregated every week, were evaluated by computing their MAPE values for every region, and then taking the median value across all regions (reported here).

Rank	Model	Median MAPE (%)
1	UMass-MechBayes	1.21
2	Karlen-pypm	1.32
3	SteveMcConnell-COVIDComplete	1.33
4	<b>ABMA</b>	<b>1.38</b>
5	YYG-ParamSearch	1.44
6	UCLA-SuEIR	1.49
7	PSI-DRAFT	1.65
8	DDS-NBDS	1.70
9	CEID-Walk	1.77
10	COVIDhub-baseline	1.81

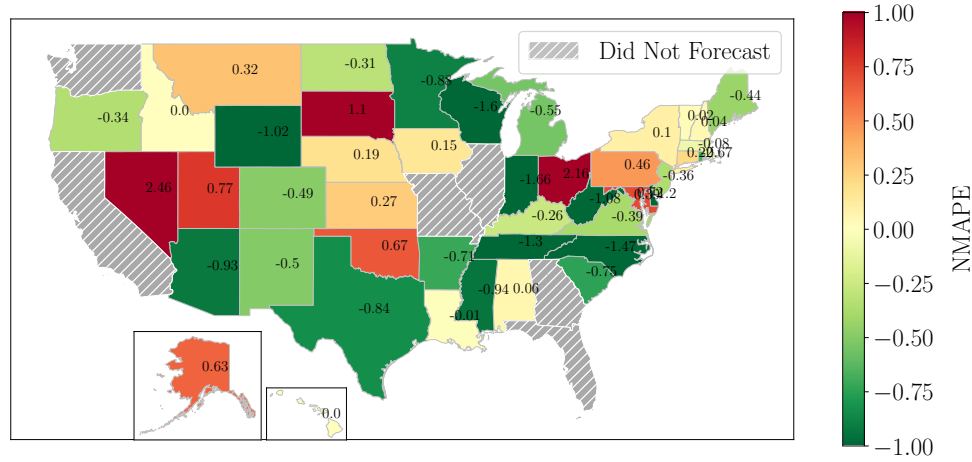
and Jharkhand, India. Our framework enables rapid forecasting with uncertainty estimates, and is extensible to other model families and to different loss functions. It also enables the optimization of hyperparameters such as fitting durations and ensemble weights. We motivated the choice of the specific compartmental model used via identifiability of the underlying parameters given the data constraints. Empirical comparison of our methods with other advanced models in the Reich Lab hub on real-world data further points to their efficacy.

The code for our framework is open-sourced<sup>7</sup>. In future work, we plan to apply it to the estimation of case burden in other infectious diseases, such as Tuberculosis, which are widespread across the developing world. One of the limitations of the current framework that we recognize is that the underlying parameters are static, necessitating periodic refitting to data. We plan to extend this modeling approach to capture time-varying yet interpretable parameters, and thus enable better forecasting.

## Acknowledgements

We would like to thank Janak Shah and the WIAI Programs team for their support with data and deployment processes. This study is made possible by the generous support of the American People through the United States Agency for International Development (USAID). The work described in this article was implemented under the TRACETB Project, managed by WIAI under the

<sup>7</sup> <https://github.com/WadhwaniAI/covid-modelling>.



**Fig. 8: Model Performance on US States.** NMAPE is the normalized MAPE (green values correspond to lower error) defined in the text (Equation 7).

terms of Cooperative Agreement Number 72038620CA00006. The contents of this manuscript are the sole responsibility of the authors and do not necessarily reflect the views of USAID or the United States Government. This work is co-funded by the Bill and Melinda Gates Foundation, Fondation Botnar and CSIR - Institute of Genomics and Integrative Biology. It was enabled by several partners through the informal COVID19 data science consortium, most notably Shreyas Shetty, Anupama Agarwal, and a group of volunteer data scientists under the umbrella group “Data Science India vs. COVID”. We thank Amazon Web Services for Cloud credits and support, the Smart Cities Mission for providing early impetus to this work, and the Brihanmumbai Municipal Corporation and Integrated Disease Surveillance Program (Jharkhand) for support while facing unprecedented challenges.

