Time-Discrete Parameter Identification Algorithms for Two Deterministic Epidemiological Models applied to the Spread of COVID-19

Benjamin Wacker[1,*] **and Jan Schlüter**[1,2]

[1]Next Generation Mobility Group, Max-Planck-Institut for Dynamics and Self-Organization, Department of Dynamics of Complex Fluids, Am Fassberg 17, D-37077 Göttingen, Germany
[2]Institute for Dynamics of Complex Systems, Faculty of Physics, Georg-August-University of Göttingen, Friedrich-Hund-Platz 1, D-37077 Göttingen, Germany
[*]Corresponding authors: Benjamin Wacker (bewa87@gmx.de), Jan Schlüter (jan.schlueter@ds.mpg.de)

ABSTRACT

The pandemic outbreak of COVID-19 threatens people worldwide. Politicians around the globe have to balance different interests. Accordingly, they make decisions which heavily impact our daily life or even curtail fundamental rights. They consequently base their judgements on advices from scientists. For these reasons, many epidemiological models have appeared and have been famous since the beginning of the twentieth century. However, scientists depend on rich data to evaluate them and their assumptions. As this pandemic outbreak has lead to large data collections by the John Hopkins University since its beginning, we take advantage of this situation and compare two deterministic epidemiological models heavily used - the Susceptible-Infectious-Recovered model (SIR model) and the Susceptible-Infectious-Recovered-Dead model (SIRD model). In contrast to other works which concentrate on future predictions, we aim to investigate the validity of assumptions with respect to available data. For that purpose, instead of using these models forward in time, we propose modified time-discrete versions of the continuous models with relaxing all transfer rates to be time-varying and develop algorithms to predict these coefficients based on the principle of "divide-and-conquer" for the inverse problems. Since Asia, Europe and North America are especially affected by this pandemic, we choose countries from these continents to illustrate our results. In the main part of this article, we escepially discuss our results for Hubei (China). As supplementary material, we illustrate our findings for the aforementioned choice of countries. We finally draw some conclusions for future applications of these models and data acquisition.

Introduction

Since its outbreak in Wuhan (China) in December 2019, the worldwide spread of COVID-19 has enormously impacted our daily life and governments around the globe are actually facing tremendous different challenges. Due to complexity of the current situation and extensive consequences of governments' decisions, politicians heavily rely on scientific advice. Thus, John Hopkins University collects epidemiological data from many countries during the last months[1]. Additionally, many biological and medical studies regarding different aspects of this new corona-virus have been rapidly appeared in scientific journals[2–8]. For example, Lauer et al. published one study of estimation of the incubation time for COVID-19[9].

Additionally, even more theoretical scientists like mathematicians have recently gained further interests in epidemiology. One of mathematical epidemiology's most groundbreaking work introduced the now well-known SIR model by Kermack and McKermack in 1927[10]. Kermack and McKendrick assumed a fixed population size and decomposed this population in three different homogeneous groups of people, namely susceptible people, infectious people and recovered people. They excluded births, deaths and deaths by disease from their model. Due to its success and simplicity, their works were reprinted in 1991[11–13]. In upcoming decades, epidemiologists and mathematicians have proposed many variants and extensions of this basic model by, for example, adding age or spatial structures[14–18].

After the outbreak of COVID-19, many scientists are recently publishing articles with emphasis on epidemic forecasts which strongly relate to mathematical modelling. Many approaches, mainly focusing on stochastic arguments, with respect to predicting forecasts of the total number of infected people have been appeared during the last weeks[19–24] or in the past[25, 26]. Recently, neural networks have been applied to forecasting[27].

Since these questions definitely deserve their current attention, current data acquisition by John Hopkins University allows us to certainly take a closer look on the parameter identification problem in such epidemiological models, so-called inverse problems[28]. Here, one aims to estimate different parameters appearing in our epidemiological models based on observed data.

Different deterministic approaches by optimisation techniques[29, 30] like least-squares methods[31, 32], variational imbedding[33] and Gaussian fitting[34] or stochastic models like time-series[35] or parameter identification by Bayesian methods[36] have been successfully used in inverse epidemiological problems. Evaluating the calculated constant or time-varying transfer rates on acquired data, we have reasonable foundations to make model assumptions plausible.

Many approaches regarding parameter estimation techniques in differential equations and especially in epidemiological models are well known[37–53]. The works by Hadeler[45, 46] and Mummert[52] use the idea of assuming all transfer rates to be time-varying in their models. However, they only apply this in a continuous setting.

In contrast to many of the aforementioned works, our main goal is the proposal of parameter identification algorithms for SIR and SIRD models which avoid optimisation techniques for the full problem and use principal techniques from applied linear algebra and applied statistics on a decoupled time-discrete problem formulation for both the SIR and the SIRD model. We start with continuous formulations of the time-continuous SIR model and the time-continuous SIRD model. Based on forward finite differences, we propose implicit time-discrete models for both model approaches. However, instead of using constant transfer coefficients as in the classical models, we allow all transfer rates in these models to be time-varying. With this proposed modification, we obtain implicit time-discrete models with time-varying transfer rates we want to identify. Given observed data for susceptible people, infectious people, recovered and dead people, we can now identify our time-varying parameters by simply solving one small linear system at each time step instead of application of optimisation techniques with respect to the full SIR or SIRD problem. Hence, we are able to avoid high-dimensional, possible non-convex problems with many local minima. If we further want the earlier parameters constant, we additionally propose an estimation based on arithmetic means, medians and standard deviations for uncertainty estimation which can readily be applied to the data.

Due to current data acquisition of John Hopkins University, we test our developed parameter identification algorithms on data of the worldwide spread of COVID-19 and evaluate their performances. Finally, we discuss some future research possibilities.

Methods

In this section, we describe our continuous and time-discrete forward SIR and SIRD models. The ideas of these models heavily base on the pioneering works by Kermack and McKendrick[10–13]. Afterwards, we propose two algorithms for solving each inverse problem to compute time-varying transfer coefficients in these models.

SIR Model

We begin with describing the SIR model. Before stating equations, we give some information on assumptions of the model. After its formulation, we propose a time-discrete version of the model. In contrast to traditional works with constant transfer coefficients, we introduce time-varying coefficient for our parameter identification algorithm which we finally portray in this subsection. From a mathematical viewpoint, we can directly solve this inverse problem without application of regression algorithms. Instead, we apply optimisation techniques on smaller subproblems.

Assumptions
In its classical formulation[16–18], the following assumptions should be fulfilled for application of the SIR model:

- One neglects the population's age-structure, inhomogeneities with respect to spatial dependence and group behaviour. One assumes that the population consists of homogeneous pools like in simulations of different biological system[54] or chemical reactions[55];

- In addition, one divides the population into the three homogeneous pools of susceptible people (S), infectious people (I) and recovered people (R);

- Since the considered time period is small, one ignores natural births and natural and infection-related deaths within the population. Thus, the population size N is constant in time and it holds

$$N \equiv N(t) = S(t) + I(t) + R(t) \tag{1}$$

  for all times t in the respective time interval;

- Except from the transfer coefficient from susceptible people to infectious people which might depend on government's rulings in this classical formulation, all other transfer rate between these pools are constant.

For further details on assumptions, we refer the interested reader to two papers by Brauer[56, 57]. We illustrate a sketch of this model in Figure 1.

