

Towards Improved Uncertainty Quantification of Stochastic Epidemic Models Using Sequential Monte Carlo

Arindam Fadikar*, Abby Stevens*, Nicholson Collier*, Kok Ben Toh[†]
Olga Morozova[‡], Anna Hotton[§], Jared Clark[¶], David Higdon[¶] and Jonathan Ozik*

*Decision and Infrastructure Sciences

Argonne National Laboratory, Lemont, IL, USA

Email: afadikar@anl.gov, stevensa@anl.gov, ncollier@anl.gov, jozik@anl.gov

[†]Department of Preventive Medicine

Northwestern University, Chicago, IL, USA

Email: bentoh@northwestern.edu

[‡]Department of Public Health Sciences

University of Chicago, Chicago, IL, USA

Email: omorozova@bsd.uchicago.edu

[§]Department of Medicine

University of Chicago, Chicago, IL, USA

Email: ahotton@bsd.uchicago.edu

[¶]Department of Statistics

Virginia Tech, Blacksburg, VA, USA

Email: cjared96@vt.edu, dhigdon@vt.edu

Abstract—Sequential Monte Carlo (SMC) algorithms represent a suite of robust computational methodologies utilized for state estimation and parameter inference within dynamical systems, particularly in real-time or online environments where data arrives sequentially over time. In this research endeavor, we propose an integrated framework that combines a stochastic epidemic simulator with a sequential importance sampling (SIS) scheme to dynamically infer model parameters, which evolve due to social as well as biological processes throughout the progression of an epidemic outbreak and are also influenced by evolving data measurement bias. Through iterative updates of a set of weighted simulated trajectories based on observed data, this framework enables the estimation of posterior distributions for these parameters, thereby capturing their temporal variability and associated uncertainties. Through simulation studies, we showcase the efficacy of SMC in accurately tracking the evolving dynamics of epidemics while appropriately accounting for uncertainties. Moreover, we delve into practical considerations and challenges inherent in implementing SMC for parameter estimation within dynamic epidemiological settings, areas where the substantial computational capabilities of high-performance computing resources can be usefully brought to bear.

Index Terms—sequential Monte-Carlo, computational epidemiology, uncertainty quantification

This material is based upon work supported by the National Science Foundation under Grants 2200234 and CCF-1918770, the U.S. Department of Energy, Office of Science, under contract number DE-AC02-06CH11357 and the Bio-preparedness Research Virtual Environment (BRaVE) initiative. This research was completed with resources provided by the Laboratory Computing Resource Center at Argonne National Laboratory.

I. INTRODUCTION

Computational epidemiologic models have become essential tools for public health decision-making, enabling rapid responses during times of crisis and uncertainty [21]. While the COVID-19 pandemic highlighted the possibilities enabled by computational epidemiological modeling, such as testing the impact of different interventions [16], it also revealed key weaknesses in modeling workflows [6]. Models are imperfect representations of the true epidemiological system, and for them to be used as trusted *in silico* laboratories they need to be calibrated to empirical data, often from heterogeneous sources with varying degrees of uncertainty, while accounting for time-varying dynamics of an ongoing epidemic.

Models themselves are often computationally demanding, and calibration, which typically requires repeated runs of a model, is even more so and necessitates the use of high-performance computing (HPC) resources. Furthermore, epidemics are constantly changing and evolving – for example, disease transmission may increase or decrease with time due to social and biological processes, and observational data, such as case counts (which are often used as a proxy for underlying infection incidence), may be reported at a rate that changes as an epidemic evolves in response to interventions or behavioral changes [20].

Stochastic disease simulators, such as compartmental or agent-based models, further complicate parameter estimation

by introducing additional randomness. Unlike deterministic models, which generate a fixed response given a particular input, multiple runs of a stochastic model with the same input parameters will generate an ensemble of random realizations of output trajectories. Traditional calibration techniques [17] do not generally account for such randomness, hence so-called *trajectory-oriented* optimization approaches have been developed to include the random seed as a parameter to be estimated [9]. Calibrating to individual trajectories, rather than summary statistics of the ensemble of outputs, can capture specific population dynamics, such as mixing patterns, and enables detailed scenario modeling and forecasting.

In light of the described complexities and in an effort to advance the application of epidemiologic modeling for supporting robust public health decision making, we propose an HPC-aware framework in which stochastic simulators are calibrated sequentially against multiple updating data streams. In addition to estimating epidemic model parameters, we introduce a binomial bias model for the empirical data to account for reporting errors and additionally estimate this parameter. Using Sequential Importance Sampling (SIS, [5], [8], [22]), we calibrate simulations within a given time frame against one or more empirical data sources and select the best trajectories. As the epidemic evolves and additional data become available, we calibrate the next time frame using parameter estimates from the previous time period as priors. Our system can be decomposed into an inner and outer loop corresponding to empirical data update cadences (e.g., daily, weekly), with the outer loop moving the model forward in time and the inner loop calibrating individual time periods. Estimating parameters sequentially in time allows us to capture their time-varying dynamics, as calibrating each time period results in different parameter estimates.

By considering individual trajectories (rather than aggregations of stochastic output), we are able to store, or checkpoint, the exact state of the model, allowing models to be restarted from time-stamped stored states rather than restarting them from the beginning of an epidemic. This facilitates significant computational savings when deploying calibration algorithms in sequentially updating operational modes. Even with checkpointing facilities in place, producing sufficient parameter space samples from stochastic simulators can be computationally burdensome and thus our framework is designed to exploit the concurrency provided by HPC resources to make the computation feasible.

We demonstrate our proposed sequential inference procedure on a stochastic SEIR compartmental model that simulates COVID-19 transmission in Chicago [26]. The model parameter of interest is the transmission rate, which we vary across time intervals. We simulate empirical case counts and deaths by applying a time-varying reporting bias to individual trajectories of the model. Our goal is to infer both the time-varying transmission rate and reporting bias using our SIS framework, and we calibrate to both case counts alone and a combination of case counts and deaths.

Our framework is particularly suited for rapid public health

decision-making – it is computationally efficient, accounts for the uncertainties from various data sources, and can give up-to-date insights into the evolution of the epidemic. The rest of this paper proceeds as follows. In section II, we give an overview of related work and highlight novel elements of our system. Section III describes the specific stochastic epidemic simulator we use to demonstrate our calibration method and checkpointing procedure. Section IV formally defines the statistical model and derives the updates for our SIS procedure, and section V provides results from our experiment. The discussion can be found in section VI.

II. RELATED WORK

The development of Bayesian calibration methods for stochastic computational models, particularly for epidemic applications has seen significant progress in the recent past [2], [10], [14]. However sampling-based, or Monte Carlo-based approaches for sensitivity analysis [30] or inference [7], [11] on stochastic models is too demanding when running the simulations is expensive. In such cases, emulation – modeling the input-output response surface – has proven to be an effective approach. In some applications, where only the mean (or other summary statistics) response is modeled as a function of the input parameters [2], [10], [18], the realizations from a simulation can be averaged, preprocessed, or accounted for in the inference framework. In the case of an agent-based epidemic model where each realization may indicate a specific mixing behavior among the agents, it is not well understood what the average (or other summary) of such mixing behavior would represent [9]. This issue further hinders the ability to use a *mean* calibrated epidemic model to simulate forward in time.

