Computers in Biology and Medicine 125 (2020) 104011

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: http://www.elsevier.com/locate/compbiomed

Uncertainty quantification in epidemiological models for the COVID-19 pandemic

Leila Taghizadeh [a,*], Ahmad Karimi [a], Clemens Heitzinger [a,b]

[a] Institute of Analysis and Scientific Computing, TU Wien, Wiedner Hauptstraße 8–10, 1040, Vienna, Austria
[b] School of Mathematical and Statistical Sciences, Arizona State University, Tempe, AZ, 85287, USA

ARTICLE INFO

Keywords:
Coronavirus forecasting
Epidemic models
COVID-19 pandemic
Bayesian inversion
MCMC methods

ABSTRACT

Mathematical modeling of epidemiological diseases using differential equations are of great importance in order to recognize the characteristics of the diseases and their outbreak. The procedure of modeling consists of two essential components: the first component is to solve the mathematical model numerically, the so-called forward modeling. The second component is to identify the unknown parameter values in the model, which is known as inverse modeling and leads to identifying the epidemiological model more precisely. The main goal of this paper is to develop the forward and inverse modeling of the coronavirus (COVID-19) pandemic using novel computational methodologies in order to accurately estimate and predict the pandemic. This leads to governmental decisions support in implementing effective protective measures and prevention of new outbreaks. To this end, we use the logistic equation and the SIR (susceptible-infected-removed) system of ordinary differential equations to model the spread of the COVID-19 pandemic. For the inverse modeling, we propose Bayesian inversion techniques, which are robust and reliable approaches, in order to estimate the unknown parameters of the epidemiological models. We deploy an adaptive Markov-chain Monte-Carlo (MCMC) algorithm for the estimation of a posteriori probability distribution and confidence intervals for the unknown model parameters as well as for the reproduction number. We perform our analyses on the publicly available data for Austria to estimate the main epidemiological model parameters and to study the effectiveness of the protective measures by the Austrian government. The estimated parameters and the analysis of fatalities provide useful information for decision-makers and makes it possible to perform more realistic forecasts of future outbreaks. According to our Bayesian analysis for the logistic model, the growth rate and the carrying capacity are estimated respectively as 0.28 and 14 974. Moreover for the parameters of the SIR model, namely the transmission rate and recovery rate, we estimate 0.36 and 0.06, respectively. Additionally, we obtained an average infectious period of 17 days and a transmission period of 3 days for COVID-19 in Austria. We also estimate the reproduction number over time for Austria. This quantity is estimated around 3 on March 26, when the first recovery was reported. Then it decays to 1 at the beginning of April. Furthermore, we present a fatality analysis for COVID-19 in Austria, which is also of importance for governmental protective decision-making. According to our analysis, the case fatality rate (CFR) is estimated as 4% and a prediction of the number of fatalities for the coming 10 days is also presented. Additionally, the ICU bed usage in Austria indicates that around 2% of the active infected individuals are critical cases and require ICU beds. Therefore, if Austrian governmental protective measures would not have taken place and for instance if the number of active infected cases would have been around five times larger, the ICU bed capacity could have been exceeded.

1. Introduction

The coronavirus COVID-19 pandemic is a new infectious disease which emerged from China in fall 2019 and then spread around the world. This pandemic spreads through (micro-) droplets and its outbreak speed is very high.

The first reported case of SARS-CoV-2 was identified in Wuhan, China. The first case outside of China was reported in Thailand on January 13, 2020 [1]. Since then, this ongoing outbreak has now spread all over the world [2]. Till May 21st, this pandemic has infected around

* Corresponding author.
E-mail addresses: leila.taghizadeh@tuwien.ac.at (L. Taghizadeh), ahmad.karimi@tuwien.ac.at (A. Karimi), clemens.heitzinger@tuwien.ac.at (C. Heitzinger).

https://doi.org/10.1016/j.compbiomed.2020.104011
Received 4 July 2020; Received in revised form 11 September 2020; Accepted 17 September 2020
Available online 25 September 2020

5230000 individuals around the world and caused more than 335000 deaths. Out of more than 2780000 active cases around the world, 2% are critical patients. Source of the data is the Johns Hopkins CSSE database (https://github.com/CSSEGISandData/COVID-19).

The COVID-19 pandemic was first officially confirmed to have spread to Austria on February 25, 2020 and till May 21st more than 16400 people have been infected and 633 deaths and 833 active cases have been reported. Various databases have reported different numbers and some of them update their daily reported data even for preceding days. For instance, the Federal Ministry of Social Affairs, Health, Care and Consumer Protection, Republic of Austria (https://info.gesundheitsministerium.at), has updated the number of deaths till May 21 to 657. However, here we have used the Johns Hopkins CSSE database. Fig. 1 displays daily confirmed cases as well as total cumulative count of confirmed, active and fatality cases in Austria. By removing deaths and recoveries from total cases, we obtain the "currently infected cases" or "active cases" (cases still awaiting an outcome).

Infected people need breathing assistance and a large number of them require medical treatment in an intensive care unit (ICU). Countries which are affected by COVID-19 attempt to keep the daily number of cases below the capacity of their health care system. In order to avert the disastrous inundation of hospitals, the virus must be kept from spreading fast. To this end, countries have been implementing protective measures such as closing schools, canceling mass gatherings, working from home (home office), self-quarantine, self-isolation, avoiding crowds, social distancing, wearing protection masks, etc.

In this work, we propose Bayesian inference for the analysis of the COVID-19 data in order to estimate the crucial unknown quantities of the pandemic models. We use an adaptive MCMC method to find the probability distributions and confidence intervals of the epidemiological models parameters using the Austrian infection data. We use this analysis for the prediction of the duration of the epidemic in Austria as well as the total number of infected people and fatalities till the end of the epidemic. The model validation shows a very good agreement between the computational and measurement data of infections in Austria which proves the reliability and the accuracy of the predictions. This is of great importance for making governmental decisions in implementing the measures in order to prevent the spread of the virus.

This paper is organized as follows: Section 2 presents the logistic and SIR (susceptible-infected-removed) epidemiological models and introduces their unknown parameters. Section 3 is devoted to the Bayesian analysis as the inversion method which is proposed for quantifying the uncertain model parameters in the epidemiological models. Numerical results of the forward and inverse epidemic models including the quantification of uncertain parameters of the model validation using the measurement data, pandemic forecast, fatality analysis and the effect of governmental protective measures are presented in Section 4. Finally, conclusions are drawn in Section 5.

2. Mathematical models for COVID-19

Predictive mathematical models are essential for the quantitative understanding of epidemics and for supporting decision-makers in order to implement the most effective and protective measures. Many mathematical models for the spread of infectious diseases [3–6] and in particular for the novel COVID-19 [7–11] have been presented and analyzed. Here we start with the logistic equation as a preliminary model for epidemics and continue with the SIR model and its extensions [12–17].

2.1. The logistic model

The logistic equation is a nonlinear ordinary differential equation, which is used for modeling population growth. This ODE is also well-known as logistic growth model and is given by

$$y'(t) = \alpha y(t)\left(1 - \frac{y(t)}{\beta}\right), \quad y(0) = y_0, \tag{1}$$

where $y_0 \neq 0$ is the initial population size (initial number of confirmed cases), y denotes population size (total accumulated confirmed cases) and t time. Furthermore, α and β are respectively the growth rate (infection rate) and the carrying capacity (maximum number of confirmed cases), which are positive constants.