Let further $[t_1, t_M]$ be the respective time interval with start time t_1 and end time t_M. Later, since our constant time data step amounts to one day, we consider a sequence $\{t_j\}_{j=1}^{M}$ of time points regarding our time-discrete model versions.

Figure 1. Sketch of the SIR model. We divide the people into three homogeneous pools of susceptible people (P), infectious people (I) and recovered people (R). The arrows portray flows between these pools with respective transfer rates.

Continuous Formulation

As an abbreviation for the time derivative of a function f, we simply write $f'(t) := \dfrac{\mathrm{d}f(t)}{\mathrm{d}t}$. The classical continuous formulation of the SIR model reads

$$
\left.\begin{aligned}
S'(t) &= -\alpha(t) \cdot \frac{S(t) \cdot I(t)}{N} \\
I'(t) &= \alpha(t) \cdot \frac{S(t) \cdot I(t)}{N} - \beta \cdot I(t) \\
R'(t) &= \beta \cdot I(t)
\end{aligned}\right\}
\tag{2}
$$

with appropriate initial conditions $S(t_1) = S_1 > 0$, $I(t_1) = I_1 > 0$ and $R(t_1) = R_1 \geq 0$. In the forward model, we aim to seek solutions $S, I, R\colon [t_1, t_M] \longrightarrow \mathbb{R}$ that fulfill (2) in a meaningful sense almost everywhere on $[t_1, t_M]$. This means that solutions could be strongly differentiable in a classical or weakly differentiable in a Lebesgue-measurable manner[58].

Time-Discrete Formulation

For abbreviation, we shortly write $f_j := f(t_j)$ for all $j \in \{1, \ldots, M\}$. We apply finite differences such that the approximation $f'(t_{j+1}) \approx \dfrac{f_{j+1} - f_j}{\Delta t} = f_{j+1} - f_j$ holds for all $j \in \{1, \ldots, M-1\}$ since $\Delta t = 1$ with respect to our data and for ease of notation. In addition, we apply an implicit scheme for forward modelling in contrast to Dehning et al. in their stochastic SIR model[19]. To state our modified time-discrete version of the SIR model (2), we relax our conditions and allow time-dependence of the transfer coefficient β. Consequently, application of finite differences leads to

$$
\left.\begin{aligned}
S_{j+1} - S_j &= -\alpha_{j+1} \cdot \frac{S_{j+1} \cdot I_{j+1}}{N} \\
I_{j+1} - I_j &= \alpha_{j+1} \cdot \frac{S_{j+1} \cdot I_{j+1}}{N} - \beta_{j+1} \cdot I_{j+1} \\
R_{j+1} - R_j &= \beta_{j+1} \cdot I_{j+1}
\end{aligned}\right\}
\tag{3}
$$

as our time-discrete version of (2) for all $j \in \{1, \ldots, M-1\}$ with initial conditions $S_1 > 0$, $I_1 > 0$ and $R_1 \geq 0$ as stated before.

Time-Discrete Inverse Parameter Identification SIR Problem

Why do we proceed with (3) although the transfer coefficient β is generally assumed to be constant in time? If β satisfied this condition, we must solve a regression problem due to

$$
R_{j+1} - R_j = \beta \cdot I_{j+1}
\tag{4}
$$

for all $j \in \{1, \ldots, M-1\}$ to obtain β and possibly use all data points. Depending on the cost function defined for this high-dimensional, probably non-convex minimization problem, this normally yields an optimisation problem with many local extrema if we take also add the other subproblems to our investigation. Since we want to avoid this, we use our transformed system (3).

In contrast to the forward model (3), we now seek the transfer coefficients α_{j+1} and β_{j+1} for all $j \in \{1, \ldots, M-1\}$ provided given data S_j, I_j, R_j for all $j \in \{1, \ldots, M\}$. Now, we observe from (3) that

$$
\begin{aligned}
I_{j+1} - I_j &= \alpha_{j+1} \cdot \frac{S_{j+1} \cdot I_{j+1}}{N} - \beta_{j+1} \cdot I_{j+1} \\
&= -(S_{j+1} - S_j) - (R_{j+1} - R_j)
\end{aligned}
\tag{5}
$$

is valid for all $j \in \{1, \ldots M-1\}$. By (5), we reduce (3) and obtain

$$
\begin{pmatrix} S_{j+1} - S_j \\ R_{j+1} - R_j \end{pmatrix} = \begin{pmatrix} -\dfrac{S_{j+1} \cdot I_{j+1}}{N} & 0 \\ 0 & I_{j+1} \end{pmatrix} \cdot \begin{pmatrix} \alpha_{j+1} \\ \beta_{j+1} \end{pmatrix}
\tag{6}
$$

for all $j \in \{1, \ldots, M-1\}$. Hence, we will solve a linear system for all $j \in \{1, \ldots, M-1\}$ if the system matrix is invertible, i.e. $S_{j+1} \neq 0$ and $S_{j+1} \neq 0$. If we attain $S_{j+1} = 0$ and $I_{j+1} \neq 0$ for some $j \in \{1, \ldots, M-1\}$, we must set $\alpha_{j+1} = 0$, solely solve $\beta_{j+1} = \dfrac{R_{j+1} - R_j}{I_{j+1}}$ and we can conclude that we reach the end of the disease's spread. In case of $S_{j+1} \neq 0$ and $I_{j+1} = 0$, the same conclusion applies, but both transfer parameters reach $\alpha_{j+1} = 0$ and $\beta_{j+1} = 0$.

Preprocessing of Population Data

Our input data consists of cumulative infected people $\left(\left\{ \widetilde{I}_j \right\}_{j=1}^{M} \right)$, cumulative recovered people $\left(\left\{ \widetilde{R}_j \right\}_{j=1}^{M} \right)$ and cumulative dead people $\left(\left\{ \widetilde{D}_j \right\}_{j=1}^{M} \right)$ for M time points. Hence, we necessarily must process these data first to develop an algorithm for our time-discrete inverse parameter identification SIR problem (6). Our procedure reads

$$\left. \begin{array}{l} R_j = \widetilde{R}_j + \widetilde{D}_j \\ I_j = \widetilde{I}_j - \widetilde{R}_j \\ S_j = N - I_j - R_j \end{array} \right\} \tag{7}$$

for all $j \in \{1, \ldots, M\}$.

SIR Parameter Identification Algorithm

We summarize our procedure in Table 1.

Input:	- Population size N
	- Cumulative infected people $\left\{ \widetilde{I}_j \right\}_{j=1}^{M}$ at M time points
	- Cumulative recovered people $\left\{ \widetilde{R}_j \right\}_{j=1}^{M}$ at M time points
	- Cumulative dead people $\left\{ \widetilde{D}_j \right\}_{j=1}^{M}$ at M time points
Step 1:	- Process input data according to (7) for all $j \in \{1, \ldots, M\}$
Step 2:	- Solve time-discrete inverse parameter identification SIR problem according to (6) for all $j \in \{1, \ldots, M-1\}$ For that purpose, distinguish the following four cases: - 1) If $S_{j+1} \neq 0$ and $I_{j+1} \neq 0$, set $\alpha_{j+1} = \dfrac{N \cdot (S_j - S_{j+1})}{S_{j+1} \cdot I_{j+1}}$ and $\beta_{j+1} = \dfrac{R_{j+1} - R_j}{I_{j+1}}$ - 2) If $S_{j+1} = 0$ and $I_{j+1} \neq 0$, set $\alpha_{j+1} = 0$ and $\beta_{j+1} = \dfrac{R_{j+1} - R_j}{I_{j+1}}$ - 3) If $S_{j+1} \neq 0$ and $I_{j+1} = 0$, set $\alpha_{j+1} = 0$ and $\beta_{j+1} = 0$ - 4) If $S_{j+1} = 0$ and $I_{j+1} = 0$, set $\alpha_{j+1} = 0$ and $\beta_{j+1} = 0$
Output:	- Time-varying transfer coefficient sequence $\left\{ \alpha_j \right\}_{j=2}^{M}$ for all $j \in \{2, \ldots, M\}$
	- Time-varying transfer coefficient sequence $\left\{ \beta_j \right\}_{j=2}^{M}$ for all $j \in \{2, \ldots, M\}$

Table 1. Algorithmic summary of SIR Parameter Identification Algorithm. Input data, algorithmic steps and output data are enlisted in this table.