Understanding epidemic dynamics in the face of incomplete or biased data and imperfect computational models has long been of interest to the scientific community, and research in this space has proliferated since the start of the COVID-19 pandemic. Many have studied the impact of incomplete or biased data on the estimation of infection rates (e.g., [12], [24]) and the effective reproductive number R_t [13], [15], [33], [34]. While most existing studies consider mechanistic models, such as compartmental or agent-based models, attempts to bridge traditional epidemiology with more recent statistical methods have shown promise in addressing data gaps [23], [25], [29], [32]. Similar to our work, [31] estimate time-varying epidemic parameters and case reporting errors using a Sequential Monte-Carlo calibration approach; however, they cite computation as a major bottleneck in their analysis, which we address with checkpointing and exploiting the concurrency in high-performance computing systems. Furthermore, our work seeks to extend on other efforts to operationalize epidemiological modeling systems, such as [28] who developed real-time influenza forecasts using advanced data assimilation techniques.

III. STOCHASTIC DISEASE SIMULATOR

We utilize a stochastic SEIR compartmental model published by [26] that simulates the transmission of SARS-CoV-2

and the progression of COVID-19 disease states to illustrate our proposed sequential inference procedure. This model incorporates various symptom statuses, including asymptomatic, presymptomatic, mild, and severe, as well as outcomes such as hospitalization, ICU admission for critical illness, and mortality.

A. Basic setup

At the onset of the epidemic, the entire population is considered susceptible (S). Susceptible individuals become exposed and infected by the virus at a rate determined by the transmission rate parameter and the number of infectious and susceptible individuals in the population. Following a latent period, exposed individuals (E) transition to the infectious stage without symptoms, categorized as either completely asymptomatic (As) or presymptomatic (P). Upon completing the incubation period, presymptomatic individuals may develop either mild (Sm) or severe (Ss) symptoms. Asymptomatic individuals or those with mild symptoms eventually recover (R), while those with severe symptoms require hospitalization (H) after spending some time in the Ss compartment. Individuals requiring hospital care may either recover without complications or progress to critical illness (C) necessitating ICU admission. Those in critical condition may recover by transitioning back to post-ICU hospitalization (Hp) or succumb to the illness (D). All exposed individuals remain undetected until they transition to a symptomatic state. Upon entering asymptomatic, presymptomatic, mild, or severe states, a fraction of the individuals are detected after a certain period, leading to isolation and reduced infectiousness. A schematic representation of the model is provided in Figure 1. Model parameters, such as the incubation period, disease severity and symptomaticity, and lengths of hospital and ICU stays, are determined based on existing literature and observed data from Chicago. The transmission rate, probability of detection, and disease recovery rate are dynamically adjusted over time to better align with daily ICU and hospital census as well as reported deaths in Chicago. Further details regarding the model and its parameters, along with the source code, can be accessed from the GitHub repository [27].

B. Checkpointing

We have implemented checkpointing in the model such that the current state of the model (i.e., the number of persons in each state, the future state transition events, the current simulated time, etc.) can be serialized to a file and then restarted from that file, continuing from that serialized state. More importantly, when restarting from a checkpoint, the serialized state can be updated from user-input, parameterizing the model along a new trajectory. The following parameters can be specified to create a new trajectory: 1) the random seed; 2) the fraction of persons moving from E to P; 3) fraction of persons moving from P to Sm; 4) infectiousness of persons with symptomatic versus asymptomatic infections; 5) infectiousness of persons with detected versus undetected infections; 6) rate of persons moving from S to E. For

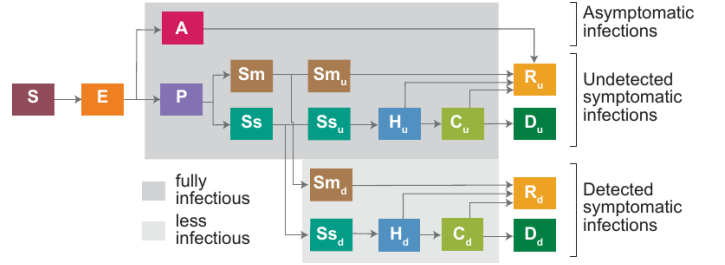


Fig. 1. **SEIR model to simulate SARS-CoV-2 transmission.** S=Susceptible, E=Exposed, A=Asymptomatic, P=Presymptomatic, Sm=Mild symptomatic, Ss=Severe symptomatic, H=Requires hospitalization, C=Critically ill (ICU), D=Death, R=Recovered. Subscripts u and d denote if the infection is undetected or detected. Asymptomatic and presymptomatic infections may also be detected, which would result in less infectiousness.

illustrations in this paper, we treat the last parameter as the calibration parameter and set the other parameters to values obtained from prior study.

Checkpointing plays a pivotal role within the proposed inference framework, with detailed insights provided in subsequent sections. Here, we briefly explore its significance and advantages within sequential Monte Carlo algorithms. Within a sequential inference paradigm, where model parameters evolve over time, the simulation model must be executed iteratively to accommodate these parameter updates. For instance, when simulating an epidemic, the dynamic nature of the process demands continuous adjustment of model parameters with respect to time. By preserving the detailed state of the model at intermediate time points through checkpointing, we enable the seamless updating and application of parameters during forward simulation. This preserves the historical data and obviates the need to restart the simulation from the epidemic's onset.

IV. STATISTICAL MODEL FORMULATION

The application involves a stochastic compartmental disease simulator (i.e., a computer model) that takes a d dimensional input $\theta = (\theta_1, \dots, \theta_d)$ and produces vector output $\eta_{1:T} = (\eta_1, \dots, \eta_t)$, indexed by time t . The outputs are characteristics of the disease dynamics, such as the number of infectious individuals $\eta_{1:T}^i$ and deaths $\eta_{1:T}^d$, some of which are (partially) observable in the real world. In the case of stochastic simulation, each model run also requires the specification of a random seed s to constitute a one-to-one mapping between input (θ, s) to $\eta_{1:T}$. We note that this level of input specification is one of the novel ingredients of our calibration framework. Trajectory outputs from an epidemic simulation, even at the same input values represent specific histories and dynamics of the disease, which are otherwise lost when represented via summary statistics. We assume that observed data include the reported number of cases – $y_{1:T}^c$ and deaths – $y_{1:T}^d$ during a disease outbreak. These data series are used to calibrate the stochastic disease simulation, i.e., to estimate $\theta | (y_{1:T}^c, y_{1:T}^d)$.