The solution to the logistic model equation is

$$y(t) = \frac{\beta y_0}{y_0 + (\beta - y_0)e^{-\alpha t}},$$

which can be rewritten as

$$y(t) = \frac{\beta}{1 + Ae^{-\alpha t}},$$

where

$$A = \frac{\beta - y_0}{y_0}.$$

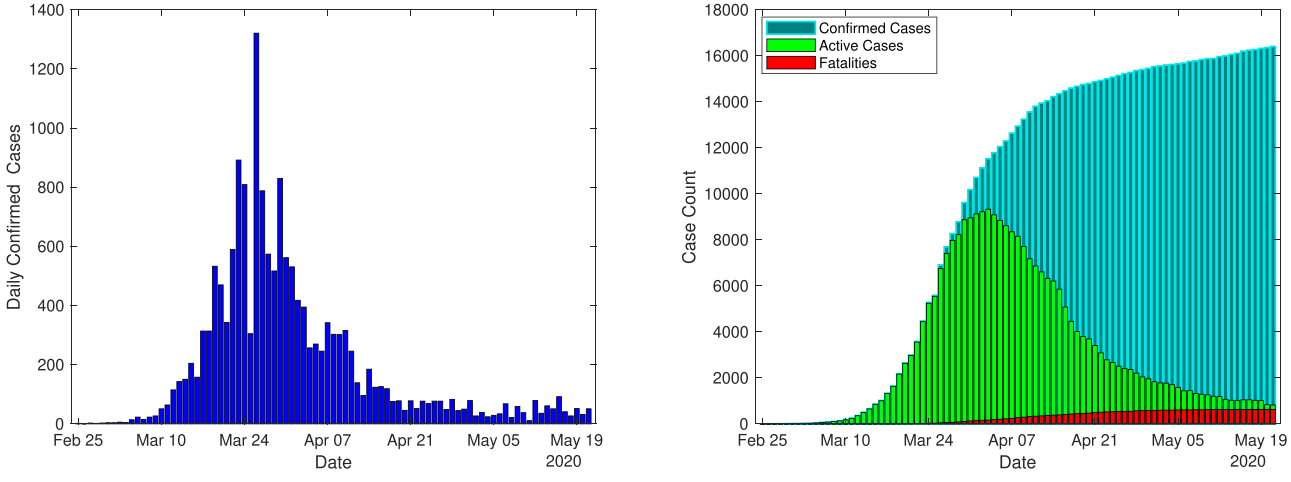

Fig. 1. (Left) Daily confirmed count of coronavirus infected cases and (right) total cumulative count of confirmed infected, active and fatality cases (till May 21st) in Austria.

The inflection point represents the time that maximal rate of confirmed cases (growth rate) occurs. The inflection point of the logistic function is calculated as

$$t^* = \frac{\ln(A)}{\alpha},$$

where the estimated number of infected people is $\beta/2$. However, there are generalizations of the naïve logistic function, and we will present numerical results based on them in Section 4.2.

2.2. The SIR model

The susceptible-infected-removed (SIR) model is an epidemiological model that computes the number of people infected with a contagious disease in a closed population over time. The Kermack-McKendrick model [18–20] is one of the SIR models, which is defined by the system

$$\frac{dS}{dt} = -\frac{\beta IS}{N}, \tag{2a}$$

$$\frac{dI}{dt} = \frac{\beta IS}{N} - \gamma I, \tag{2b}$$

$$\frac{dR}{dt} = \gamma I \tag{2c}$$

of ordinary differential equations, where β and γ are the infection and recovery rates, respectively. The model consists of three components: S for the number of susceptible, I for the number of infectious, and R for the number of recovered or deceased (or immune) individuals. Furthermore, N denotes the constancy of population, i.e.

$$S(t) + I(t) + R(t) = N. \tag{3}$$

Moreover, the dynamics of the infectious class depends on the basic reproduction number, which is defined as

$$R_0 := \frac{\beta}{\gamma}.$$

If the reproduction number is high, the probability of a pandemic is high, too. This number is also used to estimate the herd immune threshold (HIT). If the reproduction number multiplied by the percentage of susceptible individuals is equal to 1, it shows an equilibrium state and thus the number of infectious people is constant.

Additionally, the recovery period is defined by

$$t_1 = \frac{1}{\gamma}$$

and describes the average days to recover from infection. The transmission period in the sense of the average days to transmit the infection to a person is defined by

$$t_2 = \frac{1}{\beta}.$$

However in a population with vital dynamics, new births can provide more susceptible individuals to the population, which sustain an epidemic or allow new introductions to spread in the population. Taking the vital dynamics into account, the SIR model is extended to

$$\frac{dS}{dt} = \mu N - \frac{\beta IS}{N} - \nu S, \tag{4a}$$

$$\frac{dI}{dt} = \frac{\beta IS}{N} - \gamma I - \nu I, \tag{4b}$$

$$\frac{dR}{dt} = \gamma I - \nu R, \tag{4c}$$

where μ and ν denote the birth and death rates, respectively. To maintain a constant population, we assume $\mu = \nu$ is the natural mortality rate.

3. Bayesian inversion for the model parameters

We propose Bayesian inversion methods, in which probabilities are used as a general concept to represent the uncertainty in the model parameters in order to solve the backward/inverse problem of COVID-19, i.e., the problem of accurate estimation of the epidemiological model parameters as well as the reproduction ratio. Bayesian inference in the context of the statistical inversion theory is based on Bayes' Theorem and represents the uncertainty probabilistically by defining a probability distribution over the possible values of the parameters and uses sample data to update this distribution. Bayesian analysis, in contrast to traditional inverse methods, is a robust inversion technique for determining parameters, yields the (a posteriori) probability distribution, and has the advantage of updating the prior knowledge about the unknown quantity using the measurement/observation data, giving confidence intervals for the unknowns instead of providing a single estimate. We have already successfully applied Bayesian inversion techniques to various PDE models in engineering and medicine in order to identify parameters (see for instance Ref. [21–24]).

3.1. Bayesian analysis

As mentioned, the Bayesian inversion approach is a robust and reliable technique to quantify the uncertain parameters of the epidemic models. In fact, the solution of the inverse problem is the posterior density that best reflects the distribution of the parameter based on the observations. As the observations or measurements are subject to noise, and the observational noise, i.e., the error e due to modeling and measurement, is unbiased and i.i.d. (independent and identically distributed), it can be represented by random variables as

$$Y = f(Q) + e, \tag{5}$$

where e is a mean-zero random variable and Y is a given random variable representing observed data or measurements, for which we have a model $f(Q)$ (observation operator) dependent on a random variable Q with realizations $q = Q(\omega)$ representing parameters to be estimated [25].

Assume a given probability space (Ω, F, P), where Ω is the set of elementary events (sample space), F a σ-algebra of events, and P a probability measure. Furthermore assume that all the random variables are absolutely continuous.

Bayes' Theorem in terms of probability densities can be written as

$$\pi(q|y) = \frac{\pi_0(q)\pi(y|q)}{\pi(y)} \tag{6}$$

with

$$\pi(y) := \int_{\mathbb{R}^p} \pi_0(q)\pi(y|q)\mathrm{d}q \neq 0, \tag{7}$$

where the unknown parameters $q = (q_1, \ldots, q_p) \in \mathbb{R}^p$ and the observed data y are realizations of the random variables Q and Y, respectively. Furthermore, $\pi_0(q)$, $\pi(q|y)$, and $\pi(y|q)$ are the probability density functions of the prior, posterior, and (data) sampling distributions, respectively. A probability density function is density of a continuous random variable, which is used to specify the probability of the random variable falling within a particular range of values. The density $\pi(y|q)$ of the data provides information from the measurement data to update the prior knowledge, and it is well-known as the likelihood density function. The goal of Bayesian inversion is to estimate the posterior probability density function $\pi(q|y)$, which reflects the uncertainty about the quantity of interest q using measurement data y.

Equation (6) gives the posterior density and summarizes our beliefs

Table 1
Estimated confidence intervals and mean of Markov chains for the parameters of the logistic model using Bayesian inversion method for Austria.

Parameter	Description	Estimated mean	Confidence interval 95%
α	Growth rate	0.28	[0.23, 0.33]
β	Carrying capacity	14974	[12703, 17244]

Table 2
Estimated and actual number of the infected cases in Austria at the inflection point of the logistic model.

Inflection point t^*	Infected cases at inflection point	
Estimated	Estimated	Actual
March 27	7486	7697

Table 3
Estimated confidence intervals and means of Markov chains for the parameters of the SIR model using Bayesian inversion method for Austria.

Parameter	Description	Estimated mean	Confidence interval 95%
β	Transmission rate	0.36	[0.32, 0.39]
γ	Recovery rate	0.06	[0.03, 0.09]

about q after we have observed y. Therefore, Bayes' Theorem for inverse problems can be stated as follows.