We use GNU Octave to implement this algorithm[59].

Possible Parameter Reduction by Simple Statistical Analysis

Since often the transfer coefficient β is assumed to be constant in time, we further need to process our time-varying transfer coefficient sequence $\left\{ \beta_j \right\}_{j=2}^{M}$. For this purpose, we simply calculate the arithmetic mean

$$\overline{\beta} = \sum_{j=2}^{M} \beta_j \tag{8}$$

and respectively the median $\widehat{\beta}$. With (8), we obtain the standard deviation

$$\sigma_\beta = \sqrt{\frac{1}{M-2} \cdot \sum_{j=2}^{M} \left(\beta_j - \overline{\beta} \right)^2}. \tag{9}$$

This finally allows us to calculate confidence intervals of the form $\left[\bar{\beta}-2\cdot\sigma_\beta,\bar{\beta}+2\cdot\sigma_\beta\right]$ or $\left[\bar{\beta}-\sigma_\beta,\bar{\beta}+\sigma_\beta\right]$ based on standard deviations.

Optimisation Procedure for Time-Varying Parametric Coefficients

Since social behaviour even changes doe to severe epidemics before established rulings by governments, Chowell et al.[60], Fisman et al.[61], Althaus[62] and Brauer[63] suggested application of a time-varying contact rate (exponential decaying)

$$\alpha(t) = \delta\cdot\exp(-\varepsilon\cdot t) \tag{10}$$

with positive constants δ and ε which adapts to given data. For that purpose, we need to identify a proper cost function. Since we do not want to discuss optimisation techniques in details (we refer interested readers to the book of Nocedal and Wright[64]), we simply use Nelder-Mead-algorithms[65] as implemented in GNU Octave's function FMINSEARCH[59] due to its simplicity and its ability to capture local extrema of non-differentiable functions. However, one still has to deal with problems of ending in local extrema[64], but this is out of the scope of this work.

Now, our scoring function for minimization of the desired parameters δ and ε, given data points (t_j, α_j) for all $j \in \{2,\dots,M\}$, reads

$$g(\delta,\varepsilon) = \sum_{j=2}^{M}\left|\delta\cdot\exp(-\varepsilon\cdot t_j) - \alpha_j\right|. \tag{11}$$

In case of Hubei(China), we additionally consider Gaussian recovering rates

$$\beta(t) = a\cdot\exp\left(-\left(\frac{t-b}{c}\right)^2\right) \tag{12}$$

with real constants a, b and c. We proceed similarly to the case of exponentially decaying contact rates and we want to shortly note that an additional uncertainly quantification analysis by the Fisher-Information matrix (i.e. Hessian matrix) becomes possible by applying a two times continuously differentiable function like the l^2-norm or mollified versions of the above used l^1-norm. However, we relinquish it here.

SIRD Model

We start with a description of the SIRD model. Before giving its mathematical formulation, we report typical assumptions of the model. Afterwards, we present a time-discrete version of the model. Again, we use time-varying coefficients for our parameter identification algorithm in contrast to constant ones.

Assumptions

All assumptions of the SIR model apply. Additionally, we extend the pool of recovered people (R) by an additional homogeneous pool of dead people (D) and for ease of notion, we still count dead people (D) for total population size N because we estimate $D \ll N$.

Figure 2. Sketch of the SIRD model. We divide the people into three homogeneous pools of susceptible people (P), infectious people (I), recovered people (R) and dead people (D). The arrows portray flows between these pools with respective transfer rates.

A sketch of this model is given in Figure 2.

Continuous Formulation

We have the classical continuous formulation

$$\left.\begin{aligned}
S'(t) &= -\alpha(t)\cdot\frac{S(t)\cdot I(t)}{N} \\
I'(t) &= \alpha(t)\cdot\frac{S(t)\cdot I(t)}{N} - \beta\cdot I(t) - \gamma\cdot I(t) \\
R'(t) &= \beta\cdot I(t) \\
D'(t) &= \gamma\cdot I(t)
\end{aligned}\right\} \tag{13}$$

of the SIRD model with appropriate initial conditions $S(t_1) = S_1 > 0$, $I(t_1) = I_1 > 0$, $R(t_1) = R_1 \geq 0$ and $D(t_1) = D_1 \geq 0$. In the forward model, we again seek functions $S, I, R, D \colon [t_1, t_M] \longrightarrow \mathbb{R}$ which satisfy (13) in a suitable sense.

Time-Discrete Formulation

Equivalently to the case of the modified time-discrete SIR model, we generalize the constant transfer rates and allow them to be time-varying. Accordingly, this yields

$$
\left.
\begin{aligned}
S_{j+1} - S_j &= -\alpha_{j+1} \cdot \frac{S_{j+1} \cdot I_{j+1}}{N} \\
I_{j+1} - I_j &= \alpha_{j+1} \cdot \frac{S_{j+1} \cdot I_{j+1}}{N} - \beta_{j+1} \cdot I_{j+1} - \gamma_{j+1} \cdot I_{j+1} \\
R_{j+1} - R_j &= \beta_{j+1} \cdot I_{j+1} \\
D_{j+1} - D_j &= \gamma_{j+1} \cdot I_{j+1}
\end{aligned}
\right\}
\tag{14}
$$

for our modified time-discrete SIRD model for all $j \in \{1, \ldots, M-1\}$ with initial conditions $S_1 > 0$, $I_1 > 0$, $R_1 \geq 0$ and $D_1 \geq 0$.

Time-Discrete Inverse Parameter Identification SIRD Problem

We proceed similarly to the time-discrete inverse parameter identification SIR problem. At first, we notice

$$
I_{j+1} - I_j = -\left(S_{j+1} - S_j\right) - \left(R_{j+1} - R_j\right) - \left(D_{j+1} - D_j\right)
\tag{15}
$$

for all $j \in \{1, \ldots, M-1\}$. We reduce (14) by (15) and get

$$
\begin{pmatrix} S_{j+1} - S_j \\ R_{j+1} - R_j \\ D_{j+1} - D_j \end{pmatrix}
=
\begin{pmatrix}
-\dfrac{S_{j+1} \cdot I_{j+1}}{N} & 0 & 0 \\
0 & I_{j+1} & 0 \\
0 & 0 & I_{j+1}
\end{pmatrix}
\cdot
\begin{pmatrix} \alpha_{j+1} \\ \beta_{j+1} \\ \gamma_{j+1} \end{pmatrix}
\tag{16}
$$

for all $j \in \{1, \ldots, M-1\}$. The validity of the same argument as in the time-discrete inverse parameter identification SIR problem holds.