The basic observation model for each output type is written as:

$$\mathbf{y}_{1:T} = \boldsymbol{\eta}_{1:T}(\boldsymbol{\theta}, s) + \boldsymbol{\varepsilon}_T, \quad \boldsymbol{\varepsilon}_T \sim MVN(0, \Sigma_T^2) \quad (1)$$

i.e., we assume that the empirical counts are noisy observations of the simulated trajectories at an unknown $(\boldsymbol{\theta}, s)$. The likelihood $l(\mathbf{y}_{1:T}|\boldsymbol{\theta}, s)$ following the model is simply a multivariate Gaussian with mean $\boldsymbol{\eta}_{1:T}(\boldsymbol{\theta}, s)$ and a diagonal covariance matrix. The posterior $\pi(\boldsymbol{\theta}, s|\mathbf{y}_{1:T})$ is then proportional to $l(\mathbf{y}_{1:T}|\boldsymbol{\theta}, s) \times \pi(\boldsymbol{\theta}, s)$, where, $\pi(\boldsymbol{\theta}, s)$ denotes the joint prior on $\boldsymbol{\theta}$ and s . This posterior can then be studied by repeatedly sampling from it via common Monte-Carlo techniques such as MCMC, importance sampling, and accept-reject methods. One can produce summary statistics and credible intervals of $\boldsymbol{\theta}$ and future predictions using the samples from the posterior distribution. In this paper, we adopt a sequential Monte Carlo (SMC) approach based on importance sampling.

A. Bias model

When calibrating a computer model, a bias term is often used to represent the misalignment between the true physical process and the computer simulation, which is often limited by our understanding of the true underlying system or the fidelity of the numerical solver. This is more commonly referred to as *discrepancy* and is used as an additive or multiplicative factor to [1]. Similarly, in the case of an epidemic application, there are sources of discrepancy such as inaccurate reporting of cases and reporting lag, which can potentially affect the estimation of the model parameters and their uncertainties if ignored. Next we describe our proposed bias model with a focus on epidemic applications.

In the early days of an epidemic when data collection systems are often at an immature stage, the case counts can suffer from a lack of reporting. To capture this *missing* count process, we use a binomial model to represent the observed counts as a fraction of the true unobserved counts. The basic underlying assumption is that for each occurrence of an event, it gets observed with probability $0 < \rho_t < 1, t \in \{1, 2, \dots\}$. Then, the total number of reported counts at time t , $\mathbf{y}_t$ follows a binomial distribution with probability ρ_t . We write, $\mathbf{y}_t \sim \text{Binomial}(\mathbf{y}_t^{\text{true}}, \rho_t)$. In practice, both $\mathbf{y}_t^{\text{true}}$ and ρ_t are unknown and our goal is to infer $\mathbf{y}_t^{\text{true}}$. The observation model in [1] for each time index t can now be rewritten as:

$$\mathbf{y}_t = \boldsymbol{\eta}_t^{\text{obs}}(\boldsymbol{\theta}, s, \rho_t) + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim N(0, \sigma_t^2), 0 < \rho_t < 1 \quad (2)$$

where, $\boldsymbol{\eta}_t^{\text{obs}}(\boldsymbol{\theta}, s, \rho_t) \sim \text{Binomial}(\boldsymbol{\eta}_t(\boldsymbol{\theta}, s), \rho_t)$, i.e., the simulated observed counts $\boldsymbol{\eta}_t^{\text{obs}}$ is a binomial random variable depending on the true simulated counts $\boldsymbol{\eta}_t$ and the unknown probability ρ_t . The posterior is accordingly updated by modifying the likelihood that reflects the updated mean of the

Gaussian distribution.

$$l(\mathbf{y}_{1:T}|\boldsymbol{\theta}, s, \rho_t) = \frac{1}{\sqrt{2\pi}|\Sigma_T|^{1/2}} \times \exp \left\{ -\frac{1}{2} (\mathbf{y}_{1:T} - \boldsymbol{\eta}_{1:T}^{\text{obs}}(\boldsymbol{\theta}, s, \rho_t))^T \Sigma_T^{-1} (\mathbf{y}_{1:T} - \boldsymbol{\eta}_{1:T}^{\text{obs}}(\boldsymbol{\theta}, s, \rho_t)) \right\} \quad (3)$$

Together with the prior specification for $\boldsymbol{\theta}$, s , and ρ_t , the posterior can be fully obtained and sampled from. For the rest of the paper, we assume that ρ_t does not change significantly within a relatively shorter time window, i.e., $\rho_t = \rho$.

B. Sequential Importance Sampling (SIS)

Importance sampling is a Monte Carlo technique used to estimate properties of a probability distribution, typically used when sampling directly from that distribution is difficult or inefficient. Instead of directly sampling from the target distribution, which might be computationally expensive or intractable, importance sampling generates samples from a different, easier-to-sample-from distribution called the “proposal distribution.” These samples are then weighted according to how likely they would have been generated from the target distribution. The key idea is to bias the sampling towards regions of the target distribution where the function being integrated or the property being estimated is large, thereby reducing the variance of the estimate. By adjusting for the differences in probabilities between the target and proposal distributions, importance sampling can provide more accurate estimates than direct sampling, especially in high-dimensional or complex scenarios.

Sequential importance sampling (SIS, [5], [8]) is an extension of importance sampling that is particularly useful in scenarios where the target distribution changes over time, such as in an epidemic process. In SIS, instead of generating all samples at once from the target posterior distribution, samples are generated sequentially over time, with each sample’s importance weight being updated as new information becomes available. This allows SIS to adaptively track changes in the target distribution. The typical workflow of SIS involves iteratively generating samples from the proposal distribution, updating their importance weights based on the likelihood of the observed data given the sample, and then resampling a subset of these samples according to their weights to prevent degeneracy (i.e., where only a few samples have significant weights). This process is repeated as new data arrive, allowing SIS to provide estimates of the target distribution that are conditioned on the observed data sequence. The steps to draw samples from $p(\mathbf{x}_{0:t} | \mathbf{y}_{1:t})$, which represents the probability of the state variables $\mathbf{x}_{0:t}$ given the observed data sequence $\mathbf{y}_{1:t}$ are as follows:

- **Initialization:** Start with initial samples and weights drawn from the proposal distribution:

$$\mathbf{x}_{0:t}^{(i)} \sim q(\mathbf{x}_{0:t}) \quad \text{and} \quad w_t^{(i)}$$

- **Sequential Sampling and Weight Update:** At each time step t , perform:

1) *Sampling*:

$$x_t^{(i)} \sim q(x_t | x_{t-1}^{(i)})$$

2) *Weight Update*:

$$w_t^{(i)} \propto w_{t-1}^{(i)} \frac{p(y_t | x_t^{(i)})}{q(x_t^{(i)} | x_{t-1}^{(i)})}$$

• **Normalization**:

$$w_t^{(i)} = \frac{w_t^{(i)}}{\sum_{j=1}^N w_t^{(j)}}$$

• **Resampling**: Resample $x_t^{(i)}$ with probabilities given by the weights $w_t^{(i)}$ to construct samples from the target distribution.