## References

1. Balabdaoui, F., Mohr, D.: Age-stratified discrete compartment model of the COVID-19 epidemic with application to Switzerland. *Scientific Reports* **10**, 21306 (2020)
2. Banerjee, A., Merugu, S., Dhillon, I., Ghosh, J.: Clustering with Bregman divergences. *Journal of Machine Learning Research* **6** (2004)
3. Bannur, N., Shah, V., Raval, A., White, J.: Synthetic data generation for improved COVID-19 epidemic forecasting. *medRxiv* (2020), <https://www.medrxiv.org/content/early/2020/12/07/2020.12.04.20243956>
4. Barndorff-Nielsen, O.: *Information and Exponential Families in Statistical Theory*. Probability and Statistics Series, Wiley (1978)
5. Bergstra, J., Bardenet, R., Bengio, Y., Kégl, B.: Algorithms for hyper-parameter optimization. In: Shawe-Taylor, J., Zemel, R., Bartlett, P., Pereira, F., Weinberger,

- K. (eds.) *Advances in Neural Information Processing Systems*. vol. 24. Curran Associates, Inc. (2011)
6. Bergstra, J., Yamins, D., Cox, D.: Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures. In: Dasgupta, S., McAllester, D. (eds.) *Proceedings of the 30th International Conference on Machine Learning*. *Proceedings of Machine Learning Research*, vol. 28, pp. 115–123. PMLR, Atlanta, Georgia, USA (2013)
7. Bertozzi, A.L., Franco, E., Mohler, G., Short, M.B., Sledge, D.: The challenges of modeling and forecasting the spread of COVID-19. *Proceedings of the National Academy of Sciences* **117**(29), 16732–16738 (2020)
8. Bettencourt, L., Ribeiro, R.: Real time Bayesian estimation of the epidemic potential of emerging infectious diseases. *PLOS ONE* **3**(5), 1–9 (2008)
9. Bhardwaj, A., Ou, H.C., Chen, H., Jabbari, S., Tambe, M., Panicker, R., Raval, A.: Robust lock-down optimization for COVID-19 policy guidance. 2020 AAAI Fall Symposium on AI for Social Good, AI4SG 2020 **2884** (2020)
10. Capaldi, A., Behrend, S., Berman, B., Smith, J., Wright, J., Lloyd, A.: Parameter estimation and uncertainty quantification for an epidemic model. *Mathematical biosciences and engineering* **9**, 553–576 (2012)
11. Center for Systems Science and Engineering at Johns Hopkins University: COVID-19 data repository (2020), <https://github.com/CSSEGISandData/COVID-19/>
12. Chowell, G.: Fitting dynamic models to epidemic outbreaks with quantified uncertainty: A primer for parameter uncertainty, identifiability, and forecasts. *Infectious Disease Modelling* **2**(3), 379–398 (2017)
13. Covid Act Now: America’s COVID warning system (2020), <https://covidactnow.org>
14. Covid19India: Coronavirus in india: Latest map and case count (2020), <https://www.covid19india.org/>
15. Deva, A., Shingi, S., Tiwari, A., Bannur, N., Jain, S., White, J., Raval, A., Merugu, S.: Interpretability of epidemiological models: The curse of non-identifiability. In: *AI for Public Health Workshop at ICLR*. arXiv (2021)
16. Diaconis, P., Ylvisaker, D.: Conjugate priors for exponential families. *Ann. Statist.* **7**(2), 269–281 (Mar 1979)
17. Elder, B., Dukic, V., Dwyer, G.: Uncertainty in predictions of disease spread and public health responses to bioterrorism and emerging diseases. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 15693–7 (11 2006)
18. Fang, Y., Nie, Y., Penny, M.: Transmission dynamics of the COVID-19 outbreak and effectiveness of government interventions: A data-driven analysis. *Journal of Medical Virology* **92** (03 2020)
19. Ferguson, N., Laydon, D., Nedjati-Gilani, G., Imai, N., Ainslie, K., Baguelin, M., Bhatia, S., Boonyasiri, A., Cucunubá, Z.M., Cuomo-Dannenburg, G., Dighe, A., Dorigatti, I., Fu, H., Gaythorpe, K., Green, W., Hamlet, A., Hinsley, W., Okell, L., van Elsland, S., Ghani, A.: Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand (2020)
20. Funk, S., Camacho, A., Kucharski, A., Lowe, R., Eggo, R., Edmunds, W.: Assessing the performance of real-time epidemic forecasts: A case study of Ebola in the Western Area region of Sierra Leone, 2014–15. *PLoS computational biology* **15**(2), e1006785 (2019)
21. Giordano, G., Blanchini, F., Bruno, R., Colaneri, P., Filippa, A., Matteo, A., Colaneri, M.: Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. *Nature Medicine* **26**, 1–6 (2020)

