



Published in final edited form as:

IEEE Access. 2020 ; 8: 175244–175264. doi:10.1109/access.2020.3019922.

Adaptive Bayesian Learning and Forecasting of Epidemic Evolution—Data Analysis of the COVID-19 Outbreak

DOMENICO GAGLIONE¹ [Member, IEEE], PAOLO BRACA¹ [Senior Member, IEEE], LEONARDO MARIA MILLEFIORI¹ [Member, IEEE], GIOVANNI SOLDI¹, NICOLA FORTI¹, STEFANO MARANO² [Senior Member, IEEE], PETER K. WILLET³ [Fellow, IEEE], KRISHNA R. PATTIPATI³ [Life Fellow, IEEE]

¹NATO STO Centre for Maritime Research and Experimentation (CMRE), 19126 La Spezia, Italy

²Dipartimento di Ingegneria dell'Informazione ed Elettrica e Matematica Applicata (DIEM), University of Salerno, 84084 Fisciano, Italy

³Department of Electrical and Computer Engineering, University of Connecticut, Storrs, CT 06269-4157, USA

Abstract

Since the beginning of 2020, the outbreak of a new strain of Coronavirus has caused hundreds of thousands of deaths and put under heavy pressure the world's most advanced healthcare systems. In order to slow down the spread of the disease, known as COVID-19, and reduce the stress on healthcare structures and intensive care units, many governments have taken drastic and unprecedented measures, such as closure of schools, shops and entire industries, and enforced drastic social distancing regulations, including local and national lockdowns. To effectively address such pandemics in a systematic and informed manner in the future, it is of fundamental importance to develop mathematical models and algorithms to predict the evolution of the spread of the disease to support policy and decision making at the governmental level. There is a strong literature describing the application of Bayesian sequential and adaptive dynamic estimation to surveillance (tracking and prediction) of objects such as missiles and ships; and in this article, we transfer some of its key lessons to epidemiology. We show that we can reliably estimate and forecast the evolution of the infections from daily — and possibly uncertain — publicly available information provided by authorities, e.g., daily numbers of infected and recovered individuals. The proposed method is able to estimate infection and recovery parameters, and to track and predict the epidemiological curve with good accuracy when applied to real data from Lombardia region in Italy, and from the USA. In these scenarios, the mean absolute percentage error computed after the lockdown is on average below 5% when the forecast is at 7 days, and below 10% when the forecast horizon is 14 days.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see <https://creativecommons.org/licenses/by/4.0/>

Corresponding author: Domenico Gaglione (domenico.gaglione@cmre.nato.int).

The associate editor coordinating the review of this manuscript and approving it for publication was Derek Abbott <http://orcid.org/0000-0002-0945-2674>.

Keywords

SARS-CoV-2; Bayesian sequential estimation; ensemble forecasting; compartmental model; pandemic tracking; pandemic prediction

I. INTRODUCTION

A. MOTIVATION AND BACKGROUND

Beginning in early December 2019, Chinese health authorities have been detecting and monitoring an increasing number of pneumonia cases in the city of Wuhan, a province of Hubei. The pneumonia, later named COVID-19, is caused by a new strain of Coronavirus, and is technically referred to as the severe acute respiratory syndrome Coronavirus 2 (SARS-CoV-2) [1]. As of August 23, 2020, more than 23 million people worldwide have been infected, and over 800 thousand have died. In March 2020, a series of events have pushed many governments to take extraordinary social measures. These events include the lack of effective cures and vaccines, the exponentially increasing number of individuals requiring recovery in intensive care units, and the announcement by the World Health Organization (WHO) of a Coronavirus pandemic on March 11, 2020 [2]. The adopted measures included closure of schools, universities, shops, industries, public and cultural places, the prohibition of mass gatherings, travel bans, and extreme social distancing, including local and national lockdowns. The main aim of these measures is to slow down the infection rate and alleviate the pressure on healthcare systems, in order to ensure care to all individuals stricken by the virus. Indeed, after the adoption of these measures, most countries have seen a decrease in the daily numbers of infected individuals. In order to prevent another exponential rise in infections as the restrictive measures are progressively relaxed, it is of crucial importance to develop mathematical models and algorithms to track and forecast the evolution of the infection with acceptable accuracy, which can help authorities to make informed and timely decisions. Improving our ability to model and forecast is also of paramount importance to better address future pandemic outbreaks [3].

The algorithm proposed in this article builds on the concept of compartmental epidemiological models, which assume that a given population is divided into a fixed number of compartments. Each compartment represents an epidemic state that an individual can occupy. The flow dynamics from one compartment to another are modeled as a set of stochastic differential equations that we discretize according to the discrete nature of the available data, i.e., daily update on the number of infected, recovered, dead, etc. In the standard SIR model, proposed in the pioneering study on mathematical theory of epidemics by Kermack and McKendrick [4], it is assumed that the entire population, e.g., of a city, a region, or a nation, is constant and divided into three compartments (population subgroups), namely, susceptible (S), infected (I), and recovered (R) individuals. Moreover, it is assumed that an infected individual infects a susceptible one at a given rate β [5]. Once infected, the individual is removed from the compartment of susceptibles and enters the infected compartment. Each infected person runs through the course of the disease, and eventually is removed from the number of those who are still infected either by recovery or death, thus exiting the system at “recovery” rate γ ; the recovered people are considered permanently

immune.¹ The ratio β/γ is called the contact ratio, and represents the mean number of people the infected individual comes in contact with.

The SIR model is simple, yet very successful and useful in practice. Over the years, several more sophisticated extensions have been proposed to account for more compartments and other salient aspects of the epidemics. For example, a person who comes in contact with an infected individual and contracts the infection might not develop the symptoms immediately but only with a certain delay, called incubation period; in the case of COVID-19, this delay is around 3-15 days with a median of 5.2 days [6]. The SEIR model accounts for this circumstance by adding a further compartment that represents exposed — but not yet contagious — people [7], [8]. That is, susceptible individuals who contract the virus, pass to the exposed compartment (E) before evident symptoms appear and the person is confirmed as infected. The SEIRQ model is a further extension that also accounts for quarantined people [9]. Restriction measures are directly taken into account in the recently proposed SIR-X model [10] that, introducing an additional mechanism, removes susceptibles from the transmission process when the measures become effective.

A critical epidemiological characteristic for the pandemic potential of an emergent respiratory virus is represented by the undocumented, but infectious, cases. In contrast with the documented infectious cases, they often experience mild, limited, or no symptoms at all, and therefore, since they are generally not tested, remain undetected. These are the so-called asymptomatic cases in the context of COVID-19 pandemic. Based on their contagiousness and numbers, they can expose a far greater portion of the population to the virus than would otherwise occur. Li *et al.* [11] present a model-inference framework to estimate the contagiousness and proportion of undocumented infections in China before and after the lockdown in Wuhan.

Most of the compartmental models described so far consider the disease spread inside a unique and single population: a city, a region, a nation. In contrast, metapopulation models add a further spatial dimension, by interpreting the population as a network of multiple spatially separated subpopulations (nodes), e.g., multiple cities in the same region; the connections from one subpopulation to another are represented by movements (“diffusion”) of persons. Such interconnections represent contacts such as commuting to work, second homes, or national and international travels. In such a scenario, the diffusion of the infection is not only caused by the contacts among susceptibles and infectious people within each subpopulation, but also by the spatial interactions among the different subpopulations [12]-[15]. Li *et al.* [11] utilize a stochastic metapopulation model to simulate the spatiotemporal dynamics among 375 Chinese cities. The spatial spread of COVID-19 across cities is captured by the daily number of travelers from a city to another during the Spring Festival before the lockdown. Chinazzi *et al.* [16] model both the domestic (within Wuhan) and the international spread of the Coronavirus epidemic. The effects of the travel bans imposed in the city of Wuhan and the international travel ban adopted by several countries in early February 2020 are estimated. To model the international spread

¹At the time of writing, permanent immunity of recovered COVID-19 patients is uncertain. However, it is a bedrock feature of the SIR epidemiological model.

of the COVID-19 outbreak, the authors employ the stochastic global epidemic and mobility model. This metapopulation model is integrated with real-world data and relies on a network wherein each node represents a subpopulation located near major transportation hubs, e.g., airports; there are more than 3200 subpopulations, in roughly 200 different countries and territories. The degree of connection among subpopulations is represented by the number of people traveling daily among them. Within each subpopulation, there exist four states of the compartmental model, i.e., susceptible, latent (similar to exposed), infectious, and removed. The model generates an ensemble of possible epidemic scenarios described by the number of newly generated infections, time of disease onset in each subpopulation, and the number of traveling infection carriers.

B. CONTRIBUTIONS AND PAPER ORGANIZATION

Most of the aforementioned epidemic models assume that relevant model parameters, e.g., the infection rate β and the recovery rate γ , are time-invariant, and several approaches have been proposed in the literature for tuning or estimating them [17]-[19]. However, the sudden imposition of restriction measures — and their subsequent relaxations — means that a *static* stochastic model is inappropriate. Moreover, even in the absence of dramatic restriction measures, there is no doubt that a time-varying model for the key epidemic parameters would better reflect the ground-truth.²

The main contribution of this article is to propose a Bayesian sequential learning and forecasting framework of the epidemic curve based on the data that authorities provide on a daily basis, e.g., number infected and number recovered. We leverage our recent research on unknown covariance matrix estimation [20] and self-tuning multisensor multitarget tracking [21]. Indeed, similarly to the target tracking problem, where the objective is to automatically detect the time instants when a target sharply maneuvers in order to improve the overall tracking performance, here we aim to closely track the epidemic curve and the model parameters in order to provide reliable and accurate forecast of the contagion. Adapting ideas and tools from those works, our approach to Bayesian sequential learning and forecasting of epidemic evolution is as follows. First, the model parameters are assumed to take on values from rich but prespecified finite sets; their time-evolution is modeled by Markov chains. Second, in order to capture the effects of mitigation strategies (e.g., mobility restriction, lockdown, wearing masks, and social distancing), the marginal posterior distributions of both the variable states (number of infected and recovered people) and model parameters (infection rate β and recovery rate γ), are calculated at each time by means of recursive prediction and update formulae. Finally, we develop an efficient implementation of the proposed method based on mixture models and provide a concrete example of application using the stochastic SIR model. The proposed method is validated on real datasets acquired during the recent COVID-19 outbreak in the Lombardia region, Italy, and in the USA. As we shall see, even adopting the simple stochastic SIR model, we obtain superior forecast accuracy when compared to prediction algorithms that use time-invariant parameter models. The approach developed in this article is general enough to be applied

²This applies, e.g., to the influenza virus, as its infection rate periodically increases and decreases depending on the season.

to more-sophisticated stochastic epidemiological models [8]-[10], including more-complex metapopulation models [11], [16]; these extensions are left for future investigations.

The remainder of the paper is organized as follows. In Section II, we describe a general Bayesian adaptive framework that can be tailored to any discrete-time epidemiological model, and in Section III we propose an implementation thereof based on the use of mixture models. In Section IV, we develop the mixture-based Bayesian sequential approach in the context of the stochastic SIR model. Section V presents results using synthetic as well as real data. Finally, in Section VI we provide some conclusions and possible directions for future investigations.

C. NOTATION

Vectors are denoted by boldface lower-case letters (e.g., \mathbf{a}), matrices by boldface upper-case letters (e.g., \mathbf{A}), and sets by calligraphic letters (e.g., \mathcal{A}). The transpose is written as $(\cdot)^T$. We write $\text{diag}(a_1, \dots, a_N)$ for an $N \times N$ diagonal matrix with diagonal entries a_1, \dots, a_N , \mathbf{I}_N for the $N \times N$ identity matrix, $\mathbf{1}_N$ for the N -dimensional vector of all ones, and $\mathbf{0}$ for the zero vector. $E[\cdot]$ denotes statistical expectation, and $\mathcal{P}(\cdot)$ refers to both the probability density function (pdf) of a continuous random variable or vector and the probability mass function (pmf) of a discrete random variable or vector; the difference will be clear from the context. $\mathcal{N}(\boldsymbol{\mu}, \mathbf{C})$ indicates a Gaussian distributed vector with mean $\boldsymbol{\mu}$ and covariance matrix \mathbf{C} , and $\mathcal{U}(a, b)$ represents a uniformly distributed variable between a and b . Finally, $\mathcal{N}(\mathbf{x}; \boldsymbol{\mu}, \mathbf{C})$ refers to a multivariate Gaussian pdf of random vector \mathbf{x} with mean $\boldsymbol{\mu}$ and covariance matrix \mathbf{C} .

II. PROPOSED ALGORITHM

We present a sequential Bayesian framework that, at each time interval, jointly computes the posterior distribution of S unknown time-varying states and of M unknown time-varying parameters. These unknown quantities are inferred at times t_k , with $k \in \{1, 2, \dots\}$, using noisy observations (e.g., information on the number of infected, discharged COVID-19 patients from the hospitals, dead). We assume that the time interval $\Delta t \triangleq t_k - t_{k-1}$ between consecutive observations is one day, unless otherwise stated. We denote $\mathbf{x}_k \triangleq [x_{1,k}, \dots, x_{S,k}]^T \in \mathcal{X} \subseteq \mathbb{R}^S$ as the state vector comprising the S epidemic states $x_{s,k}$ at time t_k (e.g., numbers of infected and recovered individuals), and $\boldsymbol{\theta}_k \triangleq [\theta_{1,k}, \dots, \theta_{M,k}]^T \in \mathcal{Q} \subseteq \mathbb{R}^M$ as the parameter vector comprising the M model parameters $\theta_{m,k}$ at time t_k (e.g., infection and recovery rates).

The objective of the proposed algorithm is twofold: to estimate, at each time k , the epidemic state vector \mathbf{x}_k and the model parameter vector $\boldsymbol{\theta}_k$; and, at a fixed time k , to forecast the epidemic evolution up to time $k + K$ with associated uncertainty in the form of prediction variance. Both tasks are based on the past and present observations. Hereafter, Section II-A describes the dynamic and observation models, and Section II-B and Section II-C present the proposed Bayesian sequential estimation and forecasting tasks, respectively. The reader who is already familiar with dynamic estimation of a *hybrid* state, that is, comprising the state of the dynamic model and nuisance parameters [20]-[22], might skip ahead to Section II-C.

A. DYNAMIC AND OBSERVATION MODELS

The dynamic model that describes the evolution of the epidemic is formally expressed as

$$\mathbf{x}_k = \mathbf{f}(\boldsymbol{\theta}_k, \boldsymbol{\theta}_{k-1}, \mathbf{x}_{k-1}; \mathbf{u}_k), \quad (1)$$

where \mathbf{u}_k is a random vector — whose dimension depends on S , M , and $\mathbf{f}(\cdot)$ — with known distribution modeling the stochastic variation of the epidemic state in the time interval Δt . Note that function $\mathbf{f}(\cdot)$ might embed additional known (either time-varying or time-invariant) parameters. We assume that, conditioned on $\boldsymbol{\theta}_k$, $\boldsymbol{\theta}_{k-1}$, and \mathbf{x}_{k-1} , the state vector \mathbf{x}_k is independent of the previous states and parameters, that is,

$$\begin{aligned} \mathcal{P}(\mathbf{x}_k | \boldsymbol{\theta}_k, \boldsymbol{\theta}_{k-1}, \mathbf{x}_{k-1}, \boldsymbol{\theta}_{k-2}, \mathbf{x}_{k-2}, \dots, \boldsymbol{\theta}_1, \mathbf{x}_1) \\ = \mathcal{P}(\mathbf{x}_k | \boldsymbol{\theta}_k, \boldsymbol{\theta}_{k-1}, \mathbf{x}_{k-1}). \end{aligned} \quad (2)$$

Given appropriate initial conditions, the pdf in (2) is fully determined by the dynamic model in (1) and the statistics of \mathbf{u}_k . Moreover, we assume that the parameter vector evolves according to a first-order Markov model fully described by the transition pdf $\mathcal{P}(\boldsymbol{\theta}_k | \boldsymbol{\theta}_{k-1})$, assumed known. From these assumptions, it follows that the adopted Bayesian framework is a hierarchical Markov model (an event-driven dynamic process): firstly, the parameter vector evolves according to the Markov model described by $\mathcal{P}(\boldsymbol{\theta}_k | \boldsymbol{\theta}_{k-1})$; then the state vector evolves, given the current parameter vector as well as the previous state and parameter vectors, according to (1). Furthermore, it is easy to verify that

$$\begin{aligned} \mathcal{P}(\mathbf{x}_k, \boldsymbol{\theta}_k | \mathbf{x}_{k-1}, \boldsymbol{\theta}_{k-1}, \mathbf{x}_{k-2}, \boldsymbol{\theta}_{k-2}, \dots, \mathbf{x}_1, \boldsymbol{\theta}_1) \\ = \mathcal{P}(\mathbf{x}_k, \boldsymbol{\theta}_k | \mathbf{x}_{k-1}, \boldsymbol{\theta}_{k-1}), \end{aligned}$$

that is, the joint evolution of \mathbf{x}_k and $\boldsymbol{\theta}_k$ follows a first-order Markov model.

The observation vector at time k is denoted by $\mathbf{z}_k \in \mathbb{R}^B$, and consists of up-to-date information on the state of the epidemic at time k . To take into account the randomness unavoidably present in real-world measurements, it is assumed that this information is *uncertain*, i.e., affected by noise (e.g., due to data collection errors, biases, holidays), and is modeled as

$$\mathbf{z}_k = \mathbf{h}(\mathbf{x}_k, \boldsymbol{\theta}_k; \mathbf{v}_k), \quad (3)$$

where \mathbf{v}_k is a random vector with known distribution and whose dimension depends on S , M , B , and $\mathbf{h}(\cdot)$. The model in (3) and the statistics of \mathbf{v}_k determine the likelihood $\mathcal{P}(\mathbf{z}_k | \mathbf{x}_k, \boldsymbol{\theta}_k)$.

For convenience, we define the vector $\mathbf{z}_{1:k} \triangleq [\mathbf{z}_1^\top, \dots, \mathbf{z}_k^\top]^\top$ containing all the observations up to time k , that is, past and present observations.

B. BAYESIAN SEQUENTIAL ESTIMATION

The basic principles of Bayesian sequential estimation are now recalled. The reader is referred to [23] for further details. In the Bayesian setting, the estimation of state and

parameters amounts to calculating the posterior pdf $\mathcal{P}(\mathbf{x}_k | \mathbf{z}_{1:k})$ of the state vector \mathbf{x}_k , and the posterior pdf $\mathcal{P}(\boldsymbol{\theta}_k | \mathbf{z}_{1:k})$ of the parameter vector $\boldsymbol{\theta}_k$, respectively. The minimum mean square estimators (MMSEs) of \mathbf{x}_k and $\boldsymbol{\theta}_k$ are given by [23, Ch. 4]

$$\hat{\mathbf{x}}_k^E \triangleq \int \mathbf{x}_k \mathcal{P}(\mathbf{x}_k | \mathbf{z}_{1:k}) d\mathbf{x}_k, \quad (4)$$

and

$$\hat{\boldsymbol{\theta}}_k^E \triangleq \int \boldsymbol{\theta}_k \mathcal{P}(\boldsymbol{\theta}_k | \mathbf{z}_{1:k}) d\boldsymbol{\theta}_k, \quad (5)$$

respectively.³ We further note that, using the law of total probability, the pdf $\mathcal{P}(\mathbf{x}_k | \mathbf{z}_{1:k})$ can be expressed as

$$\begin{aligned} \mathcal{P}(\mathbf{x}_k | \mathbf{z}_{1:k}) &= \int \mathcal{P}(\mathbf{x}_k, \boldsymbol{\theta}_k | \mathbf{z}_{1:k}) d\boldsymbol{\theta}_k \\ &= \int \mathcal{P}(\mathbf{x}_k | \boldsymbol{\theta}_k, \mathbf{z}_{1:k}) \mathcal{P}(\boldsymbol{\theta}_k | \mathbf{z}_{1:k}) d\boldsymbol{\theta}_k. \end{aligned} \quad (6)$$

Thus, the estimation problem boils down to calculation of the posterior pdfs $\mathcal{P}(\mathbf{x}_k | \boldsymbol{\theta}_k, \mathbf{z}_{1:k})$ and $\mathcal{P}(\boldsymbol{\theta}_k | \mathbf{z}_{1:k})$. In the following, we show how they can be obtained sequentially through the implementation of recursive update and prediction steps.