C. Full Bayesian inference

Having reviewed all the necessary components, we provide a comprehensive outline of the sequential inference scheme applied to our epidemic system. To illustrate, let's consider a scenario resembling an actual epidemic. Imagine conducting a simulation over a total duration of T weeks, during which it is compared against real-world data in discrete time windows, and model parameters are adjusted accordingly. We will describe this process for two such intervals, the first window and any window thereafter. Let $[1, t_1]$, $t_1 > 0$ be the first time window and the m -th time window be denoted as $[t_{(m-1)} + 1, t_m]$.

Following the model in (2), the likelihood for the full T -variate observation $\mathbf{y}_{1:T} = (y_1, \dots, y_T)$, can be decomposed into two independent components according to the window sizes. Furthermore, we assume that both case counts and the number of deaths are observable from the epidemic process and can be used to constrict the parameter tuple (θ, s, ρ) . The posterior of (θ, s, ρ) conditioning on the full observation is as follows:

$$\begin{aligned} & \pi(\theta, s, \rho | \mathbf{y}_{1:T}^c, \mathbf{y}_{1:T}^d) \\ & \propto l(\mathbf{y}_{1:T}^c | \theta, s, \rho) \times l(\mathbf{y}_{1:T}^d | \theta, s, \rho) \times \pi(\theta, s, \rho) \\ & \propto \text{MVN}(\mathbf{y}_{1:T}^c | \boldsymbol{\eta}_{1:T}^c(\theta, s, \rho), \Sigma_t^c) \times \\ & \quad \text{MVN}(\mathbf{y}_{1:T}^d | \boldsymbol{\eta}_{1:T}^d(\theta, s, \rho), \Sigma_t^d) \times \\ & \quad \pi(\theta) \times \pi(s) \times \pi(\rho) \end{aligned} \quad (4)$$

SIS is used to sample from this posterior and the posterior predictions for $\mathbf{y}_{(T+1):T'}$ can be constructed by sampling from the Gaussian distribution with mean $\boldsymbol{\eta}(\hat{\theta}, \hat{s}, \hat{\rho})$ and covariance matrix $\hat{\Sigma}$.

1) Inference in time window $[1 : t_1]$

The posterior for the first time window is trivially obtained from (4), by replacing T by t_1 . To sample from this posterior, we choose the prior distribution to be the same as the proposal distribution for importance sampling. In practice, this choice amounts to sampling from the prior distributions, running the simulations, and computing the importance weights by evaluating the likelihood function. At the end of the resampling

Algorithm 1 Pseudo-code for single window IS

Require: N (total simulation budget)

- 1: Sample: $(\theta_i, s_i, \rho_i) \sim \pi(\theta, s, \rho)$, $i = 1 : N$.
- 2: Run simulation: $\boldsymbol{\eta}_{1:t_1}^c(\theta_i, s_i, \rho_i), \boldsymbol{\eta}_{1:t_1}^d(\theta_i, s_i, \rho_i)$.
- 3: Compute weight:

$$w_i^{(t_1)} \propto l(\mathbf{y}_{1:t_1}^c | \theta_i, s_i, \rho_i) \times l(\mathbf{y}_{1:t_1}^d | \theta_i, s_i, \rho_i)$$

- 4: Resample: Sample $(\theta_1, s_1, \rho_1), \dots, (\theta_N, s_N, \rho_N)$ with probabilities proportional to $w_i^{(t_1)}$.
-

step, the posterior epidemic trajectories are already available from the pool of simulation runs.

2) Inference in time window $[t_{(m-1)} + 1, t_m]$

To sample from $\pi(\theta, s, \rho | \mathbf{y}_{1:t_m}^c, \mathbf{y}_{1:t_m}^d)$, i.e., the posterior conditioning on the full trajectory until time t_m , we make use of the posterior samples from $\pi(\theta, s, \rho | \mathbf{y}_{1:t_{(m-1)}}^c, \mathbf{y}_{1:t_{(m-1)}}^d)$ as described in IV-B. For the sake of simplicity, we show the sequential update based on the conditional probability of the full posterior on the first two consecutive time windows, $[1, t_1]$ and $[t_1 + 1, t_2]$.

$$\begin{aligned} \pi(\theta, s, \rho | \mathbf{y}_{1:t_2}^c, \mathbf{y}_{1:t_2}^d) &= \frac{l(\mathbf{y}_{1:t_2}^c, \mathbf{y}_{1:t_2}^d | \theta, s, \rho) \pi(\theta, s, \rho)}{\pi(\mathbf{y}_{1:t_2}^c, \mathbf{y}_{1:t_2}^d)} \\ &= \frac{l(\mathbf{y}_{1:t_1}^c, \mathbf{y}_{1:t_1}^d | \theta, s, \rho) \pi(\theta, s, \rho)}{\pi(\mathbf{y}_{1:t_1}^c, \mathbf{y}_{1:t_1}^d)} \times \\ & \quad \frac{l(\mathbf{y}_{(t_1+1):t_2}^c, \mathbf{y}_{(t_1+1):t_2}^d | \mathbf{y}_{1:t_1}^c, \mathbf{y}_{1:t_1}^d, \theta, s, \rho)}{\pi(\mathbf{y}_{(t_1+1):t_2}^c, \mathbf{y}_{(t_1+1):t_2}^d | \mathbf{y}_{1:t_1}^c, \mathbf{y}_{1:t_1}^d)} \\ & \propto \pi(\theta, s, \rho | \mathbf{y}_{1:t_1}^c, \mathbf{y}_{1:t_1}^d) \times \\ & \quad l(\mathbf{y}_{(t_1+1):t_2}^c, \mathbf{y}_{(t_1+1):t_2}^d | \mathbf{y}_{1:t_1}^c, \mathbf{y}_{1:t_1}^d, \theta, s, \rho) \end{aligned} \quad (5)$$

With the above formulation, samples from the intermediate posterior in the time window $1 : t_1$ can now be used to update importance weights $w^{(t_2)}$ for the target posterior based on full t_2 -variate observation. Replacing t_1 and t_2 by $t_{(m-1)}$ and t_m for $m > 1$ will give the formula for the general time window $[t_{(m-1)} + 1, t_m]$.

3) Estimating time-varying epidemic parameters

Apart from improving sampling efficiency, another rationale behind sequential calibration is to adapt to time-varying model parameters denoted as θ_t . Epidemic dynamics are subject to changing conditions, necessitating adjustments in model parameters over time. By calibrating the simulation model to real-world data within shorter time intervals sequentially, it becomes feasible to capture how model parameters evolve over time.