22. Hethcote, H.: The Mathematics of Infectious Diseases. *SIAM Rev.* **42**(4), 599–653 (2000)
23. IHME COVID-19 health service utilization forecasting team, Murray, C.J.: Forecasting the impact of the first wave of the COVID-19 pandemic on hospital demand and deaths for the USA and European Economic Area countries. *medRxiv* (2020), <https://www.medrxiv.org/content/early/2020/04/26/2020.04.21.20074732>
24. Jones, D.: A taxonomy of global optimization methods based on response surfaces. *Journal of Global Optimization* **21**(4), 345–383 (2001)
25. Kaggle: COVID19 global forecasting (2020), <https://www.kaggle.com/c/covid19-global-forecasting-week-1>
26. Kermack, W., McKendrick, A.: A contribution to the mathematical theory of epidemics. *Proc. R. Soc. Lond. A* **115**(4), 700–721 (1927)
27. Mandal, S., Bhatnagar, T., Arinaminpathy, N., Agarwal, A., Chowdhury, A., Murhekar, M.V., Gangakhedkar, R.R., Sarkar, S.: Prudent public health intervention strategies to control the coronavirus disease 2019 transmission in India: A mathematical model-based approach. *Indian Journal of Medical Research* **151**(2), 190–199 (2020)
28. Massonis, G., Banga, J., Villaverde, A.: Structural identifiability and observability of compartmental models of the COVID-19 pandemic. *Annual Reviews in Control* **51** (12 2020)
29. Moghadas, S., Shoukat, A., Fitzpatrick, M., Wells, C., Sah, P., Pandey, A., Sachs, J., Wang, Z., Meyers, L., Singer, B., Galvani, A.: Projecting hospital utilization during the COVID-19 outbreaks in the United States. *Proceedings of the National Academy of Sciences* **117**, 202004064 (2020)
30. Paggi, M.: Simulation of Covid-19 epidemic evolution: are compartmental models really predictive? (2020), <https://arxiv.org/abs/2004.08207>
31. Philipson, T.: Chapter 33 economic epidemiology and infectious diseases. *Handbook of Health Economics*, vol. 1, pp. 1761–1799. Elsevier (2000)
32. Reichlab, University of Massachusetts, Amherst: COVID-19 Forecast Hub (2020), <https://github.com/reichlab/covid19-forecast-hub>
33. Roosa, K., Chowell, G.: Assessing parameter identifiability in compartmental dynamic models using a computational approach: Application to infectious disease transmission models. *Theoretical Biology and Medical Modelling* **16** (2019)
34. Samui, P., Mondal, J., Khajanchi, S.: A mathematical model for covid-19 transmission dynamics with a case study of india. *Chaos, Solitons & Fractals* **140**, 110173 (2020)
35. Serhani, M., Labbardi, H.: Mathematical modeling of COVID-19 spreading with asymptomatic infected and interacting peoples. *Journal of Applied Mathematics and Computing* **66** (2020)
36. Tabataba, F., Chakraborty, P., Ramakrishnan, N., Venkatramanan, S., Chen, J., Lewis, B., Marathe, M.: A framework for evaluating epidemic forecasts. *BMC Infectious Diseases* **17** (2017)
37. Thakkar, N., Burstein, R., Hu, H., Selvaraj, P., Klein, D.: Social distancing and mobility reductions have reduced COVID-19 transmission in King County, WA (2020), [https://web.archive.org/web/20200817192227id/https://covid.idmod.org/data/Social\\_distancing\\_mobility\\_reductions\\_reduced\\_COVID\\_Seattle.pdf](https://web.archive.org/web/20200817192227id/https://covid.idmod.org/data/Social_distancing_mobility_reductions_reduced_COVID_Seattle.pdf), Institute for Disease Modeling
38. Walker, P., Whittaker, C., Watson, O., Baguelin, M., Winskill, P., Hamlet, A., Djaafara, B., Cucunubá, Z.M., Olivera, D., Green, W., Thompson, H., Nayagam,