Theorem 1 (Bayes' Theorem for inverse problems [25,26]). Let $\pi_0(q)$ be the prior probability density function of the realizations q of the random parameter Q. Let y be a realization or measurement of the random observation variable Y. Then the posterior density of Q given the measurements y is

$$\pi(q|y) = \frac{\pi_0(q)\pi(y|q)}{\pi(y)} = \frac{\pi_0(q)\pi(y|q)}{\int_{\mathbb{R}^p}\pi_0(q)\pi(y|q)\mathrm{d}q}. \tag{8}$$

Computing the integral appearing in Bayes' Theorem 1 is costly especially if the parameter space $\mathbb{R}^p$ is high-dimensional. Another problem with quadrature rules is that they require a relatively good knowledge of the support of the probability distribution, which is usually part of the information that we seek [25,26]. In Section 3.2 we shortly discuss the algorithms for Bayesian estimation, which do not require evaluations of the integral and that are used to achieve the numerical results for the nonlinear model equation.

3.2. Markov-chain Monte-Carlo (MCMC) methods

Markov-chain Monte-Carlo methods are a class of Monte-Carlo methods with the general idea of constructing Markov chains whose stationary distribution is the posterior density [25]. The Metropolis-Hastings algorithm is an MCMC algorithm to draw samples from a desired distribution by building a Markov-chain of accepted values (out of proposed values) for the unknown parameter as a posteriori distribution. In this algorithm, the first state of the chain q_0 is given and the new state q_k, $k = 1, 2, \ldots, N$, of the chain is constructed based on the previous state q_{k-1}. To this end, a new value q^* is proposed using the proposal density function $J(q^*|q_{k-1}) = N(q_{k-1}, \sigma_P^2)$, where σ_P is the proposal covariance. Admissibility of this proposed value is tested by means of calculating the acceptance ratio $\alpha(q^*|q_{k-1})$, which is defined by

$$\alpha(q^*|q_{k-1}) = \min\left(1, \frac{\pi(q^*|y)}{\pi(q_{k-1}|y)}\cdot\frac{J(q_{k-1}|q^*)}{J(q^*|q_{k-1})}\right), \tag{9}$$

where $\pi(q|y)$ and J are respectively posterior and the proposal distributions. Applying Bayes' Theorem of inverse problems, we calculate α as

$$\alpha(q^*|q_{k-1}) = \min\left(1, \frac{\pi(y|q^*)\pi_0(q^*)}{\pi(y|q_{k-1})\pi_0(q_{k-1})}\cdot\frac{J(q_{k-1}|q^*)}{J(q^*|q_{k-1})}\right), \tag{10}$$

where $J(q_{k-1}|q^*) = J(q^*|q_{k-1})$ for symmetric proposal functions and $\pi_0(q)$ is a given prior distribution. Furthermore, $\pi(y|q)$ is the likelihood distribution which is defined by

$$\pi(y|q) = N(y, \sigma_L^2) = \frac{1}{(2\pi\sigma_L^2)^{n/2}} e^{-S_q/2\sigma_L^2}, \tag{11}$$

where σ_L is the likelihood covariance, $S_q := \sum_{i=1}^{n}(y_i - f_i(q))^2$ is the sum of squares error and $f(q)$ denotes the parameter-dependent model response. If the proposed value is admissible, it is accepted as q_k, otherwise the old value is kept. The mechanism of acceptance and the evolution of the chain are clearly described in Algorithm 1, which is an adaptive MCMC algorithm and will be explained in the next subsection. More details can be found for example in Ref. [27–30] and we refer the reader especially to [25, Chapter 8].

Although the convergence speed is determined by the choice of a good proposal distribution, at least tens or hundreds of thousands of samples are necessary to converge to the target distribution. Choosing

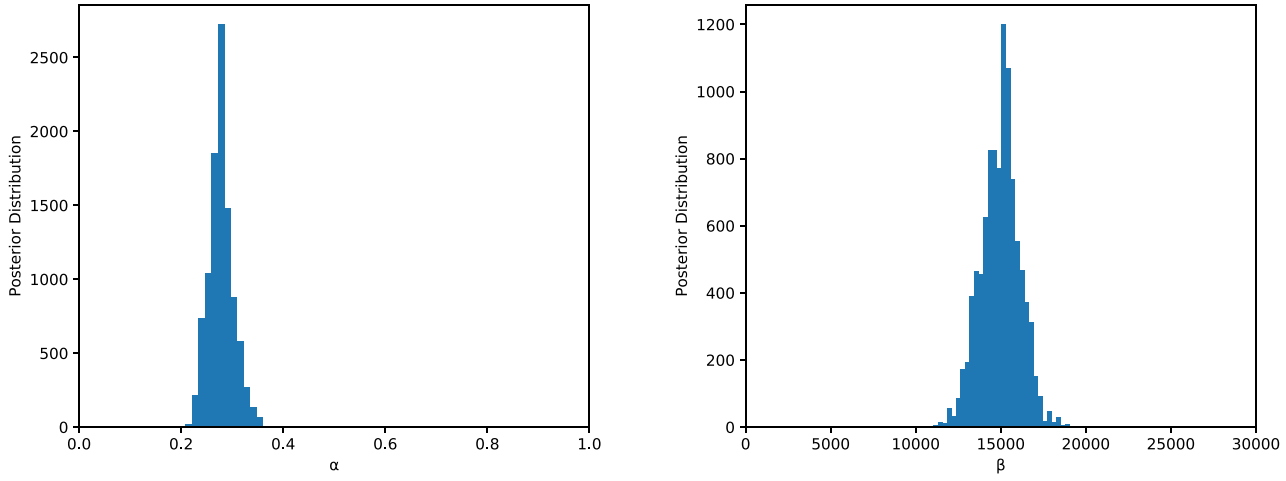

Fig. 2. The marginal histograms of posterior distribution for the two quantities of interest in the logistic model, namely α and β.