Preprocessing of Population Data

Similarly to the case of the time-discrete inverse parameter identification SIR problem, we need to process the same given input data of cumulative infected people $\left(\left\{ \widetilde{I_j} \right\}_{j=1}^{M} \right)$, cumulative recovered people $\left(\left\{ \widetilde{R_j} \right\}_{j=1}^{M} \right)$ and cumulative dead people $\left(\left\{ \widetilde{D_j} \right\}_{j=1}^{M} \right)$ for M time points. Consequently, we compute

$$
\left.
\begin{aligned}
D_j &= \widetilde{D_j} \\
R_j &= \widetilde{R_j} \\
I_j &= \widetilde{I_j} - R_j - D_j \\
S_j &= N - I_j - R_j - D_j
\end{aligned}
\right\}
\tag{17}
$$

for all $j \in \{1, \ldots, M\}$.

SIRD Parameter Identification Algorithm

We summarize our procedure in Table 2 and use GNU Octave to implement this algorithm[59].

Parameter Reduction by Simple Statistical Analysis

We proceed as in the corresponding description of the time-discrete parameter identification SIR model.

Optimisation Procedure for Time-Varying Parametric Coefficients

Again, we apply the same procedure as for the time-discrete parameter identification SIR model.

Input:	- Population size N
	- Cumulative infected people $\left\{ \widetilde{I}_j \right\}_{j=1}^{M}$ at M time points
	- Cumulative recovered people $\left\{ \widetilde{R}_j \right\}_{j=1}^{M}$ at M time points
	- Cumulative dead people $\left\{ \widetilde{D}_j \right\}_{j=1}^{M}$ at M time points
Step 1:	- Process input data according to (17) for all $j \in \{1,\ldots,M\}$
Step 2:	- Solve time-discrete inverse parameter identification SIR problem according to (16) for all $j \in \{1,\ldots,M-1\}$
	For that purpose, distinguish the following four cases:
	- 1) If $S_{j+1} \neq 0$ and $I_{j+1} \neq 0$, set $\alpha_{j+1} = \dfrac{N \cdot (S_j - S_{j+1})}{S_{j+1} \cdot I_{j+1}}$, $\beta_{j+1} = \dfrac{R_{j+1} - R_j}{I_{j+1}}$ and $\gamma_{j+1} = \dfrac{D_{j+1} - D_j}{I_{j+1}}$
	- 2) If $S_{j+1} = 0$ and $I_{j+1} \neq 0$, set $\alpha_{j+1} = 0$, $\beta_{j+1} = \dfrac{R_{j+1} - R_j}{I_{j+1}}$ and $\gamma_{j+1} = \dfrac{D_{j+1} - D_j}{I_{j+1}}$
	- 3) If $S_{j+1} \neq 0$ and $I_{j+1} = 0$, set $\alpha_{j+1} = 0$, $\beta_{j+1} = 0$ and $\gamma_{j+1} = 0$
	- 4) If $S_{j+1} = 0$ and $I_{j+1} = 0$, set $\alpha_{j+1} = 0$, $\beta_{j+1} = 0$ and $\gamma_{j+1} = 0$
Output:	- Time-varying transfer coefficient sequence $\left\{ \alpha_j \right\}_{j=2}^{M}$ for all $j \in \{2,\ldots,M\}$
	- Time-varying transfer coefficient sequence $\left\{ \beta_j \right\}_{j=2}^{M}$ for all $j \in \{2,\ldots,M\}$
	- Time-varying transfer coefficient sequence $\left\{ \gamma_j \right\}_{j=2}^{M}$ for all $j \in \{2,\ldots,M\}$

Table 2. Algorithmic summary of SIR Parameter Identification Algorithm. Input data, algorithmic steps and output data are enlisted in this table.
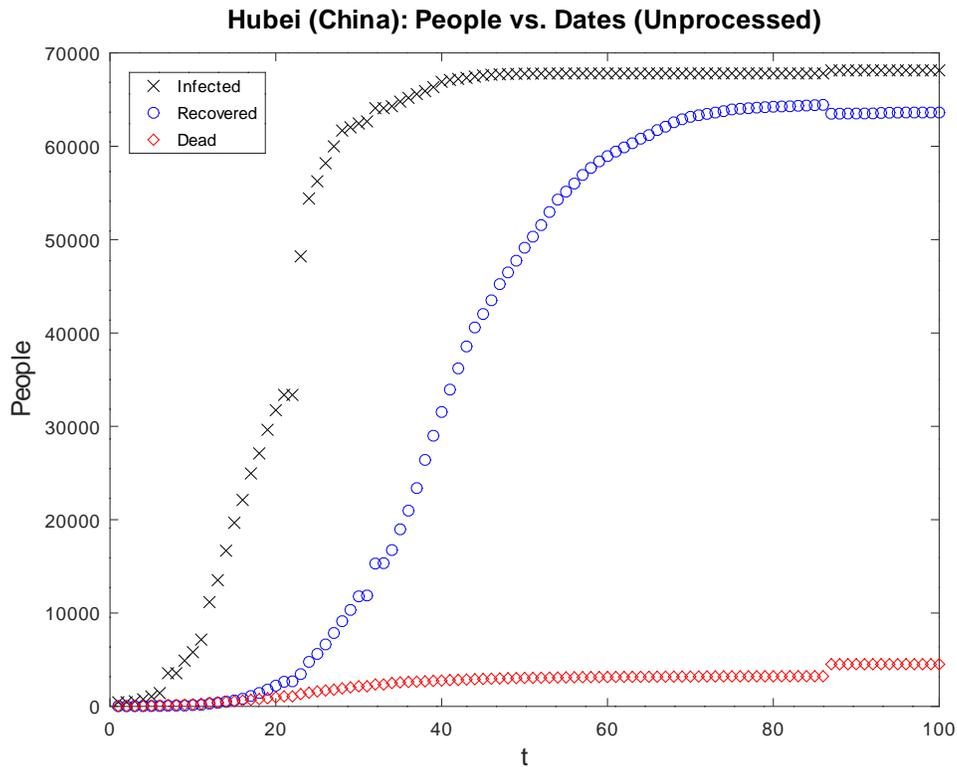

Figure 3. Plots of the unprocessed data from Hubei (China). Crosses represent cumulative cases of reported infected people, circles stand for cumulative cases of reported (estimated) cases of recovered people and diamonds portray cumulative cases of reported (estimated) dead people.

Results

Here, we consider our results for Hubei (China) because the worldwide spread of COVID-19 started there. Hence, the longest timed data are available in this case[1]. In this case, the population size reads $N = 59170000$ as stated in our additional information. We consider a time period ranging from 22 January, 2020 ($t = 1$) to 30 April, 2020 ($t = 100$).

In Figure 3, we plot cumulative cases of reported infected people, cumulative cases of reported recovered people and cumulative cases of reported dead people. Apparently, there are two jumps around day 35 and day 85 of the COVID-19 epidemic which possibly occur due to changes in Chinese reports. Especially, the last jump values are excluded for calculation of means and medians. For details, we refer to our code availability statement.

In all different subplots of Figure 4, we depict our results of Hubei (China) for our proposed time-discrete parameter identification SIR algorithm. We relinquish integrating the standard deviations of arithmetic mean and median in our plots for our estimates of β. In all subplots of Figure 5, we illustrate our results of Hubei (China) for our time-discrete parameter identification SIRD algorithm. Especially, the last jump values are excluded for calculation of means and medians. For details, we refer to our code availability statement.