- S., Ainslie, K., Bhatia, S., Bhatt, S., Boonyasiri, A., Boyd, O., Brazeau, N., Cattarino, L., Ghani, A.: The impact of COVID-19 and strategies for mitigation and suppression in low- and middle-income countries. *Science* **369**, eabc0035 (2020)
39. Wilder, B., Charpignon, M., Killian, J., Ou, H.C., Mate, A., Jabbari, S., Perrault, A., Desai, A., Tambe, M., Majumder, M.: Modeling between-population variation in COVID-19 dynamics in Hubei, Lombardy, and New York City. *Proceedings of the National Academy of Sciences of the United States of America* **117** (2020)

## A Appendix

Here, we outline additional modeling and implementation details required to reproduce the reported results.

### A.1 Parameter Searchspace

Table 4 below lists the parameter searchspace and hyperparameter ranges used in the experiments described in the main text.

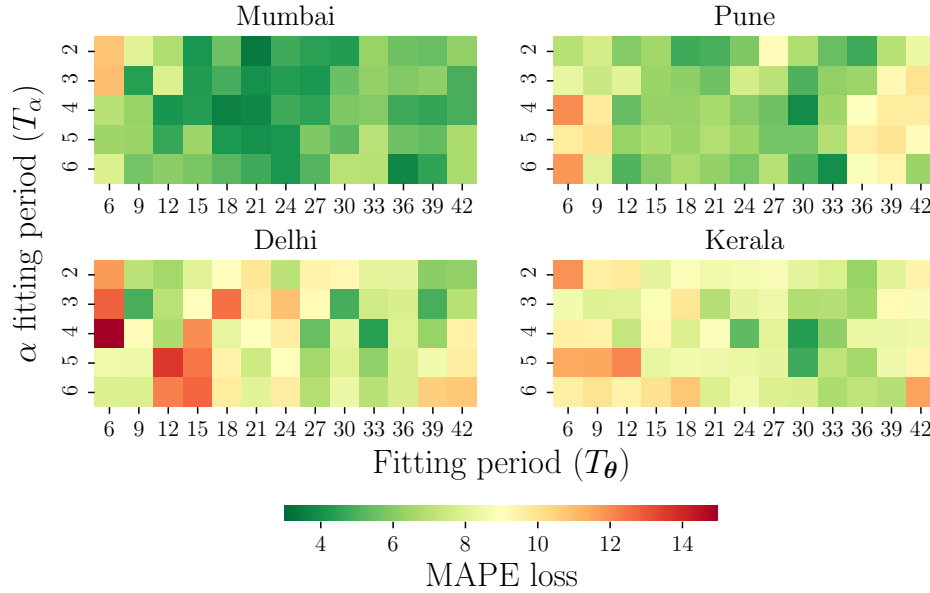
**Table 4: Parameter and hyperparameter ranges for all experiments.** (a-b) Simulated smoothing data and spikes; (c-d) Reich Lab comparison experiments; (e-f) ABMA approach applied to Mumbai data; (g-h) MCMC approach applied to Mumbai data. The “init” subscript refers to the corresponding initial value. The four loss weight values are for the four series C, A, R, D respectively.