V. RESULTS

A. Simulation setup

To illustrate the sequential inference on the disease trajectory, an example from the Covid-age stochastic simulator is used as the ground truth empirical data. We use the transmission rate as the only model parameter that is to be estimated from the empirical data and this parameter is allowed to vary over time, indicating an evolving epidemic. To make the problem of estimation of this time-varying parameter tractable, we assume that the value changes at discrete time points, called horizons, and the boundaries of the time windows for inference may or may not coincide with the horizons. To simulate the ground truth, we set the value of the transmission rate parameter at 0.3 for days 0 to 33, 0.27 for days 34 to 47, 0.25 from days 48 to 61, and 0.4 for days 62 onward. The selection of these values is qualitatively driven by the initial trends in COVID-19, where heightened awareness and social interventions lead to a decline in transmission rates over time, while also considering the potential emergence of new variants. The outputs from the simulator are the number of daily infections and deaths. The case counts from the simulator are treated as the actual number of cases, which is unobserved in reality. Hence, following the model in (2), a bias is applied to the actual case counts to simulate observed case counts. The parameter of the binomial bias model is also set as a time-varying parameter with values of 0.6 for days 0 to 33, 0.7 for days 34 to 47, 0.85 for days 48 to 61, and 0.8 for days 62 and beyond. These values are chosen to mimic the improvement in data collection and reporting over time during a pandemic.

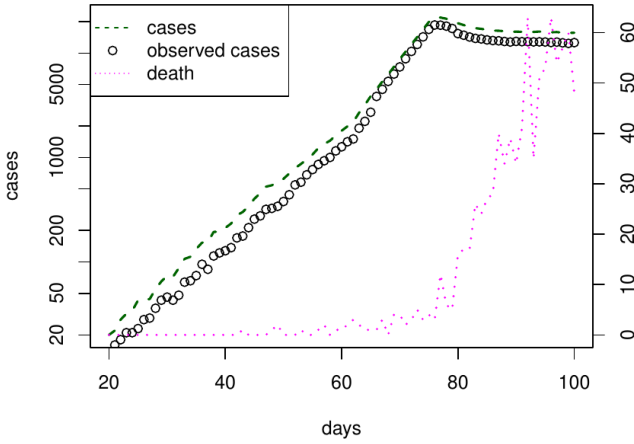


Fig. 2. A simulated ground-truth using SEIR Covid model

The sequential inference is performed in four time windows that reflect the changes in the epidemic behavior and the bias pattern. The steps to perform inference in each time window are the same, i.e., draw samples from the prior distribution of the 2-d parameter (θ, ρ) space, run simulations, compute the importance weight, and finally resample from the prior according to weights to construct the posterior. Except for the

first window, the posterior samples from the previous window serve as the prior samples for the current window, as illustrated in Figure 4 and Figure 5. For the first time window, we use a uniform prior for the transmission rate parameter θ and a beta prior for the binomial bias model parameter ρ . At each unique parameter combination of (θ, ρ) the simulation is replicated 20 times.

B. Calibrating to case counts

The first calibration experiment relies solely on the *reported* case counts. Illustrated in Figure 3, this experiment demonstrates a single-step calibration employing importance sampling. Uniform priors over the interval $(0.1, 0.5)$ for θ and a $\beta(4, 1)$ prior for ρ are considered, yielding a set of 25,000 prior samples. It's important to note that in the absence of prior information, an independent product prior is assumed for (θ, ρ) . Since at each combination of (θ, ρ) the simulation is repeated 20 times, this generates a total of 500,000 unique trajectories. To control variability between replicates, the same set of random seeds is employed to generate the 20 realizations from the stochastic simulation. The left panel in Figure 3 displays all 500,000 prior trajectories in grey. Using a Gaussian likelihood on square-root transformed counts with $\sigma_t = 1$, a sample of size 10,000 is then drawn with replacement from the 500,000 prior samples. This selection is based on probabilities proportional to the importance weights, thereby constructing the posterior distribution of the model parameters and output. The purple trajectories depicted in Figure 3 represent the posterior model output, while the empirical histograms depict the posterior distributions on model parameters. It's worth noting that the posterior on ρ exhibits less influence compared to that on θ , partly due to the utilization of strong informative priors on ρ .

Following the inference procedure described in IV-C, Figure 4 illustrates the outcome of a comprehensive sequential calibration performed across four successive time intervals. Following the initial calibration stage (days 20-33), posterior models are preserved (i.e., checkpointed) to enable their forward simulation with updated model parameter values. Specifically, a uniform distribution centered around each posterior value of the model parameter is employed to generate prior samples for subsequent time intervals. Additionally, employing an informative prior, such as an asymmetric interval for the uniform distribution, may enhance the efficiency of importance sampling if such information is accessible, for example, if it is expected that the identification of cases is improving. We adopt a symmetric interval in the uniform prior for θ and an asymmetric interval for ρ , with a higher density toward the higher value of ρ , reflecting the reduced reporting error in later epidemic stages. Following the acquisition of calibrated parameters, we can deduce the actual number of unobserved cases in reality. The top-left plot in Figure 4 shows the simulated reported cases as black circles, alongside posterior trajectories in grey. The cyan and pink ribbons denote 50% and 90% credible intervals based on the posterior trajectories, respectively. By marginalizing out ρ from the joint posterior of

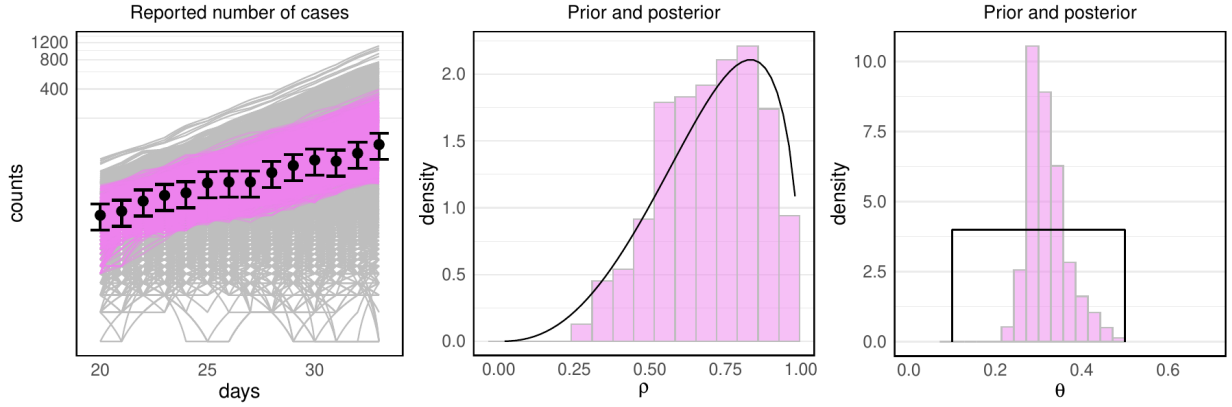


Fig. 3. **Calibration using only reported case counts within the first time window** – (left) The simulated ground truth is shown by black dots which are used to find the posterior trajectories (in purple) that are consistent with the ground truth from the prior trajectories (in grey), (center) the prior density (in black) and posterior of the ρ parameter in the binomial bias model, (right) the prior and posterior distribution of simulation input θ .