1) UPDATE STEP—Let us assume that we know the pdf of the parameter vector at time k given the observations up to time $k-1$, i.e., $\mathcal{P}(\boldsymbol{\theta}_k | \mathbf{z}_{1:k-1})$, and the pdf of the state vector at time k given the parameter vector at time k and the observations up to time $k-1$, i.e., $\mathcal{P}(\mathbf{x}_k | \boldsymbol{\theta}_k, \mathbf{z}_{1:k-1})$. Then, when a new observation \mathbf{z}_k becomes available, the parameter vector pdf and the state vector pdf are updated through Bayes' rule as

$$\begin{aligned} \mathcal{P}(\boldsymbol{\theta}_k | \mathbf{z}_{1:k}) &= \frac{\mathcal{P}(\mathbf{z}_k, \boldsymbol{\theta}_k | \mathbf{z}_{1:k-1})}{\int \mathcal{P}(\mathbf{z}_k, \boldsymbol{\theta}'_k | \mathbf{z}_{1:k-1}) d\boldsymbol{\theta}'_k} \\ &= \frac{\mathcal{P}(\mathbf{z}_k | \boldsymbol{\theta}_k, \mathbf{z}_{1:k-1}) \mathcal{P}(\boldsymbol{\theta}_k | \mathbf{z}_{1:k-1})}{\int \mathcal{P}(\mathbf{z}_k | \boldsymbol{\theta}'_k, \mathbf{z}_{1:k-1}) \mathcal{P}(\boldsymbol{\theta}'_k | \mathbf{z}_{1:k-1}) d\boldsymbol{\theta}'_k}, \end{aligned} \quad (7)$$

and

$$\begin{aligned} \mathcal{P}(\mathbf{x}_k | \boldsymbol{\theta}_k, \mathbf{z}_{1:k}) &= \frac{\mathcal{P}(\mathbf{z}_k, \mathbf{x}_k | \boldsymbol{\theta}_k, \mathbf{z}_{1:k-1})}{\mathcal{P}(\mathbf{z}_k | \boldsymbol{\theta}_k, \mathbf{z}_{1:k-1})} \\ &= \frac{\mathcal{P}(\mathbf{z}_k | \mathbf{x}_k, \boldsymbol{\theta}_k, \mathbf{z}_{1:k-1}) \mathcal{P}(\mathbf{x}_k | \boldsymbol{\theta}_k, \mathbf{z}_{1:k-1})}{\mathcal{P}(\mathbf{z}_k | \boldsymbol{\theta}_k, \mathbf{z}_{1:k-1})} \\ &= \frac{\mathcal{P}(\mathbf{z}_k | \mathbf{x}_k, \boldsymbol{\theta}_k) \mathcal{P}(\mathbf{x}_k | \boldsymbol{\theta}_k, \mathbf{z}_{1:k-1})}{\mathcal{P}(\mathbf{z}_k | \boldsymbol{\theta}_k, \mathbf{z}_{1:k-1})}, \end{aligned} \quad (8)$$

³The superscript E stands for *estimate*, and is used to distinguish the estimate from the *forecast*, later identified by the superscript F.

respectively, where the last equality of (8) exploits the assumption that the observation at time k is conditionally independent of all the previous observations, given the state and parameter vectors at time k , i.e., $\mathcal{P}(\mathbf{z}_k | \mathbf{x}_k, \boldsymbol{\theta}_k, \mathbf{z}_{1:k-1}) = \mathcal{P}(\mathbf{z}_k | \mathbf{x}_k, \boldsymbol{\theta}_k)$. Using the same assumption and the law of total probability, the pdf $\mathcal{P}(\mathbf{z}_k | \boldsymbol{\theta}_k, \mathbf{z}_{1:k-1})$ appearing in (7) is calculated as

$$\begin{aligned} & \mathcal{P}(\mathbf{z}_k | \boldsymbol{\theta}_k, \mathbf{z}_{1:k-1}) \\ &= \int \mathcal{P}(\mathbf{z}_k, \mathbf{x}_k | \boldsymbol{\theta}_k, \mathbf{z}_{1:k-1}) d\mathbf{x}_k \\ &= \int \mathcal{P}(\mathbf{z}_k | \mathbf{x}_k, \boldsymbol{\theta}_k, \mathbf{z}_{1:k-1}) \mathcal{P}(\mathbf{x}_k | \boldsymbol{\theta}_k, \mathbf{z}_{1:k-1}) d\mathbf{x}_k \\ &= \int \mathcal{P}(\mathbf{z}_k | \mathbf{x}_k, \boldsymbol{\theta}_k) \mathcal{P}(\mathbf{x}_k | \boldsymbol{\theta}_k, \mathbf{z}_{1:k-1}) d\mathbf{x}_k. \end{aligned} \quad (9)$$

2) PREDICTION STEP—In the prediction step, we assume that the posterior pdf $\mathcal{P}(\boldsymbol{\theta}_k | \mathbf{z}_{1:k})$ in (7) and the posterior pdf $\mathcal{P}(\mathbf{x}_k | \boldsymbol{\theta}_k, \mathbf{z}_{1:k})$ in (8) are known, and derive the pdfs $\mathcal{P}(\boldsymbol{\theta}_{k+1} | \mathbf{z}_{1:k})$ and $\mathcal{P}(\mathbf{x}_{k+1} | \boldsymbol{\theta}_{k+1}, \mathbf{z}_{1:k})$. The former, i.e., the pdf of the predicted parameter vector at time $k+1$, is obtained using the law of total probability and the Markovian assumption as follows:

$$\begin{aligned} \mathcal{P}(\boldsymbol{\theta}_{k+1} | \mathbf{z}_{1:k}) &= \int \mathcal{P}(\boldsymbol{\theta}_{k+1}, \boldsymbol{\theta}_k | \mathbf{z}_{1:k}) d\boldsymbol{\theta}_k \\ &= \int \mathcal{P}(\boldsymbol{\theta}_{k+1} | \boldsymbol{\theta}_k) \mathcal{P}(\boldsymbol{\theta}_k | \mathbf{z}_{1:k}) d\boldsymbol{\theta}_k. \end{aligned} \quad (10)$$

Analogously, using the law of total probability, the pdf of the predicted state vector at time $k+1$ is given by

$$\begin{aligned} & \mathcal{P}(\mathbf{x}_{k+1} | \boldsymbol{\theta}_{k+1}, \mathbf{z}_{1:k}) \\ &= \int \mathcal{P}(\mathbf{x}_{k+1}, \boldsymbol{\theta}_k | \boldsymbol{\theta}_{k+1}, \mathbf{z}_{1:k}) d\boldsymbol{\theta}_k \\ &= \int \mathcal{P}(\mathbf{x}_{k+1} | \boldsymbol{\theta}_{k+1}, \boldsymbol{\theta}_k, \mathbf{z}_{1:k}) \mathcal{P}(\boldsymbol{\theta}_k | \boldsymbol{\theta}_{k+1}, \mathbf{z}_{1:k}) d\boldsymbol{\theta}_k. \end{aligned} \quad (11)$$

The first term within the integral (11), i.e., $\mathcal{P}(\mathbf{x}_{k+1} | \boldsymbol{\theta}_{k+1}, \boldsymbol{\theta}_k, \mathbf{z}_{1:k})$, is calculated using the law of total probability and assuming that \mathbf{x}_{k+1} is conditionally independent of $\mathbf{z}_{1:k}$ given $\boldsymbol{\theta}_{k+1}$, $\boldsymbol{\theta}_k$, and \mathbf{x}_k , i.e., $\mathcal{P}(\mathbf{x}_{k+1} | \boldsymbol{\theta}_{k+1}, \boldsymbol{\theta}_k, \mathbf{x}_k, \mathbf{z}_{1:k}) = \mathcal{P}(\mathbf{x}_{k+1} | \boldsymbol{\theta}_{k+1}, \boldsymbol{\theta}_k, \mathbf{x}_k)$. Thus,

$$\begin{aligned} & \mathcal{P}(\mathbf{x}_{k+1} | \boldsymbol{\theta}_{k+1}, \boldsymbol{\theta}_k, \mathbf{z}_{1:k}) \\ &= \int \mathcal{P}(\mathbf{x}_{k+1}, \mathbf{x}_k | \boldsymbol{\theta}_{k+1}, \boldsymbol{\theta}_k, \mathbf{z}_{1:k}) d\mathbf{x}_k \\ &= \int \mathcal{P}(\mathbf{x}_{k+1} | \boldsymbol{\theta}_{k+1}, \boldsymbol{\theta}_k, \mathbf{x}_k) \mathcal{P}(\mathbf{x}_k | \boldsymbol{\theta}_{k+1}, \boldsymbol{\theta}_k, \mathbf{z}_{1:k}) d\mathbf{x}_k \\ &= \int \mathcal{P}(\mathbf{x}_{k+1} | \boldsymbol{\theta}_{k+1}, \boldsymbol{\theta}_k, \mathbf{x}_k) \mathcal{P}(\mathbf{x}_k | \boldsymbol{\theta}_k, \mathbf{z}_{1:k}) d\mathbf{x}_k. \end{aligned} \quad (12)$$

The last step follows from the Markov property of the parameter vector. Indeed, the pdf $\mathcal{P}(\mathbf{x}_k | \boldsymbol{\theta}_{k+1}, \boldsymbol{\theta}_k, \mathbf{z}_{1:k})$ can be calculated as

$$\begin{aligned}
& \mathcal{P}(\mathbf{x}_k | \boldsymbol{\theta}_{k+1}, \boldsymbol{\theta}_k, \mathbf{z}_{1:k}) \\
&= \frac{\mathcal{P}(\mathbf{x}_k, \boldsymbol{\theta}_{k+1}, \boldsymbol{\theta}_k | \mathbf{z}_{1:k})}{\mathcal{P}(\boldsymbol{\theta}_{k+1}, \boldsymbol{\theta}_k | \mathbf{z}_{1:k})} \\
&= \frac{\mathcal{P}(\boldsymbol{\theta}_{k+1} | \mathbf{x}_k, \boldsymbol{\theta}_k, \mathbf{z}_{1:k}) \mathcal{P}(\mathbf{x}_k, \boldsymbol{\theta}_k | \mathbf{z}_{1:k})}{\mathcal{P}(\boldsymbol{\theta}_{k+1} | \boldsymbol{\theta}_k, \mathbf{z}_{1:k}) \mathcal{P}(\boldsymbol{\theta}_k | \mathbf{z}_{1:k})} \\
&= \frac{\mathcal{P}(\boldsymbol{\theta}_{k+1} | \mathbf{x}_k, \boldsymbol{\theta}_k, \mathbf{z}_{1:k}) \mathcal{P}(\mathbf{x}_k | \boldsymbol{\theta}_k, \mathbf{z}_{1:k}) \mathcal{P}(\boldsymbol{\theta}_k | \mathbf{z}_{1:k})}{\mathcal{P}(\boldsymbol{\theta}_{k+1} | \boldsymbol{\theta}_k, \mathbf{z}_{1:k}) \mathcal{P}(\boldsymbol{\theta}_k | \mathbf{z}_{1:k})} \\
&= \frac{\mathcal{P}(\boldsymbol{\theta}_{k+1} | \boldsymbol{\theta}_k) \mathcal{P}(\mathbf{x}_k | \boldsymbol{\theta}_k, \mathbf{z}_{1:k}) \mathcal{P}(\boldsymbol{\theta}_k | \mathbf{z}_{1:k})}{\mathcal{P}(\boldsymbol{\theta}_{k+1} | \boldsymbol{\theta}_k) \mathcal{P}(\boldsymbol{\theta}_k | \mathbf{z}_{1:k})} \\
&= \mathcal{P}(\mathbf{x}_k | \boldsymbol{\theta}_k, \mathbf{z}_{1:k}).
\end{aligned}$$

The second term within the integral (11), i.e., $\mathcal{P}(\boldsymbol{\theta}_k | \boldsymbol{\theta}_{k+1}, \mathbf{z}_{1:k})$, is obtained using Bayes' rule and exploiting again the Markov property of the parameter vector, that is,

$$\begin{aligned}
\mathcal{P}(\boldsymbol{\theta}_k | \boldsymbol{\theta}_{k+1}, \mathbf{z}_{1:k}) &= \frac{\mathcal{P}(\boldsymbol{\theta}_{k+1}, \boldsymbol{\theta}_k | \mathbf{z}_{1:k})}{\mathcal{P}(\boldsymbol{\theta}_{k+1} | \mathbf{z}_{1:k})} \\
&= \frac{\mathcal{P}(\boldsymbol{\theta}_{k+1} | \boldsymbol{\theta}_k, \mathbf{z}_{1:k}) \mathcal{P}(\boldsymbol{\theta}_k | \mathbf{z}_{1:k})}{\mathcal{P}(\boldsymbol{\theta}_{k+1} | \mathbf{z}_{1:k})} \\
&= \frac{\mathcal{P}(\boldsymbol{\theta}_{k+1} | \boldsymbol{\theta}_k) \mathcal{P}(\boldsymbol{\theta}_k | \mathbf{z}_{1:k})}{\mathcal{P}(\boldsymbol{\theta}_{k+1} | \mathbf{z}_{1:k})},
\end{aligned} \tag{13}$$

where $\mathcal{P}(\boldsymbol{\theta}_{k+1} | \mathbf{z}_{1:k})$ is given by (10).

C. FORECASTING

As the model we consider is nonlinear, the forecast of the epidemic evolution is assessed numerically through a methodology known as ensemble forecasting [24]-[28], that consists of generating a collection of possible evolutions of the epidemic — given the state and parameter vectors estimated so far — and provide a single mean forecast with associated uncertainty. Let us assume that the latest available observation is \mathbf{z}_k , and that the posterior pdfs $\mathcal{P}(\boldsymbol{\theta}_k | \mathbf{z}_{1:k})$ and $\mathcal{P}(\mathbf{x}_k | \mathbf{z}_{1:k}) = \int \mathcal{P}(\mathbf{x}_k | \boldsymbol{\theta}_k, \mathbf{z}_{1:k}) \mathcal{P}(\boldsymbol{\theta}_k | \mathbf{z}_{1:k}) d\boldsymbol{\theta}_k$ (cf. (6)) are known; the proposed ensemble forecasting approach is a Monte-Carlo technique that samples the posterior distribution of \mathbf{x}_k and $\boldsymbol{\theta}_k$, and evolves these sampled initial vectors up to time $k + K$, where K is the forecast horizon. Specifically, let $\tilde{\mathbf{x}}_k^{(j)}$ and $\tilde{\boldsymbol{\theta}}_k^{(j)}$ be the j^{th} state vector and parameter vector samples extracted from $\mathcal{P}(\mathbf{x}_k | \mathbf{z}_{1:k})$ and $\mathcal{P}(\boldsymbol{\theta}_k | \mathbf{z}_{1:k})$, respectively, where $j \in \mathcal{J} \triangleq \{1, \dots, J\}$ and J is the ensemble size. The sampled state and parameter vectors are then allowed to evolve⁴ according to the state vector forecast transition distribution $\mathcal{P}_{\mathbf{F}}(\tilde{\mathbf{x}}_{k'}^{(j)} | \tilde{\boldsymbol{\theta}}_{k'}^{(j)}, \tilde{\boldsymbol{\theta}}_{k'-1}^{(j)}, \tilde{\mathbf{x}}_{k'-1}^{(j)})$ and the parameter vector forecast transition distribution $\mathcal{P}_{\mathbf{F}}(\tilde{\boldsymbol{\theta}}_{k'}^{(j)} | \tilde{\boldsymbol{\theta}}_{k'-1}^{(j)})$, respectively, for $k' \in \{k+1, \dots, k+K\}$. We observe that these transition distributions can be equal to the transition

⁴By “evolving” is meant a simple procedure of recursively (in time) generating random variables according to a transition distribution; colloquially, this is a process of “rolling the dice”.

distributions used within the Bayesian sequential estimation procedure described in the previous section, that is, $\mathcal{P}(\mathbf{x}_k | \boldsymbol{\theta}_k, \boldsymbol{\theta}_{k-1}, \mathbf{x}_{k-1})$ and $\mathcal{P}(\boldsymbol{\theta}_k | \boldsymbol{\theta}_{k-1})$, respectively, or can be suitably devised to improve the forecast performance. Finally, defining the ensemble state matrix as $\tilde{\mathbf{X}}_k \triangleq [\tilde{\mathbf{x}}_k^{(1)}, \dots, \tilde{\mathbf{x}}_k^{(J)}] \in \mathcal{X}^{1 \times J} \subseteq \mathbb{R}^{S \times J}$, and the ensemble parameter matrix as $\tilde{\boldsymbol{\Theta}}_k \triangleq [\tilde{\boldsymbol{\theta}}_k^{(1)}, \dots, \tilde{\boldsymbol{\theta}}_k^{(J)}] \in \mathcal{Q}^{1 \times J} \subseteq \mathbb{R}^{M \times J}$, the mean of the epidemic state and model parameter at any time step $k' \in \{k+1, \dots, k+K\}$ can be calculated as sample means of $\tilde{\mathbf{X}}_{k'}$ and $\tilde{\boldsymbol{\Theta}}_{k'}$, respectively; that is [29],

$$\hat{\mathbf{x}}_{k'}^F \triangleq \frac{1}{J} \tilde{\mathbf{X}}_{k'} \cdot \mathbf{1}_J, \quad \hat{\boldsymbol{\theta}}_{k'}^F \triangleq \frac{1}{J} \tilde{\boldsymbol{\Theta}}_{k'} \cdot \mathbf{1}_J. \quad (14)$$

Higher order moments, such as sample covariance matrices, can also be computed [29].

III. MIXTURE MODEL IMPLEMENTATION

This section describes a mixture model implementation — similar to the approach proposed in [20] — of the Bayesian sequential estimation procedure presented in Section II-B. For computational efficiency, the first step is the discretization of the parameter vector $\boldsymbol{\theta}_k = [\theta_{1,k}, \dots, \theta_{M,k}]^T$, such that each element $\theta_{m,k}$, $m \in \{1, \dots, M\}$, takes on values from a finite set $\mathcal{D}_m \triangleq \{\vartheta_1^{(m)}, \dots, \vartheta_{D_m}^{(m)}\}$. It follows that $\boldsymbol{\theta}_k \in \mathcal{D}$, where $\mathcal{D} \triangleq \mathcal{D}_1 \times \dots \times \mathcal{D}_M$ is the discretized finite set with cardinality $D \triangleq \prod_{m=1}^M D_m$. We note that all the expressions in Section II-B remain valid, provided that integrals $\int d\boldsymbol{\theta}$ are replaced with summations $\sum_{\boldsymbol{\theta}}$; e.g., the MMSE estimator of $\boldsymbol{\theta}_k$ in (5) is rewritten as

$$\hat{\boldsymbol{\theta}}_k^E \triangleq \sum_{\boldsymbol{\theta}_k \in \mathcal{D}} \boldsymbol{\theta}_k \mathcal{P}(\boldsymbol{\theta}_k | \mathbf{z}_{1:k}). \quad (15)$$

The key aspect of the formulation is to model the pdf $\mathcal{P}(\mathbf{x}_k | \boldsymbol{\theta}_k, \mathbf{z}_{1:k-1})$ as a mixture of N components, that is,

$$\mathcal{P}(\mathbf{x}_k | \boldsymbol{\theta}_k, \mathbf{z}_{1:k-1}) = \sum_{n=1}^N w_k^{(n, \boldsymbol{\theta}_k)} \mathcal{P}(\mathbf{x}_k | N_k = n, \boldsymbol{\theta}_k, \mathbf{z}_{1:k-1}), \quad (16)$$

where $w_k^{(n, \boldsymbol{\theta}_k)} \triangleq \mathcal{P}(N_k = n | \boldsymbol{\theta}_k, \mathbf{z}_{1:k-1})$ and $\mathcal{P}(\mathbf{x}_k | N_k = n, \boldsymbol{\theta}_k, \mathbf{z}_{1:k-1})$ are weight and pdf of the n^{th} component, respectively. The auxiliary variable $N_k \in \{1, \dots, N\}$ models the switch between the N mixture components; hereafter, for notational convenience, we will simply use n to denote that N_k takes on the value n , i.e., $N_k = n$. In the next two subsections, exploiting the development presented in Section II-B1 and Section II-B2, we provide the expressions for the sequential update and prediction of the state and parameter vectors according to the mixture model.