Param	Range	Range	Range	Range	Prop. var.
$\mathcal{R}_0$	[0.6, 1.2]	[0.7, 1.5]	[0.7, 1.2]	[0.7, 1.2]	0.1
$T_{\text{inc}}$	[4, 6]	[4, 5]	[4, 5]	[3, 5]	0.3
$T_{\text{inf}}$	[3, 4]	[3, 4]	[3, 4]	[3, 5]	0.1
$T_{\text{recov}}$	[0, 40]	[49, 50]	[0, 40]	[0, 40]	1.25
$T_{\text{fatal}}$	[0, 40]	[0, 100]	[0, 40]	[0, 40]	1.25
$P_{\text{fatal}}$	[0, 0.5]	[0, 0.1]	[0, 0.1]	[0, 0.2]	0.01
$E_{\text{active\_ratio}}$	[0, 2]	[0, 2]	[0, 2]	[0, 2]	0.1
$I_{\text{active\_ratio}}$	[0, 2]	[0, 1]	[0, 1]	[0, 1]	0.1
(a) Synthetic		(c) US	(e) ABMA	(g) MCMC	
Hyper Param	Value	Value	Hyper Param	Value	Value
$R_{\text{init}}$	[50k, 100k]	$n$		3000	3000
$A_{\text{init}}$	[50k, 100k]	$T_{\theta}$		30	28
$D_{\text{init}}$	7500	$T_{\alpha}$		3	3
Spike length	[7, 21]	Loss weights		0,0,0,1	0.25,0.25, 0.25,0.25,
		$\alpha$ range		[0.1, 10]	0.25,0.25
(b) Synthetic		(d) US	(f) ABMA	(h) MCMC	

### A.2 Hyperparameter Optimisation

Since we wished to assess the robustness of optimal hyperparameters to choice of location, we performed experiments for several regions of India that were highly impacted by the pandemic: Mumbai, Pune, Delhi and Kerala. To facilitate comparisons, we used case count data from public sources [14], across two different





**Fig. 9: Heatmap of MAPE values on test periods as a function of  $T_\theta$  and  $T_\alpha$ .** MAPE is computed on a four week evaluation period and averaged across experiments for two time periods. The optimal values of  $(T_\theta, T_\alpha)$  are – Mumbai: (21, 2), Pune: (30, 4), Delhi: (33, 4), Kerala: (30, 4), Overall: (30, 3). “Overall” corresponds to the lowest mean MAPE across all four regions.

stages of the pandemic: our evaluations were carried out on case counts from July 1 to 28, 2020 and November 1 to 28, 2020. Fitting was carried out over durations prior to these dates. The evaluation metric used was the MAPE value on the entire evaluation period.

Due to the prohibitively large search space, a complete grid search over the hyperparameter choices was not feasible. We thus ordered the experiments and performed them sequentially, optimising one or more parameters in each step and fixing them in subsequent experiments to optimise other parameters.

We first performed a grid search over combinations of  $T_\theta$  and  $T_\alpha$  (Figure 9) to find the best choices. Within the selected fitting period, we found that  $n=3000$  trials were sufficient for convergence of sample mean parameters, and thus chose this value of  $n$ . While the application loss function of interest is the MAPE loss, we also explored two other loss functions during fitting as well as  $\alpha$  estimation: root mean square error (RMSE) and root mean square logarithm error (RMSLE). The MAPE value on the evaluation period with the three different configurations for fitting loss was 5.50 (MAPE), 5.96 (RMSE) and 15.81 (RMSLE). Based on the results, the loss function selected was MAPE for both fitting and estimating  $\alpha$ .