Discussion

In media, we normally see non-decreasing curves which report cumulative cases of infected, dead or recovered people. However, we note that we need those values at a specific time in our aforementioned SIR and SIRD models. Hence, the grouped plots of Figures 4 and 5 differ in general from the ones we notice in Figure 3 or in media.

Regarding all transfer rates, we note that they should be non-negative for all times. However, due to changes in Chinese reports, we observe statistical outliers around day 85. Thus, we exclude those data from calculate which react sensitively on such phenomena like arithmetic means.

If we closely look at the contact rates in Figures 4 and 5, we are able to detect an exponentially decaying behaviour of contact rates in time. Hence, in accordance to descriptions of Brauer[63] and references therein[60–62], we can probably conclude that people change their social behaviour due to life-threatening diseases. This general trend does not only occur in data from Hubei (China), but is also observed in the analysed data in our supplementary material. However, we want to note that the contact rate heavily depends on social behaviour and if care eases up, this will generally lead to rising contact rates and to exponential growth of infected people again.

Let us shortly remark on the recovery and death rates. Surprisingly, we only notice an approximately Gaussian behaviour in recovery rates in Hubei (China) compared to all other countries in our supplementary material. Hence, we recommend to stay with arithmetic means and median for these transfer rates regarding their biological meaning[14, 63].

For all data, we have to keep in mind that these are reported cases and nobody really knows real numbers with respect to infected, dead and recovered people due to COVID-19. Hence, we state that application of such models and of our parameter identification algorithms build on profound data which need to be provided.

Due to this short time period of observed data, it is reasonable to apply SIR or SIRD models. Furthermore, as already pointed out in our discussion, time-decreasing contact rates can also be assumed at the beginning of such epidemics in general. In addition to these observations, we stress that our proposed algorithms are not only applicable to the actual COVID-19 epidemic, but to all epidemics which fulfill the assumptions of SIR and SIRD models. Therefore, it is a reasonable contribution to report about.

As future research projects, we might consider different adaption like statistical[66] or stochastic approaches by basis functions[67]. Pollicott et al. applied a deterministic continuous approach through multiple differentiation and interpolations by splines and trigonometric functions for the time-varying contact rate function $\alpha(t)$[68].

Since it is out of the scope of this work, it might be an interesting point to integrate time-delay in our parameter identification approach without applying a full optimization process as given by Chen et al[69]. In addition to that, one could evaluate extensions to fractional differential operators in time[70–72]. As a concluding remark, one can apply PDE-ODE-models to epidemiology[73] to introduce spatial or even further structures.
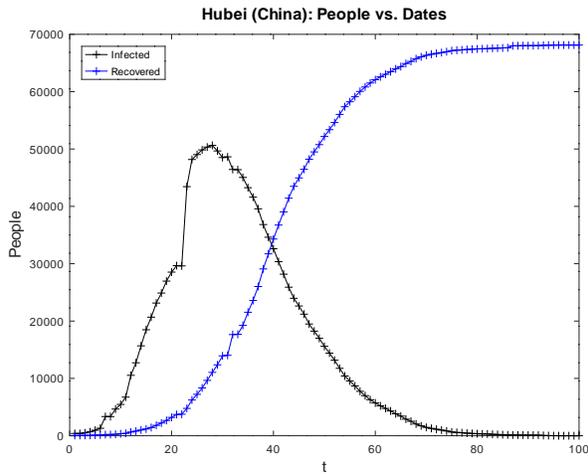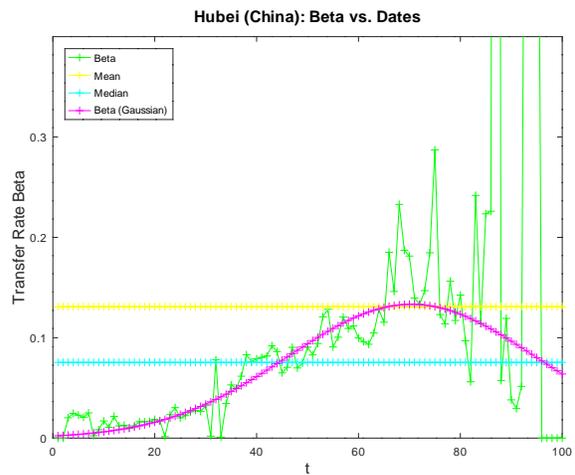
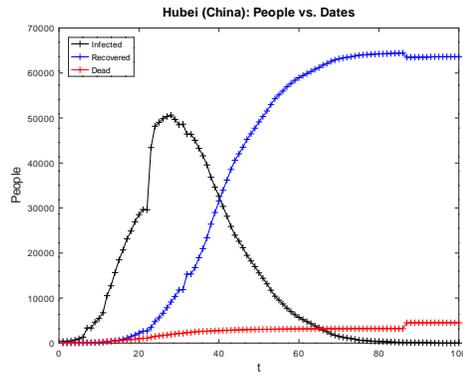**(a)** Grouped Plots
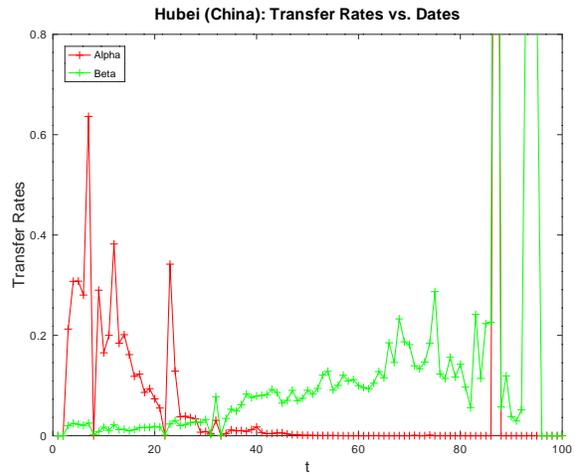
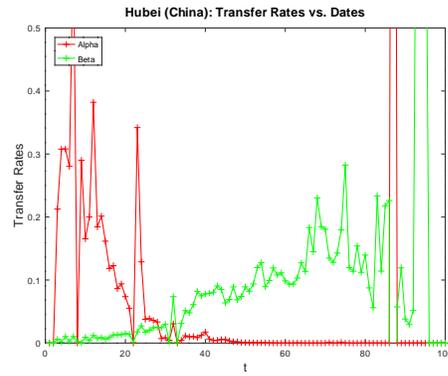**(b)** Transfer Rates
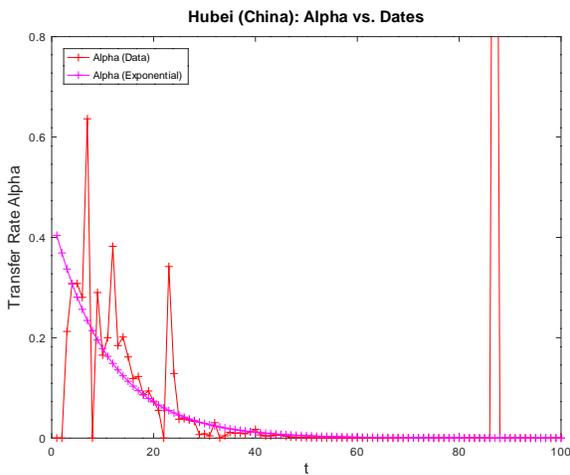
(c) Contact Rate
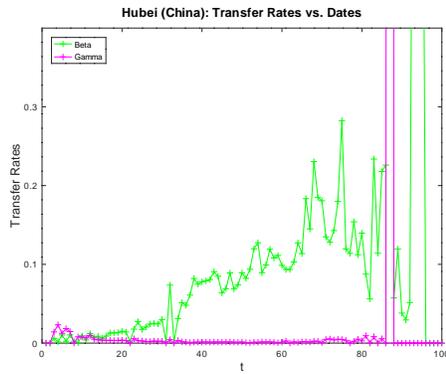
(d) Recovery Rate

