Nicola Guglielmi, Elisa Iacomini* and Alex Viguerie

Identification of time delays in COVID-19 data

Abstract

Objective: COVID-19 data released by public health authorities is subject to inherent time delays. Such delays have many causes, including delays in data reporting and the natural incubation period of the disease. We develop and introduce a numerical procedure to recover the distribution of these delays from data.
Methods: We extend a previously-introduced compartmental model with a nonlinear, distributed-delay term with a general distribution, obtaining an integrodifferential equation. We show this model can be approximated by a weighted-sum of constant time-delay terms, yielding a linear problem for the distribution weights. Standard optimization can then be used to recover the weights, approximating the distribution of the time delays. We demonstrate the viability of the approach against data from Italy and Austria.
Results: We find that the delay-distributions for both Italy and Austria follow a Gaussian-like profile, with a mean of around 11 to 14 days. However, we note that the delay does not appear constant across all data types, with infection, recovery, and mortality data showing slightly different trends, suggesting the presence of independent delays in each of these processes. We also found that the recovered delay-distribution is not sensitive to the discretization resolution.
Conclusions: These results establish the validity of the introduced procedure for the identification of time-delays in COVID-19 data. Our methods are not limited to COVID-19, and may be applied to other types of epidemiological data, or indeed any dynamical system with time-delay effects.

Keywords: compartmental models; COVID-19; delay differential equations; epidemiology.

Introduction

The outbreak of COVID-19 in 2020 and into 2021 has led to a surge in interest in the mathematical modeling of the COVID-19 epidemic, as well as the mathematical modeling of epidemics generally. These models have taken many types. A full review of the relevant literature is beyond the current work. However, some general classes of models include data-driven and machine learning approaches (Barros et al. 2022; Bhouri et al. 2021; Jha, Cao, and Oden 2020; Linka et al. 2020; Viguerie et al. 2022a; Wang et al. 2020), models based on partial differential equation (PDE) systems (Albi et al. 2022; Bertaglia and Pareschi 2021a,b; Bertrand and Pirch 2021; Grave et al. 2021; Grave and Coutinho 2021; Viguerie et al. 2020, 2021), agent-based models (Zohdi 2020), and models based on ordinary differential equation (ODE) systems. This last category is by far the most common, with such articles numbering in the thousands. Some works of this type include e.g. (Calafiore, Novara, and Possieri 2020; Choi and Ki 2020; Ferguson et al. 2020; Gatto et al. 2020; Ivorra et al. 2020; Parolini et al. 2021; Piccolomini and Zama 2020; Remuzzi and Remuzzi 2020). We note that the categories listed above are not necessarily strict delimiters; indeed, many of the above cited works, as well myriad others, have characteristics of multiple of different model classes. One commonality that nearly all the presented models share is that they are, in some way, compartmental models in

*Corresponding author: Elisa Iacomini, Department of Mathematics and Computer Science, University of Ferrara, Via Nicolò Machiavelli 30, Ferrara, FE 44121, Italy, E-mail: elisa.iacomini@unife.it. https://orcid.org/0000-0002-0981-2086
Nicola Guglielmi and Alex Viguerie, Division of Mathematics, Gran Sasso Science Institute, L'Aquila, AQ, Italy,
E-mail: nicola.guglielmi@gssi.it (N. Guglielmi), alexander.viguerie@gssi.it (A. Viguerie)

the SIR family owing to the seminal work of Kermack and Mackendrick (Ogilvy Kermack and McKendrick 1927) (see also: Breda et al. 2012; Murray 2007; Iannelli and Pugliese 2015).

One of the most important aspects of COVID-19 modeling, and epidemiological modeling in general deals with the presence of *time delays* in the data. Time delays certainly manifest themselves via the natural incubation period of the disease; this is reflected in the employed models commonly incorporating an *exposed* compartment, to model the period after exposure to the virus but before symptomatic infection (Bhouri et al. 2021; Gatto et al. 2020; Linka et al. 2020; Viguerie et al. 2020) also shown in Figure 1.