A. UPDATE STEP

The posterior pdf $\mathcal{P}(\mathbf{x}_k | \boldsymbol{\theta}_k, \mathbf{z}_{1:k})$ appearing in (8) can be written through the law of total probability as

$$\mathcal{P}(\mathbf{x}_k | \boldsymbol{\theta}_k, \mathbf{z}_{1:k}) = \sum_{n=1}^N w_{k|k}^{(n, \boldsymbol{\theta}_k)} \mathcal{P}(\mathbf{x}_k | n, \boldsymbol{\theta}_k, \mathbf{z}_{1:k}). \quad (17)$$

The updated weight $w_{k|k}^{(n, \boldsymbol{\theta}_k)} = \mathcal{P}(n | \boldsymbol{\theta}_k, \mathbf{z}_{1:k})$ is calculated through Bayes's rule as

$$w_{k|k}^{(n, \boldsymbol{\theta}_k)} = \frac{\alpha_{k|k-1}^{(n, \boldsymbol{\theta}_k)} w_{k|k-1}^{(n, \boldsymbol{\theta}_k)}}{\sum_{n'=1}^N \alpha_{k|k-1}^{(n', \boldsymbol{\theta}_k)} w_{k|k-1}^{(n', \boldsymbol{\theta}_k)}}, \quad (18)$$

where the update coefficient $\alpha_{k|k-1}^{(n, \boldsymbol{\theta}_k)} \triangleq \mathcal{P}(\mathbf{z}_k | n, \boldsymbol{\theta}_k, \mathbf{z}_{1:k-1})$ can be derived from (9) assuming the conditional independence of the observation \mathbf{z}_k from the previous observations $\mathbf{z}_{1:k-1}$ and the specific mixand n , given \mathbf{x}_k and $\boldsymbol{\theta}_k$, i.e., $\mathcal{P}(\mathbf{z}_k | n, \mathbf{x}_k, \boldsymbol{\theta}_k, \mathbf{z}_{1:k-1}) = \mathcal{P}(\mathbf{z}_k | \mathbf{x}_k, \boldsymbol{\theta}_k)$. That is,

$$\begin{aligned} \alpha_{k|k-1}^{(n, \boldsymbol{\theta}_k)} &= \mathcal{P}(\mathbf{z}_k | n, \boldsymbol{\theta}_k, \mathbf{z}_{1:k-1}) \\ &= \int \mathcal{P}(\mathbf{z}_k | \mathbf{x}_k, \boldsymbol{\theta}_k) \mathcal{P}(\mathbf{x}_k | n, \boldsymbol{\theta}_k, \mathbf{z}_{1:k-1}) d\mathbf{x}_k. \end{aligned} \quad (19)$$

Using the same assumption, the posterior pdf of the n^{th} mixture component is calculated from (8) as

$$\mathcal{P}(\mathbf{x}_k | n, \boldsymbol{\theta}_k, \mathbf{z}_{1:k}) = \frac{\mathcal{P}(\mathbf{z}_k | \mathbf{x}_k, \boldsymbol{\theta}_k) \mathcal{P}(\mathbf{x}_k | n, \boldsymbol{\theta}_k, \mathbf{z}_{1:k-1})}{\alpha_{k|k-1}^{(n, \boldsymbol{\theta}_k)}}. \quad (20)$$

Then, using (7) with integrals replaced by summations, the posterior pmf of the parameter vector is obtained as

$$\mathcal{P}(\boldsymbol{\theta}_k | \mathbf{z}_{1:k}) = \frac{\mathcal{P}(\boldsymbol{\theta}_k | \mathbf{z}_{1:k-1}) \mathcal{P}(\mathbf{z}_k | \boldsymbol{\theta}_k, \mathbf{z}_{1:k-1})}{\sum_{\boldsymbol{\theta}'_k \in \mathcal{D}} \mathcal{P}(\boldsymbol{\theta}'_k | \mathbf{z}_{1:k-1}) \mathcal{P}(\mathbf{z}_k | \boldsymbol{\theta}'_k, \mathbf{z}_{1:k-1})}, \quad (21)$$

where, through the law of total probability, $\mathcal{P}(\mathbf{z}_k | \boldsymbol{\theta}_k, \mathbf{z}_{1:k-1})$ can be calculated as

$$\begin{aligned} \mathcal{P}(\mathbf{z}_k | \boldsymbol{\theta}_k, \mathbf{z}_{1:k-1}) &= \sum_{n=1}^N \mathcal{P}(\mathbf{z}_k, n | \boldsymbol{\theta}_k, \mathbf{z}_{1:k-1}) \\ &= \sum_{n=1}^N \mathcal{P}(\mathbf{z}_k | n, \boldsymbol{\theta}_k, \mathbf{z}_{1:k-1}) \mathcal{P}(n | \boldsymbol{\theta}_k, \mathbf{z}_{1:k-1}). \end{aligned}$$

Here, we recognize the update coefficient $\alpha_k^{(n, \boldsymbol{\theta}_k)} = \mathcal{P}(\mathbf{z}_k | n, \boldsymbol{\theta}_k, \mathbf{z}_{1:k-1})$ and the weight $w_k^{(n, \boldsymbol{\theta}_k)} = \mathcal{P}(n | \boldsymbol{\theta}_k, \mathbf{z}_{1:k-1})$. Therefore, the posterior pmf in (21) can be finally recast as

$$\mathcal{P}(\boldsymbol{\theta}_k | \mathbf{z}_{1:k}) = \frac{\mathcal{P}(\boldsymbol{\theta}_k | \mathbf{z}_{1:k-1}) \sum_{n=1}^N \alpha_k^{(n, \boldsymbol{\theta}_k)} w_k^{(n, \boldsymbol{\theta}_k)}}{\sum_{\boldsymbol{\theta}'_k \in \mathcal{D}} \mathcal{P}(\boldsymbol{\theta}'_k | \mathbf{z}_{1:k-1}) \sum_{n'=1}^N \alpha_k^{(n', \boldsymbol{\theta}'_k)} w_k^{(n', \boldsymbol{\theta}'_k)}}. \quad (22)$$

B. PREDICTION STEP

The pmf of the predicted parameter vector at time $k+1$, i.e., $\mathcal{P}(\boldsymbol{\theta}_{k+1} | \mathbf{z}_{1:k})$, is simply obtained by inserting (22) into (10), where the integral is replaced by summation. To derive the pdf of the predicted state vector at time $k+1$, i.e., $\mathcal{P}(\mathbf{x}_{k+1} | \boldsymbol{\theta}_{k+1}, \mathbf{z}_{1:k})$, instead, we consider the discrete version of (11), that is,

$$\begin{aligned} \mathcal{P}(\mathbf{x}_{k+1} | \boldsymbol{\theta}_{k+1}, \mathbf{z}_{1:k}) \\ = \sum_{\boldsymbol{\theta}_k \in \mathcal{D}} \mathcal{P}(\boldsymbol{\theta}_k | \boldsymbol{\theta}_{k+1}, \mathbf{z}_{1:k}) \mathcal{P}(\mathbf{x}_{k+1} | \boldsymbol{\theta}_{k+1}, \boldsymbol{\theta}_k, \mathbf{z}_{1:k}), \end{aligned}$$

and insert therein (12) and (13) to obtain

$$\begin{aligned} \mathcal{P}(\mathbf{x}_{k+1} | \boldsymbol{\theta}_{k+1}, \mathbf{z}_{1:k}) \\ = \sum_{\boldsymbol{\theta}_k \in \mathcal{D}} \frac{\mathcal{P}(\boldsymbol{\theta}_{k+1} | \boldsymbol{\theta}_k) \mathcal{P}(\boldsymbol{\theta}_k | \mathbf{z}_{1:k})}{\mathcal{P}(\boldsymbol{\theta}_{k+1} | \mathbf{z}_{1:k})} \\ \times \int \mathcal{P}(\mathbf{x}_{k+1} | \boldsymbol{\theta}_{k+1}, \boldsymbol{\theta}_k, \mathbf{x}_k) \mathcal{P}(\mathbf{x}_k | \boldsymbol{\theta}_k, \mathbf{z}_{1:k}) d\mathbf{x}_k. \end{aligned} \quad (23)$$

Finally, by using (17) into (23), the latter can be recast as

$$\begin{aligned} \mathcal{P}(\mathbf{x}_{k+1} | \boldsymbol{\theta}_{k+1}, \mathbf{z}_{1:k}) \\ = \sum_{n=1}^N \sum_{\boldsymbol{\theta}_k \in \mathcal{D}} w_{k+1|k}^{(n, \boldsymbol{\theta}_{k+1}, \boldsymbol{\theta}_k)} \mathcal{P}(\mathbf{x}_{k+1} | n, \boldsymbol{\theta}_{k+1}, \boldsymbol{\theta}_k, \mathbf{z}_{1:k}), \end{aligned}$$

where the pdf of the n^{th} component, i.e., $\mathcal{P}(\mathbf{x}_{k+1} | n, \boldsymbol{\theta}_{k+1}, \boldsymbol{\theta}_k, \mathbf{z}_{1:k})$, is

$$\begin{aligned} \mathcal{P}(\mathbf{x}_{k+1} | n, \boldsymbol{\theta}_{k+1}, \boldsymbol{\theta}_k, \mathbf{z}_{1:k}) \\ \triangleq \int \mathcal{P}(\mathbf{x}_{k+1} | \boldsymbol{\theta}_{k+1}, \boldsymbol{\theta}_k, \mathbf{x}_k) \mathcal{P}(\mathbf{x}_k | n, \boldsymbol{\theta}_k, \mathbf{z}_{1:k}) d\mathbf{x}_k, \end{aligned} \quad (24)$$

and the predicted weights are defined as

$$w_{k+1|k}^{(n, \boldsymbol{\theta}_{k+1}, \boldsymbol{\theta}_k)} \triangleq \frac{\mathcal{P}(\boldsymbol{\theta}_{k+1} | \boldsymbol{\theta}_k) \mathcal{P}(\boldsymbol{\theta}_k | \mathbf{z}_{1:k})}{\mathcal{P}(\boldsymbol{\theta}_{k+1} | \mathbf{z}_{1:k})} w_k^{(n, \boldsymbol{\theta}_k)}. \quad (25)$$

We note that in the prediction step the number of mixture components increases from N to $N \times D$, thus a suitable merging/pruning criterion is required to avoid the exponential growth of the computational complexity [30], [31].

Remark: We aim to provide a general framework for adaptive Bayesian estimation of epidemic evolution, here implemented via an efficient mixture model. Nonetheless, the same nonlinear problem could have been approached and implemented differently, for example by using an extended or unscented Kalman filter (EKF or UKF), or by means of sequential Monte Carlo (SMC) methods, e.g., particle filters [32]. These are not alternate to our approach, but can be possibly combined, as noted later in Section IV-B. However, we also observe that direct use of the aforementioned techniques — if not adequately tailored to the specific problem — may lead to poor or unexpected results. For example, the EKF approximation of the nonlinear behaviour of the system through local linearization, might fail in the presence of strong nonlinearities, leading to unreliable estimates or even to divergence. The UKF can potentially provide higher-order estimation accuracy using the unscented transform, but this usually has the effect of simply delaying the unavoidable divergence that will still happen in the case of severe process or measurement nonlinearities. On the other hand, SMC methods generally provide reliable numerical approximations to sequential nonlinear estimation problems. However, in real-world applications, where the system also depends on unknown time-varying parameters to be inferred from uncertain data, conventional particle filters could fail to detect and track the change in parameters, quickly leading to implementation issues, such as sample impoverishment. In addition, high performance of particle methods comes at the expense of increased computational demands.

Another class of methods is based on system identification/machine learning (ML) techniques. Canonical approaches here include expectation-maximization, variational Bayes methods, and the variety of nonlinear auto-regressive models with external inputs, recurrent neural networks, long short-term memory networks [33]-[36]. The non-parametric ML methods suffer from lack of explainability and causal reasoning needed for policy decisions in pandemics.

IV. ESTIMATION AND FORECASTING WITH STOCHASTIC SIR MODEL

A. DYNAMIC AND OBSERVATION MODEL

The SIR model [4], [5] subdivides the population of a community into three interacting groups: susceptible, infectious, and recovered individuals. The interactions are governed by the infection rate, usually denoted by β , that is the average rate at which an infected individual can infect a susceptible one, and by the recovery rate, generally called γ . Let P be the total population size⁵; s_k , i_k , and r_k be the normalized (to P) number of susceptible, infectious, and recovered individuals at time k , such that $s_k + i_k + r_k = 1$; and β_k and γ_k be the infection and recovery rates at time k . The discrete-time stochastic SIR system of equations is expressed as [17], [37]

⁵Note that authors generally refer to the population size as N [4], [5], [7]-[13].

$$\begin{aligned}
s_k &= s_{k-1} - \beta_{k-1} s_{k-1} i_{k-1} \Delta_t + \sigma_{1,k-1} u_{1,k}, \\
i_k &= i_{k-1} + \beta_{k-1} s_{k-1} i_{k-1} \Delta_t - \gamma_{k-1} i_{k-1} \Delta_t \\
&\quad - \sigma_{1,k-1} u_{1,k} + \sigma_{2,k-1} u_{2,k}, \\
r_k &= r_{k-1} + \gamma_{k-1} i_{k-1} \Delta_t - \sigma_{2,k-1} u_{2,k},
\end{aligned}$$

where $\sigma_{1,k} \triangleq \sqrt{P^{-1} \beta_k s_k i_k}$, $\sigma_{2,k} \triangleq \sqrt{P^{-1} \gamma_k i_k}$, and $\mathbf{u}_k \triangleq [u_{1,k}, u_{2,k}]^T \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_2 \Delta_t)$. Since s_k and i_k determine r_k , we define the state vector as $\mathbf{x}_k \triangleq [s_k, i_k]^T$, and the parameter vector as $\boldsymbol{\theta}_k \triangleq [\beta_k, \gamma_k]^T$; hence, we have $S = 2$ and $M = 2$. The dynamic model of the epidemic is then expressed as (cf. (1))

$$\mathbf{x}_k = \mathbf{f}_1(\mathbf{x}_{k-1}, \boldsymbol{\theta}_{k-1}) + \mathbf{f}_2(\mathbf{x}_{k-1}, \boldsymbol{\theta}_{k-1}) \mathbf{u}_k \quad (26)$$

where

$$\mathbf{f}_1(\mathbf{x}_k, \boldsymbol{\theta}_k) \triangleq \begin{bmatrix} (1 - \beta_k i_k \Delta_t) s_k \\ (1 + \beta_k s_k \Delta_t - \gamma_k \Delta_t) i_k \end{bmatrix},$$

and

$$\mathbf{f}_2(\mathbf{x}_k, \boldsymbol{\theta}_k) \triangleq \begin{bmatrix} \sqrt{P^{-1} \beta_k s_k i_k} & 0 \\ -\sqrt{P^{-1} \beta_k s_k i_k} & \sqrt{P^{-1} \gamma_k i_k} \end{bmatrix}.$$

From (26) it then follows that the state transition pdf (cf. (2)) is independent of the current parameter vector $\boldsymbol{\theta}_k$, i.e., $\mathcal{P}(\mathbf{x}_k | \boldsymbol{\theta}_k, \boldsymbol{\theta}_{k-1}, \mathbf{x}_{k-1}) = \mathcal{P}(\mathbf{x}_k | \boldsymbol{\theta}_{k-1}, \mathbf{x}_{k-1})$, and is distributed according to

$$\begin{aligned}
&\mathcal{P}(\mathbf{x}_k | \boldsymbol{\theta}_{k-1}, \mathbf{x}_{k-1}) \\
&= \mathcal{N}(\mathbf{x}_k; \mathbf{f}_1(\mathbf{x}_{k-1}, \boldsymbol{\theta}_{k-1}), \mathbf{F}(\mathbf{x}_{k-1}, \boldsymbol{\theta}_{k-1}) \Delta_t),
\end{aligned} \quad (27)$$

where

$$\mathbf{F}(\mathbf{x}_k, \boldsymbol{\theta}_k) = \begin{bmatrix} \sigma_{1,k}^2 & -\sigma_{1,k}^2 \\ -\sigma_{1,k}^2 & \sigma_{1,k}^2 + \sigma_{2,k}^2 \end{bmatrix}.$$

We observe the (uncertain) normalized number of infected and of recovered individuals at each time k . Therefore, $B = 2$, and the observation model (cf. (3)) is

$$\mathbf{z}_k = \mathbf{h}_1(\mathbf{x}_k) + \mathbf{v}_k,$$

where $\mathbf{v}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{R}(\mathbf{x}_k))$ models the observation uncertainty, and

$$\mathbf{h}_1(\mathbf{x}_k) \triangleq [0 \ 1]^T + \mathbf{H} \mathbf{x}_k, \quad \mathbf{H} \triangleq \begin{bmatrix} 0 & 1 \\ -1 & -1 \end{bmatrix}.$$

The covariance matrix $\mathbf{R}(\mathbf{x}_k)$ depends on the state vector at time k : since we assume that the observation “noise” accrues from the sum of uncertainties of each individual epidemic state, the variances are linear in the number of infected and recovered individuals, respectively. Hence, we define

$$\mathbf{R}(\mathbf{x}_k) \triangleq \begin{bmatrix} P^{-1}i_k & 0 \\ 0 & P^{-1}(1 - i_k - s_k) \end{bmatrix} \mathbf{R}_c, \quad (28)$$

where \mathbf{R}_c is a constant diagonal matrix. Thus, the likelihood is independent of $\boldsymbol{\theta}_k$, i.e., $\mathcal{P}(\mathbf{z}_k | \mathbf{x}_k, \boldsymbol{\theta}_k) = \mathcal{P}(\mathbf{z}_k | \mathbf{x}_k)$, and distributed according to

$$\mathcal{P}(\mathbf{z}_k | \mathbf{x}_k) = \mathcal{N}(\mathbf{z}_k; \mathbf{h}_1(\mathbf{x}_k), \mathbf{R}(\mathbf{x}_k)). \quad (29)$$

B. GAUSSIAN MIXTURE FILTER

Given the Gaussian nature of the dynamic and observation models, we adopt a Gaussian mixture implementation of the Bayesian sequential estimation. That is, we assume that the pdf of the n^{th} mixture component in (16) is Gaussian with mean $\hat{\mathbf{x}}_{k|k-1}^{(n, \boldsymbol{\theta}_k)}$ and covariance matrix $\hat{\mathbf{C}}_{k|k-1}^{(n, \boldsymbol{\theta}_k)}$, i.e.,

$$\mathcal{P}(\mathbf{x}_k | n, \boldsymbol{\theta}_k, \mathbf{z}_{1:k-1}) = \mathcal{N}\left(\mathbf{x}_k; \hat{\mathbf{x}}_{k|k-1}^{(n, \boldsymbol{\theta}_k)}, \hat{\mathbf{C}}_{k|k-1}^{(n, \boldsymbol{\theta}_k)}\right). \quad (30)$$

When the observation \mathbf{z}_k is available, mean, covariance matrix, and weight are updated. Specifically, the weight of the n^{th} mixand is updated as in (18) through the coefficient $\alpha_{k|k-1}^{(n, \boldsymbol{\theta}_k)}$; this, in turn, is calculated recalling that the likelihood in (29) is independent of $\boldsymbol{\theta}_k$, and inserting (29) and (30) into (19), that is,

$$\begin{aligned} & \alpha_{k|k-1}^{(n, \boldsymbol{\theta}_k)} \\ &= \int \mathcal{N}(\mathbf{z}_k; \mathbf{h}_1(\mathbf{x}_k), \mathbf{R}(\mathbf{x}_k)) \mathcal{N}\left(\mathbf{x}_k; \hat{\mathbf{x}}_{k|k-1}^{(n, \boldsymbol{\theta}_k)}, \hat{\mathbf{C}}_{k|k-1}^{(n, \boldsymbol{\theta}_k)}\right) d\mathbf{x}_k \\ &\approx \mathcal{N}\left(\mathbf{z}_k; \mathbf{h}_1\left(\hat{\mathbf{x}}_{k|k-1}^{(n, \boldsymbol{\theta}_k)}\right), \mathbf{R}\left(\hat{\mathbf{x}}_{k|k-1}^{(n, \boldsymbol{\theta}_k)}\right) + \mathbf{H}\hat{\mathbf{C}}_{k|k-1}^{(n, \boldsymbol{\theta}_k)}\mathbf{H}^T\right), \end{aligned} \quad (31)$$

where we made the approximation $\mathbf{R}(\mathbf{x}_k) \approx \mathbf{R}(\hat{\mathbf{x}}_{k|k-1}^{(n, \boldsymbol{\theta}_k)})$. We observe that the last step would be an equality — rather than an approximation — if the observation covariance matrix was independent of the state \mathbf{x}_k . Then, the updated pdf of the n^{th} Gaussian component is obtained by inserting (29), (30), and (31) into (20), and is equal to [30]

$$\mathcal{P}(\mathbf{x}_k | n, \boldsymbol{\theta}_k, \mathbf{z}_{1:k}) = \mathcal{N}\left(\mathbf{x}_k; \hat{\mathbf{x}}_{k|k}^{(n, \boldsymbol{\theta}_k)}, \hat{\mathbf{C}}_{k|k}^{(n, \boldsymbol{\theta}_k)}\right), \quad (32)$$

where

$$\begin{aligned}\hat{\mathbf{x}}_{k|k}^{(n, \boldsymbol{\theta}_k)} &= \hat{\mathbf{x}}_{k|k-1}^{(n, \boldsymbol{\theta}_k)} + \mathbf{K}_{k|k-1}^{(n, \boldsymbol{\theta}_k)} \left[\mathbf{z}_k - \mathbf{h}_1 \left(\hat{\mathbf{x}}_{k|k-1}^{(n, \boldsymbol{\theta}_k)} \right) \right], \\ \widehat{\mathbf{C}}_{k|k}^{(n, \boldsymbol{\theta}_k)} &= \widehat{\mathbf{C}}_{k|k-1}^{(n, \boldsymbol{\theta}_k)} - \mathbf{K}_{k|k-1}^{(n, \boldsymbol{\theta}_k)} \mathbf{H} \widehat{\mathbf{C}}_{k|k-1}^{(n, \boldsymbol{\theta}_k)}\end{aligned}$$

and

$$\mathbf{K}_{k|k-1}^{(n, \boldsymbol{\theta}_k)} = \widehat{\mathbf{C}}_{k|k-1}^{(n, \boldsymbol{\theta}_k)} \mathbf{H}^T \left[\mathbf{H} \widehat{\mathbf{C}}_{k|k-1}^{(n, \boldsymbol{\theta}_k)} \mathbf{H}^T + \mathbf{R} \left(\hat{\mathbf{x}}_{k|k-1}^{(n, \boldsymbol{\theta}_k)} \right) \right]^{-1}.$$

This is similar to a standard Kalman update, per mixture element, with essential difference that the measurement noise covariance is state-dependent.