(θ, s, ρ) , we can now formulate posterior predictive summaries and intervals on true case counts, showcased in the right plot. The bottom panel in Figure 4 illustrates the evolution of the posterior on model parameters as additional data is observed and utilized to constrain the trajectories over four time windows.

C. Calibrating to case counts and deaths

In an effort to improve our time-varying estimation of (θ, ρ) , we next integrate death counts as an additional model output. The model parameters are now constrained by both reported case counts and deaths. We do not assume any reporting bias on death counts, instead we use a Gaussian error model on the square-root counts similar to reported case counts. The updated posterior prediction, shown in Figure 5, encompasses reported cases, actual cases, and deaths. Notably, there is a reduction in uncertainty regarding reported case predictions. This trend is further reflected in the joint posterior contour plot, where high probability regions exhibit greater concentration around specific values. While our example operates under the assumption of no reporting bias in death counts, our SMC framework can easily be extended to incorporate bias in multiple data sources. It’s worth emphasizing that this framework is highly adaptable, capable of incorporating various types of likelihoods, measurement bias models, and simulators, and can effectively handle multiple data sources and time windows.

VI. DISCUSSION

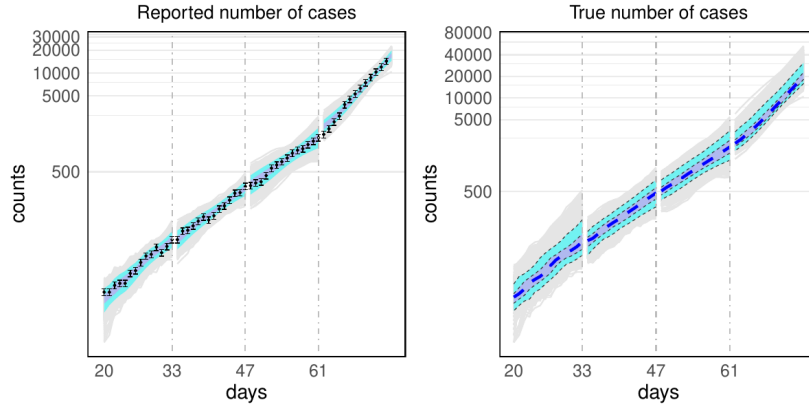
This model-based approach demonstrates the value of considering multiple types of information to constrain uncertainty when producing epidemic forecasts of difficult to observe quantities such as cases. This is accomplished via direct comparison of model outputs with surveillance data produced over the course of an epidemic. For some outputs, there is a need to account for the form of known biases and errors; the simple Bernoulli-based error model that maps modeled cases to reported cases is an example of this – we don’t specify

the actual probability a case is reported, but rather a simple, parameterized, probabilistic mechanism for reporting errors. As additional surveillance data are to be considered in the future, we’ll need to develop models that link the simulated output (or perhaps model parameters) to the actual observation to be used.

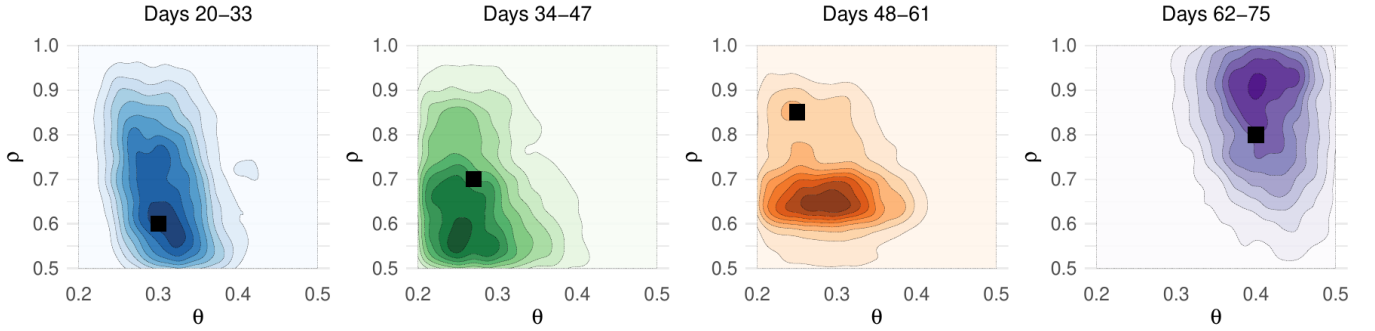
This SMC approach produces plausible epidemic trajectories/histories given the observed data. While a compartmental model is used here, the approach applies equally well to other stochastic simulation models, such as ABMs. The computational demands of ABMs will likely require better efficiency; the use of surrogates for the individual trajectories (e.g. [9]) may be required to refine this current SMC implementation. Unlike more aggregated models, these individual-based models provide a “coordinate system” that more readily maps to reality. Hence, geographically and/or demographically targeted interventions, such as closing schools or vaccinating specific age-groups in certain areas, can be more easily captured in this type of modeling. The trajectories produced from this SMC-based analysis can produce samples of plausible outcomes that allow direct, probabilistic assessment of different intervention strategies. Still, these interventions will need to be encoded into these individual-based models in order to simulate their impact on future trajectories.

We note there is still some work to be done to ensure this approach remains operational as we consider more complicated epidemic models over many time periods. Our example contained only a few parameters for each time period, and even after considering multiple time periods, the resulting parameter space was still small compared to parameter spaces typical of large, geophysical systems [1]. It’s known that SMC has difficulties in settings with large parameter spaces [3]. Therefore, as this effort advances we’ll also integrate reduced model and dimension reduction techniques [4], [35], [36] to ensure that our approach remains feasible as we consider more complex modeling and additional time periods.

Another challenge is to ensure the posterior ensemble of model trajectories remains close to the actual epidemic. Our



(a) Reported case counts (in black circles) are used to constrain the model parameters resulting in the posterior model output shown by grey lines. Purple and cyan shaded regions show 50% and 90% posterior credible intervals on the model output. The left plot shows the posterior on reported case counts and the right plot shows the posterior prediction on actual case counts that are unobserved. The blue dashed line represents the median predicted actual case counts. The vertical dashed grey lines represent the time boundaries for each calibration time window.