Such an approach is a reasonable one, though, there are several potential problems. First of all, the exposed compartment is not necessarily easily known from data, and hence must be estimated. This may be done, for example, through rule-of-thumb estimates based on the infected population, which result in high uncertainty (Albi et al. 2022; Grave and Coutinho 2021; Grave et al. 2021). Compounding this problem further are the well-known presence of asymptomatic patients, who may spread the disease while never showing appreciable symptoms, and, at least in the earlier stages of the pandemic, were almost certainly under-counted (Albi et al. 2022; Viguerie et al. 2020).

Another major problem with the exposed compartment in most SEIR-type formulations come from estimating the sojourn time of incubation period, or the length of time spent in the compartment. Under most common compartmental approaches, the sojourn time of the exposed period is implicitly assumed to follow an *exponential distribution*. In practice, this may not always be a realistic (Iannelli and Pugliese 2015) assumption.

In (Guglielmi, Iacomini, and Viguerie 2022; Viguerie et al. 2022a), the authors employed a delay-differential equation (DDE) model, in which the incubation period was modeled with a time-delay term, rather than as a separate exposed compartment. Such a modeling approach eliminated the need to estimate the exposed compartment, as well as the assumption of an exponentially-distributed sojourn time for the incubation period, instead assuming a Dirac-delta (constant) distribution. Other works incorporating DDE models into the analysis of COVID-19 include (Dell'Anna 2020; Devipriya, Dhamodharavadhani, and Selvi 2021; Kumar and Erturk 2020), while references for more general epidemiological models of this type may be found in (Brauer and Castillo-Chavez 2012; Buonomo, d'Onofrio, and Lacitignola 2008; Forde 2005; Iannelli and Pugliese 2015; Murray 2007; Takeuchi, Ma, and Beretta 2000).
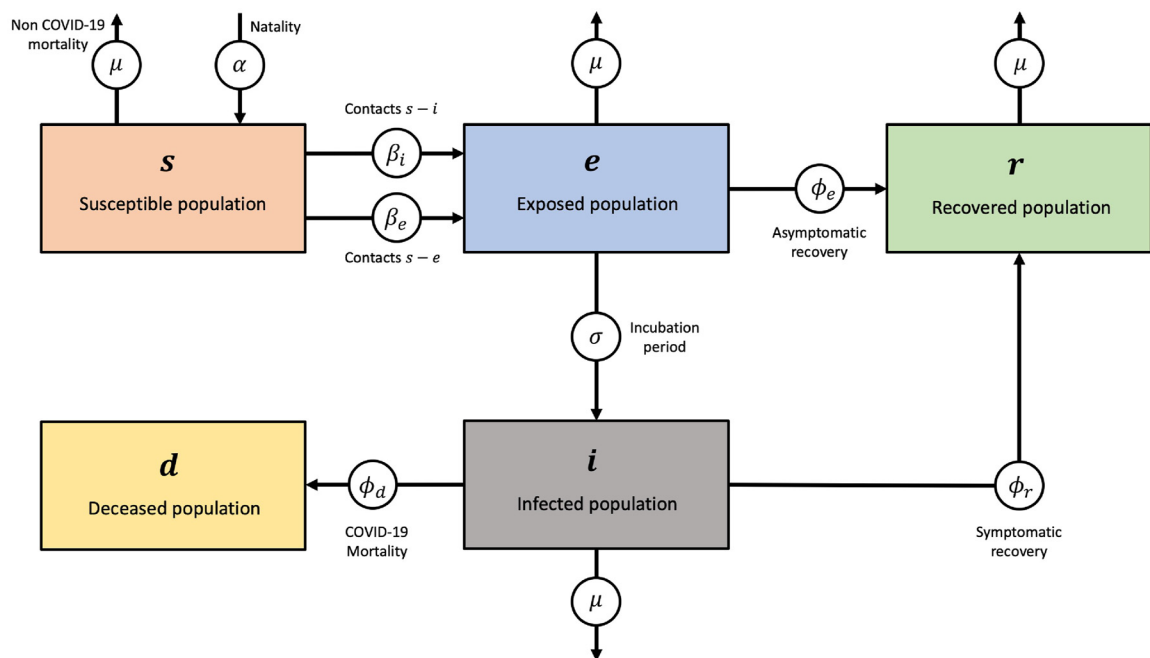
Figure 1: Flow chart describing the evolution of the various compartments in a classical SEIRD model.