Let us now consider the evolution of the mixands according to the dynamic model. The n^{th} predicted weight is computed using (25) where, assuming that infection rate β_k and recovery rate γ_k evolve independently, the parameter vector transition pmf can be written as $\mathcal{P}(\boldsymbol{\theta}_{k+1} | \boldsymbol{\theta}_k) = \mathcal{P}(\beta_{k+1} | \beta_k) \mathcal{P}(\gamma_{k+1} | \gamma_k)$; the marginal transition pmfs $\mathcal{P}(\beta_{k+1} | \beta_k)$ and $\mathcal{P}(\gamma_{k+1} | \gamma_k)$ are specified later in this section. The pdf of the n^{th} predicted mixand is obtained by recalling that the transition pdf in (27) is independent of the current parameter vector $\boldsymbol{\theta}_k$, and inserting (27) and (32) into (24). This yields

$$\begin{aligned}\mathcal{P}(\mathbf{x}_{k+1} | n, \boldsymbol{\theta}_{k+1}, \boldsymbol{\theta}_k, \mathbf{z}_{1:k}) \\ = \int \mathcal{N}(\mathbf{x}_{k+1}; \mathbf{f}_1(\mathbf{x}_k, \boldsymbol{\theta}_k), \mathbf{F}(\mathbf{x}_k, \boldsymbol{\theta}_k) \Delta_t) \\ \times \mathcal{N} \left(\mathbf{x}_k; \hat{\mathbf{x}}_{k|k}^{(n, \boldsymbol{\theta}_k)}, \widehat{\mathbf{C}}_{k|k}^{(n, \boldsymbol{\theta}_k)} \right) d\mathbf{x}_k.\end{aligned}\quad (33)$$

Given the nonlinearity of the dynamic model (26), the integral (33) cannot be computed explicitly. A viable alternative is to approximate the pdf $\mathcal{P}(\mathbf{x}_{k+1} | n, \boldsymbol{\theta}_{k+1}, \boldsymbol{\theta}_k, \mathbf{z}_{1:k}) = \mathcal{P}(\mathbf{x}_{k+1} | n, \boldsymbol{\theta}_k, \mathbf{z}_{1:k})$ as a Gaussian via moment matching, that is,

$$\mathcal{P}(\mathbf{x}_{k+1} | n, \boldsymbol{\theta}_k, \mathbf{z}_{1:k}) = \mathcal{N} \left(\mathbf{x}_{k+1}; \hat{\mathbf{x}}_{k+1|k}^{(n, \boldsymbol{\theta}_k)}, \widehat{\mathbf{C}}_{k+1|k}^{(n, \boldsymbol{\theta}_k)} \right).\quad (34)$$

The computation of $\hat{\mathbf{x}}_{k+1|k}^{(n, \boldsymbol{\theta}_k)}$ and of $\widehat{\mathbf{C}}_{k+1|k}^{(n, \boldsymbol{\theta}_k)}$ by moment matching is detailed in Appendix A. An alternative method to solve integral (33) is via the unscented transformation used within the UKF.

As described in Section III, the parameter vector $\boldsymbol{\theta}_k = [\beta_k, \gamma_k]^T$ is discretized for computational efficiency. Specifically, $\beta_k \in \mathcal{D}_1 = \{\vartheta_1^{(1)}, \dots, \vartheta_{D_1}^{(1)}\}$ and $\gamma_k \in \mathcal{D}_2 = \{\vartheta_1^{(2)}, \dots, \vartheta_{D_2}^{(2)}\}$. For concreteness, we assume that \mathcal{D}_m ($m = 1$ or 2) is an ordered set, i.e., such that for any $j, \ell \in \{1, \dots, D_m\}$ with $j < \ell$ we have $\vartheta_j^{(m)} < \vartheta_\ell^{(m)}$; and that

the elements of \mathcal{D}_m are selected to be equally spaced between $\vartheta_1^{(m)}$ and $\vartheta_{D_m}^{(m)}$. The marginal transition pmfs $\mathcal{P}(\beta_k | \beta_{k-1})$ and $\mathcal{P}(\gamma_k | \gamma_{k-1})$ are therefore fully described by the matrix $\mathbf{P}_\beta \in [0, 1]^{D_1 \times D_1}$ and matrix $\mathbf{P}_\gamma \in [0, 1]^{D_2 \times D_2}$, respectively, where

$$[\mathbf{P}_\beta]_{j,\ell} \triangleq \mathcal{P}(\beta_k = \vartheta_j^{(1)} | \beta_{k-1} = \vartheta_\ell^{(1)}), \quad (35)$$

and

$$[\mathbf{P}_\gamma]_{j,\ell} \triangleq \mathcal{P}(\gamma_k = \vartheta_j^{(2)} | \gamma_{k-1} = \vartheta_\ell^{(2)}). \quad (36)$$

We note that $\sum_{j=1}^{D_1} [\mathbf{P}_\beta]_{j,\ell} = 1$ and that $\sum_{j=1}^{D_2} [\mathbf{P}_\gamma]_{j,\ell} = 1$. Eventually, according to (4) and (6), and replacing the integral in (6) with the summation, the MMSE estimates of the normalized numbers of susceptible and infectious are

$$\begin{aligned} \hat{s}_k^E &= \sum_{\beta_k \in \mathcal{D}_1} \sum_{\gamma_k \in \mathcal{D}_2} \mathcal{P}(\beta_k, \gamma_k | \mathbf{z}_{1:k}) \\ &\quad \times \iint s_k \mathcal{P}(s_k, i_k | \beta_k, \gamma_k, \mathbf{z}_{1:k}) di_k ds_k, \end{aligned} \quad (37)$$

and

$$\begin{aligned} \hat{i}_k^E &= \sum_{\beta_k \in \mathcal{D}_1} \sum_{\gamma_k \in \mathcal{D}_2} \mathcal{P}(\beta_k, \gamma_k | \mathbf{z}_{1:k}) \\ &\quad \times \iint i_k \mathcal{P}(s_k, i_k | \beta_k, \gamma_k, \mathbf{z}_{1:k}) ds_k di_k, \end{aligned} \quad (38)$$

respectively; and, according to (15), the MMSE estimates of the infection and recovery rates are

$$\hat{\beta}_k^E = \sum_{\beta_k \in \mathcal{D}_1} \beta_k \sum_{\gamma_k \in \mathcal{D}_2} \mathcal{P}(\beta_k, \gamma_k | \mathbf{z}_{1:k}), \quad (39)$$

and

$$\hat{\gamma}_k^E = \sum_{\gamma_k \in \mathcal{D}_2} \gamma_k \sum_{\beta_k \in \mathcal{D}_1} \mathcal{P}(\beta_k, \gamma_k | \mathbf{z}_{1:k}), \quad (40)$$

respectively. Note that the estimates for the parameters $\hat{\beta}_k^E$ and $\hat{\gamma}_k^E$ are updated automatically based on the attractiveness (measured in terms of the relative likelihoods) of the estimates that assume them. Finally, the prior pmfs of the infection and recovery rates at time $k=0$ are set to $\mathcal{P}(\beta_0) = \mathcal{N}(\beta_0; \bar{\beta}_0, \sigma_{\bar{\beta}}^2)$ and $\mathcal{P}(\gamma_0) = \mathcal{N}(\gamma_0; \bar{\gamma}_0, \sigma_{\bar{\gamma}}^2)$; the prior pdf of the n^{th} Gaussian component at time $k=0$ is $\mathcal{P}(\mathbf{x}_0 | n, \boldsymbol{\theta}_0) = \mathcal{P}(\mathbf{x}_0 | n) = \mathcal{N}(\mathbf{x}_0; \bar{\mathbf{x}}_0^{(n)}, \bar{\mathbf{C}}_0^{(n)})$.

A detailed statement of the proposed Gaussian mixture filter for the Bayesian estimation of the epidemic evolution with the stochastic SIR model is provided in Algorithm 1.

C. FORECASTING

The forecasting is as described in Section II-C. Let us assume that \mathbf{z}_k is the most recent available observation, and that the posterior pdf $\mathcal{P}(\mathbf{x}_k | \mathbf{z}_{1:k})$ and pmf $\mathcal{P}(\boldsymbol{\theta}_k | \mathbf{z}_{1:k})$ are known.

The j^{th} sample state vector extracted from the posterior pdf $\mathcal{P}(\mathbf{x}_k | \mathbf{z}_{1:k})$ is $\tilde{\mathbf{x}}_k^{(j)} = [\tilde{s}_k^{(j)}, \tilde{i}_k^{(j)}]^T$, $j \in \mathcal{J}$, and evolves according to the state vector forecast transition distribution $\mathcal{P}_{\mathbf{F}}(\tilde{\mathbf{x}}_k^{(j)} | \tilde{\boldsymbol{\theta}}_k^{(j)}, \tilde{\boldsymbol{\theta}}_{k-1}^{(j)}, \tilde{\mathbf{x}}_{k-1}^{(j)})$; we assume that this forecast transition distribution coincides with that used within the sequential Bayesian estimation procedure (cf. (27)), i.e., we assume that the sampled state vector $\tilde{\mathbf{x}}_k^{(j)}$ evolves according to the dynamic model in (26).

Concerning the parameter vectors, in order to obtain samples $\tilde{\boldsymbol{\theta}}_k^{(j)} = [\tilde{\beta}_k^{(j)}, \tilde{\gamma}_k^{(j)}]^T$, $j \in \mathcal{J}$, from the infinite set \mathcal{Q} — rather than the discrete finite set \mathcal{D} —, the posterior pmf $\mathcal{P}(\boldsymbol{\theta}_k | \mathbf{z}_{1:k})$ is approximated with a suitable continuous distribution; here, $\mathcal{P}(\boldsymbol{\theta}_k | \mathbf{z}_{1:k})$ is approximated with a bivariate Gaussian pdf and samples $\tilde{\boldsymbol{\theta}}_k^{(j)}$ are extracted from it. Then, these sampled parameter vectors are allowed to evolve according to the parameter vector forecast transition distribution $\mathcal{P}_{\mathbf{F}}(\tilde{\boldsymbol{\theta}}_k^{(j)} | \tilde{\boldsymbol{\theta}}_{k-1}^{(j)}) = \mathcal{P}_{\mathbf{F}}(\tilde{\beta}_k^{(j)} | \tilde{\beta}_{k-1}^{(j)}) \mathcal{P}_{\mathbf{F}}(\tilde{\gamma}_k^{(j)} | \tilde{\gamma}_{k-1}^{(j)})$, where we assumed the infection and recovery rates to change independently. We recall that the inverse of the recovery rate expresses the average time that an individual takes to move from the group of infected (I) to the group of recovered (R) people; in our model, the latter includes both those discharged from hospitals, and those for whom the infection was fatal. Even though it is likely that the recovery rate will change during the forecast period (due to, e.g., reporting delays or the application of different criteria used to declare an individual recovered), there are no prior information that would suggest when and how this will happen; it is therefore reasonable to assume the recovery rate to be constant and deterministic during the forecast period. This equals to set the recovery rate forecast transition distribution to

$$\mathcal{P}_{\mathbf{F}}(\tilde{\gamma}_k^{(j)} | \tilde{\gamma}_{k-1}^{(j)}) = \delta(\tilde{\gamma}_k^{(j)} - \tilde{\gamma}_{k-1}^{(j)}),$$

where $\delta(\cdot)$ is the Dirac delta. The infection rate, instead, models the interaction between people, and it is therefore affected by the restriction measures. Therefore, once its time evolution is captured, it is reasonable to assume — in the absence of further knowledge — that it keeps the same trend linearly. That is, the infection rate samples $\tilde{\beta}_k^{(j)}$, $j \in \mathcal{J}$, are assumed to evolve according to

$$\tilde{\beta}_{k'}^{(j)} = \tilde{\beta}_{k'-1}^{(j)} + \dot{\beta}_k \Delta_t + d_{L,k},$$

for $k' \in \{k+1, \dots, k+K\}$, where $\dot{\beta}_k$ is a constant slope, and $d_{L,k} \sim \mathcal{N}(0, \sigma_{L,k}^2)$; hence, the infection rate forecast transition distribution is

$$\mathcal{P}_{\mathbf{F}}(\tilde{\beta}_k^{(j)} | \tilde{\beta}_{k-1}^{(j)}) = \mathcal{N}(\tilde{\beta}_k^{(j)}; \tilde{\beta}_{k-1}^{(j)} + \dot{\beta}_k \Delta t, \sigma_{L,k}^2).$$

Appendix B provides details on the estimation of the slope $\dot{\beta}_k$ over the time interval $[k - L, k]$, on the selection of L , with $L_{\min} \leq L \leq L_{\max}$, and on the computation of $\sigma_{L,k}^2$.

Eventually, defining the ensemble state and parameter matrices as $\tilde{\mathbf{X}}_k = [\tilde{\mathbf{s}}_k, \tilde{\mathbf{I}}_k]$ and $\tilde{\Theta}_k = [\tilde{\beta}_k, \tilde{\gamma}_k]$, respectively, where $\tilde{\mathbf{s}}_k \triangleq [\tilde{s}_k^{(1)}, \dots, \tilde{s}_k^{(J)}]^T$, $\tilde{\mathbf{I}}_k \triangleq [\tilde{i}_k^{(1)}, \dots, \tilde{i}_k^{(J)}]^T$, $\tilde{\beta}_k \triangleq [\tilde{\beta}_k^{(1)}, \dots, \tilde{\beta}_k^{(J)}]^T$, and $\tilde{\gamma}_k \triangleq [\tilde{\gamma}_k^{(1)}, \dots, \tilde{\gamma}_k^{(J)}]^T$, the mean of the epidemic states and model parameters at any time step $k' \in \{k+1, \dots, k+K\}$ are calculated as (cf. (14))

$$\hat{\mathbf{s}}_{k'}^F \triangleq J^{-1} \tilde{\mathbf{s}}_{k'}^T \cdot \mathbf{1}_J, \hat{i}_{k'}^F \triangleq J^{-1} \tilde{\mathbf{I}}_{k'}^T \cdot \mathbf{1}_J, \hat{\beta}_{k'}^F \triangleq J^{-1} \tilde{\beta}_{k'}^T \cdot \mathbf{1}_J, \text{ and } \hat{\gamma}_{k'}^F \triangleq J^{-1} \tilde{\gamma}_{k'}^T \cdot \mathbf{1}_J.$$

The steps of the proposed forecasting algorithm with stochastic SIR model are detailed in Algorithm 2.

V. FORECAST PERFORMANCE ANALYSIS: SYNTHETIC DATA AND REAL COVID-19 OUTBREAK DATA

We present numerical results obtained with the sequential estimation and forecasting algorithm described in Section IV. In Section V-A the algorithm is applied to synthetic data, while real data from the recent COVID-19 outbreak are considered in Section V-B.

A. SYNTHETIC DATA EXPERIMENT

The effectiveness of the proposed algorithm is validated in two simulated epidemic scenarios involving a community of $P = 10^6$ individuals. The simulations span 80 days, during which the infection rate changes as shown in Fig. 1. The variations of the infection rate model the effects of the restriction measures established by the authorities: in the first scenario, the epidemic outbreak is controlled by long-term soft restriction measures that cause a slow, yet consistent, decrease in the infection rate; in the second scenario, an initial strict lockdown is then followed by a relaxation of the restriction measures that leads to a slight increase in the infection rate. The recovery rate is fixed and set to $\gamma = 0.1$. The initial state of the epidemic is described by the normalized numbers of susceptible, infected, and recovered individuals at time $k = 0$, that are $s_0 = 1 - i_0 - r_0$, $i_0 = 20/P$, and $r_0 = 1/P$. The infection rate is discretized with $D_1 = 25$ values between $\vartheta_1^{(1)} = 0$ and $\vartheta_{D_1}^{(1)} = 0.6$; the recovery rate with $D_2 = 10$ values between $\vartheta_1^{(2)} = 0$ and $\vartheta_{D_2}^{(2)} = 0.2$. The prior pmf of the infection rate has mean $\bar{\beta}_0 = 0.4$ in the first scenario, and $\bar{\beta}_0 = 0.3$ in the second scenario; the standard deviation is $\sigma_\beta = 0.1$ for both the scenarios. Mean and standard deviation of the prior pmf of the recovery rate are $\bar{\gamma}_0 = 0.1$ and $\sigma_\gamma = 0.04$, respectively, for both the scenarios. The transition matrix for the parameter β_k defined in (35) is set to $[\mathbf{P}_\beta]_{\ell\ell} = 0.9$ for $\ell \in \{1, \dots, D_1\}$, $[\mathbf{P}_\beta]_{\ell-1,\ell} = [\mathbf{P}_\beta]_{\ell,\ell-1} = 0.05$ for $\ell \in \{2, \dots, D_1 - 1\}$, and $[\mathbf{P}_\beta]_{2,1} = [\mathbf{P}_\beta]_{D_1-1,D_1} = 0.1$. The transition matrix for the parameter γ_k defined in (36) is analogously set to $[\mathbf{P}_\gamma]_{\ell\ell} = 0.99$ for $\ell \in \{1, \dots, D_2\}$, $[\mathbf{P}_\gamma]_{\ell-1,\ell} = [\mathbf{P}_\gamma]_{\ell,\ell-1}$

$= 0.005$ for $\ell \in \{2, \dots, D_2 - 1\}$, and $[\mathbf{P}_\gamma]_{2,1} = [\mathbf{P}_\gamma]_{D_2-1,D_2} = 0.01$. The number of mixture components is $N = 5$, and mean and covariance matrix of the prior pdf of the n^{th} component are

$$\bar{\mathbf{x}}_0^{(n)} = \begin{bmatrix} s_0 \\ i_0 \end{bmatrix} + \begin{bmatrix} -1 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \varepsilon_1^{(n)} \\ \varepsilon_2^{(n)} \end{bmatrix}$$

and $\bar{\mathbf{C}}_0^{(n)} = \bar{\mathbf{C}}_0 = \mathbf{I}_2 i_0$, respectively, where $\varepsilon_1^{(n)} \sim \mathcal{U}(-i_0/5, i_0/5)$ and

$\varepsilon_2^{(n)} \sim \mathcal{U}(-r_0/5, r_0/5)$. Finally, the observation noise covariance matrix in (28) is $\mathbf{R}_c = 50 \mathbf{I}_2$. As for the forecasting, the ensemble size is $J = 2 \times 10^4$, and the minimum and maximum numbers of points used to estimate the slope of the infection rate are $L_{\min} = 5$ and $L_{\max} = 14$, respectively.

Fig. 2 shows the infection and recovery rates estimated in the first scenario over the 80 days, along with their 90% confidence intervals. Analogously, Fig. 3 shows the estimated infection and recovery rates in the second scenario. The results demonstrate the capability of proposed Bayesian sequential estimation algorithm to closely follow the time variation of the infection rates even in the presence of abrupt fluctuations, as well as to accurately estimate the recovery rate.

In turn, the accuracy of the proposed algorithm allows one to reliably forecast the epidemic evolution. Fig. 4 presents the estimation and forecast on the infection rate and of the number of infected in the first scenario; we assume that the latest available observation is on day $k = 44$ — so that the estimation stops on this day —, and the forecast is up to day $k = 80$. The forecast of the number of infected individuals well represents the evolution of the epidemic, suggesting a peak between days 55 and 65. Furthermore, we observe how both the true infection rate and the true number of infected is always enveloped within the 90% confidence interval, showing the high reliability of the proposed algorithm. Finally, in Fig. 5, we show the forecast of the epidemic evolution in the second scenario. Here, the estimation is performed up to day $k = 57$; the capability of the proposed algorithm to accurately estimate the large variation in the infection rate and forecast its future average value, allows one to forecast the evolution of the number infected, even though a further small variation of the infection rate will start at $k = 60$.

B. REAL DATA: COVID-19 OUTBREAK

This section presents the results obtained with the proposed estimation and forecasting algorithm when applied to real data obtained from the recent COVID-19 outbreak. The focus is on two very different areas in terms of population and interactions: Lombardia region in Italy, and the USA.

1) LOMBARDIA REGION, ITALY—Official data on the COVID-19 epidemic outbreak in Italy are made available from Protezione Civile on a daily basis [38]. This includes many entries, both nationwide and per region, as the total number of cases, total number of current positive cases, new positive cases per day, number of hospitalised patients, number of tests

performed, number of discharged COVID-19 patients from the hospitals, and number of deaths. Here, we focus on the data from Lombardia region, the centre of Italy's COVID-19 outbreak, whose population is $P \approx 10^7$ people. We used the normalized (to P) total number of current positive cases as number of infected i_k , and the normalized sum of number of discharged patients and number of deaths as the number of recovered r_k . These are reported in Fig. 6 and refer to the period between February 24, 2020, and June 30, 2020. The figure also shows the beginning of the lockdown established by the Italian government on March 8, 2020. Furthermore, we observe that, on May 6, the number of infected and number of recovered individuals present large steps, which hardly reflect physical reality. These steps are due to the fact that the numbers reported on May 6 include not only data referring to that day, but also data collected on previous days, and, erroneously, not reported in the correct day [38].

The setting of the Bayesian sequential estimation and forecasting algorithm is as described in Section V-A, except that the smallest and largest values used for the discretization of the infection rate are $\vartheta_1^{(1)} = 0$ and $\vartheta_{D_1}^{(1)} = 0.4$, respectively; the smallest and largest values used for the discretization of the recovery rate are $\vartheta_1^{(2)} = 0$ and $\vartheta_{D_2}^{(2)} = 0.1$, respectively; mean and standard deviation of their prior pmfs are $\bar{\beta}_0 = 0.3$ and $\sigma_\beta = 0.07$, and $\bar{\gamma}_0 = 0.06$ and $\sigma_\gamma = 0.02$; and the observation noise covariance matrix is $\mathbf{R}_c = 100 \mathbf{I}_2$. The initial state of the epidemic is given by the normalized numbers of susceptible, infected, and recovered on February 24, that are $s_0 = 1 - i_0 - r_0$, $i_0 = 166/P$, and $r_0 = 6/P$.

The estimated infection and recovery rates are shown in Fig. 7. The decrease in the infection rate, which represents the slowdown of the epidemic, clearly reflects the restriction measures established on March 8. The recovery rate, instead, decreases up to May 6, when it then shows a slight increase. The reduction of the recovery rate balances the decrease in the infection rate; indeed, up to May 6, the number of infected is still growing, which suggests that the infection rate is greater than the recovery rate, i.e., $\beta > \gamma$. After May 6, this trend changes.