Figure 4. Results for our time-discrete parameter identification SIR model. (a) Plots of processed data for infected people and (cumulative) recovered people at time t. (b) Plots of both time-varying transfer rates $\alpha(t)$ and $\beta(t)$ from our identification algorithm. (c) Plots of time-varying contact rate $\alpha(t)$ and a parametric approximation with a decaying exponential function as a model. (d) Plots of time-varying recovery rate $\beta(t)$, its mean and median and a parametric approximation with a Gaussian function.
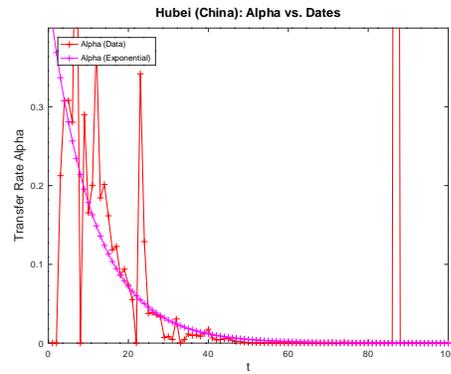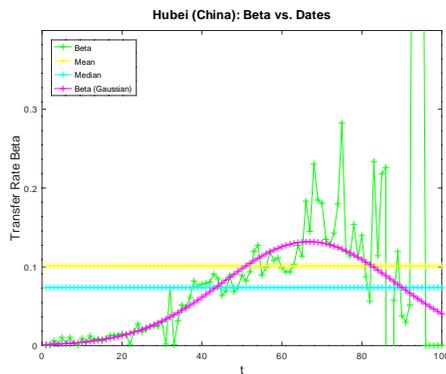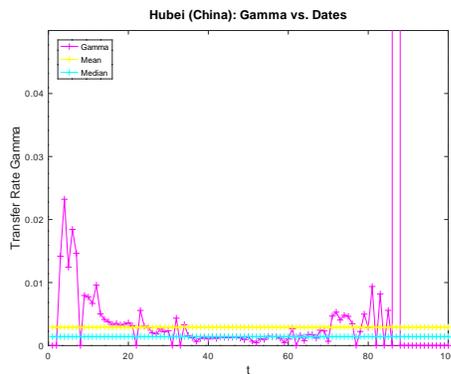
(a) Grouped Plots

(b) Transfer Rates α and β

(c) Transfer Rates β and γ

(d) Contact Rate

(e) Recovery Rate

(f) Death Rate

Figure 5. Results for our time-discrete parameter identification SIRD model. (a) Plots of processed data for infected people, (cumulative) recovered people and (cumulative) dead people at time t. (b) Plots of the time-varying transfer rates $\alpha(t)$ and $\beta(t)$ from our identification algorithm. (c) Plots of the time-varying transfer rates $\beta(t)$ and $\gamma(t)$ from our identification algorithm. (d) Plots of time-varying contact rate $\alpha(t)$ and a parametric approximation with a decaying exponential function. (e) Plots of time-varying recovery rate $\beta(t)$, its mean and median and a parametric approximation with a Gaussian function. (f) Plots of time-varying death rate $\gamma(t)$, its mean and its median.

References

1. Dong, E., Du, H. & Gardner, L. An interactive web-based dashboard to track COVID-19 in real time. *The Lancet - Infectious Diseases*, https://dx.doi.org/10.1016/S1473-3099(20)30120-1 (2020).

2. Zhou, P. et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**:270–273, https://dx.doi.org/10.1038/s41586-020-2012-7 (2020).

3. Chen, H. et al. Clinical characteristics and intrauterine vertical transmission potential of COVID-19 infection in nine pregnant women: a retrospective review of medical records. *The Lancet* **395**(10226):809–815, https://dx.doi.org/10.1016/S0140-6736(20)30360-3 (2020).

4. Xu, Z. et al. Pathological findings of COVID-19 associated with acute respiratory distress syndrome. *The Lancet - Respiratory Medicine* **8**(4):420–422, https://dx.doi.org/10.1016/S2213-2600(20)30076-X (2020).

5. Wang, H. & Zhang, L. Risk of COVID-19 for patients with cancer. *The Lancet - Oncology* **21**(4):E181, https://doi.org/10.1016/S1470-2045(20)30149-2 (2020).

6. Liu, Y. et al. Aerodynamic analysis of SARS-CoV-2 in two Wuhan hospitals. *Nature*, https://dx.doi.org/10.1038/s41586-020-2271-3 (2020).

7. Poyiadji, N. et al. COVID-19-associated Acute Hemorrhagic Necrotizing Encephalopathy: CT and MRI Features. *Radiology*, https://dx.doi.org/10.1148/radiol.2020201187 (2020).

8. Grillet, F. et al. Acute Pulmonary Embolism Associated with COVID-19 Pneumonia Detected by Pulmonary CT Angiography. *Radiology*, https://dx.doi.org/10.1148/radiol.2020201544 (2020).

9. Lauer, S.A. et al. The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application. *Annals of Internal Medicine*, https://dx.doi.org/10.7326/M20-0504 (2020).

10. Kermack, W. & McKendrick, A. A Contribution to the Mathematical Theory of Epidemics. *Proceedings of the Royal Society London A* **115**:700–721 (1927).

11. Kermack, W. & McKendrick, A. Contributions to the mathematical theory of epidemics - I. *Bulletin of Mathematical Biology* **53**(1-2):33-55, https://dx.doi.org/10.1007/BF02464423 (1991).

12. Kermack, W. & McKendrick, A. Contributions to the mathematical theory of epidemics - II. The problem of endemicity. *Bulletin of Mathematical Biology* **53**(1-2):57–87, https://dx.doi.org/10.1007/BF02464424 (1991).

13. Kermack, W. & McKendrick, A. Contributions to the mathematical theory of epidemics - III. Further studies of the problem of endemicity. *Bulletin of Mathematical Biology* **53**(1-2):89–118, https://dx.doi.org/BF02464425 (1991).

14. Hethcote, H.H. The Mathematics of Infectious Diseases. *SIAM Review* **42**(4):599–653, https://dx.doi.org/10.1137/S0036144500371907 (2000).

15. Murray, J.D. *Mathematical Biology II: Spatial Models and Medical Applications*, In: Interdisciplinary Applied Mathematics, Vol. **18**, https://dx.doi.org/10.1007/b98869 (Springer, New York, 2003).

16. Brauer, F. & Castillo-Chavez, C. *Mathematical Models in Population Biology and Epidemiology*, In: Texts in Applied Mathematics, Vol. **40**, https://dx.doi.org/10.1007/978-1-4614-1686-9 (Springer, New York, 2012).

17. Martcheva, M. *An Introduction to Mathematical Epidemiology*, In: Texts in Applied Mathematics, Vol. **61**, https://dx.doi.org/10.1007/978-1-4899-7612-3 (Springer, New York, 2015).