Aside from the presence of incubation periods, delays in data may result from other sources: additional time from initial onset of symptoms to when a person decides to be tested, the availability of testing, the speed of test result processing, testing center capacity, and other factors may have a large influence on when a given case is properly identified in the data (Sarnaglia et al. 2021). The presence of such lags is important, particularly when attempting to model measured data, as such processes are intrinsic to the data (Bastos et al. 2019). Hence, when attempting to properly evaluate different intervention strategies, for instance, such delays must be considered in the modeling process. A model designed to fit measured parameters, while not taking into account the delays in measurement, may fail provide a wholly accurate description of the dynamics. In this sense, we may view such delays as no longer time from exposure to *infection*, but rather time from exposure to *identification*.

Quantifying both the magnitude and distribution of such time delays, considering the effects of both identification lag-time as well as the natural incubation period, is thus important for evaluating intervention efficacy and allocation of resources, and is the focus of the present work. In order to accomplish this task, we first introduce, as a forward problem, an SIRD (susceptible-infected-recovered-deceased) model, in which the time from infection to identification is not assumed to follow any particular distribution. At the continuous level, this is accomplished via a convolution with an unknown probability distribution function. We then approximate the continuous probability distribute function with a discrete approximation, reducing the convolution term to a weighted sum of constant time-delays. In this way, we obtain a linear problem in terms of the distribution weights, enabling us to employ an optimization procedure against measured data to identify the weights, thus approximating the distribution of the time delays present in the data. We note that the introduced method, while applied in the current work to COVID-19, is general in nature and such a technique may be employed in any situation in which a dynamical system and its associated measurements may exhibit time-delayed dynamics whose distribution is not known.

The article is outlined as follows. We first introduce the employed delay-differential equation model and discuss some of its characteristics at the continuous level. We then introduce its discrete analogue and describe the optimization process for the identification of the time delay distribution. We then validate the proposed technique on an interval of data for the COVID-19 outbreaks in Italy and Austria demonstrating its effectiveness for different cases under different modeling assumptions. Finally, we conclude with a summary of our main findings and possible directions for future research in this area.

Delay differential equation model

The delay differential model introduced in (Guglielmi, Iacomini, and Viguerie 2022) reads as follows:

$$\partial_t s(t) = \alpha n(t) - \left(1 - \frac{A}{n(t)}\right)\beta_e s(t)i(t) - \left(1 - \frac{A}{n(t)}\right)\beta_i s(t)i(t - \sigma) - \mu s(t) + \nabla \cdot \left(n(t)\,v_s \nabla s(t)\right) \tag{1}$$

$$\partial_t i(t) = \left(1 - \frac{A}{n(t)}\right)\beta_e s(t)i(t) + \left(1 - \frac{A}{n(t)}\right)\beta_i s(t)i(t - \sigma) - \phi_d i(t - \sigma) - \phi_r i(t - \sigma)$$
$$- \mu i(t) + \nabla \cdot \left(n(t)v_i \nabla i(t)\right) \tag{2}$$

$$\partial_t r(t) = \phi_r i(t - \sigma) - \mu r(t) + \nabla \cdot \left(n(t)\,v_r \nabla r(t)\right) \tag{3}$$

$$\partial_t d(t) = \phi_d i(t - \sigma). \tag{4}$$

Here, $s(t), i(t), r(t), d(t)$ represent respectively the susceptible, infected, recovered and deceased compartments at time t. The parameter σ represents the time delay term which models the incubation period of the disease, as well as potential reporting delays. For a comprehensive description of the parameters and variables of the model, with the units of measure, we refer to Table 1. Moreover the living population is given by $n = s + i + r + d$.

Table 1: Relevant parameters and variables, with symbols and unit of measure. Note that s, i, r, d are the variables of the system, while the others are parameters.