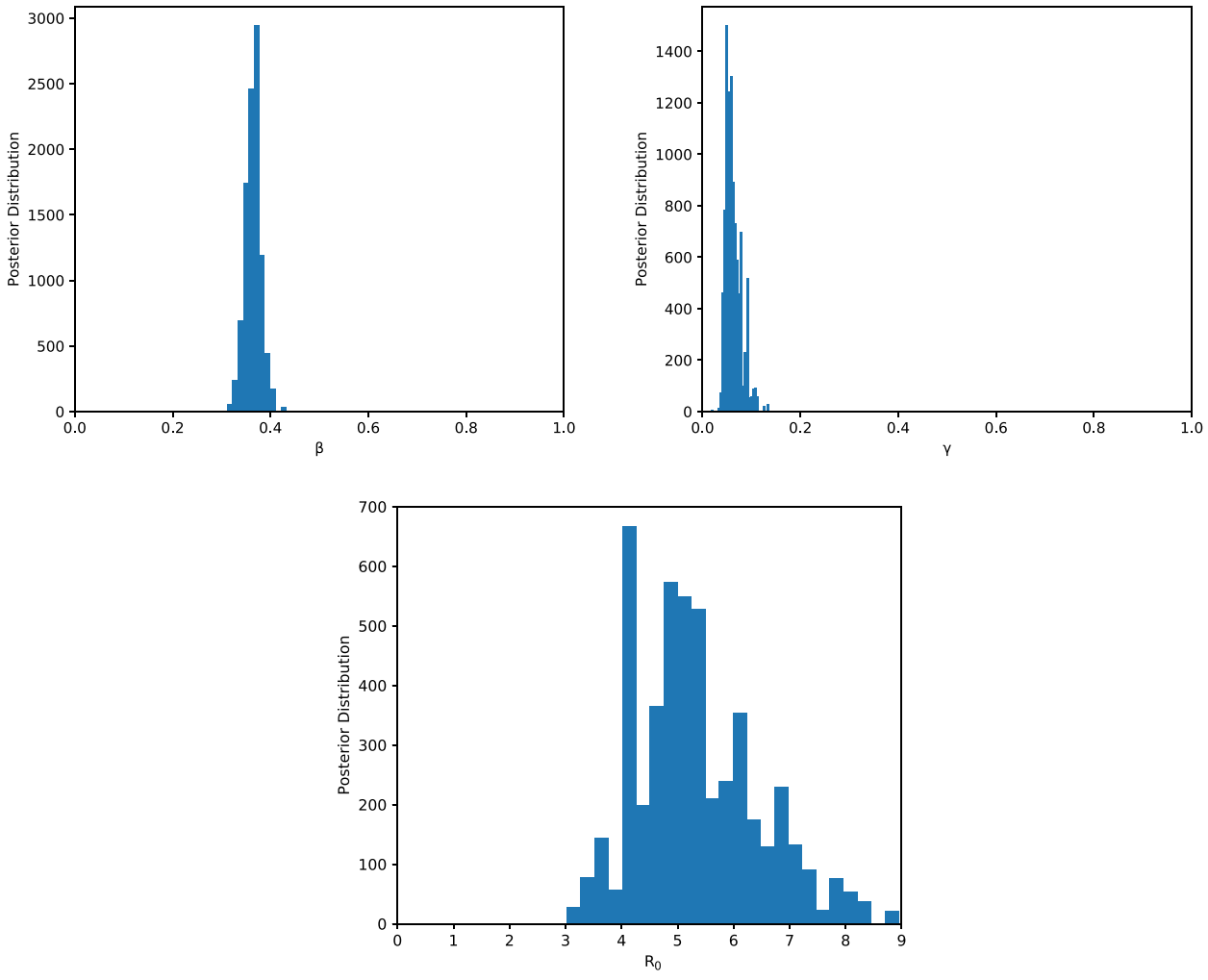

Fig. 3. The marginal histograms of posterior distribution for the three quantities of interest in the SIR model, namely β, γ and R_0.

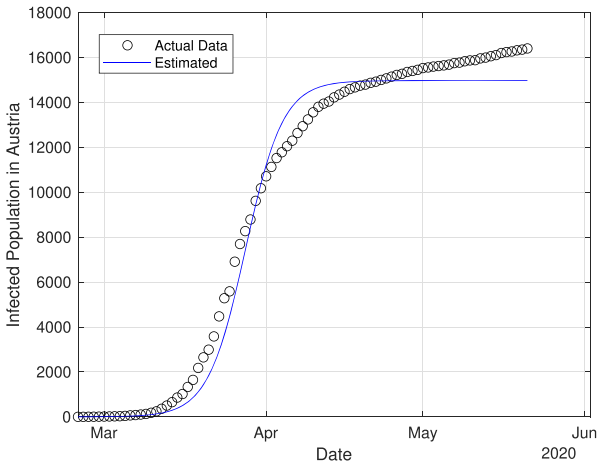

Fig. 4. Estimated total cumulative count of coronavirus confirmed infected cases in Austria using Bayesian analysis for the logistic model versus actual or measured infected population.

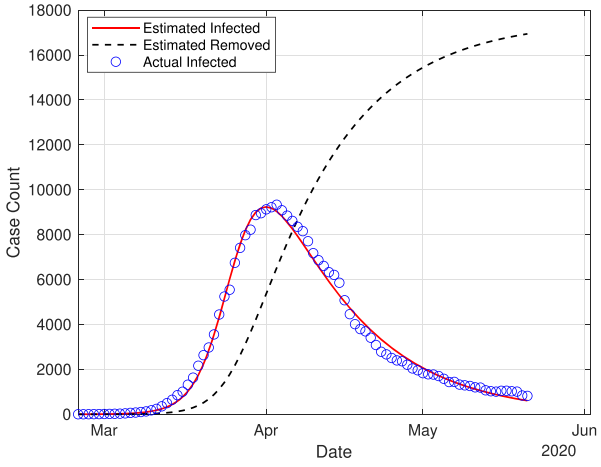

Fig. 5. Estimated total count of coronavirus infected and recovered cases using Bayesian analysis for the SIR model as well as actual confirmed active cases in Austria.

the optimal proposal scaling is a crucial issue and affects the MCMC results; if the covariance of the proposal distribution is too small, the generated Markov chain moves too slowly, and if it is too large, the proposals are rejected. Hence, optimal proposal values should be found to avoid both extremes, which leads to adaptive MCMC methods [31–33]. In the following section, we will consider an adaptive algorithm that helps sample from potentially complicated distributions.

3.3. Delayed-rejection adaptive-Metropolis (DRAM) algorithm

Searching for a good proposal value can be done manually through trial and error, but this becomes intractable in high dimensions. Therefore, adaptive algorithms that find optimal proposal scales automatically are advantageous. The delayed-rejection adaptive-Metropolis (DRAM) algorithm is an efficient adaptive MCMC algorithm [32]. It is based on the combination of two powerful ideas to modify the Markov-chain Monte-Carlo method, namely adaptive Metropolis (AM) [34,35] and delayed-rejection (DR) [36,37], which are used as global and local adaptive algorithms, respectively. AM finds an optimal proposal scale and updates the proposal covariance matrix, while DR updates the proposal value when q^* is rejected.

The basic idea of the DR algorithm is that, if the proposal q^* is rejected, delayed rejection (DR) provides an alternative candidate q^{**} as a second-stage move rather than just retaining the previous value q_{k-1}. This process is called delayed rejection, which can be done for one or many stages. Furthermore, the acceptance probability of the new candidate(s) is calculated. Therefore, in the DR process, the previous state of the chain is updated using the optimal parameter scale or proposal covariance matrix that has been calculated via the AM algorithm.

The AM algorithm is a global adaptive strategy, where a recursive relation is used to update the proposal covariance matrix. In this algorithm, we take the Gaussian proposal centered at the current state of the chain q_k and update the chain covariance matrix at the k-th step using

$$V_k = s_p \mathrm{Cov}(q_0, q_1, \ldots, q_{k-1}) + \varepsilon I_p, \tag{12}$$

where s_p is a design parameter and depends only on the dimension p of the parameter space. This parameter is specified as $s_p := 2.38^2/p$ as the common choice for Gaussian targets and proposals [38], as it optimizes the mixing properties of the Metropolis-Hastings search in the case of Gaussians. Furthermore, I_p denotes the p-dimensional identity matrix, and $\varepsilon \geq 0$ is a very small constant to ensure that V_k is not singular, and in most cases it can be set to zero [32].

The adaptive Metropolis algorithm employs the recursive relation

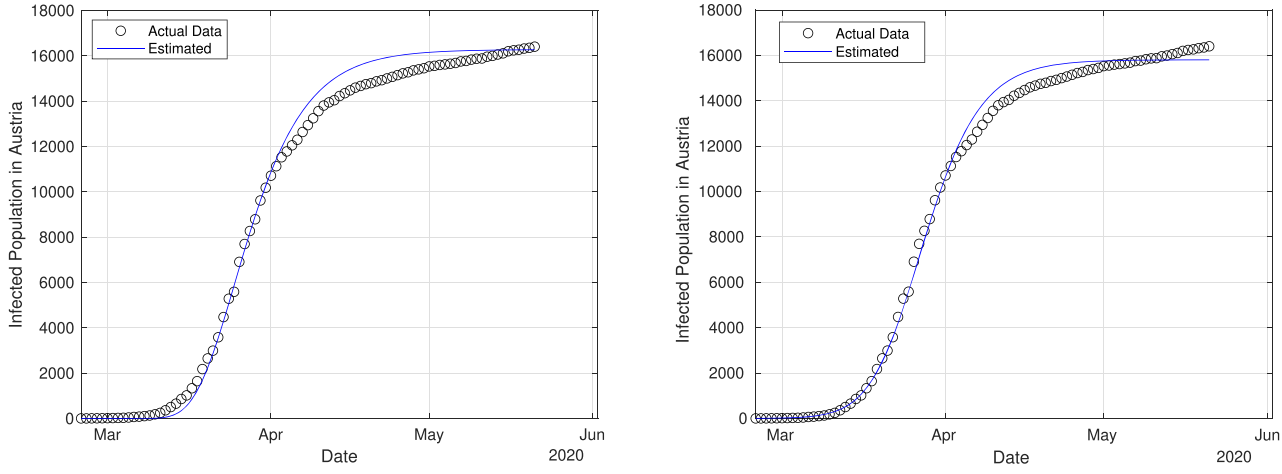

Fig. 6. Estimated total cumulative count of infected cases using the Gompertz function (left) and a generalized logistic function (right) versus the observed infected population in Austria.