Fig. 8 reports the forecasts of the epidemic evolution assessed every five days in the time period between April 13 and June 7, and Table 1 presents the mean absolute percentage errors (MAPEs) calculated for each forecast and for different forecast horizons, that is, 3, 7, and 14 days. We note that the forecasts made on April 13, 18, and 23, follow the future observations well, with an average MAPE below 3% at a forecast horizon of 7 days. On April 28 and May 3 the forecasts are not reliable, since the future observations are not contained within the 90% confidence interval. However, this poor performance is related to the inaccurate data provided later on May 6; indeed, the next forecasts made from May 8, to June 7, present again low MAPEs, with an average of 3.49 %, 4.24 %, and 6.1 %, at forecast horizons of 3, 7, and 14 days, respectively. Neglecting the forecasts whose horizon includes May 6 (marked with an asterisk in Table 1), the average MAPE from April 13, to June 7, is 3.6% for forecasts at 7 days, and below 6% when the forecast horizon is 14 days.

The proposed algorithm is compared with two alternative curve-fitting approaches. The first one, hereafter named SIR-fit, employs a nonlinear least squares fitting algorithm that,

using the number of infected and recovered individuals, computes the *best*⁶ time-invariant infection and recovery rates of the deterministic SIR model. These *best* rates are then used to forecast the evolution of the epidemic. The second curve-fitting approach follows the same methodology applied on a more-sophisticated recently proposed generalized SEIR (GSEIR) model [39], for this reason hereafter called GSEIR-fit. The GSEIR model consists of seven compartments — three more compartments than those in the standard SEIR model, i.e., insusceptible, quarantined, and death — and six parameters. Table 2 compares the average MAPEs obtained with the proposed algorithm, the SIR-fit, and the GSEIR-fit. The comparison is made averaging the MAPEs over two different time intervals. The first interval is from March 4, i.e., the 10th day since the beginning of the data collection, to June 16; the second interval is from April 1, i.e., approximately three weeks after the lockdown, to June 16. The proposed algorithm clearly outperforms the SIR-fit for all the forecast horizons. The GSEIR-fit, instead, presents a single lower average MAPE over the interval from March 4 to June 16 when the forecast horizon is 14 days; however, when the interval from April 1 to June 16 is considered, the proposed algorithm outperforms the GSEIR-fit in all the cases. This confirms the benefit of sequentially estimating the time-varying model parameters, in order to have reliable and accurate forecasts.

2) UNITED STATES OF AMERICA—Since the beginning of the COVID-19 epidemic outbreak, the Johns Hopkins University (JHU) has tracked the evolution of the contagion and made the collected data publicly available [40], [41]. The repository includes the total number of cases, the number of deaths, and the number of discharged COVID-19 patients from the hospitals from the USA and other countries at different levels of details, i.e., for the country as a whole and, when available, for single states and regions. Here, we use the overall dataset from the USA, whose population is $P \approx 329.8 \cdot 10^6$ people. As for the experiment made on the dataset from Lombardia, the normalized (to P) sum of number of discharged patients and number of deaths is used as the number of recovered r_k ; the normalized number of infected i_k is then given by the normalized difference between the total number of cases and the number of recovered. These are reported in Fig. 9 and refer to the period between March 1, 2020, and July 31, 2020.

The setting of the Bayesian sequential estimation and forecasting algorithm is unchanged, except that for the discretization and initialization of the parameters, and the initialization of the epidemic state. Specifically, the smallest and largest values used for the discretization of the infection rate are $\vartheta_1^{(1)} = 0$ and $\vartheta_{D_1}^{(1)} = 0.5$, respectively; the smallest and largest values used for the discretization of the recovery rate are $\vartheta_1^{(2)} = 0$ and $\vartheta_{D_2}^{(2)} = 0.05$, respectively; mean and standard deviation of their prior pmfs are $\bar{\beta}_0 = 0.35$ and $\sigma_\beta = 0.08$, and $\bar{\gamma}_0 = 0.015$ and $\sigma_\gamma = 0.008$; and the observation noise covariance matrix is $\mathbf{R}_c = 2000 \mathbf{I}_2$. The initial state of the epidemic is given by the normalized numbers of susceptible, infected, and recovered on March 1, that are $s_0 = 1 - i_0 - r_0$, $i_0 = 22/P$, and $r_0 = 8/P$.

⁶*Best* is meant in the least-squares sense.

Fig. 10 shows the estimated infection and recovery rates. From the second half of March and through April the infection rate decreases, presumably due to the restriction measures established by each single State. Here, we cannot mark a specific date as the beginning of the lockdown; nevertheless, it is reasonable to assume that about three out of four US citizens were under some form of lockdown by early April [42]. Around April 30, the estimated recovery rate shows a slight increase followed by an abrupt decrease. On that day, 35 thousand new recovered (i.e., hospital releases plus deaths) individuals were reported against a decrease of infected individuals of only 6 thousand (cf. Fig. 9); this results in a sudden increase of the recovery rate. Large numbers of new recovered individuals are also reported on May 22 and July 4, that are, 53 thousand and 104 thousand, respectively; however, these are better balanced by the numbers of people leaving the infected group, that are, 29 thousand and 58 thousand, respectively, thus not significantly affecting the estimates of the infection and recovery rates.

Overall, the estimated recovery rate is roughly 0.01, which translates into an average number of 100 days that an individual takes to move from the group of infected (I) to the group of recovered (R). Although the recovery duration seems overestimated, it is worth highlighting that this is an aggregate estimate of the recovery rate from multiple States, which therefore suffers from multiple different reporting delays, as well as from the different criteria used to declare an individual as fully recovered. It underscores the need for the USA to provide timely and consistent data, similar to that just analysed in Lombardia, if public health policy is to be driven by reliable estimation and prediction.

Forecasts of the epidemic evolution evaluated every five days in the time period between May 6 and June 30 are reported in Fig. 11. Table 3 presents the MAPEs calculated for each forecast and for different forecast horizons, that is, 3, 7, and 14 days.

We note that the forecasts made on May 6, 11, 16, and 21, present the highest MAPEs; these forecasts, however, are negatively affected by the abrupt decrease of the recovery rate that follows April 30, as described above. The forecasts made from May 26, to June 30, instead, present MAPEs that are always below 3% at forecast horizons of 3 and 7 days, and always below 4% when the forecast horizon is 14 days. Overall, the results in Table 3 confirm those obtained with the data from Lombardia: the average MAPE for forecast horizons of 3, 7, and 14 days, is, respectively, 2.35 %, 3.03 %, and 4.16 %. Table 4 compares the average MAPEs obtained with the proposed algorithm, with those obtained with the SIR-fit and the GSEIR-fit curve-fitting approaches. The proposed algorithm consistently outperforms both the curve-fitting approaches, in both the considered time intervals. Moreover, we observe a significant improvement of the MAPE for the proposed algorithm when considering the interval from April 1 to July 17, compared to the interval from March 10 to July 17, which again confirms the benefit of sequentially estimating the time-varying model parameters in pursuance of reliable and accurate forecasts.

C. LIMITATIONS AND EXTENSIONS

The proposed analysis presents some limitations that may lead to future extensions and these are worth exploring. First, the considered stochastic SIR model may be broadened to include the fraction of undocumented (or asymptomatic) and quarantined infected individuals.

Undocumented infections usually include mild or asymptomatic cases that go undetected and, hence, based on their proportion and contagiousness, can potentially increase the spread of the disease. The portion of undocumented infectious cases is suspected to be a critical epidemiological characteristic that is not easy to quantify. Most of the available evidence on asymptomatic SARS-CoV-2 infections, reviewed and summarised in [43] for different circumscribed cohorts, suggests that this is a significant factor in the fast progression of the COVID-19 pandemic. However, the difficulty in quantification of undocumented cases is largely due to the imperfection of the data available, which does not accurately reflect a large, representative sample of the general population. Moreover, in order to distinguish asymptomatic and presymptomatic cases, longitudinal data — that is, repeated observations of the individuals over time — should be available.

Another possible extension may be to separate people who are confirmed infected and home-quarantined into a dedicated epidemic compartment. In addition, the recovered compartment typical of the SIR model may be separated into two distinct recovery and death compartments in the detection phase so that the available data on reported cases can be taken into account separately. Furthermore, the considered stochastic SIR model describes the spread of the disease inside a single and confined population, e.g., a city, a region (Lombardia, Italy), a nation (USA). However, the approach presented in this article could be extended to more complicated metapopulation models, which introduce a further spatial travel/diffusion dimension in the dynamic and observation models. In particular, the population could be represented as a network of multiple spatially separated subpopulations nodes, such as multiple cities in the same region or nation. The interconnections among different populations represents the diffusion of people, thus contributing to the disease spread among the different subpopulations. All these limitations can be addressed in future studies because they mainly affect data analysis, and do not restrict the application and the effectiveness of the proposed approach to learning and forecasting the evolution of the critical epidemiological characteristics.

VI. CONCLUSIONS

The recent worldwide epidemic outbreak, due to a new strain of Coronavirus, has intensified research into novel mathematical models and algorithms that are able to reliably estimate and predict the epidemiological curve of the infection. In this article, we proposed a Bayesian sequential estimation and forecasting algorithm that, based on the information that authorities provide on a daily basis, is able to estimate the state of the epidemic and the parameters of the underlying model, as well as to forecast the evolution of the epidemiological curve. We developed an efficient implementation of the above-mentioned Bayesian framework, specifically tailored to the stochastic SIR model of pandemic evolution. The proposed algorithm is validated using synthetic data simulating two epidemic scenarios, and on real data acquired during the recent COVID-19 outbreak in the Lombardia region, Italy, and in the USA. Results show that the mean absolute percentage error computed after the lockdown is on average below 5% when the forecast is at 7 days, and below 10% when the forecast horizon is 14 days. Moreover, the described Bayesian framework outperforms curve-fitting approaches that use deterministic epidemiological models (e.g., SIR and GSEIR), particularly when a clear change of model parameters occur,

e.g., a decrease of the infection rate following the lockdown. Finally, accurate and timely data collection, especially on recovered individuals, hospitalizations, intensive care unit admissions, and intubations, is essential for reliable model-based decisions.

There exists an enormous amount of very recent literature related to the forecast of COVID-19 pandemic evolution, part of which has been reviewed at the beginning of this article. The analysis of this literature makes clear the effectiveness of model-based approaches, over less structured data-centric methodologies. In this respect, one lesson learned by the present study is that accurate epidemic modeling requires accurate estimation of time-varying parameters, such as the infection rate β . This is obviously true in the presence of abrupt changes of the underlying physical situation (e.g., adoption of drastic countermeasures) but, more interestingly, it is by no means limited to these extreme situations. One consequence is that, once the epidemic is under control, small variations in the estimated β may be used as a sensible proxy for the incipient detection of possible pandemic recrudescence.

Acknowledgments

The work of Domenico Gaglione, Paolo Braca, Leonardo Maria Millefiori, Giovanni Soldi, and Nicola Forti was supported by the Data Knowledge and Operational Effectiveness (DKOE) program, sponsored by the North Atlantic Treaty Organization (NATO) Allied Command Transformation (ACT). The work of Peter K. Willett was supported in part by Air Force Office of Scientific Research (AFOSR) under Contract FA9500-18-1-0463 and in part by Centre for Maritime Research and Experimentation (CMRE). The work of Krishna R. Pattipati was supported in part by the U.S. Office of Naval Research, in part by the U.S. Naval Research Laboratory under Grant N00014-18-1-1238 and Grant N00173-16-1-G905, and in part by the Space Technology Research Institutes from National Aeronautics and Space Administration's (NASA's) Space Technology Research Grants Program under Grant 80NSSC19K1076.

APPENDIX A

MOMENT MATCHING

This section reports the expressions, derived by moment matching, used to compute $\hat{\mathbf{x}}_{k+1|k}^{(n, \boldsymbol{\theta}_k)}$ and $\hat{\mathbf{C}}_{k+1|k}^{(n, \boldsymbol{\theta}_k)}$ in (34). Let $\hat{\mathbf{x}}_{k|k}^{(n, \boldsymbol{\theta}_k)} \triangleq [\hat{s}_{k|k}^{(n, \boldsymbol{\theta}_k)}, \hat{t}_{k|k}^{(n, \boldsymbol{\theta}_k)}]^T$ and

$$\hat{\mathbf{C}}_{k|k}^{(n, \boldsymbol{\theta}_k)} \triangleq \begin{bmatrix} \left(\boldsymbol{\Sigma}_{s,k}^{(n)}\right)^2 & \rho_k^{(n)} \boldsymbol{\Sigma}_{s,k}^{(n)} \boldsymbol{\Sigma}_{i,k}^{(n)} \\ \rho_k^{(n)} \boldsymbol{\Sigma}_{s,k}^{(n)} \boldsymbol{\Sigma}_{i,k}^{(n)} & \left(\boldsymbol{\Sigma}_{i,k}^{(n)}\right)^2 \end{bmatrix}$$

be mean and covariance matrix of the n^{th} posterior mixand $\mathcal{P}(\mathbf{x}_k | n, \boldsymbol{\theta}_k, \mathbf{z}_{1:k})$. We aim to approximate the n^{th} predicted mixand $\mathcal{P}(\mathbf{x}_{k+1} | n, \boldsymbol{\theta}_k, \mathbf{z}_{1:k})$ to be Gaussian with mean

$$\hat{\mathbf{x}}_{k+1|k}^{(n, \boldsymbol{\theta}_k)} \triangleq \mathbb{E}[\mathbf{x}_{k+1} | n, \boldsymbol{\theta}_k, \mathbf{z}_{1:k}] = \left[\hat{s}_{k+1|k}^{(n, \boldsymbol{\theta}_k)}, \hat{t}_{k+1|k}^{(n, \boldsymbol{\theta}_k)} \right]^T$$

and covariance matrix

$$\begin{aligned}
\widehat{\mathbf{C}}_{k+1|k}^{(n, \boldsymbol{\theta}_k)} &\triangleq \mathbb{E}[(\mathbf{x}_{k+1} - \widehat{\mathbf{x}}_{k+1|k})(\mathbf{x}_{k+1} - \widehat{\mathbf{x}}_{k+1|k})^T | n, \boldsymbol{\theta}_k, \mathbf{z}_{1:k}] \\
&= \mathbb{E}[\mathbf{x}_{k+1} \mathbf{x}_{k+1}^T | n, \boldsymbol{\theta}_k, \mathbf{z}_{1:k}] - \widehat{\mathbf{x}}_{k+1|k} \widehat{\mathbf{x}}_{k+1|k}^T \\
&= \begin{bmatrix} \left(\boldsymbol{\Sigma}_{s,k+1}^{(n)}\right)^2 & \rho_{k+1}^{(n)} \boldsymbol{\Sigma}_{s,k+1}^{(n)} \boldsymbol{\Sigma}_{i,k+1}^{(n)} \\ \rho_{k+1}^{(n)} \boldsymbol{\Sigma}_{s,k+1}^{(n)} \boldsymbol{\Sigma}_{i,k+1}^{(n)} & \left(\boldsymbol{\Sigma}_{i,k+1}^{(n)}\right)^2 \end{bmatrix}.
\end{aligned}$$

To simplify the notation, we hereafter omit the index n and the dependence on the parameter vector $\boldsymbol{\theta}_k$. Following simple algebraic calculations, the components of the mean vector are obtained as

$$\widehat{s}_{k+1|k} = (1 - \beta_k \widehat{\gamma}_k | k \Delta_t) \widehat{s}_{k|k} - \beta_k \Delta_t \rho_k \boldsymbol{\Sigma}_{s,k} \boldsymbol{\Sigma}_{i,k},$$

and

$$\widehat{i}_{k+1|k} = (1 + \beta_k \widehat{s}_k | k \Delta_t - \gamma_k \Delta_t) \widehat{i}_{k|k} + \beta_k \Delta_t \rho_k \boldsymbol{\Sigma}_{s,k} \boldsymbol{\Sigma}_{i,k}.$$

We note that the products $\widehat{i}_{k|k} \widehat{s}_{k|k}$ induce nonlinearity in their predicted values, and hence require extra calculation for the associated uncertainties; indeed, the components of the covariance matrix are

$$\begin{aligned}
\boldsymbol{\Sigma}_{i,k+1}^2 &= (1 - \gamma_k \Delta_t)^2 (\widehat{i}_{k|k}^2 | k + \boldsymbol{\Sigma}_{i,k}^2) \\
&\quad + \beta_k^2 \Delta_t^2 \left[\widehat{s}_{k|k}^2 | k \widehat{i}_{k|k}^2 | k + \widehat{s}_{k|k}^2 | k \boldsymbol{\Sigma}_{i,k}^2 + \widehat{i}_{k|k}^2 | k \boldsymbol{\Sigma}_{s,k}^2 \right. \\
&\quad \left. + (1 + 2\rho_k^2) \boldsymbol{\Sigma}_{s,k}^2 \boldsymbol{\Sigma}_{i,k}^2 + 4\rho_k \widehat{s}_{k|k} | k \widehat{i}_{k|k} | k \boldsymbol{\Sigma}_{s,k} \boldsymbol{\Sigma}_{i,k} \right] \\
&\quad + 2(1 - \gamma_k \Delta_t) \beta_k \Delta_t \left[\widehat{s}_{k|k} | k (\widehat{i}_{k|k}^2 | k + \boldsymbol{\Sigma}_{i,k}^2) \right. \\
&\quad \left. + 2\rho_k \widehat{i}_{k|k} | k \boldsymbol{\Sigma}_{s,k} \boldsymbol{\Sigma}_{i,k} \right] + P^{-1} \Delta_t \left[\beta_k (\widehat{s}_{k|k} | k \widehat{i}_{k|k} | k \right. \\
&\quad \left. + \rho_k \boldsymbol{\Sigma}_{s,k} \boldsymbol{\Sigma}_{i,k}) + \gamma_k \widehat{i}_{k|k} | k \right] - \widehat{i}_{k+1|k}^2,
\end{aligned}$$

$$\begin{aligned}
\boldsymbol{\Sigma}_{s,k+1}^2 &= \widehat{s}_{k|k}^2 | k + \boldsymbol{\Sigma}_{s,k}^2 + \beta_k^2 \Delta_t^2 \left[\widehat{s}_{k|k}^2 | k \widehat{i}_{k|k}^2 | k \right. \\
&\quad \left. + \widehat{s}_{k|k}^2 | k \boldsymbol{\Sigma}_{i,k}^2 + \widehat{i}_{k|k}^2 | k \boldsymbol{\Sigma}_{s,k}^2 + (1 + 2\rho_k^2) \boldsymbol{\Sigma}_{s,k}^2 \boldsymbol{\Sigma}_{i,k}^2 \right. \\
&\quad \left. + 4\rho_k \widehat{s}_{k|k} | k \widehat{i}_{k|k} | k \boldsymbol{\Sigma}_{s,k} \boldsymbol{\Sigma}_{i,k} \right] \\
&\quad - 2\beta_k \Delta_t \left[\widehat{i}_{k|k} | k (\widehat{s}_{k|k}^2 | k + \boldsymbol{\Sigma}_{s,k}^2) + 2\rho_k \widehat{s}_{k|k} | k \boldsymbol{\Sigma}_{s,k} \boldsymbol{\Sigma}_{i,k} \right] \\
&\quad + P^{-1} \beta_k \Delta_t (\widehat{s}_{k|k} | k \widehat{i}_{k|k} | k + \rho_k \boldsymbol{\Sigma}_{s,k} \boldsymbol{\Sigma}_{i,k}) - \widehat{s}_{k+1|k}^2,
\end{aligned}$$

and

$$\begin{aligned}
& \rho_{k+1} \Sigma_{s,k+1} \Sigma_{i,k+1} \\
&= (1 - \gamma_k \Delta t - P^{-1} \beta_k \Delta t) (\hat{s}_k | k \hat{i}_k | k + \rho_{k,n} \Sigma_{s,k} \Sigma_{i,k}) \\
&\quad - (1 - \gamma_k \Delta t) \beta_k \Delta t \left[\hat{s}_k | k (\hat{i}_k^2 | k + \Sigma_{i,k}^2) \right. \\
&\quad \left. + 2\rho_k \hat{i}_k | k \Sigma_{s,k} \Sigma_{i,k} \right] + \beta_k \Delta t \left[\hat{i}_k | k (\hat{s}_k^2 | k + \Sigma_{s,k}^2) \right. \\
&\quad \left. + 2\rho_k \hat{s}_k | k \Sigma_{s,k} \Sigma_{i,k} \right] - \beta_k^2 \Delta t^2 \left[\hat{s}_k^2 | k \hat{i}_k^2 | k + \hat{s}_k^2 | k \Sigma_{i,k}^2 \right. \\
&\quad \left. + \hat{i}_k^2 | k \Sigma_{s,k}^2 + (1 + 2\rho_k^2) \Sigma_{s,k}^2 \Delta_{i,k}^2 \right. \\
&\quad \left. + 4\rho_k \hat{s}_k | k \hat{i}_k | k \Sigma_{s,k} \Sigma_{i,k} \right] - \hat{s}_{k+1} | k \hat{i}_{k+1} | k.
\end{aligned}$$

APPENDIX B

ADAPTIVE ESTIMATION OF THE INFECTION RATE SLOPE

Let us assume that the infection rate in the time interval $[k-L, k]$ is linearly changing according to the following model

$$\beta_{k-\ell+1} = \beta_{k-\ell} + \dot{\beta}_k \Delta t + \omega_{k,\ell}, \quad (41)$$

for $\ell \in \{1, \dots, L\}$, where $\omega_{k,\ell} \sim \mathcal{N}(0, \sigma_\omega^2)$ and $\dot{\beta}_k$ is a constant slope. An estimate of the slope can be obtained through the simple linear regression estimator [44] using the