Parameter/variable	Name	Units
s	Susceptible individuals	Persons
i	Infected individuals	Persons
r	Recovered individuals	Persons
d	Deceased individuals	Persons
β	Contact rate	Persons^{-1} · Days^{-1}
ϕ_r	Recovery rate	Days^{-1}
ϕ_d	Mortality rate	Days^{-1}
$\sigma_j, j = 1 : k$	Time-delay j	Days
$w_j, j = 1 : k$	Time-delay weight j	Dimensionless

We assume the following:

1. That the time scales considered are sufficiently short such that we do not need to consider new births or non-COVID-19 deaths;
2. That the population is sufficiently well-mixed in space, such that spatial dynamics (including the Allee effect on the transmission terms, see Viguerie et al. 2021) can be ignored;
3. That the time scales considered are sufficiently short such that we do not need to consider waning immunity;
4. That recovery and mortality rates follow an exponential distribution;
5. That the model depends on time-delays, arising from both natural (incubation period) and non-natural (measurement and reporting delays, and similar) causes, whose distribution is unknown.

We acknowledge that some of these assumptions are simplifications; however, for the scope of the current work, we feel they are nonetheless reasonable. Under these assumptions, our model reads:

$$\dot{s}(t) = -\beta s(t) \int_{\sigma_{\min}}^{\sigma_{\max}} i(t - \sigma) w(\sigma) \mathrm{d}\sigma \tag{5}$$

$$\dot{i}(t) = \beta s(t) \int_{\sigma_{\min}}^{\sigma_{\max}} i(t - \sigma) w(\sigma) \mathrm{d}\sigma - \phi_d \int_{\sigma_{\min}}^{\sigma_{\max}} i(t - \sigma) w(\sigma) \mathrm{d}\sigma - \phi_r \int_{\sigma_{\min}}^{\sigma_{\max}} i(t - \sigma) w(\sigma) \mathrm{d}\sigma \tag{6}$$

$$\dot{r}(t) = \phi_r \int_{\sigma_{\min}}^{\sigma_{\max}} i(t - \sigma) w(\sigma) \mathrm{d}\sigma \tag{7}$$

$$\dot{d}(t) = \phi_d \int_{\sigma_{\min}}^{\sigma_{\max}} i(t - \sigma) w(\sigma) \mathrm{d}\sigma. \tag{8}$$

where $\sigma_{\min}$ and $\sigma_{\max}$ denote the extrema of the time delays. Note that the presence of the convolution term:

$$\int_{\sigma_{\min}}^{\sigma_{\max}} i(t - \sigma) w(\sigma) \mathrm{d}\sigma \tag{9}$$

where $w(\sigma) \geq 0$ is a probability density function describing the distribution of the time-delay dependence present in the model. By definition,

$$\int_{\sigma_{\min}}^{\sigma_{\max}} w(\sigma) \mathrm{d}\sigma = 1. \tag{10}$$

Under the assumption (with $\delta(\cdot)$ a Dirac-δ distribution):

$$w(\sigma) = \delta(\sigma - \hat{\sigma}) \tag{11}$$

for a certain constant $\hat{\sigma} \in [\sigma_{\min}, \sigma_{\max}]$, one obtains the model introduced in (Guglielmi, Iacomini, and Viguerie 2022). In this sense, (5)–(8) can be considered a generalization of the model introduced in (Guglielmi, Iacomini, and Viguerie 2022).

The function w is unknown in general, and its identification in terms of the available measured data is the focus of the present work. To render the problem numerically tractable, we replace $w(\sigma)$ in (5)–(8) with a piecewise-constant approximation, letting $\{\sigma_j\}$ denote a partition of the interval, assumed uniform for simplicity: $[\sigma_{\min}, \sigma_{\max}]$ and $\Delta = \sigma_j - \sigma_{j-1}$:

$$w(\sigma) \approx \sum_{j=1}^{k} \widetilde{w}_j \chi_j(\sigma), \tag{12}$$

where the χ_j are simple functions, defined such that:

$$\chi_j = \begin{cases} 1 & \text{if } \sigma \in [\sigma_{j-1}, \sigma_j) \\ 0 & \text{else.} \end{cases} \tag{13}$$

From the mean value theorem, for each $[\sigma_{j-1}, \sigma_j)$, there exists a $\widetilde{\sigma}_j \in [\sigma_j, \sigma_{j-1})$ such that:

$$w(\widetilde{\sigma}_j) = \frac{1}{\Delta} \int_{\sigma_{j-1}}^{\sigma_j} w(\sigma) \mathrm{d}\sigma. \tag{14}$$