$$V_{k+1} := \frac{k-1}{k}V_k + \frac{s_p}{k}\left(k\overline{q}_{k-1}\overline{q}_{k-1}^{\top} - (k+1)\overline{q}_k\overline{q}_k^{\top} + q_k q_k^{\top}\right)$$

to update the proposal covariance matrix, where the sample mean $\overline{q}_k$ is calculated recursively by

$$\overline{q}_k = q_k + \frac{k}{k+1}\left(\overline{q}_{k-1} - q_k\right).$$

A second-stage candidate q^{**} is chosen using the proposal function

$$J_2(q^{**}|q_{k-1}, q^*) := N\left(q_{k-1}, \gamma_2^2 V_k\right), \tag{13}$$

where V_k is the covariance matrix produced by the adaptive algorithm (AM) as the covariance of the first-stage and $\gamma_2 < 1$ is a constant. The probability of accepting the second-stage candidate, having started at q_{k-1} and rejected q^*, is

$$\alpha_2(q^{**}|q_{k-1}, q^*) := \min\left(1, \frac{\pi(q^{**}|y)J(q^*|q^{**})(1-\alpha(q^*|q^{**}))}{\pi(q_{k-1}|y)J(q^*|q_{k-1})(1-\alpha(q^*|q_{k-1}))}\right), \tag{14}$$

where α is the acceptance probability (15) in the non-adaptive approach. The acceptance probability is computed so that reversibility of the posterior Markov chain is preserved (for more details see for example [25, §8.6]). The DRAM technique is summarized in Algorithm 1.

Algorithm 1. The DRAM algorithm

Initialization:
Choose the first state of the chain q_0 such that $\pi_0(q_0) > 0$.
Choose the number N_{samples} of samples or iterations.
Choose the parameter ε.
Choose the initial proposal covariance matrix V_0 (diagonal or symmetric).
Choose the factor γ (often $\gamma := 1/5$) for the second-stage proposal distribution.
for $k = 1 : N_{\text{samples}}$ **do**

1. (Adaptivity:) The covariance matrix V_k in the k-th step is updated by (12).
2. A first-stage proposal q^* is generated from $J(q^*|q_{k-1}) := N(q_{k-1}, V_k)$.
3. The new value q^* is accepted with probability

$$\alpha(q^*|q_{k-1}) = \min\left(1, \frac{\pi(q^*|y)}{\pi(q_{k-1}|y)} \cdot \frac{J(q_{k-1}|q^*)}{J(q^*|q_{k-1})}\right).$$

4. If the new state is accepted, we set $q_k = q^*$. Otherwise:
   (a) (Delayed rejection:) A second-stage proposal q^{**} is generated from proposal density

$$J_2(q^{**}|q_{k-1}, q^*) := N(q_{k-1}, \gamma_2^2 V_k),$$

   where V_k is the adapted covariance matrix.
   (b) The new value q^{**} is accepted with probability

$$\alpha_2(q^{**}|q_{k-1}, q^*) := \min\left(1, \frac{\pi(q^{**}|y)J(q^*|q^{**})(1-\alpha(q^*|q^{**}))}{\pi(q_{k-1}|y)J(q^*|q_{k-1})(1-\alpha(q^*|q_{k-1}))}\right).$$

   (c) If the new state is accepted, we set $q_k := q^{**}$, otherwise $q_k := q_{k-1}$.

end for

4. Numerical results

In this section, we present simulation results of Bayesian inversion and the adaptive MCMC method (see Algorithm 1) for the two epidemic models, namely the logistic and the SIR models, using the data of the COVID-19 outbreak in Austria. The results include model parameter estimation, model validation and outbreak forecasting.

4.1. Parameter estimation

According to Bayesian analysis, the unknown parameters of the logistic and SIR models using the data of COVID-19 outbreak in Austria were found and summarized in Table 1 and Table 3, respectively. These tables show the confidence intervals for the models parameters as well as the mean of the obtained Markov chains in the Bayesian inference.

Furthermore, Tables 2 and 4 include temporal quantities such as inflection time of the outbreak estimated using the Bayesian inference for the logistic and SIR models. According to our analysis, March 27 is estimated as the inflection point of the outbreak in Austria, when the maximal rate of confirmed cases (growth rate) occurs. The estimated total number of confirmed cases till the inflection point agrees with measured data (see Table 2).

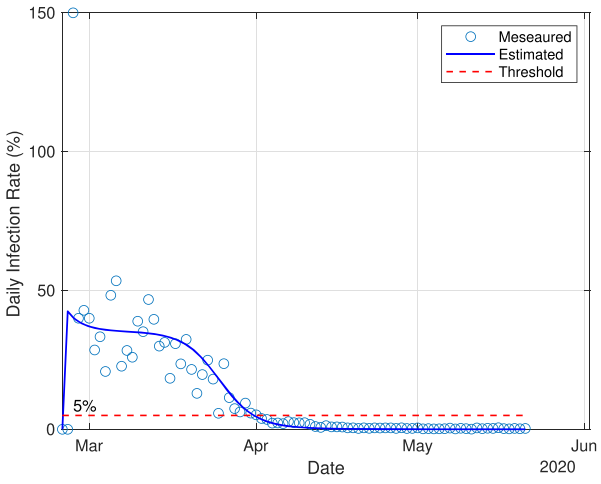

Fig. 7. Infection rate in Austria.

Fig. 2 and Fig. 3 illustrate marginal histograms of posteriori distribution for the quantities of interest respectively in the logistic and SIR models using Bayesian analysis.

4.2. Model validation

Here, we aim to validate the logistic and SIR models for forecasting the COVID-19 outbreak by comparing the Bayesian simulation results and the actual data. Fig. 4 illustrates the actual number of infected individuals in Austria till now, as well as the estimated number of infected people according to the Bayesian inversion for the logistic equation.

Fig. 5 displays a similar estimation using the Bayesian inference for the SIR model, which shows a very good agreement between the measurements and the simulation.

Fig. 6 illustrates the number of infected people estimated using the Gompertz function and a generalized logistic function. The Gompertz function is a sigmoid function which describes population growth and it is defined by

$$G(t) : = ae^{-be^{-ct}}, \tag{18}$$

where a is an asymptote, and b and c are the negative growth rates. The

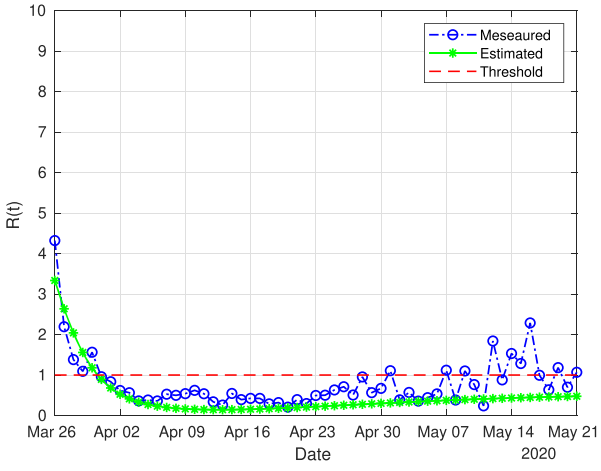

Fig. 8. Reproduction number in Austria.

Table 4
Estimated temporal quantities for Austria using the SIR model.

Quantity	Description	Average Estimation (days)
t_1	Infectious period	16.7
t_2	Transmission period	2.8

second derivative of $G(t)$ gives the point of maximum rate $t = \ln(b)/c$ in which the growth of daily cases will start to decrease. In contrast to the logistic function, the Gompertz function does not have a symmetrical first derivative at the inflection point.