Algorithm 1 Epidemic Estimation with Stochastic SIR Model

Input: $k, \mathbf{z}_{1:k} = [\mathbf{z}_1^T, \dots, \mathbf{z}_k^T], N, \mathcal{D}, \bar{\beta}_0, \sigma_{\beta}^2, \bar{\gamma}_0, \sigma_{\gamma}^2, \{\hat{\mathbf{x}}_0^{(n)}, \hat{\mathbf{C}}_0^{(n)}\}_{n=1}^N$

Output: $\{\mathcal{P}(\mathbf{x}_{k'} | \boldsymbol{\theta}_{k'}, \mathbf{z}_{1:k'}), \mathcal{P}(\boldsymbol{\theta}_{k'} | \mathbf{z}_{1:k'}), \hat{s}_{k'}^E, \hat{r}_{k'}^E, \hat{\beta}_{k'}^E, \hat{\gamma}_{k'}^E\}_{k'=1}^k$

INITIALIZATION

1: $k' \leftarrow 1$

2: **for** $\boldsymbol{\theta}_{k'} = [\beta_{k'}, \gamma_{k'}]^T \in \mathcal{D}$ **do**

3: $\mathcal{P}(\boldsymbol{\theta}_{k'} | \mathbf{z}_{1:k'-1}) \leftarrow \mathcal{N}(\beta_{k'}; \bar{\beta}_0, \sigma_{\beta}^2) \mathcal{N}(\gamma_{k'}; \bar{\gamma}_0, \sigma_{\gamma}^2)$

4: **for** $n = 1$ **to** N **do**

5: $w_{k'|k'-1}^{(n, \boldsymbol{\theta}_{k'})} \leftarrow N^{-1}$

6: $\hat{\mathbf{x}}_{k'|k'-1}^{(n, \boldsymbol{\theta}_{k'})} \leftarrow \hat{\mathbf{x}}_0^{(n)}$

7: $\hat{\mathbf{C}}_{k'|k'-1}^{(n, \boldsymbol{\theta}_{k'})} \leftarrow \hat{\mathbf{C}}_0^{(n)}$

8: **end for**

9: **end for**

10: **for** $k' = 1$ **to** k **do**

UPDATE

11: **for** $\boldsymbol{\theta}_{k'} \in \mathcal{D}$ **do**

12: **for** $n = 1$ **to** N **do**

13: $\alpha_{k'|k'-1}^{(n, \boldsymbol{\theta}_{k'})} \leftarrow \mathcal{N}\left(\mathbf{z}_{k'}; \mathbf{h}_1\left(\hat{\mathbf{x}}_{k'|k'-1}^{(n, \boldsymbol{\theta}_{k'})}\right), \mathbf{R}\left(\hat{\mathbf{x}}_{k'|k'-1}^{(n, \boldsymbol{\theta}_{k'})}\right) + \mathbf{H}\hat{\mathbf{C}}_{k'|k'-1}^{(n, \boldsymbol{\theta}_{k'})}\mathbf{H}^T\right)$

14: $w_{k'|k'}^{(n, \boldsymbol{\theta}_{k'})} \leftarrow \frac{\alpha_{k'|k'-1}^{(n, \boldsymbol{\theta}_{k'})} w_{k'|k'-1}^{(n, \boldsymbol{\theta}_{k'})}}{\sum_{n'=1}^N \alpha_{k'|k'-1}^{(n', \boldsymbol{\theta}_{k'})} w_{k'|k'-1}^{(n', \boldsymbol{\theta}_{k'})}}$

15: $\hat{\mathbf{x}}_{k'|k'}^{(n, \boldsymbol{\theta}_{k'})} \leftarrow \hat{\mathbf{x}}_{k'|k'-1}^{(n, \boldsymbol{\theta}_{k'})} + \mathbf{K}_{k'|k'-1}^{(n, \boldsymbol{\theta}_{k'})} \left[\mathbf{z}_{k'} - \mathbf{h}_1\left(\hat{\mathbf{x}}_{k'|k'-1}^{(n, \boldsymbol{\theta}_{k'})}\right) \right]$

16: $\hat{\mathbf{C}}_{k'|k'}^{(n, \boldsymbol{\theta}_{k'})} \leftarrow \hat{\mathbf{C}}_{k'|k'-1}^{(n, \boldsymbol{\theta}_{k'})} - \mathbf{K}_{k'|k'-1}^{(n, \boldsymbol{\theta}_{k'})} \mathbf{H} \hat{\mathbf{C}}_{k'|k'-1}^{(n, \boldsymbol{\theta}_{k'})}$

17: **end for**

18: $\mathcal{P}(\mathbf{x}_{k'} | \boldsymbol{\theta}_{k'}, \mathbf{z}_{1:k'}) \leftarrow \sum_{n=1}^N w_{k'|k'}^{(n, \boldsymbol{\theta}_{k'})} \mathcal{N}\left(\mathbf{x}_{k'}; \hat{\mathbf{x}}_{k'|k'}^{(n, \boldsymbol{\theta}_{k'})}, \hat{\mathbf{C}}_{k'|k'}^{(n, \boldsymbol{\theta}_{k'})}\right)$

19: $\mathcal{P}(\boldsymbol{\theta}_{k'} | \mathbf{z}_{1:k'}) \leftarrow \frac{\mathcal{P}(\boldsymbol{\theta}_{k'} | \mathbf{z}_{1:k'-1}) \sum_{n=1}^N \alpha_{k'|k'-1}^{(n, \boldsymbol{\theta}_{k'})} w_{k'|k'-1}^{(n, \boldsymbol{\theta}_{k'})}}{\sum_{\boldsymbol{\theta}_{k'} \in \mathcal{D}} \mathcal{P}(\boldsymbol{\theta}_{k'} | \mathbf{z}_{1:k'-1}) \sum_{n'=1}^N \alpha_{k'|k'-1}^{(n', \boldsymbol{\theta}_{k'})} w_{k'|k'-1}^{(n', \boldsymbol{\theta}_{k'})}}$

20: **end for**

21: Compute $\hat{s}_{k'}^E, \hat{r}_{k'}^E, \hat{\beta}_{k'}^E$, and $\hat{\gamma}_{k'}^E$ according to (37) – (40)

PREDICTION

22: **for** $\boldsymbol{\theta}_{k'+1} \in \mathcal{D}$ **do**

23: $\mathcal{P}(\boldsymbol{\theta}_{k'+1} | \mathbf{z}_{1:k'}) \leftarrow \sum_{\boldsymbol{\theta}_{k'} \in \mathcal{D}} \mathcal{P}(\boldsymbol{\theta}_{k'+1} | \boldsymbol{\theta}_{k'}) \mathcal{P}(\boldsymbol{\theta}_{k'} | \mathbf{z}_{1:k'})$

24: $n' \leftarrow 1$

25: **for** $\boldsymbol{\theta}_{k'} \in \mathcal{D}$ **do**

26: **for** $n = 1$ **to** N **do**

27: $w_{k'+1|k'}^{(n', \boldsymbol{\theta}_{k'+1})} = w_{k'+1|k'}^{(n, \boldsymbol{\theta}_{k'})} \leftarrow \frac{\mathcal{P}(\boldsymbol{\theta}_{k'+1} | \boldsymbol{\theta}_{k'}) \mathcal{P}(\boldsymbol{\theta}_{k'} | \mathbf{z}_{1:k'})}{\mathcal{P}(\boldsymbol{\theta}_{k'+1} | \mathbf{z}_{1:k'})} w_{k'|k'}^{(n, \boldsymbol{\theta}_{k'})}$

28: $\hat{\mathbf{x}}_{k'+1|k'}^{(n', \boldsymbol{\theta}_{k'+1})} \leftarrow \hat{\mathbf{x}}_{k'+1|k'}^{(n, \boldsymbol{\theta}_{k'})}$ (via moment matching)

29: $\hat{\mathbf{C}}_{k'+1|k'}^{(n', \boldsymbol{\theta}_{k'+1})} \leftarrow \hat{\mathbf{C}}_{k'+1|k'}^{(n, \boldsymbol{\theta}_{k'})}$ (via moment matching)

30: $n' \leftarrow n' + 1$

31: **end for**

$L + 1$ most recent MMSE estimates of the infection rate, that is,

$$\hat{\beta}_{L,k} = \frac{\sum_{\ell=0}^L (\ell - L/2) (\hat{\beta}_{k-L+\ell}^E - \ell - (L+1)^{-1} \sum_{\ell'=0}^L \hat{\beta}_{k-\ell'}^E)}{\Delta_t \sum_{\ell=0}^L (\ell - L/2)^2}.$$

The estimator $\hat{\beta}_{L,k}$ is unbiased, and its variance is

$$\sigma_{L,k}^2 = \frac{\hat{\sigma}_{\omega,L}^2}{\Delta_t^2 \sum_{\ell=0}^L (\ell - L/2)^2},$$

where $\hat{\sigma}_{\omega,L}^2$ is an estimate of σ_{ω}^2 , i.e., the variance of the noise term $\omega_{k,\beta}$ given by

$$\hat{\sigma}_{\omega,L}^2 = \frac{\sum_{\ell=1}^L (\hat{\beta}_{k-\ell}^E - \ell + 1 - \hat{\beta}_{k-\ell}^E - \Delta_t \hat{\beta}_{L,k})^2}{L-1}.$$

In practice, the slope $\hat{\beta}_k$ is piecewise constant, that is, constant over subintervals of the entire interval $[k-L, k]$. Therefore, we aim to find the value L^* , $L_{\min} \leq L^* \leq L_{\max}$, — hence, the subinterval $[k-L^*, k]$ — such that the linear model with constant slope in (41) is a valid assumption for the time evolution of the infection rate. Let \bar{L} be the candidate value for L^* , and let us define the auxiliary variable $y_{\bar{L},k}$ as

$$y_{\bar{L},k} \triangleq \frac{(\hat{\beta}_k^E - \hat{\beta}_{k-1}^E - \hat{\beta}_{\bar{L},k}^E \Delta_t)^2}{\hat{\sigma}_{\omega,L}^2}. \quad (42)$$

It is easy to verify that, if the linear model (41) is valid and $\hat{\beta}_{\bar{L},k}$ represents an accurate estimate of the constant slope over the interval $[k-\bar{L}, k]$, then $y_{\bar{L},k}$ is a central chi-square distributed variable with one degree of freedom, i.e., $y_{\bar{L},k} \sim \chi_1^2(0)$; otherwise, $y_{\bar{L},k}$ is a non-central chi-square distributed variable with one degree of freedom and non-centrality parameter $\nu > 0$, i.e., $y_{\bar{L},k} \sim \chi_1^2(\nu)$. The problem of finding L^* can therefore be recast as a hypothesis test: the null hypothesis (H_0) states that the linear assumption is valid and $\hat{\beta}_{\bar{L},k}$ is a good estimate of the constant slope; the alternate hypothesis (H_1) states that the linear hypothesis with constant slope is invalid. The hypothesis test can be formally written as

$$\begin{cases} H_0: y_{\bar{L},k} \sim \chi_1^2(0), \\ H_1: y_{\bar{L},k} \sim \chi_1^2(\nu), \end{cases}$$

and the decision rule is

$$y_{\bar{L},k} \underset{H_1}{\overset{H_0}{\leq}} T_H.$$

The threshold T_H is obtained as

$$T_H = F_{\chi_1^2(0)}^{-1}(1 - P_{fa}),$$

where P_{fa} is the probability to reject hypothesis H_0 when it is true and $F_{\chi_1^2(0)}^{-1}(\cdot)$ is the inverse cumulative density function of a central chi-square variable with one degree of freedom. In order to select L^\star , we employ the following iterative procedure: if with \bar{L} hypothesis H_0 is accepted, and with $\bar{L} - 1$ hypothesis H_0 is rejected, then $L^\star = \bar{L}$; otherwise, \bar{L} is decreased and the test is repeated. This iterative procedure starts with $\bar{L} = L_{\max}$. Note that, alternatively, the threshold could be obtained by fixing the probability to accept hypothesis H_0 when it is false; in such a case, the non-centrality parameter ν is approximated with the numerator of (42).

Algorithm 2 Epidemic Forecasting with Stochastic SIR Model

Input: $k, \mathcal{P}(\mathbf{x}_k | \mathbf{z}_{1:k}), \mathcal{P}(\boldsymbol{\theta}_k | \mathbf{z}_{1:k}), J, L, K, P, \{\hat{\beta}_{k'}^E\}_{k'=k-L}^k$

Output: $\{\hat{s}_{k'}^F, \hat{t}_{k'}^F, \hat{\beta}_{k'}^F, \hat{\gamma}_{k'}^F\}_{k'=k+1}^{k+K}$

INITIALIZATION

- 1: Draw J samples $\tilde{\mathbf{x}}_k^{(j)} = [\tilde{s}_k^{(j)}, \tilde{t}_k^{(j)}]^T$ from $\mathcal{P}(\mathbf{x}_k | \mathbf{z}_{1:k})$
- 2: Draw J samples $\tilde{\boldsymbol{\theta}}_k^{(j)} = [\tilde{\beta}_k^{(j)}, \tilde{\gamma}_k^{(j)}]^T$ from $\mathcal{P}(\boldsymbol{\theta}_k | \mathbf{z}_{1:k})$
- 3: $\hat{\beta}_{L,k} \leftarrow \frac{\sum_{\ell'=0}^L (\ell-L/2) (\hat{\beta}_{k-L+\ell}^E - (L+1)^{-1} \sum_{\ell'=0}^L \hat{\beta}_{k-\ell}^E)}{\Delta_t \sum_{\ell'=0}^L (\ell-L/2)^2}$
- 4: $\hat{\sigma}_{\omega,L}^2 \leftarrow \frac{\sum_{\ell'=1}^L (\hat{\beta}_{k-\ell}^E + 1 - \hat{\beta}_{k-\ell}^E - \Delta_t \hat{\beta}_{L,k})^2}{L-1}$
- 5: $\hat{\sigma}_{L,k}^2 \leftarrow \frac{\hat{\sigma}_{\omega,L}^2}{\Delta_t^2 \sum_{\ell'=0}^L (\ell-L/2)^2}$

FORECASTING

6: **for** $j = 1$ **to** J **do**

7: **for** $k' = k+1$ **to** $k+K$ **do**

STATE EVOLUTION

- 8: Draw $\lambda = [\lambda_1, \lambda_2]^T \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_2 \Delta_t)$
- 9: $\sigma_{1,k'-1}^{(j)} \leftarrow \sqrt{P^{-1} \tilde{\beta}_{k'-1}^{(j)} - 1 \tilde{s}_{k'-1}^{(j)} - 1 \tilde{t}_{k'-1}^{(j)}}$
- 10: $\sigma_{2,k'-1}^{(j)} \leftarrow \sqrt{P^{-1} \tilde{\gamma}_{k'-1}^{(j)} - 1 \tilde{t}_{k'-1}^{(j)}}$
- 11: $\tilde{s}_{k'}^{(j)} \leftarrow \tilde{s}_{k'-1}^{(j)} - \tilde{\beta}_{k'-1}^{(j)} \tilde{s}_{k'-1}^{(j)} - 1 \tilde{t}_{k'-1}^{(j)} \Delta_t + \sigma_{1,k'-1}^{(j)} \lambda_1$
- 12: $\tilde{t}_{k'}^{(j)} \leftarrow \tilde{t}_{k'-1}^{(j)} + \tilde{\beta}_{k'-1}^{(j)} \tilde{s}_{k'-1}^{(j)} - 1 \tilde{t}_{k'-1}^{(j)} \Delta_t$
 $\quad \quad \quad - \tilde{\gamma}_{k'-1}^{(j)} \tilde{t}_{k'-1}^{(j)} \Delta_t - \sigma_{1,k'-1}^{(j)} \lambda_1 + \sigma_{2,k'-1}^{(j)} \lambda_2$

PARAMETERS EVOLUTION

- 13: Draw $\eta \sim \mathcal{N}(0, \hat{\sigma}_{L,k}^2)$
- 14: $\tilde{\beta}_{k'}^{(j)} \leftarrow \tilde{\beta}_{k'-1}^{(j)} + \hat{\beta}_{L,k} \Delta_t + \eta$
- 15: $\tilde{\gamma}_{k'}^{(j)} \leftarrow \tilde{\gamma}_{k'-1}^{(j)}$

ENSEMBLE MEAN

- 16: $\hat{s}_{k'}^F \leftarrow J^{-1} \sum_{j=1}^J \tilde{s}_{k'}^{(j)}$
- 17: $\hat{t}_{k'}^F \leftarrow J^{-1} \sum_{j=1}^J \tilde{t}_{k'}^{(j)}$
- 18: $\hat{\beta}_{k'}^F \leftarrow J^{-1} \sum_{j=1}^J \tilde{\beta}_{k'}^{(j)}$
- 19: $\hat{\gamma}_{k'}^F \leftarrow J^{-1} \sum_{j=1}^J \tilde{\gamma}_{k'}^{(j)}$
- 20: **end for**
- 21: **end for**

Biography



DOMENICO GAGLIONE (Member, IEEE) received the B.Sc. and M.Sc. degrees (*summa cum laude*) in telecommunications engineering from the Università degli Studi di Napoli “Federico II”, Naples, Italy, in 2011 and 2013, respectively, and the Ph.D. degree from the Department of Electronic and Electrical Engineering, University of Strathclyde, Glasgow, U.K., in 2017. Since 2016, he has been a Research Assistant with the University of Strathclyde. He joined the NATO STO Centre for Maritime Research & Experimentation (CMRE), La Spezia, Italy, in 2017 as a Junior Scientist. His research interests include statistical signal processing with emphasis on state estimation, data fusion, and multisensor multitarget tracking. He was a recipient of the Best Student Paper Award at the 2015 IEEE International Radar Conference (RadarCon), Arlington, VA, USA, and of the NATO STO Scientific Achievement Award in 2020 as a member of the Data Knowledge and Operational Effectiveness (DKOE) Research Group, CMRE for “Advances in Artificial Intelligence and Information Fusion for Maritime Situational Awareness.”



PAOLO BRACA (Senior Member, IEEE) received the Laurea degree (*summa cum laude*) in electronic engineering and the Ph.D. degree (Hons.) in information engineering from the University of Salerno, Italy, in 2006 and 2010, respectively. In 2009, he was a Visiting Scholar with the Department of Electrical and Computer Engineering, University of Connecticut, Storrs, CT, USA. From 2010 to 2011, he was a Postdoctoral Associate with the University of Salerno. In 2011, he joined the NATO Science & Technology Organization Centre for Maritime Research and Experimentation (CMRE), where he is currently a Senior Scientist with the Research Department and a Project Manager of the Data Knowledge and Operational Effectiveness (DKOE) Program. Furthermore, he led a number of research projects funded by the EU Horizon 2020 programme, by the U.S. Office of Naval Research (ONR), and by DRDC. He conducts research in the general area of statistical signal processing with emphasis on detection and estimation theory, wireless sensor network, multi-agent algorithms, target tracking and data fusion, adaptation and learning over graphs, and radar (sonar) signal processing. He has coauthored more than 100 publications in international scientific journals and conference proceedings. He was awarded with the National Scientific Qualification to function as an Associate and a Full Professor in Italian Universities, in 2017 and 2018, respectively. He is in the technical committee of the major international conferences in the field of signal processing and data fusion. He was a recipient of the Best Student Paper Award (first runner-up) at the 12th

International Conference on Information Fusion, in 2009, and the NATO STO Scientific Achievement Award 2017 for its contribution to “Development and Demonstration of Networked Autonomous ASW” from the NATO Chief Scientist. He coauthored the paper received the Best Paper Award (first runner-up) at the Sensor Signal Processing for Defence (SSPD), in 2019. He was also a recipient of the NATO STO Scientific Achievement Award, in 2020, as a Team Leader for the “Advances in Artificial Intelligence and Information Fusion for Maritime Situational Awareness.” He serves as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING, the IEEE TRANSACTIONS ON AEROSPACE AND ELECTRONIC SYSTEMS, *ISIF Journal of Advances in Information Fusion*, *EURASIP Journal on Advances in Signal Processing*, and *IET Radar, Sonar & Navigation*. In 2017, he was a Lead Guest Editor of the Special Issue “Sonar Multi-Sensor Applications and Techniques” in *IET Radar, Sonar & Navigation*. He served as an Associate Editor for the *IEEE Signal Processing Magazine* (E-Newsletter), from 2014 to 2016.



LEONARDO MARIA MILLEFIORI (Member, IEEE) received the B.Sc. degree in aerospace information engineering and the M.Sc. degree (*summa cum laude*) in communication engineering with focus on radar systems and remote sensing from the Sapienza University of Rome, Italy, in 2010 and 2013, respectively. He was a Visiting Researcher with the NATO Science and Technology Organization Center for Maritime Research and Experimentation (CMRE), La Spezia, where he joined the Research Department as a Research Scientist in 2014. His research interests include target motion modeling, statistical learning and signal processing, and target tracking and data fusion. He was a recipient of the NATO STO Scientific Achievement Award, in 2020, as a member of the Data Knowledge and Operational Effectiveness (DKOE) Research Group, CMRE for “Advances in Artificial Intelligence and Information Fusion for Maritime Situational Awareness.”