We then define the $\widetilde{w}_j$ in (12) such that:

$$\widetilde{w}_j = w(\widetilde{\sigma}_j). \tag{15}$$

We then have:

$$
\begin{aligned}
\int_{\sigma_{\min}}^{\sigma_{\max}} i(t - \sigma) w(\sigma) \mathrm{d}\sigma &\approx \int_{\sigma_{\min}}^{\sigma_{\max}} i(t - \sigma) \left(\sum_{j=1}^{k} \widetilde{w}_j \chi_j(\sigma) \right) \mathrm{d}\sigma \\
&= \sum_{j=1}^{k} \widetilde{w}_j \int_{\sigma_{\min}}^{\sigma_{\max}} i(t - \sigma) \chi_j(\sigma) \mathrm{d}\sigma \\
&= \sum_{j=1}^{k} \widetilde{w}_j \int_{\sigma_{j-1}}^{\sigma_j} i(t - \sigma) \mathrm{d}\sigma \\
&= \sum_{j=1}^{k} (\widetilde{w}_j \Delta) \frac{1}{\Delta} \int_{\sigma_{j-1}}^{\sigma_j} i(t - \sigma) \mathrm{d}\sigma \\
&\approx \sum_{j=1}^{k} (\widetilde{w}_j \Delta) i(t - \sigma_j),
\end{aligned} \tag{16}
$$

where the final line follows from the simple approximation:

$$\frac{1}{\Delta} \int_{\sigma_{j-1}}^{\sigma_j} i(t - \sigma) \mathrm{d}\sigma \approx i(t - \sigma_j). \tag{17}$$

Hence, by defining

$$w_j = \Delta \cdot \widetilde{w}_j, \tag{18}$$

we obtain the approximation:

$$\int_{\sigma_{\min}}^{\sigma_{\max}} i(t - \sigma)w(\sigma)\mathrm{d}\sigma \approx \sum_{j=1}^{k} w_j i(t - \sigma_j). \tag{19}$$

Note that:

$$\sum_{j=1}^{k} w_j = 1. \tag{20}$$

To see this, observe from (14), (18):

$$\sum_{j=1}^{k} w_j = \sum_{j=1}^{k} \Delta \cdot \widetilde{w}_j = \sum_{j=1}^{k} \Delta \cdot \left(\frac{1}{\Delta} \int_{\sigma_{j-1}}^{\sigma_j} w(\sigma)\mathrm{d}\sigma \right) = \sum_{j=1}^{k} \int_{\sigma_{j-1}}^{\sigma_j} w(\sigma)\mathrm{d}\sigma = \int_{\sigma_{\min}}^{\sigma_{\max}} w(\sigma)\mathrm{d}\sigma = 1. \tag{21}$$

From the preceding arguments, (16), replacing $w(\sigma)$ in (5)–(8) with its piecewise approximation (12) gives the following system of ordinary delayed-differential equations:

$$\dot{s}(t) = -\beta s(t) \sum_{j=1}^{k} w_j i(t - \sigma_j), \tag{22}$$

$$\dot{i}(t) = \beta s(t) \sum_{j=1}^{k} w_j i(t - \sigma_j) - (\phi_d + \phi_r) \sum_{j=1}^{k} w_j i(t - \sigma_j), \tag{23}$$

$$\dot{r}(t) = \phi_r \sum_{j=1}^{k} w_j i(t - \sigma_j), \tag{24}$$

$$\dot{d}(t) = \phi_d \sum_{j=1}^{k} w_j i(t - \sigma_j). \tag{25}$$

It was shown in (Guglielmi, Iacomini, and Viguerie 2022) that, for $k = 1$ (and thus only one time-delay $\hat{\sigma}$), the model (22)–(25), exhibits stable behavior provided that:

$$\phi_d + \phi_r < \frac{\pi}{2\sigma}, \tag{26}$$

which, for the more general case, easily extends to:

$$\phi_d + \phi_r < \frac{\pi}{2\sigma_j}, \tag{27}$$

for each j.