As another more complex generalization of logistic method, we apply the model

$$P(t) : = \frac{M}{(1 + e^{-at+b})^{\alpha}}, \tag{19}$$

where a is growth rate and M is the carrying capacity. Furthermore, b is multiplication of the growth rate and unknown inflection time. Moreover, $\alpha = 1/\nu$, where $\nu > 0$ affects near which asymptote maximum growth occurs.

The comparison shows a better agreement between the measured data and both of the Gompertz estimation and the generalized logistic function than the naive logistic function (see Figs. 4 and 6).

Quantifying the uncertain parameters such as the reproduction number leads to calculate the average number of days to recover from the infection and gives useful information about properly and accurately implementing protective measures in order to prevent the spread of the virus. Furthermore, the parameter identification in the epidemic model makes it possible to predict the length of the pandemic, the number of infected individuals and the fatality rate.

In Fig. 7, the actual and estimated infection rates are depicted.

This rate is defined by

$$\text{infection rate} : = \frac{\Delta I_n}{I_{n-1}}, \quad n \in \{2, 3, \ldots\}, \tag{20}$$

where $\Delta I_n = I_n - I_{n-1}$, I_n and I_{n-1} are infected population of consequent times (e.g. in days or weeks), which are obtained using the estimated infection from the SIR model. Fig. 8 shows the estimated and actual reproduction number $R(t)$ at day t for Austria, i.e. the average number of people someone infected at time t would infect over their entire infectious lifespan. In order to calculate $R(t)$, we apply the formula

$$R(t) = \frac{I_{\text{inc}}(t)}{\sum_{\tau=1}^{t_1^*} \omega(t) I_{\text{inc}}(t-\tau)}, \tag{21}$$

where $I_{\text{inc}}(t)$ is the number of incident cases at time t and $t_1^* = 17$ is the average estimated infectious period in days (see Table 4). The normal distributed function ω specifies the so-called infectivity profile during the infectious period [39,40]. The first recovery in Austria was reported on March 26, where the reproduction number is estimated around 3. This estimated quantity decays to 1 in the beginning of April, and since then remains below 1 (till the time of writing this paper on May 21, 2020). The measured $R(t)$ is below 1 in April, then oscillates around 1 for the rest of time frame in this study. In this figure, the threshold of $R = 1$ is also displayed in the sense that there is no immediate public health emergency any more when the reproduction number is below this threshold.

4.3. Fatality analysis

Social-distancing and other protective measures started in Austria around March 16, 2020 in order to slow down the outbreak and consequently to prevent an increase in fatalities by keeping the cases that require hospitalization below the capacity of the healthcare system. In Fig. 9, the daily fatalities in Austria as well as the relative change in

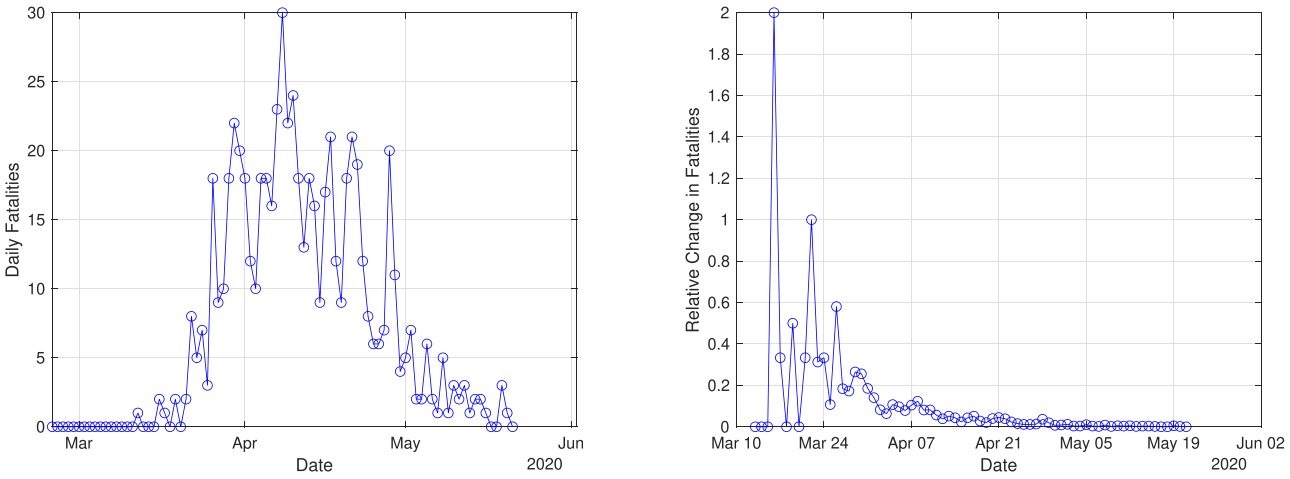

Fig. 9. Fatalities in Austria including the daily fatalities (left) and the relative change in fatalities (right).

fatalities are illustrated.

Applying the fatality ratio as well as confirmed infected cases, we present a fatality analysis which is of importance for governmental protective decision-making. In epidemiology, a case fatality rate (CFR) is the proportion of deaths from a certain disease compared to the total number of people diagnosed with the disease for a certain period of time. Fig. 10 depicts the case fatality rate (CFR), which is defined by

$$\text{CFR} := \frac{\text{fatalities}}{\text{fatalities} + \text{recoveries}}, \tag{22}$$

and we call it the true CFR here, in contrast to the naive CFR defined by

$$\text{naive CFR} := \frac{\text{fatalities}}{\text{infections}}. \tag{23}$$

The straightforward calculations using the recorded data in Austria show that both CFR and naive CFR converge to the same value of $\text{CFR}^* = 0.04$ (Fig. 10).

We can roughly predict the fatalities using the confirmed infections, a shift, and the CFR. The shift is approximately equal to the time between infection and death/recovery (currently average estimated to be 17 days (see Table 4)) minus the incubation time (currently estimated to be 5–6 days [41]) minus 1 day for testing and reporting (see Fig. 11). The average time between infection and death is reported approximately 17.8 days in Ref. [42]. The estimator of fatalities in Austria is defined by

$$\text{fatality cases} := \text{CFR}^* \times \text{confirmed infected cases (shifted by 10 days)}.$$

Fig. 11 shows a good agreement between the estimated fatalities and true values in Austria.

4.4. The impact of governmental protective measures

Here, in order to study the effect of the protective measures implemented by the Austrian government, we compare the infection rate and the infected population in different time intervals with and without implementing the measures. Although public health measures were in place from March 16 to control the spread of COVID-19, Austrians started to practice social distancing in advance. Table 5 shows the weekly infection rate in Austria and how it decays in subsequent weeks. The comparison between the estimated infection rates in subsequent weeks before and after implementing the protective measures highlights the importance and effectiveness of the measures such as social-distancing and lock-down in controlling and slowing down the spread of COVID-19.

The main goal of protective measures and lock-down is to "flatten the curve", i.e., to decrease the infection rate so that the healthcare system is

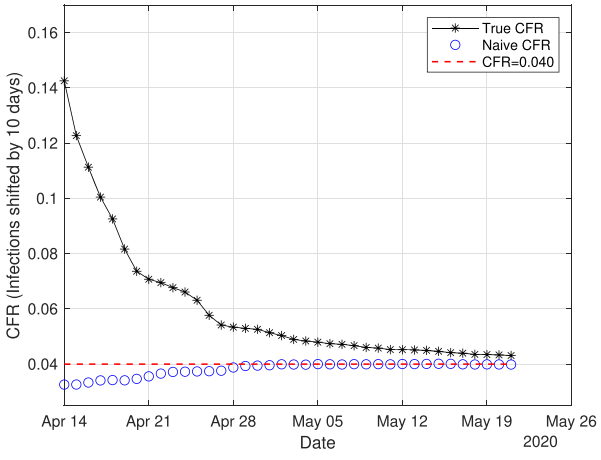

Fig. 10. Case fatality rate (CFR) in Austria.