### A.3 SEIARD Model Dynamics

The dynamical equations governing the transitions in SEIARD model ( Figure 2) are

$$\frac{dS}{dt} = -\beta \frac{IS}{N}; \quad \frac{dE}{dt} = \beta \frac{IS}{N} - \sigma E; \quad \frac{dI}{dt} = \sigma E - \gamma I, \quad (8)$$

$$\frac{dA_{\text{recov}}}{dt} = (1 - P_{\text{fatal}}) \cdot \gamma I - \frac{A_{\text{recov}}}{T_{\text{recov}}}, \quad (9)$$

$$\frac{dA_{\text{fatal}}}{dt} = P_{\text{fatal}} \cdot \gamma I - \frac{A_{\text{fatal}}}{T_{\text{fatal}}}, \quad (10)$$

$$\frac{dR}{dt} = \frac{A_{\text{recov}}}{T_{\text{recov}}}; \quad \frac{dD}{dt} = \frac{A_{\text{fatal}}}{T_{\text{fatal}}}, \quad (11)$$

where  $N$  is the city population,  $\beta$ ,  $\sigma$ , and  $\gamma$  are the standard epidemiological parameters for a SEIR model,  $P_{\text{fatal}}$  is the transition probability to the mortality branch, and  $T_{\text{recov}}$  and  $T_{\text{fatal}}$  are timescales that govern transitions out of the  $A_{\text{recov}}$  and  $A_{\text{fatal}}$  compartments. Variables  $S$ ,  $E$ ,  $I$ ,  $A_{\text{recov}}$ ,  $A_{\text{fatal}}$ ,  $R$ ,  $D$  denote the populations of the similarly named compartments.

### A.4 MCMC Implementation Details

Let  $X[t] = [X_h[t]]_{h \in \mathcal{H}}$  be a multivariate time series with  $X_h[t]$  denoting the  $h^{\text{th}}$  compartment time-series, and  $\mathcal{H}$  be the set of indices of components. Let the fitting period be given by  $[t_i, t_j]$ . The key components of our MCMC-within-Gibbs sampling are:

- *Likelihood function.* We assume a likelihood function of the form

$$P(X \mid \boldsymbol{\theta}, s) = \prod_{h \in \mathcal{H}} \prod_{t=t_i+1}^{t_j} \mathcal{N}(\hat{z}_{h,\boldsymbol{\theta}}[t] \mid z_h[t], s), \quad (12)$$

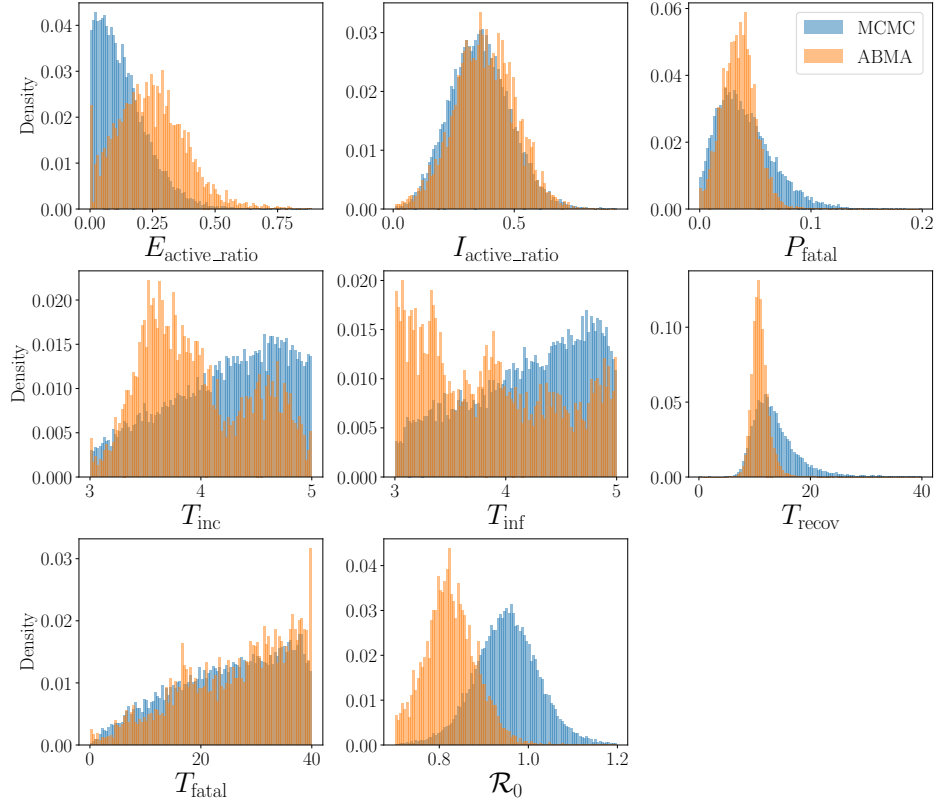
where  $\mathcal{N}(z \mid \mu, \sigma^2)$  denotes the Normal distribution pdf with mean  $\mu$  and variance  $\sigma^2$  following an appropriate conjugate prior. Further,  $z_h[t] = \log(X_h[t]) - \log(X_h[t-1])$ , and  $\hat{z}_{h,\boldsymbol{\theta}}[t]$  is the forecast equivalent of  $z_h[t]$ .