The significance of replacing (5)–(8) with the above approximation (22)–(25) is that we now have a problem which is linear in w_j. This allows us to make use of an optimization procedure to identify the w_j, allowing us to estimate $w(\sigma)$. Therefore, we are able to numerically estimate the dependence of the data on general time-delays, assuming only that $w(\sigma)$ is sufficiently smooth over the relevant interval. In contrast, many common modeling approaches assume, for example, that such delays are exponentially distributed (as in most standard compartment models), or that the delays are constant (as in most basic delay-differential equation models). We remark while we consider only distributions ω assumed to not vary in time in the present work, the above framework can nonetheless accommodate time-dependent ω, provided that such time-dependence is sufficiently smooth.

Optimization procedure and methods

Denote the measured susceptible, infected, recovered, and deceased (from COVID-19) populations as $\widetilde{s}, \widetilde{i}, \widetilde{r}, \widetilde{d}$, respectively. We then define the following optimization problem: For a given $\beta, \phi_r, \phi_d, \sigma_j, j = 1{:}k$ and $\boldsymbol{w} = [w_1, w_2, \ldots, w_k]^T$:

$$\underset{\boldsymbol{w}}{\arg\min} \quad (i - \widetilde{i})^2 + (d - \widetilde{d})^2 \tag{28a}$$

subject to

$$\sum_{j=1}^{k} w_j = 1 \tag{28b}$$

$$0 \le w_j \le 1 \; \forall j, \quad s, i, r, d \text{ solve } (22){-}(25). \tag{28c}$$

The novelty of the proposed method above lies in the approximation (12) of $w(\sigma)$, allowing us to consider the nonlinear convolution term (9) as a weighted sum of constant-delay terms. Through this approximation, it is possible to obtain an optimization problem for the time delay that is linear in w_j. This is important, as attempting to optimize for $w(\sigma)$ directly on (5)–(8) gives a highly nonlinear optimization problem that is likely untractable. This method of identifying a general distribution by approximating it as a weighted sum of constant time delays, then solving an optimization problem for the weights is, to the authors' knowledge, novel. We note that this approach is highly general and is not restricted to models of COVID-19 or epidemiological models in general. Indeed, one may use an analogous approach to identify general sojourn times in compartmental models. Moreover, the accuracy and the detailed description of the delay depend on how k is chosen. This allows this method to be very applicable for other models as well.

We note that the model (22)–(25) depends heavily not only on the approximate delay-distribution w_j, but also on the contact rate β, recovery rate ϕ_r, and mortality rate ϕ_d. These parameters are also, in general, not known. In order to obtain more robust estimates for the w_j through the optimization procedure defined above, we resort to a Monte Carlo method. The idea of this class of algorithms is to repeat random sample from the distribution function of the random variable, and perform several runs of the simulation, one for each sample. The solution will be given by the mean of all the obtained result. For a detailed description of Monte-Carlo methods we refer to (Pareschi and Toscani 2013).

Here, the Monte Carlo method consists of considering the β, ϕ_r, ϕ_d as independent random variables sampled from a Gaussian distributions. The mean of the respective distributions is obtained via an empirical parameter fitting, with the standard deviation of each distribution taken to be 5 % the value of the mean. This value was chosen based on sensitivity analyses. We then run the optimization procedure (28a)–(28c) a total of 1,000 times, corresponding to 1,000 different parameter samplings. We then identify frequency with which different w_j appear in the optimization, as well as the mean value of each w_j over the sampled trial.

We note that the result optimization procedure (16a)–(16d) depends on the choice of input data. In general, we consider the infection and deceased data to be the most reliable such data and is emphasized in our calculations. As the easiest quantity to measure, the deceased compartment is assumed the most accurate. Further, the as the rate of new infections is the driving point behind policy measures, it is considered the second most accurate measure for our purposes. However, one may modify (16a) to include other compartments, or different subsets of the compartment set.

In terms of implementation, there are several important points to note. We define the optimization procedure using the MATLAB routine `fmincon`. In order to solve the delay differential equation system (22)–(25), we use the MATLAB function `dde23`. We note that `dde23` uses adaptive grid points, and hence the temporal points of the solution obtained using `dde23` do not correspond to those in our dataset in general. To circumvent this, we use `deval` to evaluate the `dde23` solution at the desired temporal points.

Results

For our measured data, we consider the second wave of COVID-19 outbreak in Italy and Austria, as reported by the newspaper Il Sole 24 ORE (Coronavirus 2021) and AGES[1] respectively. The choice of these two Countries is mainly due to the availability of data as current number of positive cases and daily updates on recovered people, which was the most difficult data to find.