18. Iannelli, M. & Milner, F. *The Basic Approach to Age-Structured Population Dynamics - Models, Methods and Numerics*, In: Lectures Notes on Mathematical Modelling in the Life Sciences, https://dx.doi.org/10.1007/978-94-024-1146-1 (Springer, 2017).

19. Dehning, J. et al. Inferring COVID-19 spreading rates and potential change points for number forecasts. https://arxiv.org/abs/2004.01105 (2020).

20. Fanelli, D. & Piazza, F. Analysis and forecast of COVID-19 spreading in China, Italy and France. *Chaos, Solitons & Fractals*, https://dx.doi.org/10.1016/j.chaos.2020.109761 (2020).

21. He, S., Tang, S. & Rong, L. A discrete stochastic model of the COVID-19 outbreak: Forecast and control. *Mathematical Biosciences and Engineering* **17**(4):2792–2804, https://dx.doi.org/10.3934/mbe.2020153 (2020).

22. Flaxman, S. et al. Report 13: Estimating the number of infections and the impact of non-pharmaceutical interventions on COVID-19 in 11 European countries. https://www.imperial.ac.uk/media/imperial-college/medicine/mrc-gida/2020-03-30-COVID19-Report-13.pdf (2020).

23. Flaxman, S. et al. Estimating the number of infections and the impact of non-pharmaceutical interventions on COVID-19 in European countries: technical description update. https://arxiv.org/abs/2004.11342 (2020).

24. Maier, B.F. & Brockmann, D. Effective containment explains sub-exponential growth in confirmed cases of recent COVID-19 outbreak in Mainland China. *Science*, https://dx.doi.org/10.1126/science.abb4557 (2020).

25. Gray, A. et al. A Stochastic Differential Equation SIS Epidemic Model. *SIAM Journal of Applied Mathematics* **71**(3):876–902, https://dx.doi.org/10.1137/10081856X (2011).

26. Osthus, D. et al. Forecasting seasonal influenza with a state-space SIR model. *Annals of Applied Statistics* **11**(1):202–224, https://dx.doi.org/10.1214/16-AOAS1000 (2017).

27. Al-Qaness, M.A.A., Ewees, A.A., Fan, H. & Abd El Aziz, M. Optimization Method for Forecasting Confirmed Cases of COVID-19 in China. *Journal of Clinical Medicine* **9**(3):674, https://dx.doi.org/10.3390/jcm9030674 (2020).

28. Tarantola, A. *Inverse problem theory and methods for parameter estimation*, In: Other Titles in Applied Mathematics, Vol. **89**, https://dx.doi.org/10.1137/1.9780898717921 (SIAM, Philadelphia, 2005).

29. Peng, L. et al. Epidemic analysis of COVID-19 in China by dynamical modeling. https://arxiv.org/abs/2002.06563 (2020).

30. Cantó, B., Coll, C. & Sánchez, E. Estimation of parameters in a structured SIR model. *Advances in Difference Equations* **33**, https://dx.doi.org/10.1186/s13662-017-1078-5 (2017).

31. Jia, J. et al. Modeling the Control of COVID-19: Impact of Policy Interventions and Meteorological Factors. https://arxiv.org/abs/2003.02985 (2020).

32. Bastos, S.B. & Cajueiro, D.O. Modeling and forecasting the early evolution of the Covid-19 pandemic in Brazil. https://arxiv.org/abs/2003.14288 (2020).

33. Marinov, T.T., Marinova, R.S., Omojola, J., Jackson, M. Inverse problem for coefficient identification in SIR epidemic models. *Computers & Mathematics with Applications* **67**(12):2218–2227, https://dx.doi.org/10.1016/j.camwa.2014.02.002 (2014).

34. Barmparis, G.D. & Tsironis, G.P. Estimating the infection horizon of COVID-19 in eight countries with a data-driven approach. *Chaos, Solitons & Fractals*, https://dx.doi.org/10.1016/j.chaos.2020.109842 (2020).

35. Bjørnstad, O.N., Finkenstädt, B.F. & Grenfell, B.T. Dynamics of Measles Epidemics: Estimating Scaling of Transmission Rates Using a Time Series SIR Model. *Ecological Monographs* **72**(2):169–184, https://dx.doi.org/10.1890/0012-9615(2002)072[0169:DOMEES]2.0.CO;2 (2002).

36. Coelho, F.C., Torres Codeço, C. & Gomes, M.G.M. A Bayesian Framework for Parameter Estimation in Dynamical Models. *PLoS One* **6**(5): e19616, https://dx.doi.org/10.1371/journal.pone.0019616 (2011).

37. Andreasen, V. The Final Size of an Epidemic and Its Relation to the Basic Reproduction Number. *Bulletin of Mathematical Biology* **73**(10):2305–2321, https://dx.doi.org/10.1007/s11538-010-9623-3 (2011).

38. Arino, J. et al. A Final Size Relation for Epidemic Models. *Mathematical Biosciences and Engineering* **4**(2):159–175, https://dx.doi.org/10.3934/mbe.2007.4.159 (2007).

39. Capistrán, M.A., Moreles, M.A. & Lara, B. Parameter Estimation of Some Epidemic Models. The Case of Recurrent Epidemics Caused by Respiratory Syncytial Virus. *Bulletin of Mathematical Biology* **71**:1890–1901, https://dx.doi.org/10.1007/s11538-009-9429-3 (2009).

40. Chowell, G. et al. Modelling the transmission dynamics of acute haemorrhagic conjunctivitis: application to the 2003 outbreak in Mexico. *Statistic in Medicine* **25**(11):1840–1857, https://dx.doi.org/10.1002/sim.2352 (2006).

41. Chowell, G. et al. Estimation of the reproduction number of dengue fever from spatial epidemic data. *Mathematical Biosciences* **208**(2):571–589, https://dx.doi.org/10.1016/j.mbs.2006.11.011 (2007).

42. Diekmann, O., Heesterbeek, J.A.P. & Metz, J.A.J. On the Definition and the Computation of the Basic Reproduction Ratio R_0 in Models for Infectious Diseases in Heterogeneous Populations. *Journal of Mathematical Biology* **28**:365–382, https://dx.doi.org/10.1007/BF00178324 (1990).

43. Evans, N.D. et al. The structural identifiability of the susceptible infected recovered model with seasonal forcing. *Mathematical Biosciences* **194**(2):175–197, https://dx.doi.org/10.1016/j.mbs.2004.10.011 (2005).

44. Grassly, N.C. & Fraser, C. Seasonal infectious disease epidemiology. *Proceedings of the Royal Society B: Biological Sciences* **273**:2541–2550, https://dx.doi.org/10.1098/rspb.2006.3604 (2006).

45. Hadeler, K.P. Parameter identification in epidemic models. *Mathematical Biosciences* **229**: 185–189, `https://dx.doi.org/10.1016/j.mbs.2010.12.004` (2011).

46. Hadeler, K.P. Parameter Estimation in Epidemic Models: Simplified Formulas. *Canadian Applied Mathematics Quarterly* **19**(4):343–356, `http://www.math.ualberta.ca/ami/CAMQ/pdf_files/vol_19/19_4/19_4c.pdf` (2011).

47. Hooker, G., Ellner, S.P., De Vargas Roditi, L. & Earn, D.J.D. Parameterizing state-space models for infectious disease dynamics by generalized profiling: measles in Ontario. *Journal of the Royal Society: Interface* **8**:961–974, `https://dx.doi.org/10.1098/rsif.2010.0412` (2011).