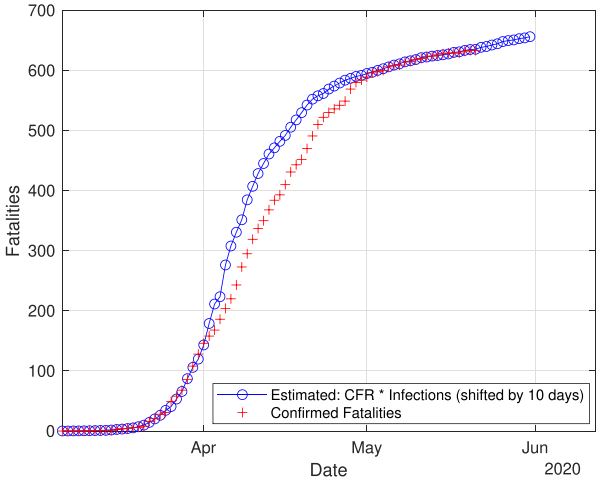

Fig. 11. Prediction of number of fatalities in Austria.

Table 5
Infection rate (see Equation (20)) and total cumulative infected population at the end of different time intervals with and without implementing the protective measures in Austria.

	without measures	with measures			
Time interval	March before 18th	March 18–24	March 25–31	April 1–7	April after 7th
Infection rate	7.30	3.97	1.93	1.2	≈ 1
Infected population	1332	5283	10180	12639	17000

kept from becoming overwhelmed with too many critical cases at the same time. In countries where the counterfactual scenario i.e., no public health interventions is applied, for instance in Sweden, the ICU demand is estimated to be almost 20 times higher than the intensive care capacity in the country and a much larger number of deaths is predicted [43]. However according to Institute for Health Metrics and Evaluation (IHME) database (http://www.healthdata.org/), France, Italy, and Spain, where lock-downs were enacted, also experienced problems due to the limited numbers of ICU beds. In early April, these countries experienced an overload in their health care systems.

The fatality forecast in Section 4.3 is valid as long as protective measures are in place, otherwise the number of fatalities will increase due to a large number of infected people and the limit in the capacity of intensive care unit (ICU) beds as the number of intensive cases increases dramatically. According to the website of the Federal Ministry of Social Affairs, Health, Care and Consumer Protection of Austria, around 700 ICU beds have been available on average for COVID-19 patients since the beginning of the pandemic. Fig. 12 illustrates the average number of available ICU beds for COVID-19 patients and the reported number of occupied ICU beds in Austria. In early April around 26% and in mid-May around 5% of all available intensive care beds for COVID-19 patients were occupied by these patients. This amount of ICU beds usage shows that approximately 2% of the active infected individuals were critical cases and required ICU beds. If Austrian governmental protective measures would not have taken place and the reproduction number would have remained as before the measures, active infected cases would have increased dramatically. In the case the number of active infected cases would have been around five times larger, the ICU bed capacity could have been exceeded according to the rate of around 2% critical cases in Austria.

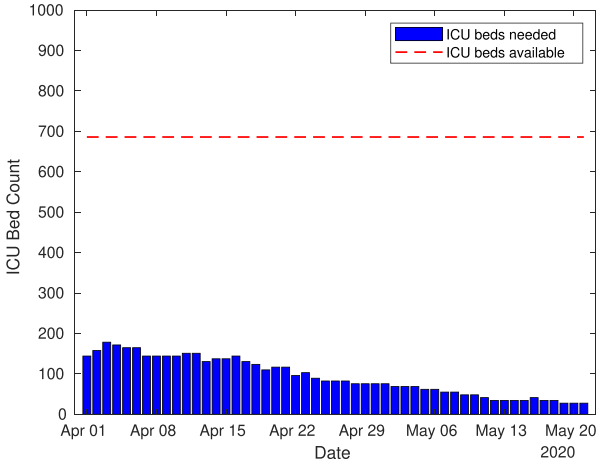

Fig. 12. The average number of available ICU beds for COVID-19 patients vs. the reported number of occupied ICU beds in Austria.

5. Conclusions

In this work, we developed an adaptive Bayesian inversion for epidemiological models, namely the logistic and the SIR models, in order to solve the inverse problem of estimating unknown quantities for the novel coronavirus COVID19. Quantifying the uncertainties in these models is essential since it leads to describe the characteristics of the epidemics on one hand and accurately forecasting the pandemic on the other hand. The proposed inversion recipe is robust and yields probability distributions and confidence intervals for the unknown parameters of the epidemic models including the growth rate of the outbreak and transmission and recovery rates as well as the reproduction number, whose quantification is crucial for decision-makers.

We applied our methodology to the publicly available data for Austria to estimate the main epidemiological model parameters and to present a fatality analysis, all of which are of great importance for the government and decision-makers to adopt the most efficient and effective protective measures in order to prevent human and economic damages. We also validated the presented models by comparing the simulated and measured data, whose results show a very good agreement.

Based on Bayesian analysis for the logistic model, the means of the growth rate α and the carrying capacity β are estimated respectively as 0.28 and 14974. Furthermore for these parameters, 95% confidence intervals of $[0.23, 0.33]$ and 12703, 17244 are obtained. Moreover for the parameters of the SIR model, namely the transmission rate β and recovery rate γ the means of 0.36 and 0.06 as well as the 95% confidence intervals of $[0.32, 0.39]$ and $[0.03, 0.09]$ are inferred. Additionally, we obtained the infectious period of 17 days and transmission period of 3 days for COVID-19 in Austria. The first recovery in Austria was reported on March 26, where the reproduction number is estimated around 3. This estimated quantity decays to 1 in the beginning of April, and since then remains below 1 (till the time of writing this paper on May 21, 2020). The measured $R(t)$ is below 1 in April, then oscillates around 1 for the rest of time frame in this study.

Analyzing data of infected, recovered and death cases, we obtained that the case fatality rate (CFR) has converged to the value 4%. This estimation makes it possible to forecast the fatalities in the coming 10 days. According to our analysis, the total number of death in Austria is estimated as 633 in May 21, which perfectly matches the measured data, according to the Johns Hopkins CSSE database.

Furthermore, we estimated the infection rate for consequent weeks starting from before implementing the protective measures, which shows a significant decay after the measures are in place. Moreover, the ICU bed usage shows that approximately 2% of the active infected individuals were critical cases and needed ICU beds. If Austrian governmental protective measures would not have taken place and the reproduction number would remain as before the measures, active infected cases would increase dramatically. In the case that the number of active infected cases was around five times larger, the ICU bed capacity could have been exceeded according to the rate of around 2% critical cases in Austria. These results indicate the impact of the measures such as social distancing and lock-down in controlling the spread of COVID-19.

Declaration of competing interest

None Declared.

Acknowledgments

The authors acknowledge support by FWF (Austrian Science Fund) START project no. Y660 PDE *Models for Nanotechnology*.

References

[1] David S. Hui, Esam I. Azhar, Tariq A. Madani, Francine Ntoumi, Richard Kock, Osman Dar, et al., The continuing 2019-nCoV epidemic threat of novel coronaviruses to global health–the latest 2019 novel coronavirus outbreak in Wuhan, China, Int. J. Infect. Dis. 91 (2020) 264.

[2] World Health Organization (WHO), Coronavirus Disease 2019 (COVID-19) Situation Report - 97, WHO, 2020.

[3] Roy M. Anderson, B. Anderson, Robert M. May, Infectious Diseases of Humans: Dynamics and Control, Oxford University Press, 1992.

[4] Odo Diekmann, Johan Andre Peter Heesterbeek, Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation, 5, John Wiley & Sons, 2000.

[5] Herbert W. Hethcote, The mathematics of infectious diseases, SIAM Rev. 42 (4) (2000) 599–653.

[6] Fred Brauer, Carlos Castillo-Chavez, Carlos Castillo-Chavez, Mathematical Models in Population Biology and Epidemiology, 2, Springer, 2012.

[7] Tao Zhou, Quanhui Liu, Zimo Yang, Jingyi Liao, Kexin Yang, Wei Bai, et al., Preliminary prediction of the basic reproduction number of the Wuhan novel coronavirus 2019-nCoV, J. Evid. Base Med. 13 (1) (2020) 3–7. Wiley Online Library.