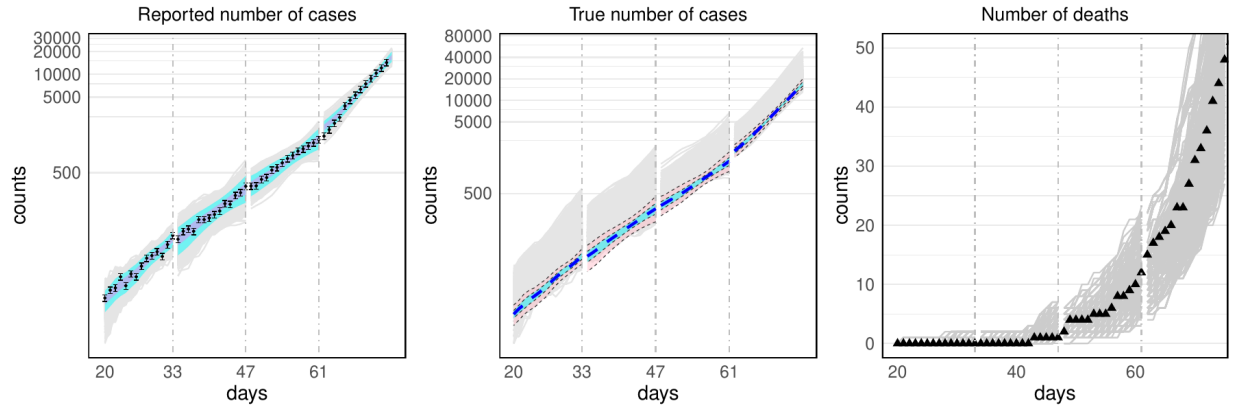
(b) 2d contour plot of the joint posterior of model parameters (θ, ρ) for four time windows respectively for which sequential calibration is performed. Black square denotes the true value of the parameters using which the ground truth is generated.

Fig. 4. Sequential calibration using only reported case counts across four time windows

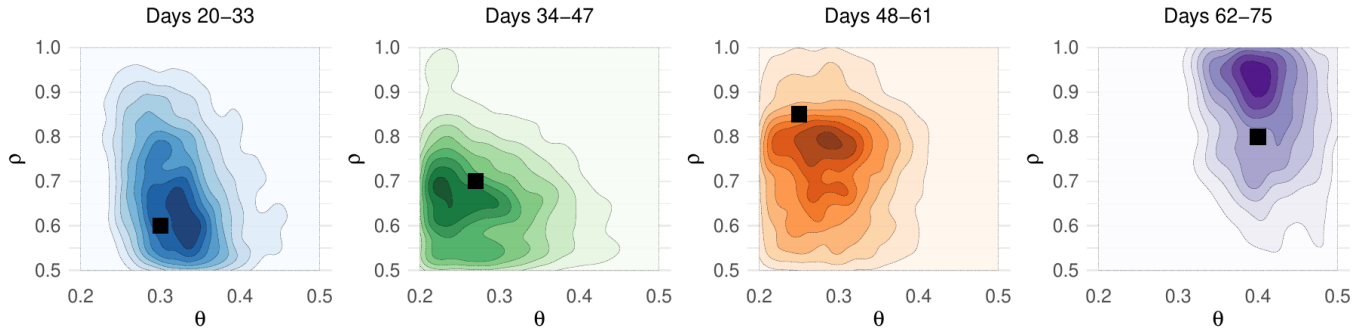
work uses stochastic simulations, which alleviates the potential issue of posterior weights concentrating on just a few draws from the importance distribution. However, if even the most highly weighted trajectories don't track reality, the SMC will produce unreliable predictions. Allowing the parameters to vary over time and allowing for error models to link model output to reality makes it easier for our model to produce trajectories that align with reality. Also, producing larger numbers of trajectories in our importance sampler will combat this potential challenge. Hence there will be a need for experimental platforms (e.g., [6], [19]) to diagnose such problems and develop potential fixes. A fuller consideration of these issues is relegated to future work.

REFERENCES

- [1] Jeffrey L Anderson and Stephen L Anderson. A monte carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Monthly weather review*, 127(12):2741–2758, 1999.
- [2] Ioannis Andrianakis, Ian R Vernon, Nicky McCreesh, Trevelyan J McKinley, Jeremy E Oakley, Rebecca N Nsubuga, Michael Goldstein, and Richard G White. Bayesian history matching of complex infectious disease models using emulation: a tutorial and a case study on hiv in uganda. *PLOS Comput Biol*, 11(1):e1003968, 2015.
- [3] Thomas Bengtsson, Peter Bickel, and Bo Li. Curse-of-dimensionality revisited: Collapse of the particle filter in very large scale systems. In *Probability and statistics: Essays in honor of David A. Freedman*, volume 2, pages 316–335. Institute of Mathematical Statistics, 2008.
- [4] Peter Benner, Albert Cohen, Mario Ohlberger, and Karen Willcox. *Model reduction and approximation: theory and algorithms*. Computational Science and Engineering. SIAM, 2017.
- [5] Olivier Cappé, Simon J Godsill, and Eric Moulines. An overview of existing methods and recent advances in sequential monte carlo. *Proceedings of the IEEE*, 95(5):899–924, 2007.
- [6] Nicholson Collier, Justin M. Wozniak, Abby Stevens, Yadu Babuji, Mickaël Binois, Arindam Fadikar, Alexandra Würth, Kyle Chard, and Jonathan Ozik. Developing distributed high-performance computing capabilities of an open science platform for robust epidemic analysis. In *2023 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW)*, pages 868–877, 2023.
- [7] Jean-Marie Cornuet, Filipe Santos, Mark A Beaumont, Christian P Robert, Jean-Michel Marin, David J Balding, Thomas Guillemaud, and Arnaud Estoup. Inferring population history with diy abc: a user-friendly approach to approximate bayesian computation. *Bioinformatics*, 24(23):2713–2719, 2008.
- [8] Arnaud Doucet, Nando De Freitas, Neil James Gordon, et al. *Sequential Monte Carlo methods in practice*, volume 1. Springer, 2001.
- [9] Arindam Fadikar, Nicholson Collier, Abby Stevens, Jonathan Ozik, Mickaël Binois, and Kok Ben Toh. Trajectory-oriented optimization of stochastic epidemiological models. In *Proceedings of the Winter Simulation Conference, WSC '23*, page 1244–1255. IEEE Press, 2024.
- [10] Arindam Fadikar, Dave Higdon, Jiangzhuo Chen, Bryan Lewis, Srin-



(a) Reported case counts (in black circles) and deaths (in black triangles) are used to constrain the model parameters resulting in the posterior model output shown by grey lines. Purple and cyan shaded regions show 50% and 90% posterior credible intervals on the model output. From left to right, the plots show the posterior on reported case counts, actual case counts that are unobserved, and death counts respectively. The blue dashed line in the middle plot represents the median predicted actual case counts. The vertical dashed grey lines represent the time boundaries for each calibration time window.



(b) 2d contour plot of the joint posterior of model parameters (θ, ρ) for four time windows respectively for which sequential calibration is performed. Black square denotes the true value of the parameters using which the ground truth is generated.