- *Proposal distribution.* At iteration  $k$ , we generate the samples from the proposal distribution for accept-reject step as

$$\boldsymbol{\theta} \sim Q(\boldsymbol{\theta}_{k-1}, \Sigma_{\text{prop}}, \boldsymbol{\theta}_{\min}, \boldsymbol{\theta}_{\max}),$$

where  $\boldsymbol{\theta}_{k-1}$  is the parameter vector chosen at  $k-1$ , and  $Q(\cdot)$  is the pdf of a multivariate truncated Gaussian with the parameter range  $[\boldsymbol{\theta}_{\min}, \boldsymbol{\theta}_{\max}]$  and covariance matrix  $\Sigma_{\text{prop}}$  as listed in Table 4(g) (columns: Range and Prop. var).

We further assume that  $s$ , the variance of the normal likelihood, has the conjugate prior  $s \sim \text{InvGamma}(u, v)$ . Thus, it is straightforward to show, by multiplying the Normal likelihood with the prior, that if the MCMC chain



**Fig. 10: Empirical distributions of ABMA and MCMC samples.** We observe large distributional overlap for all parameters.

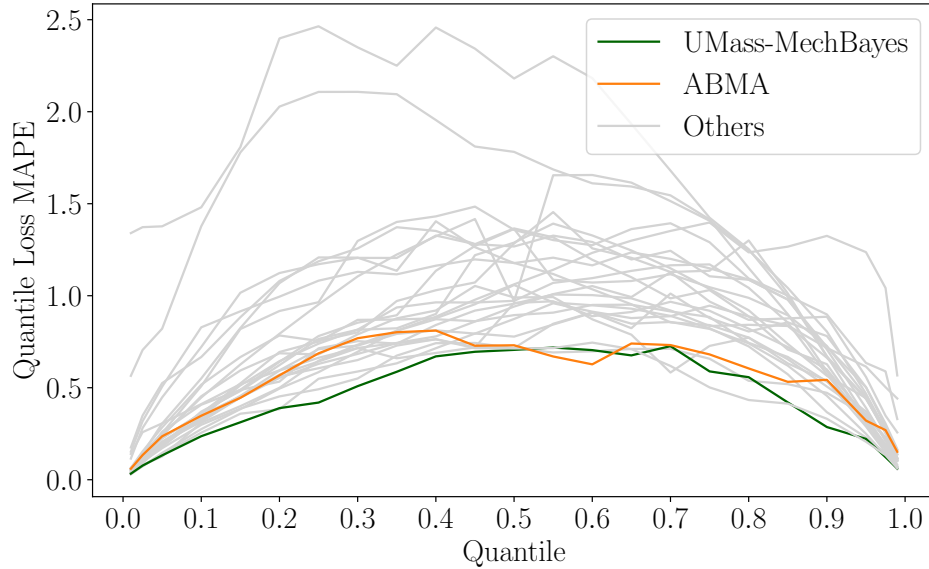
has sampled parameters  $\theta_k$ , the sample  $s_k \sim P(s_k \mid \theta_k, X)$  is also drawn from an Inverse Gamma distribution with parameters

$$u_k = u + 2(t_j - t_i - 1), \quad v_k = v + \sum_{h \in \mathcal{H}} \sum_{t=t_i+1}^{t_j} \frac{(z_h[t] - \hat{z}_{h, \theta_k}[t])^2}{2}.$$