48. Lange, A. Reconstruction of Disease Transmission Rates: Applications to Measles, Dengue and Influenza. *Journal of Theoretical Biology* **400**:138–153, `https://dx.doi.org/10.1016/j.jtbi.2016.04.017` (2016).

49. Li, J. & Lou, Y. Characteristics of an epidemic outbreak with a large initial infection size. *Journal of Biological Dynamics* **10**(1):366–378, `https://dx.doi.org/10.1080/17513758.2016.1205223` (2016).

50. Ma, J. & Earn, D.J.D. Generality of the Final Size Formula for an Epidemicof a Newly Invading Infectious Disease. *Bulletin of Mathematical Biology* **68**:679–702, `https://dx.doi.org/10.1007/s11538-005-9047-7` (2006).

51. Magal, P. & Webb, G. The parameter identification problem for SIR epidemic models: identifying unreported cases. *Journal of Mathematical Biology* **77**:1629–1648, `https://dx.doi.org/10.1007/s00285-017-1203-9` (2018).

52. Mummert, A. Studying the recovery procedure for the time-dependent transmission rate(s) in epidemic models. *Journal of Mathematical Biology* **67**:483–507, `https://dx.doi.org/10.1007/s00285-012-0558-1` (2013).

53. Van den Driessche, P. & Watmough, J. Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. *Mathematical Biosciences* **180**:29–48, `https://dx.doi.org/10.1016/S0025-5564(02)00108-6` (2002).

54. Bassingthwaighte, J.B., Butterworth, E., Jardine, B. & Raymond, G.M. Compartment Modeling in the Analysis of Biological Systems. In: Reisfeld, B. & Mayeno, A. (Editors) *Computational Toxicology*. Methods in Molecular Biology (Methods and Protocol), Vol. **929**, `https://dx.doi.org/10.1007/978-1-62703-050-2_17` (Humana Press, Totowa, NJ, 2012).

55. Jourdan, N. et al. Compartmental Modelling in chemical engineering: A critical review. *Chemical Engineering Science*, **210**: 115196, `https://doi.org/10.1016/j.ces.2019.115196` (2019).

56. Brauer, F. The Kermack-McKendrick epidemic model revisited. *Mathematical Biosciences* **198**:119–131, `https://dx.doi.org/10.1016/j.mbs.2005.07.006` (2005).

57. Brauer, F. Some simple epidemic models. *Mathematical Biosciences and Engineering* **3**(1):1–15, `https://dx.doi.org/10.3934/mbe.2006.3.1` (2006).

58. Evans, L.C. *Partial Differential Equations*, In: Graduate Studies in Mathematics, Vol. **19** (American Mathematical Society, Providence, Rhode Island, United States, 2008).

59. Eaton, J.W., Bateman, D., Hauberg, S. & Wehbring, R. GNU Octave Version 5.1.0 Manual: a High-Level Interactive Language for Numerical Computations. `https://www.gnu.org/software/octave/doc/v5.1.0/` (2019).

60. Chowell, G. et al. The basic reproductive number of Ebola and the effects of public health measures: the cases of Congo and Uganda. *Journal of Theoretical Biology* **229**(1):119–126, `https://doi.org/10.1016/j.jtbi.2004.03.006` (2004).

61. Fisman, D.N., Hauck, T.S., Tuite, A.R. & Greer, A.L. An IDEA for Short Term Outbreak Projection: Nearcasting Using the Basic Reproduction Number. *PLoS One* **8**(12):e83622, `https://dx.doi.org/10.1371/journal.pone.0083622` (2013).

62. Althaus, C.L. Estimating the Reproduction Number of Ebola Virus (EBOV) During the 2014 Outbreak in West Africa. *PLoS Current Outbreaks*, `https://dx.doi.org/10.1371/currents.outbreaks.91afb5e0f279e7f29e7056095255b288` (2014).

63. Brauer, F. The Final Size of a Serious Epidemic. *Bulletin of Mathematical Biology* **81**:869–877, `https://dx.doi.org/10.1007/s11538-018-00549-x` (2019).

64. Nocedal, J. & Wright, S. *Numerical Optimization*, `https://dx.doi.org/10.1007/978-0-387-40065-5` (Springer, New York, 2006).

65. Nelder, J.A. & Mead, R. A Simplex Method for Function Minimization, *The Computer Journal* **7**(4):308–313, https://dx.doi.org/10.1093/comjnl/7.4.308 (1965).

66. Liang, H. & Wu, H. Parameter Estimation for Differential Equation Models Using a Framework of Measurement Error in Regression Models. *Journal of the American Statistical Association* **103**:1570–1583, https://dx.doi.org/10.1198/016214508000000797 (2008).

67. Ramsay, J.O., Hooker, G., Campbell, D. & Cao, J. Parameter estimation for differential equations: a generalized smoothing approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **69**(5):741–796, https://dx.doi.org/10.1111/j.1467-9868.2007.00610.x (2007).

68. Pollicott, M., Wang, H. & Weiss, H.H. Extracting the time-dependent transmission rate from infection data via solution of an inverse ODE problem. *Journal of Biological Dynamics* **6**(2):509–523, https://dx.doi.org/10.1080/17513758.2011.645510 (2012).

69. Chen, Y., Cheng, J., Jiang, Y. & Lia, K. A time delay dynamical model for outbreak of 2019-nCov and the parameter identification. *Journal of Inverse and Ill-Posed Problems* **28**(2):243–250, https://dx.doi.org/10.1515/jiip-2020-0010 (2020).

70. Tang, B. Dynamics for a fractional-order predator-prey model with group defense. *Scientific Reports* **10**: 4906, https://doi.org/10.1038/s41598-020-61468-3 (2020).

71. Macheras, P. & Illiadis, A. *Modeling in Biopharmaceutics, Pharmacokinetics and Pharmacodynamics - Homogeneous and Heterogeneous Approaches*, In: Interdisciplinary Applied Mathematics, Vol. **30**, https://dx.doi.org/10.1007/978-3-319-27598-7 (Springer, Cham, 2016).

72. Diethelm, K. *The Analysis of Fractional Differential Equations - An Application-Oriented Exposition Using Differential Operators of Caputo Type*, In: Lecture Notes in Mathematics, Vol. **2004**, https://dx.doi.org/10.1007/978-3-642-14574-2 (Springer, Berlin, 2010).

73. David, J.F., Iyaniwura, S.A., Ward, M.J. & Brauer, F. A novel approach to modelling the spatial spread of airborne diseases: an epidemic model with indirect transmission. *Mathematical Biosciences and Engineering* **17**(4):3294–3328, https://dx.doi.org/10.3934/mbe.2020188 (2020).

Author Contributions Statement

B.W. mainly conceived and designed the research. B.W. and J.S. searched and processed data. B.W. contributed continuous and time-discrete methods. B.W. implemented numerical algorithms. Both authors analysed data, performed data representation, discussed and interpreted the results. Both authors drafted the manuscript.

Competing Interests

We declare no competing interests.

Correspondence

Correspondence to Benjamin Wacker (bewa87@gmx.de) or Jan Schlüter (jan.schlueter@ds.mpg.de).

Additional Information

Code availability: https://github.com/bewa87/2020-COVID-19-ParameterIdentification
COVID-19 data: https://data.humdata.org/dataset/novel-coronavirus-2019-ncov-cases (John Hopkins University)
Data (population sizes (2018)): https://de.statista.com/