GIOVANNI SOLDI received the master’s degree in applied mathematics from the University of Milan, in 2011, and the Ph.D. degree in signal processing from Télécom ParisTech, in 2016. Since 2016, he has been a Scientist with the Centre for Maritime Research and Experimentation (CMRE), La Spezia, Italy. His current research interests include statistical signal processing with a focus on state estimation, data fusion, and multisensor-multitarget tracking techniques. He was a recipient of the NATO STO Scientific Achievement Award, in 2020, as a member of the Data Knowledge and Operational

Effectiveness (DKOE) Research Group, CMRE for “Advances in Artificial Intelligence and Information Fusion for Maritime Situational Awareness.”



NICOLA FORTI received the M.S. degree in automation engineering and the Ph.D. degree in information engineering from the University of Florence, Italy, in 2013 and 2016, respectively. From 2015 to 2016, he was a Visiting Researcher with the Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA, USA. He is currently a Scientist with the Research Department, NATO Science & Technology Organization—Centre for Maritime Research and Experimentation (CMRE). His current research interests include state estimation and control theory, statistical signal processing, machine learning, target tracking, and data fusion. He was a recipient of the NATO STO Scientific Achievement Award, in 2020, as a member of the Data Knowledge and Operational Effectiveness (DKOE) Research Group, CMRE for “Advances in Artificial Intelligence and Information Fusion for Maritime Situational Awareness.”



STEFANO MARANO (Senior Member, IEEE) received the Laurea degree in electronic engineering and the Ph.D. degree in electronic engineering and computer science from the University of Naples, Italy, in 1993 and 1997, respectively. He has held visiting positions at the Physics Department, University of Wales, College of Cardiff, and also at the ECE Department, University of California at San Diego, in 1996 and 2013, respectively. Since 1999, he has been with the University of Salerno, Italy, where he is currently a Professor with DIEM. His current research interests include statistical signal processing with emphasis on distributed inference, sensor networks, and information theory. He was a recipient of the IEEE TRANSACTIONS ON ANTENNAS AND PROPAGATION 1999 Best Paper Award, and coauthored the paper winning the Best Student Paper Award (2nd place) at the 12th Conference on Information Fusion in 2009. He was in the Organizing Committee of FUSION 2006 and RADARCON 2008. He is a member of the EURASIP Special Area Team in Theoretical and Methodological Trends in Signal Processing and a Guest Editor of a Special Issue of *EURASIP Journal on Advances in Signal Processing*. From 2010 to 2014, he served as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING. He also served as an Associate Editor and a Technical Editor for the IEEE TRANSACTIONS ON AEROSPACE AND ELECTRONIC SYSTEMS from 2009 to 2016.



PETER K. WILLETT (Fellow, IEEE) has been a Faculty Member with the Electrical and Computer Engineering Department, University of Connecticut, since 1986. Since 1998, he has also been a Professor. His current research interests include statistical signal processing, detection, machine learning, communications, data fusion, and tracking. He is a member of the IEEE Fellows Committee and of the Ethics Committee and Periodicals Committee, and was of the IEEE Signal Processing Society's Technical Activities and Conference Boards. He is a member of the IEEE AESS Board of Governors and was the Chair of the IEEE Signal Processing Society's Sensor-Array and Multichannel (SAM) Technical Committee. He is a Chief Editor of the *IEEE AESS Magazine*, from 2018 to 2020. He was an Editor-in-Chief of the IEEE TRANSACTIONS ON AEROSPACE AND ELECTRONIC SYSTEMS from 2006 to 2011 and the IEEE SIGNAL PROCESSING LETTERS, from 2014 to 2016. He was also an AESS Vice President for Publications from 2012 to 2014.



KRISHNA R. PATTIPATI (Life Fellow, IEEE) received the B.Tech. degree (Hons.) in electrical engineering from IIT Kharagpur, in 1975, and the M.S. and Ph.D. degrees in systems engineering from the University of Connecticut (UConn), Storrs, CT, USA, in 1977 and 1980, respectively. He was with ALPHATECH, Inc., Burlington, MA, USA, from 1980 to 1986. He has been with the Department of Electrical and Computer Engineering, UConn, where he is currently the Board of Trustees Distinguished Professor and the UTC Chair Professor in systems engineering. He is also a Co-Founder of Qualtech Systems, Inc., a firm specializing in advanced integrated diagnostics software tools (TEAMS, TEAMS-RT, TEAMS-RDS, and TEAMATE), and serves on the board for Aptima, Inc. His research interests include proactive decision support, uncertainty quantification, smart manufacturing, autonomy, knowledge representation, and optimization-based learning and inference. A common theme among these applications is that they are characterized by a great deal of uncertainty, complexity, and computational intractability. He is an Elected Fellow of the Connecticut Academy of Science and Engineering. He was selected by the IEEE Systems, Man, and Cybernetics (SMC) Society as the Outstanding Young Engineer of 1984, and received the Centennial Key to the Future Award. He was a co-recipient of the Andrew P. Sage Award for the Best SMC Transactions Paper in 1999, the Barry Carlton Award for the Best AES Transactions Paper in 2000, the 2002 and 2008 NASA Space Act Awards for A Comprehensive Toolset for Model-based Health Monitoring and Diagnosis, and Real-time Update of Fault-Test Dependencies of Dynamic Systems: A Comprehensive Toolset for Model-Based Health Monitoring and Diagnostics, and the 2003 AAUP Research Excellence Award at UConn. He received the best technical paper awards at the 1985, 1990, 1994, 2002,

2004, 2005, and 2011 IEEE AUTOTEST Conferences, and at the 1997, 2004 Command and Control Conference. He was the Vice-President for Technical Activities of the IEEE SMC Society, from 1998 to 1999, and as the Vice-President for Conferences and Meetings of the IEEE SMC Society, from 2000 to 2001. He served as the Editor-in-Chief for the IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART B, from 1998 to 2001.

REFERENCES

- [1]. WHO. (1. 2020). Novel Coronavirus—China. [Online]. Available: <http://www.who.int/csr/don/12-january-2020-novel-coronavirus-china/en/>
- [2]. WHO. (3. 2020). Coronavirus Disease (COVID-19) Pandemic. [Online]. Available: <http://www.euro.who.int/en/health-topics/health-emergencies/coronavirus-covid-19/novel-coronavirus-2019-ncov>
- [3]. Heesterbeek H et al. , “Modeling infectious disease dynamics in the complex landscape of global health,” *Science*, vol. 347, no. 6227, pp. 1–12, 3. 2015.
- [4]. Kermack WO, McKendrick AG, and Walker GT, “A contribution to the mathematical theory of epidemics,” *Proc. Roy. Soc. LondonA*, vol. 115, no. 772, pp. 700–721, 8. 1927.
- [5]. Bjørnstad ON, Shea K, Krzywinski M, and Altman N, “Modeling infectious epidemics,” *Nature Methods*, vol. 17, no. 5, pp. 455–456, 4. 2020. [PubMed: 32313223]
- [6]. Lauer SA, Grantz KH, Bi Q, Jones FK, Zheng Q, Meredith HR, Azman AS, Reich NG, and Lessler J, “The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: Estimation and application,” *Ann. Internal Med*, vol. 172, no. 9, pp. 577–582, 5 2020. [PubMed: 32150748]
- [7]. Yi N, Zhang Q, Mao K, Yang D, and Li Q, “Analysis and control of an SEIR epidemic system with nonlinear transmission rate,” *Math. Comput. Model*, vol. 50, no. 9, pp. 1498–1513, 11. 2009. [PubMed: 32288203]
- [8]. Hou C, Chen J, Zhou Y, Hua L, Yuan J, He S, Guo Y, Zhang S, Jia Q, Zhao C, Zhang J, Xu G, and Jia E, “The effectiveness of quarantine of Wuhan city against the corona virus disease 2019 (COVID-19): A well- mixed SEIR model analysis,” *J. Med. Virol*, vol. 92, no. 7, pp. 841–848, 4. 2020. [PubMed: 32243599]
- [9]. Hu Z, Cui Q, Han J, Wang X, Sha WEI, and Teng Z, “Evaluation and prediction of the COVID-19 variations at different input population and quarantine strategies, a case study in guangdong province, China,” *Int. J. Infectious Diseases*, vol. 95, pp. 231–240, 6. 2020. [PubMed: 32334117]
- [10]. Maier BF and Brockmann D, “Effective containment explains subexponential growth in recent confirmed COVID-19 cases in China,” *Science*, vol. 368, no. 6492, pp. 742–746, 5 2020. [PubMed: 32269067]
- [11]. Li R, Pei S, Chen B, Song Y, Zhang T, Yang W, and Shaman J, “Sub-stantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2),” *Science*, vol. 368, no. 6490, pp. 489–493, 5 2020. [PubMed: 32179701]
- [12]. Arino J and Driessche P, “Metapopulation epidemic models. A survey,” *Fields Inst. Commun*, vol. 48, pp. 1–12, 1. 2006.
- [13]. Nowzari C, Preciado VM, and Pappas GJ, “Analysis and control of epidemics: A survey of spreading processes on complex networks,” *IEEE Control Syst. Mag*, vol. 36, no. 1, pp. 26–46, 2. 2016.
- [14]. Jia JS, Lu X, Yuan Y, Xu G, Jia J, and Christakis NA, “Population flow drives spatio-temporal distribution of COVID-19 in China,” *Nature*, vol. 582, no. 7812, pp. 389–394, 4. 2020. [PubMed: 32349120]
- [15]. Kraemer MUG, Yang C-H, Gutierrez B, Wu C-H, Klein B, Pigott AM, du Plessis L, Faria NR, Li R, Hanage WP, Brownstein JS, Layan M, Vespignani A, Tian H, Dye C, Pybus OG, and Scarpino SV, “The effect of human mobility and control measures on the COVID-19 epidemic in China,” *Science*, vol. 368, no. 6490, pp. 493–497, 5 2020. [PubMed: 32213647]
- [16]. Chinazzi M, Davis JT, Ajelli M, Gioannini C, Litvinova M, Merler S, Pastore Piontti AY, Mu K, Rossi L, Sun K, Viboud C, Xiong X, Yu H, Halloran ME, Longini IM, and Vespignani A, “The

- effect of travel restrictions on the spread of the 2019 novel Coronavirus (COVID-19) outbreak,” *Science*, vol. 368, no. 6489, pp. 395–400, 4. 2020. [PubMed: 32144116]
- [17]. Skvortsov A and Ristic B, “Monitoring and prediction of an epidemic outbreak using syndromic observations,” *Math. Biosci.*, vol. 240, no. 1, pp. 12–19, 11. 2012. [PubMed: 22705339]
- [18]. Sheinson DM, Niemi J, and Meiring W, “Comparison of the performance of particle filter algorithms applied to tracking of a disease epidemic,” *Math. Biosci.*, vol. 255, pp. 21–32, 9. 2014. [PubMed: 25016201]
- [19]. Shahtori NM, Scoglio C, Pourhabib A, and Sahneh FD, “Sequential Monte Carlo filtering estimation of Ebola progression in West Africa,” in *Proc. ACC*, Boston, MA, USA, 7. 2016, pp. 1277–1282.
- [20]. Braca P, Aubry A, Millefiori LM, De Maio A, and Marano S, “Multi-class random matrix filtering for adaptive learning,” *IEEE Trans. Signal Process.*, vol. 68, pp. 359–373, 11. 2020.
- [21]. Soldi G, Meyer F, Braca P, and Hlawatsch F, “Self-tuning algorithms for multisensor-multitarget tracking using belief propagation,” *IEEE Trans. Signal Process.*, vol. 67, no. 15, pp. 3922–3937, 8. 2019.
- [22]. Bar-Shalom Y, Willett PK, and Tian X, *Tracking Data Fusion: A Handbook Algorithms*. Storrs, CT, USA: YBS Publishing, 2011.
- [23]. Poor HV, *An Introduction to Signal Detection Estimation*. New York, NY, USA: Springer, 1994.
- [24]. Lewis JM, “Roots of ensemble forecasting,” *Monthly Weather Rev.*, vol. 133, no. 7, pp. 1865–1885, 7. 2005.
- [25]. Leutbecher M and Palmer T, “Ensemble forecasting,” *J. Comput. Phys.*, vol. 227, no. 7, pp. 3515–3539, 3. 2008.
- [26]. Leutbecher M, “Ensemble size: How suboptimal is less than infinity?” *Quart. J. Roy. Meteorol. Soc.*, vol. 145, no. S1, pp. 107–128, 9. 2019.
- [27]. Ray EL and Reich NG, “Prediction of infectious disease epidemics via weighted density ensembles,” *PLoS Comput. Biol.*, vol. 14, no. 2, pp. 1–23, 2. 2018.
- [28]. Chowell G, Luo R, Sun K, Roosa K, Tariq A, and Viboud C, “Real-time forecasting of epidemic trajectories using computational dynamic ensembles,” *Epidemics*, vol. 30, pp. 1–10, 3. 2020.
- [29]. Evensen G, “The ensemble Kalman filter: Theoretical formulation and practical implementation,” *Ocean Dyn.*, vol. 53, no. 4, pp. 343–367, 11. 2003.
- [30]. Kotecha JH and Djuric PM, “Gaussian sum particle filtering,” *IEEE Trans. Signal Process.*, vol. 51, no. 10, pp. 2602–2612, 10. 2003.
- [31]. Crouse DF, Willett P, Pattipati K, and Svensson L, “A look at Gaussian mixture reduction algorithms,” in *Proc. FUSION*, Chicago, IL, USA, 7. 2011, pp. 1–8.
- [32]. Simon D, *Optimal State Estimation: Kalman, H Infinity, and Nonlinear Approaches*. Hoboken, NJ, USA: Wiley, 2006.
- [33]. Bishop C, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer, 2006.
- [34]. Goodfellow I, Bengio Y, and Courville A, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.
- [35]. Theodoridis S, *Machine Learning*. Cambridge, MA, USA: Academic Press, 2015.
- [36]. Ljung L, *System Identification*. Hoboken, NJ, USA: Wiley, 1999.
- [37]. Allen LJS, “A primer on stochastic epidemic models: Formulation, numerical simulation, and analysis,” *Infectious Disease Model.*, vol. 2, no. 2, pp. 128–142, 5 2017.
- [38]. Dati COVID-19 Italia. Accessed: Aug. 4, 2020. [Online]. Available: <https://github.com/pcm-dpc/COVID-19>
- [39]. Peng L, Yang W, Zhang D, Zhuge C, and Hong L, “Epidemic analysis of COVID-19 in China by dynamical modeling,” 2020, arXiv:2002.06563. [Online]. Available: <http://arxiv.org/abs/2002.06563>
- [40]. Dong E, Du H, and Gardner L, “An interactive Web-based dashboard to track COVID-19 in real time,” *Lancet Infectious Diseases*, vol. 20, no. 5, pp. 533–534, 5 2020. [PubMed: 32087114]
- [41]. COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University. Accessed: Aug. 4, 2020. [Online]. Available: <https://github.com/CSSEGISandData/COVID-19>

- [42]. BBC. (3. 2020). Coronavirus: Three Out of Four Americans Under Some Form of Lockdown. [Online]. Available: <https://www.bbc.com/news/world-us-canada-52103066>
- [43]. Oran DP and Topol EJ, "Prevalence of asymptomatic SARS-CoV-2 infection," *Ann. Internal Med.*, vol. 173, no. 5, pp. 362–367, 9. 2020. [PubMed: 32491919]
- [44]. Rice J, *Mathematical Statistics and Data Analysis*. Boston, MA, USA: Cengage Learning, 2006.

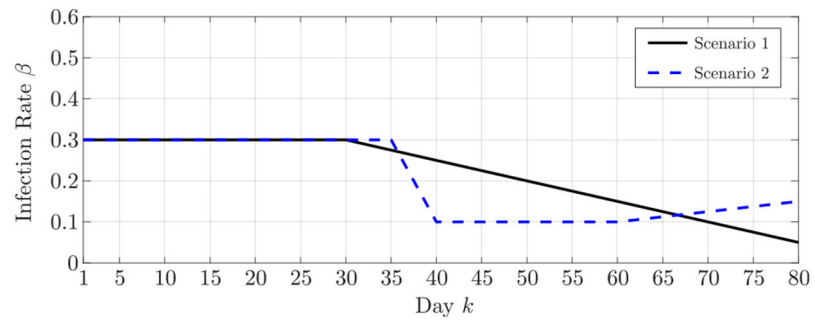


FIGURE 1.
Evolution of the infection rate in the simulated scenarios.

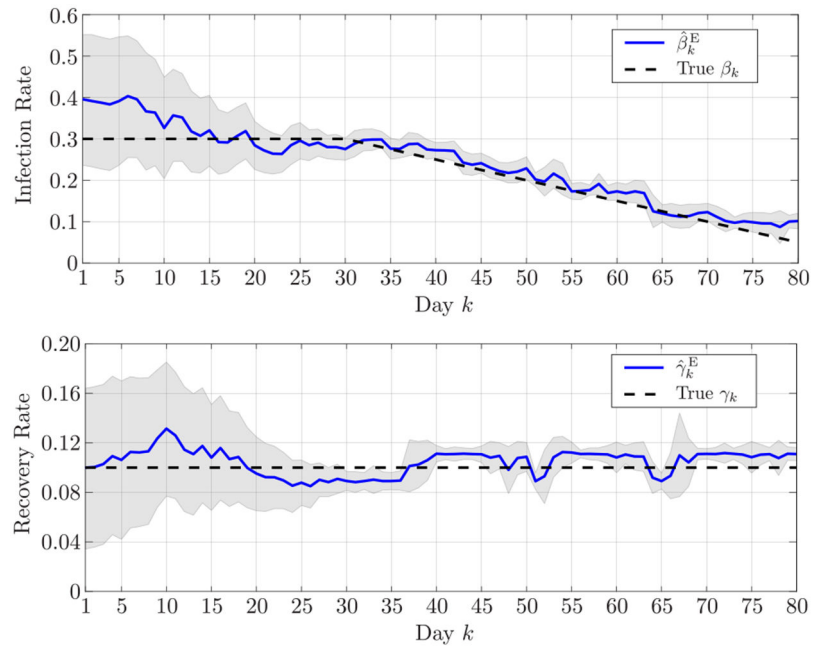


FIGURE 2. Estimated (top) infection rate and (bottom) recovery rate in the first simulated scenario. The shaded areas represent the 90% confidence interval.

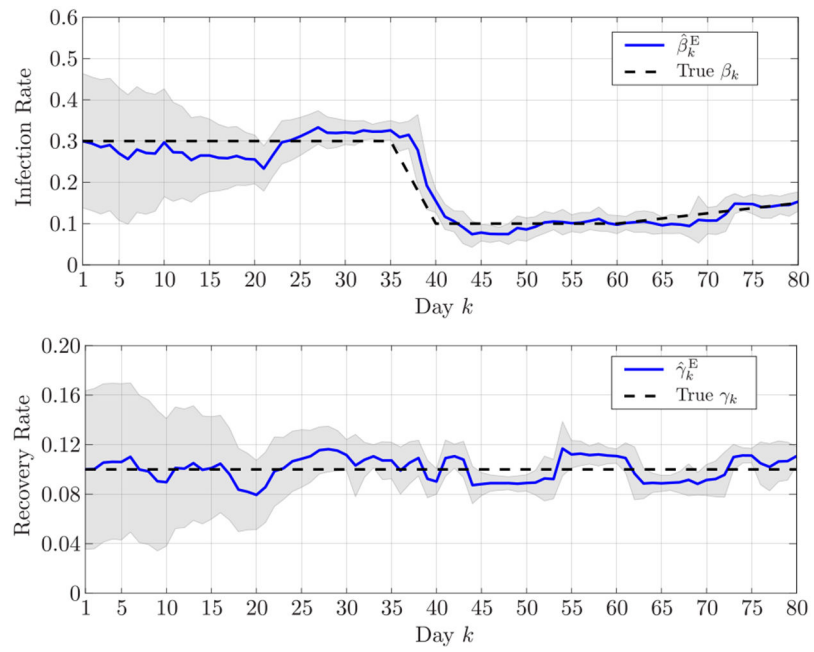


FIGURE 3. Estimated (top) infection rate and (bottom) recovery rate in the second simulated scenario. The shaded areas represent the 90% confidence interval.

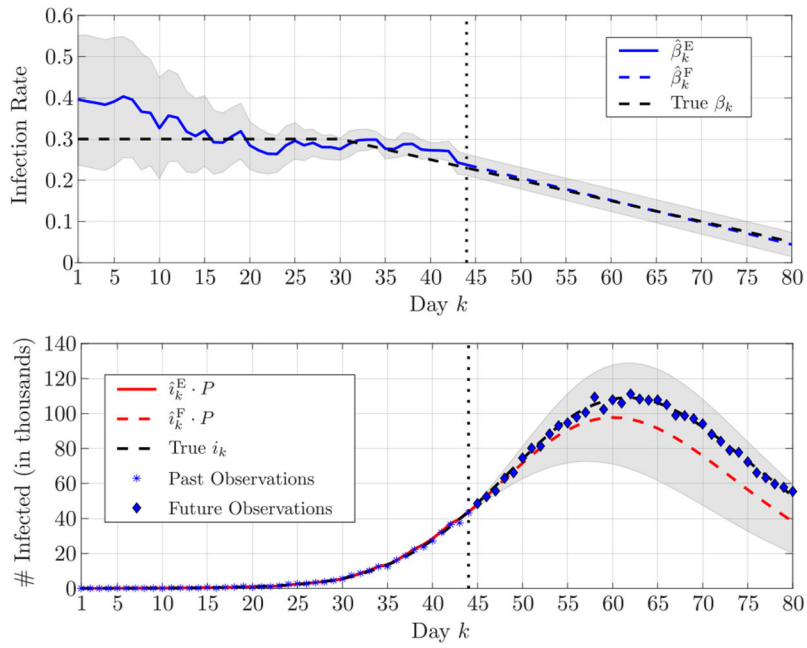


FIGURE 4.