Parameter ranges for our implementation are in Table 4. The MCMC implementation was adapted for  $X = [C, A, R, D]$  with 5 chains of length  $25k$ , a stride of 5 for sampling, a burn-in length of 50% of the chain,  $u = 40$ , and  $v = 2/700$ .

### A.5 Parameter distribution: ABMA vs. MCMC

To validate the uncertainty estimates of Section 6.3, we compared the samples collected from both methods aggregated over 10 runs of each method (Figure 10). Major distributional differences are found only in  $\mathcal{R}_0$  and  $E_{\text{active\_ratio}}$ , but even



**Fig. 11: Median Quantile MAPE Loss for 26 models from Reich Lab.** Median is taken across all states. Our model and the top ranked model in Table 3 are marked.

for these parameters there is significant overlap between the distributions sampled by the two methods. Note that distributions for the ABMA approach are weighted by the importance sampling factor  $\exp(-\alpha L(\theta))$ .

### A.6 Uncertainty Evaluation

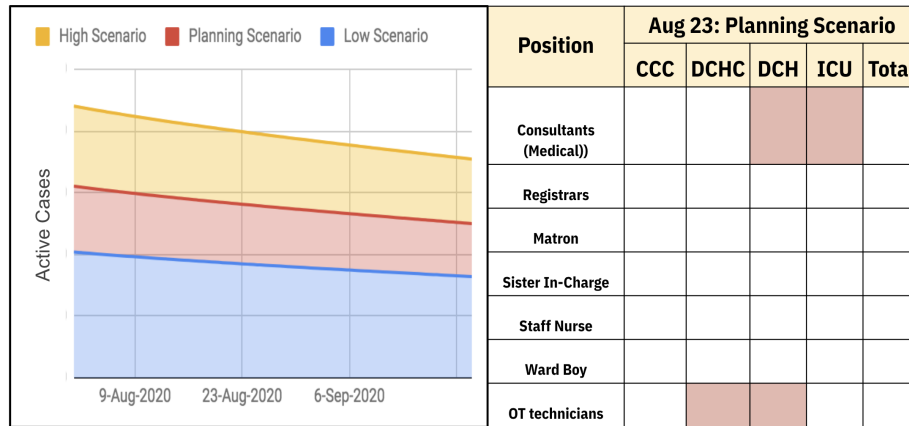
To compare the efficacy of our uncertainty estimation, we compared the forecast distributions of the models submitted to Reich Lab (Section 6.5) using the quantile loss described below:

$$Q(q, \varepsilon) = \begin{cases} q \cdot \varepsilon, & \text{if } \varepsilon \geq 0 \\ (q - 1) \cdot \varepsilon, & \text{if } \varepsilon < 0 \end{cases} \quad (13)$$

where  $\varepsilon$  is the error between the forecasted quantile  $q$  and the ground truth. In our case, the error is measured by the MAPE loss function. For every (model, quantile) tuple, we computed a median quantile loss value aggregated across regions. ABMA quantiles are generated by recomputing them daily. Figure 11 shows the variation of the median quantile loss across quantiles for all models. The loss curves of our model and the UMass-MechBayes model, the top performing model according to Table 3 are highlighted. While our model is one of the best around the 50<sup>th</sup> percentile, it exhibits higher relative loss at extreme quantiles.

### A.7 Report Template

Figure 12 shows the basic template of a sample report that was generated and shared with public health officials.



**Fig. 12: Sample report template shared with city officials.** Personnel recommendations and shortfall (red boxes) for various facilities are filled in. The facility types are CCC (COVID Care Center), DCH (Dedicated COVID Hospital), DCHC (Dedicated COVID Health Center), and ICU (Intensive Care Unit).