[8] Liangrong Peng, Wuyue Yang, Dongyan Zhang, Changjing Zhuge, Liu Hong, Epidemic Analysis of COVID-19 in China by Dynamical Modeling, 2020 arXiv preprint arXiv:2002.06563.

[9] Yu Chen, Jin Cheng, Yu Jiang, Keji Liu, A time delay dynamical model for outbreak of 2019-nCoV and the parameter identification, J. Inverse Ill-Posed Probl. 28 (2) (2020) 243–250.

[10] Igor Nesteruk, Statistics Based Predictions of Coronavirus 2019-nCoV Spreading in Mainland China, MedRxiv, 2020.

[11] Vivek Verma, Ramesh K. Vishwakarma, Anita Verma, Dilip C. Nath, Hafiz TA. Khan, Time-to-death approach in revealing chronicity and severity of COVID-19 across the world, PloS One 15 (5) (2020), e0233074.

[12] Gerardo Chowell, Nick W. Hengartner, Carlos Castillo-Chavez, Paul W. Fenimore, Jim Michael Hyman, The basic reproductive number of Ebola and the effects of public health measures: the cases of Congo and Uganda, J. Theor. Biol. 229 (1) (2004) 119–126.

[13] Huwen Wang, Zezhou Wang, Yinqiao Dong, Ruijie Chang, Chen Xu, Xiaoyue Yu, et al., Phase-adjusted estimation of the number of coronavirus disease 2019 cases in Wuhan, China, Cell Discovery 6 (1) (2020) 1–8.

[14] W. Jumpen, B. Wiwatanapataphee, Y.H. Wu, I.M. Tang, A SEIQR model for pandemic influenza and its parameter identification, Int. J. Pure Appl. Math. 52 (2) (2009) 247–265.

[15] Xueer Bai, Optimization of prognostication model about the spread of Ebola based on SIR model, in: 2016 6th International Conference on Machinery, Materials, Environment, Biotechnology and Computer, Atlantis Press, 2016.

[16] Cleo Anastassopoulou, Lucia Russo, Athanasios Tsakris, Constantinos Siettos, Data-based analysis, modelling and forecasting of the COVID-19 outbreak, PloS One 15 (3) (2020), e0230405.

[17] Giulia Giordano, Franco Blanchini, Raffaele Bruno, Patrizio Colaneri, Alessandro Di Filippo, Angela Di Matteo, et al., Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy, Nat. Med. (2020) 1–6.

[18] William Ogilvy Kermack, Anderson G. McKendrick, A contribution to the mathematical theory of epidemics, Proc. R. Soc. Lond. - Ser. A Contain. Pap. a Math. Phys. Character 115 (772) (1927) 700–721.

[19] William Ogilvy Kermack, Anderson G. McKendrick, Contributions to the mathematical theory of epidemics. ii.–the problem of endemicity, Proc. R. Soc. Lond. Ser. A Contain. Pap. Math. Phys. Character 138 (834) (1932) 55–83.

[20] William Ogilvy Kermack, Anderson G. McKendrick, Contributions to the mathematical theory of epidemics. iii.–further studies of the problem of endemicity, Proc. R. Soc. Lond. Ser. A Contain. Pap. Math. Phys. Character 141 (843) (1933) 94–122.

[21] Benjamin Stadlbauer, Andrea Cossettini, Daniel Pasterk, Paolo Scarbolo, Leila Taghizadeh, Clemens Heitzinger, et al., Bayesian estimation of physical and geometrical parameters for nanocapacitor array biosensors, J. Comput. Phys. 397 (2019) 108874.

[22] Leila Taghizadeh, Ahmad Karimi, Benjamin Stadlbauer, Wolfgang J. Weninger, Eugenijus Kaniusas, Clemens Heitzinger, Bayesian inversion for electrical-impedance tomography in medical imaging using the nonlinear Poisson–Boltzmann equation, Comput. Methods Appl. Mech. Eng. 365 (2020) 112959.

[23] Leila Taghizadeh, Ahmad Karimi, Elisabeth Presterl, Clemens Heitzinger, Bayesian inversion for a biofilm model including quorum sensing, Comput. Biol. Med. (2019) 103582.

[24] Ervin K. Lenzi, Luiz Roberto Evangelista, Leila Taghizadeh, Daniel Pasterk, Rafael S. Zola, Trifce Sandev, et al., Reliability of Poisson–Nernst–Planck anomalous models for impedance spectroscopy, J. Phys. Chem. B 123 (37) (2019) 7885–7892.

[25] Ralph C. Smith, Uncertainty Quantification: Theory, Implementation, and Applications, 12, SIAM, 2013.

[26] Jari Kaipio, Erkki Somersalo, Statistical and Computational Inverse Problems, 160, Springer Science & Business Media, 2006.

[27] Walter R. Gilks, Sylvia Richardson, David Spiegelhalter, Markov Chain Monte Carlo in Practice, Chapman & Hall, 1996.

[28] Christian P. Robert, George Cassella, Monte Carlo Statistical Methods, Springer Verlag, 1999.

[29] Malcolm Sambridge, Klaus Mosegaard, Monte Carlo methods in geophysical inverse problems, Rev. Geophys. 40 (3) (2002) 1–29.

[30] Klaus Mosegaard, Malcolm Sambridge, Monte Carlo analysis of inverse problems, Inverse Probl. 18 (3) (2002) R29.

[31] Jeffrey S. Rosenthal, et al., Optimal proposal distributions and adaptive MCMC, Handbook of Markov Chain Monte Carlo 4 (2011).

[32] Heikki Haario, Marko Laine, Antonietta Mira, Eero Saksman, DRAM: efficient adaptive MCMC, Stat. Comput. 16 (4) (2006) 339–354.

[33] Simon L. Cotter, Gareth O. Roberts, Andrew M. Stuart, David White, MCMC methods for functions: modifying old algorithms to make them faster, Stat. Sci. (2013) 424–446.

[34] Heikki Haario, Eero Saksman, Johanna Tamminen, Adaptive proposal distribution for random walk Metropolis algorithm, Comput. Stat. 14 (3) (1999) 375–396.

[35] Heikki Haario, Eero Saksman, Johanna Tamminen, et al., An adaptive Metropolis algorithm, Bernoulli 7 (2) (2001) 223–242.

[36] Luke Tierney, Antonietta Mira, Some adaptive Monte Carlo methods for Bayesian inference, Stat. Med. 18 (1718) (1999) 2507–2515.

[37] Peter J. Green, Antonietta Mira, Delayed rejection in reversible jump Metropolis–hastings, Biometrika 88 (4) (2001) 1035–1053.

[38] Andrew Gelman, Gareth O. Roberts, Walter R. Gilks, et al., Efficient Metropolis jumping rules, Bayesian Statistics 5 (599–608) (1996) 42.

[39] Christophe Fraser, Estimating individual and household reproduction numbers in an emerging epidemic, PloS One 2 (8) (2007) e758.

[40] Thomas Hotz, Matthias Glock, Stefan Heyder, Sebastian Semper, Anne Böhle, Alexander Krämer, Monitoring the Spread of COVID-19 by Estimating Reproduction Numbers over Time, 2020. ArXiv preprint arXiv:2004.08557.

[41] World Health Organization (WHO), Coronavirus Disease 2019 (COVID-19) Situation Report - 73, WHO, 2020.

[42] Robert Verity, Lucy C. Okell, Ilaria Dorigatti, Peter Winskill, Charles Whittaker, Natsuko Imai, et al., Estimates of the severity of coronavirus disease 2019: a model-based analysis, Lancet Infect. Dis. (2020). Elsevier.

[43] Henrik Sjödin, Anders F. Johansson, Åke Brännström, Zia Farooq, Hedi Katre Kriit, Annelies Wilder-Smith, et al., Covid-19 Health Care Demand and Mortality in Sweden in Response to Non-pharmaceutical (NPIs) Mitigation and Suppression Scenarios, MedRxiv, 2020, p. 20, https://doi.org/10.1101/2020.03.