Fig. 5. Sequential calibration using reported case counts and deaths across four time windows

- vasan Venkatramanan, and Madhav Marathe. Calibrating a stochastic, agent-based model using quantile-based emulation. *SIAM/ASA Journal on Uncertainty Quantification*, 6(4):1685–1706, 2018.
- [11] Thomas Flury and Neil Shephard. Bayesian inference based only on simulated likelihood: particle filter analysis of dynamic economic models. *Econometric Theory*, 27(5):933–956, 2011.
- [12] K. M. Gamado, G. Streftaris, and S. Zachary. Modelling under-reporting in epidemics. *Journal of Mathematical Biology*, 69:737–765, 2014.
- [13] K. M. Gostic, L. McGough, E. B. Baskerville, S. Abbott, K. Joshi, C. Tedijanto, and S. Cobey. Practical considerations for measuring the effective reproductive number, r_t . *PLoS Computational Biology*, 16(12):e1008409, 2020.
- [14] Daniel A Henderson, Richard J Boys, Kim J Krishnan, Conor Lawless, and Darren J Wilkinson. Bayesian emulation and calibration of a stochastic computer model of mitochondrial dna deletions in substantia nigra neurons. *Journal of the American Statistical Association*, 104(485):76–87, 2009.
- [15] G. Hettinger, D. Rubin, and J. Huang. Estimating the instantaneous reproduction number with imperfect data: A method to account for case-reporting variation and serial interval uncertainty. *arXiv preprint arXiv:2302.12078*, 2023.
- [16] Anna L. Hotton, Jonathan Ozik, Chaitanya Kaligotla, Nick Collier, Abby Stevens, Aditya S. Khanna, Margaret M. MacDonell, Cheng Wang, David J. LePoire, Young-Soo Chang, Ignacio J. Martinez-Moyano, Bogdan Mucenic, Harold A. Pollack, John A. Schneider, and Charles Macal. Impact of changes in protective behaviors and out-of-household activities by age on COVID-19 transmission and hospitalization in Chicago, Illinois. *Annals of Epidemiology*, 76:165–173, December 2022.
- [17] M. Kennedy and A. O’Hagan. Bayesian calibration of computer models (with discussion). *Journal of the Royal Statistical Society, Series B*, 63:425–464, 2001.
- [18] Jack PC Kleijnen. Kriging metamodeling in simulation: A review. *European journal of operational research*, 192(3):707–716, 2009.
- [19] Dustin Machi, Parantapa Bhattacharya, Stefan Hoops, Jiangzhuo Chen, Henning Mortveit, Srinivasan Venkatramanan, Bryan Lewis, Mandy Wilson, Arindam Fadikar, Tom Maiden, Christopher L. Barrett, and Madhav V. Marathe. Scalable epidemiological workflows to support covid-19 planning and response. In *2021 IEEE International Parallel and Distributed Processing Symposium (IPDPS)*, pages 639–650, 2021.
- [20] April R. Miller, Samin Charepoo, Erik Yan, Ryan W. Frost, Zachary J. Sturgeon, Grace Gibbon, Patrick N. Balius, Cedonia S. Thomas, Melanie A. Schmitt, Daniel A. Sass, James B. Walters, Tracy L. Flood, and Thomas A. Schmitt. Reliability of COVID-19 data: An evaluation and reflection. *PLOS ONE*, 17(11):e0251470, November 2022.
- [21] Jonathan Ozik, Justin M Wozniak, Nicholson Collier, Charles M Macal, and Mickaël Binois. A population data-driven workflow for covid-19 modeling and learning. *The International Journal of High Performance Computing Applications*, 35(5):483–499, 2021.
- [22] Iason Papaioannou, Costas Papadimitriou, and Daniel Straub. Sequential importance sampling for structural reliability analysis. *Structural Safety*, 62:66–75, September 2016.
- [23] K. V. Parag, C. A. Donnelly, and A. E. Zarebski. Quantifying the information in noisy epidemic curves. *Nature Computational Science*, 2(9):584–594, 2022.
- [24] V. E. Pitzer, M. Chitwood, J. Havumaki, N. A. Menzies, S. Perniciaro, J. L. Warren, and T. Cohen. The impact of changes in diagnostic testing practices on estimates of covid-19 transmission in the united states. *American Journal of Epidemiology*, 190(9):1908–1917, 2021.

- [25] C. Quick, R. Dey, and X. Lin. Regression models for understanding covid-19 epidemic dynamics with incomplete data. *Journal of the American Statistical Association*, 116(536):1561–1577, 2021.
- [26] Manuela Runge, Reese AK Richardson, Patrick A Clay, Arielle Bell, Tobias M Holden, Manisha Singam, Natsumi Tsuboyama, Philip Arevalo, Jane Fornoff, Sarah Patrick, et al. Modeling robust covid-19 intensive care unit occupancy thresholds for imposing mitigation to prevent exceeding capacities. *PLOS Global Public Health*, 2(5):e0000308, 2022.
- [27] Manuela et. al. Covid-Chicago Runge. <https://github.com/numalariamodeling/covid-chicago> accessed 8th July.
- [28] Jeffrey Shaman, Alicia Karspeck, Wan Yang, James Tamerius, and Marc Lipsitch. Real-time influenza forecasts during the 2012–2013 season. *Nature Communications*, 4(1):2837, December 2013. Number: 1 Publisher: Nature Publishing Group.
- [29] J. Shi, J. S. Morris, D. M. Rubin, and J. Huang. Robust modeling and inference of disease transmission using error-prone data with application to sars-cov-2. *arXiv preprint arXiv:2212.08282*, 2022.
- [30] Ilya M Sobol. Global sensitivity indices for nonlinear mathematical models and their monte carlo estimates. *Mathematics and computers in simulation*, 55(1):271–280, 2001.
- [31] A. Spannaus, T. Papamarkou, S. Erwin, et al. Inferring the spread of covid-19: the role of time-varying reporting rate in epidemiological modelling. *Scientific Reports*, 12:10761, 2022.
- [32] L. F. White, C. B. Moser, R. N. Thompson, and M. Pagano. Statistical estimation of the reproductive number from case notification data. *American Journal of Epidemiology*, 190(4):611–620, 2021.
- [33] L. F. White and M. Pagano. Reporting errors in infectious disease outbreaks, with an application to pandemic influenza a/h1n1. *Epidemiologic Perspectives & Innovations*, 7(1):1–12, 2010.
- [34] L. F. White, J. Wallinga, L. Finelli, C. Reed, S. Riley, M. Lipsitch, and M. Pagano. Estimation of the reproductive number and the serial interval in early phase of the 2009 influenza a/h1n1 pandemic in the usa. *Influenza and Other Respiratory Viruses*, 3(6):267–276, 2009.
- [35] Nathan Wycoff, Mickael Binois, and Robert B Gramacy. Sensitivity prewarping for local surrogate modeling. *Technometrics*, 64(4):535–547, 2022.
- [36] Nathan Wycoff, Mickael Binois, and Stefan M Wild. Sequential learning of active subspaces. *Journal of Computational and Graphical Statistics*, 30(4):1224–1237, 2021.