Estimation and forecasting, respectively in solid and dashed lines, of (top) the infection rate and (bottom) the number of infected individuals in the first scenario; the superscripts E and F stand for *estimate* and *forecast*, respectively. The estimation is up to $k=44$ (marked by a vertical dotted line), and the forecast is up to $k=80$. The shaded areas represent the 90% confidence interval.

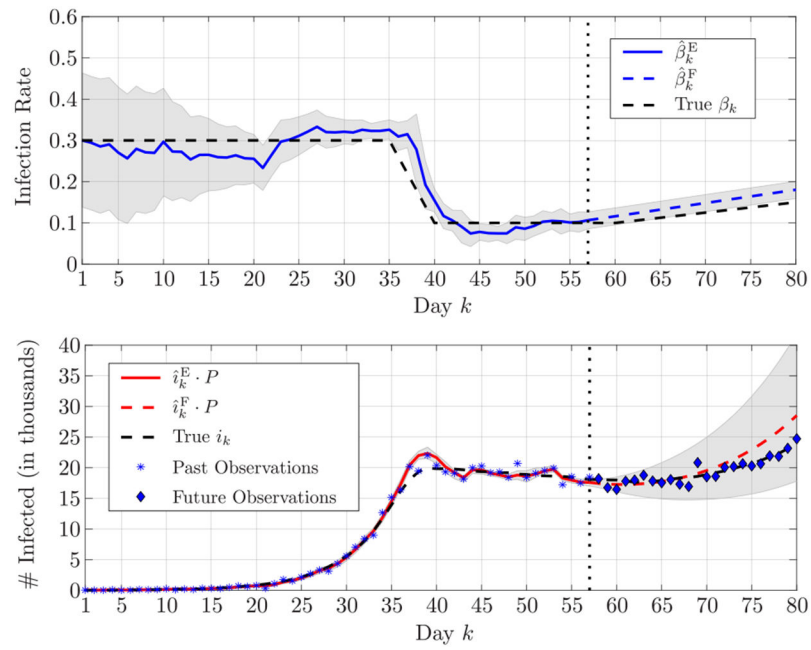


FIGURE 5.

Estimation and forecasting, respectively in solid and dashed lines, of (top) the infection rate and (bottom) the number of infected individuals in the second scenario; the superscripts E and F stand for *estimate* and *forecast*, respectively. The estimation is up to $k = 57$ (marked by a vertical dotted line), and the forecast is up to $k = 80$. The shaded areas represent the 90% confidence interval.

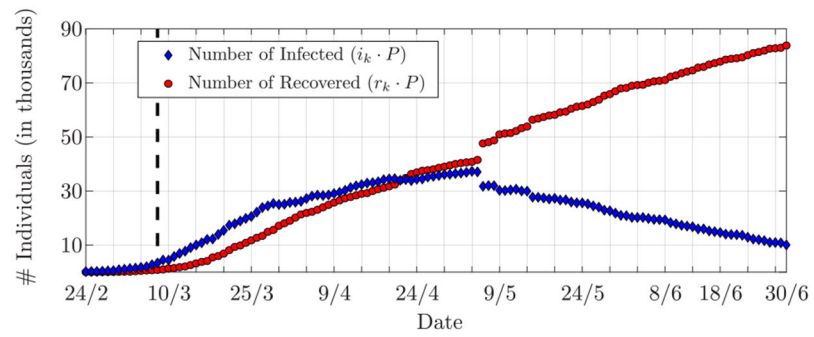


FIGURE 6. Numbers of infected and recovered (i.e., hospital releases plus deaths) individuals in Lombardia, Italy, from February 24, 2020, to June 30, 2020 (data from Protezione Civile [38]). The vertical dashed line indicates March 8, 2020, the beginning of the lockdown. The large steps on May 6 are due to an inaccurate reporting of the data, as explained in Section V-B1.

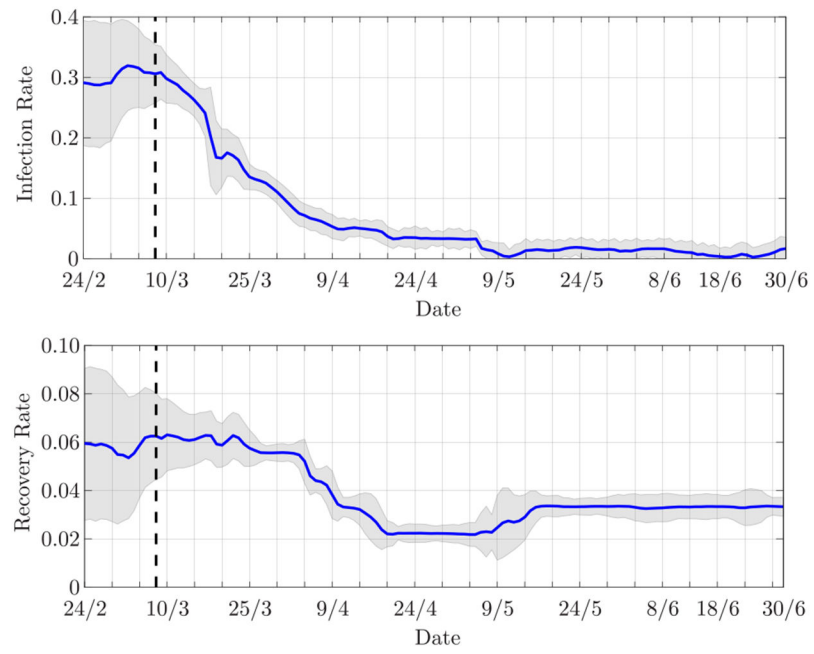


FIGURE 7. Estimated (top) infection rate and (bottom) recovery rate for Lombardia. The vertical dashed line indicates March 8, 2020, the beginning of the lockdown. The shaded areas represent the 90% confidence interval.

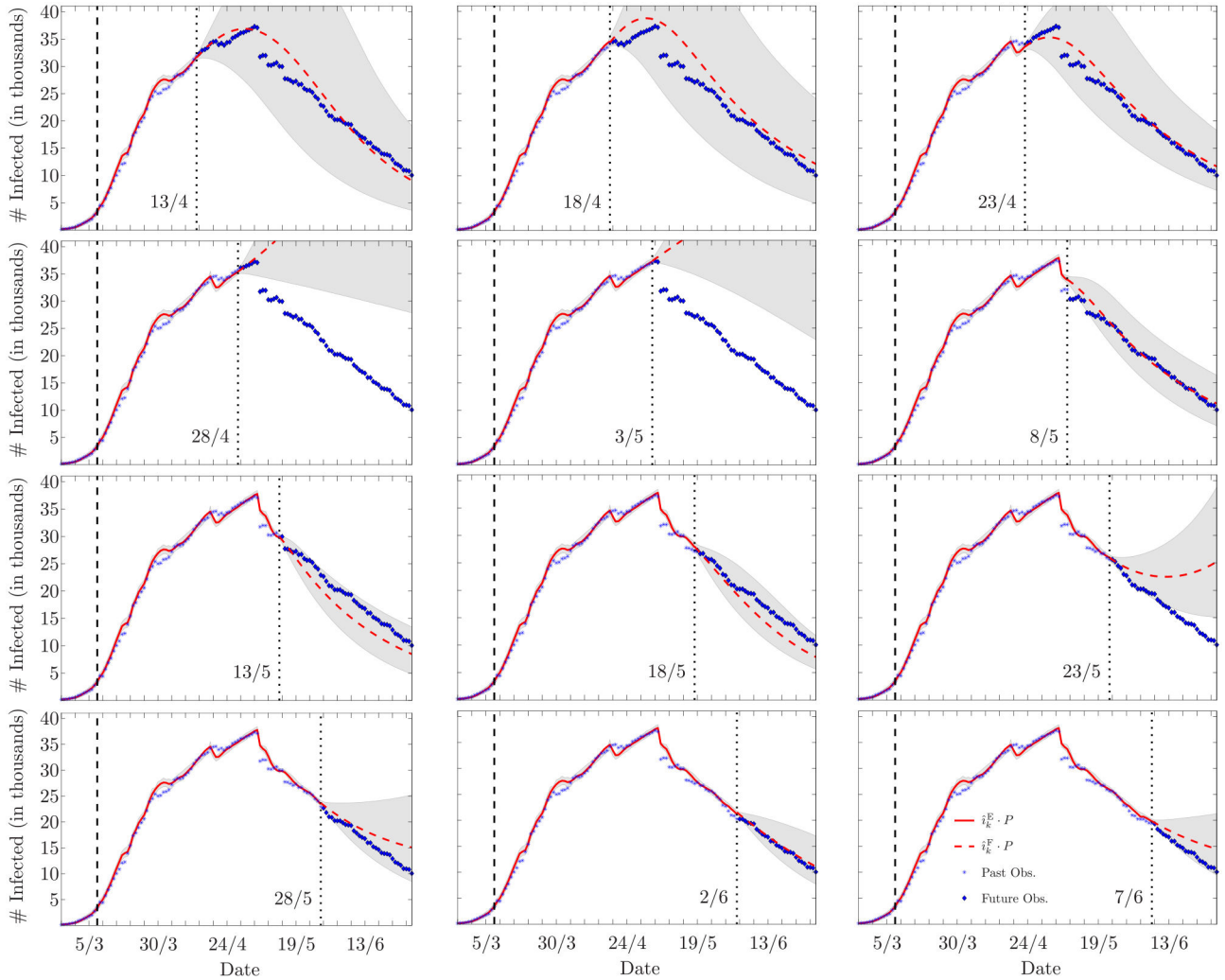


FIGURE 8. Estimation and forecasting, respectively in solid and dashed lines, of the number of infected individuals in Lombardia, Italy (legend is reported in the bottom-right corner image; the superscript E stands for *estimate*, and the superscript F stands for *forecast*). The date corresponding to the end of the estimation and the beginning of the forecast is marked by a vertical dotted line (the leftmost vertical dashed line marks March 8, the beginning of the lockdown). In all the cases, the forecast horizon is June 30. The shaded area represents the 90% confidence interval. The poor forecasts made on April 28, and May 3, relate to the inaccurate data later provided on May 6, as explained in Section V-B1.

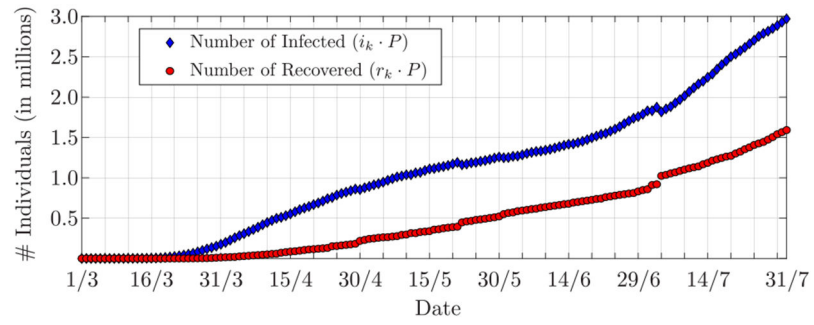


FIGURE 9. Numbers of infected and recovered (i.e., hospital releases plus deaths) individuals in the USA, from March 1, 2020, to July 31, 2020 (data from JHU [41]).

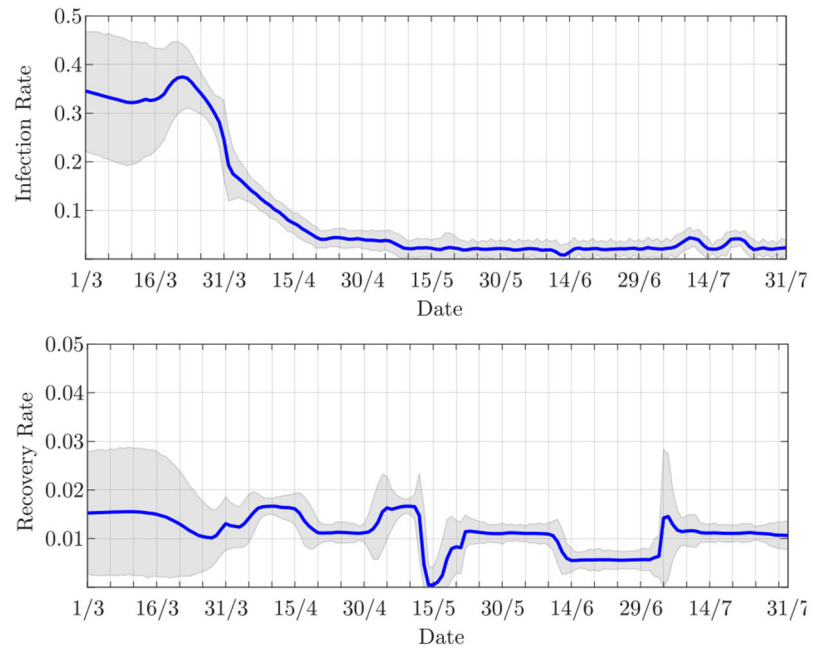


FIGURE 10. Estimated (top) infection rate and (bottom) recovery rate for the USA. The shaded areas represent the 90% confidence interval.

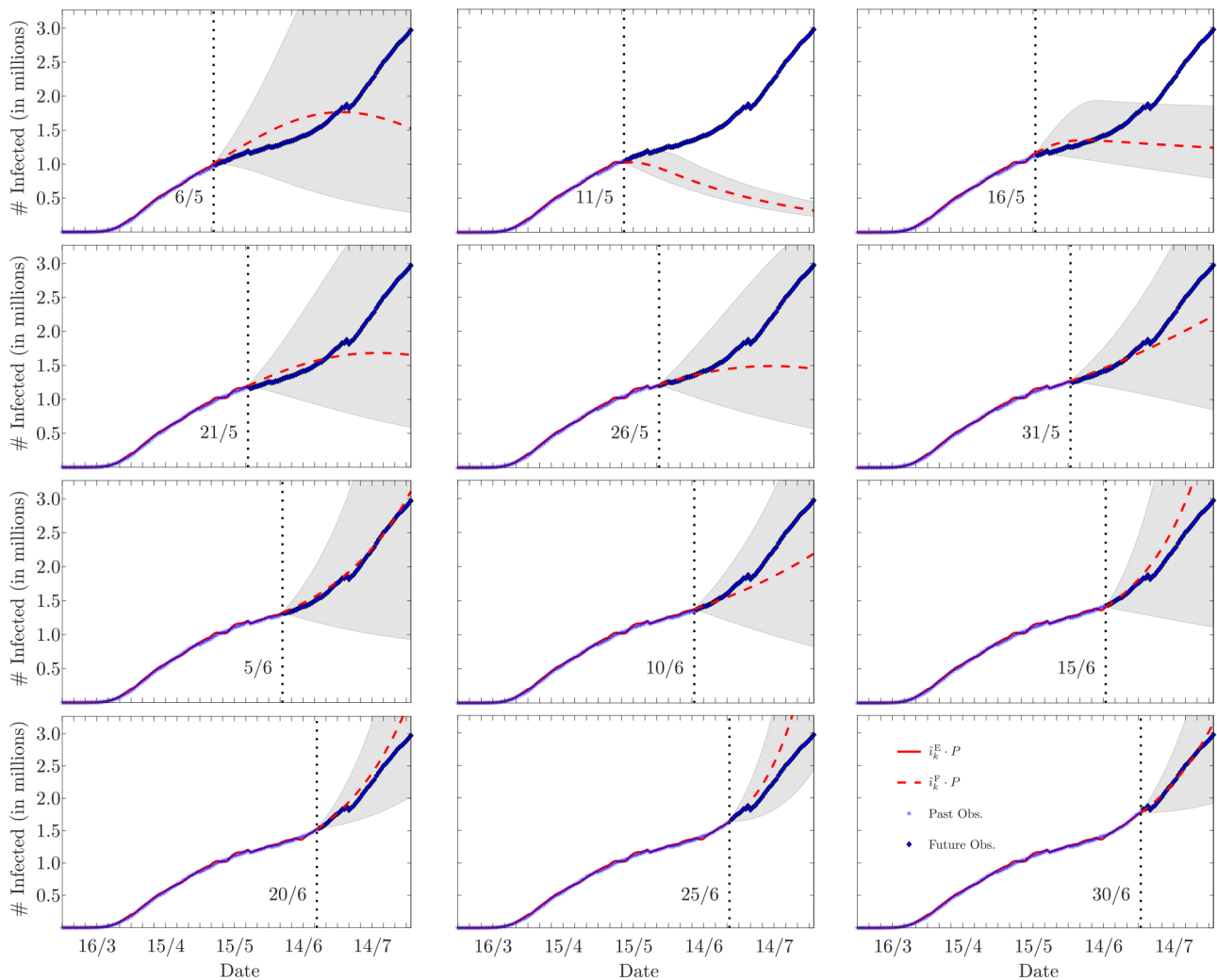


FIGURE 11.

Estimation and forecasting, respectively in solid and dashed lines, of the number of infected individuals in the USA (legend is reported in the bottom-right corner image; the superscript E stands for *estimate*, and the superscript F stands for *forecast*). The date corresponding to the end of the estimation and the beginning of the forecast is marked by a vertical dotted line. In all the cases, the forecast horizon is July 31. The shaded area represents the 90% confidence interval. The poor forecasts made on May 11 and 16, relate to the abrupt decrease of the estimated recovery rate that follows April 30 (cf. Fig. 10, bottom image).

TABLE 1.

Mean absolute percentage errors (MAPEs) of the forecasts of the epidemic evolution in Lombardia, Italy, performed at different dates and calculated for different forecast horizons, that is, 3, 7, and 14 days. The asterisk means that the forecast performed at a given date and with a given forecast horizon includes May 6, when inaccurate numbers of infected and recovered individuals were reported. The average (last row) does not take into account these cases.

FORECAST DATE	3 DAYS (%)	7 DAYS (%)	14 DAYS (%)
April 13	0.54	0.50	2.49
April 18	3.47	6.48	6.84
April 23	1.25	1.91	3.88*
April 28	0.42	1.07	14.91*
May 3	8.02*	18.44*	30.07*
May 8	9.14	7.47	6.31
May 13	2.30	2.59	6.48
May 18	0.92	2.24	3.29
May 23	0.69	3.78	9.02
May 28	4.67	5.10	4.99
June 2	2.70	1.87	1.93
June 7	4.02	6.61	10.76
AVERAGE	2.74	3.60	5.79

TABLE 2.

Average mean absolute percentage errors (MAPEs) of the forecasts of the epidemic evolution in Lombardia, Italy, obtained with the proposed algorithm, and with the SIR-fit and GSEIR-fit curve-fitting approaches, for different forecast horizons, that is, 3, 7, and 14 days. The uppermost table reports the average MAPEs computed over the interval from March 4 to June 16; the lowermost table reports the average MAPEs computed over the interval from April 1 to June 16.

ALGORITHM	3 DAYS (%)	7 DAYS (%)	14 DAYS (%)
Proposed	6.2	10.4	23.5
SIR-fit	77.5	115.3	225.8
GSEIR-fit	10.3	13.0	18.5

Average over the interval from March 4 to June 16

ALGORITHM	3 DAYS (%)	7 DAYS (%)	14 DAYS (%)
Proposed	3.3	4.9	9.4
SIR-fit	88.2	123.0	213.2
GSEIR-fit	11.6	13.6	16.8

Average over the interval from April 1 to June 16

TABLE 3.

Mean absolute percentage errors (MAPEs) of the forecasts of the epidemic evolution in the USA, performed at different dates and calculated for different forecast horizons, that is, 3, 7, and 14 days.

FORECAST DATE	3 DAYS (%)	7 DAYS (%)	14 DAYS (%)
May 6	4.31	5.73	7.44
May 11	4.55	7.25	11.80
May 16	4.84	5.97	6.87
May 21	5.01	5.61	6.71
May 26	0.61	0.98	0.96
May 31	2.58	2.58	2.87
June 5	2.29	2.86	3.49
June 10	1.09	0.90	1.25
June 15	0.26	0.70	1.51
June 20	1.15	1.04	1.52
June 25	0.40	0.49	3.93
June 30	1.11	2.28	1.52
AVERAGE	2.35	3.03	4.16

TABLE 4.

Average mean absolute percentage errors (MAPEs) of the forecasts of the epidemic evolution in the USA, obtained with the proposed algorithm, and with the SIR-fit and GSEIR-fit curve-fitting approaches, for different forecast horizons, that is, 3, 7, and 14 days. The uppermost table reports the average MAPEs computed over the interval from March 10 to July 17; the lowermost table reports the average MAPEs computed over the interval from April 1 to July 17.

ALGORITHM	3 DAYS (%)	7 DAYS (%)	14 DAYS (%)
Proposed	7.3	11.9	29.8
SIR-fit	81.6	122.4	232.4
GSEIR-fit	14.0	20.0	35.2

Average over the interval from March 10 to July 17

ALGORITHM	3 DAYS (%)	7 DAYS (%)	14 DAYS (%)
Proposed	3.4	4.5	7.2
SIR-fit	91.4	138.8	267.6
GSEIR-fit	12.1	17.0	28.7

Average over the interval from April 1 to July 17