We consider the dates ranging from September 11, 2020 to February 7, 2021. This date range was chosen to represent a 150-day span, in which different governmental policies were enacted, widespread testing was available, and before large-scale population vaccination. This choice was intentional, as we wish to avoid as many confounding factors as possible.

Italy

We consider the following distributions for the parameters

$$\beta \sim N\left(.1131/n_0, .05(.1131/n_0)\right) \tag{29}$$

$$\phi_d \sim N\left(1/940, .05(1/940)\right) \tag{30}$$

$$\phi_r \sim N(1/24, .05(1/24)), \tag{31}$$

where $N(\mu, \varsigma)$ denotes a normal distribution with mean μ and standard distribution ς and n_0 denotes the initial living population. For $t > 73$,[2] where $t_0 = 0$, we replace β by $\beta/3$ to model the introduced government restrictions, designed to curb the virus spread. The values of ϕ_d and ϕ_r remain unchanged throughout the considered interval. The mean values of these parameters were obtained via a preliminary parameter tuning. For the delays, we considered 12 possible values, from $\sigma = 2$ days to $\sigma = 35$ days, with 3 days of spacing between the different days. We also performed some experiments (not shown) using different discretizations over the same range (for instance, 10 possible values) to show the robustness of the optimization procedure, which did not appreciably change the conclusions drawn.

To examine the possible differences in delays among the different compartments, we perform the optimization four times: once considering the i and d compartments together in the objective function, and for the i, r and d compartments, considering each compartment individually. For each ensemble, we report the minimum, maximum, and mean L^2 error over the simulation ensemble when compared to the measured data, as well as the mean value and frequency of the different weights w_j corresponding to the delays σ_j.
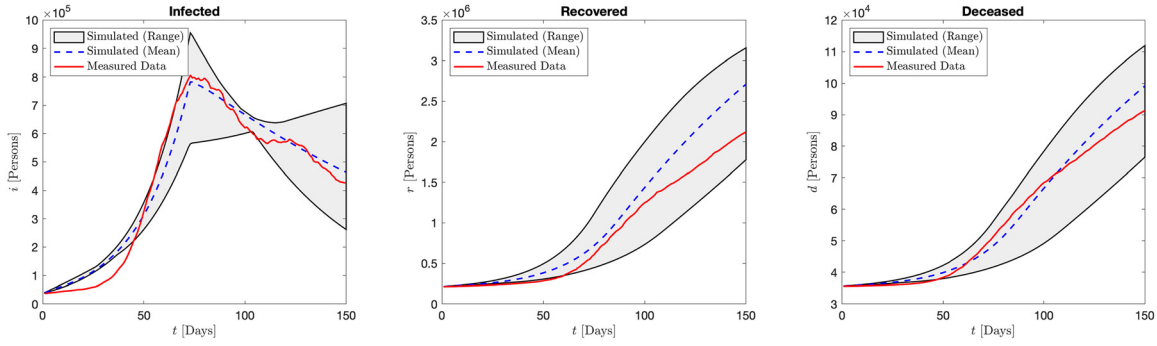
The aggregated results of the simulations in terms of the infected, recovered, and deceased compartments, when optimizing over the i and d compartments are shown in Figure 2. We see good qualitative agreement with the measured and simulated data across the three compartments of interest, with the major relevant trends being captured. In Table 2, we report the mean, minimum, and maximum of each error value computed over the simulation ensemble. We see particularly strong agreement in the susceptible and deceased compartments, reasonable agreement in the infected compartment (especially considering its rapid-changing trajectory), and somewhat worse performance in the recovered compartment. Given that the r compartment was not included in the optimization, this is unsurprising.

Turning our attention to the influence of time delays, we report the mean and frequency of the different weights w_j in Figure 3. We see a heavy concentration of weights values, for both frequency and mean, in the range of $\sigma = 8$ to $\sigma = 17$ days. This is consistent with what has been reported in general media, as the incubation

---

1 https://covid19-dashboard.ages.at/.

2 $t = 73$ coincides with the effective introduction of the curfew measures and limitations for bars, shops and restaurants and sports facilities and events.
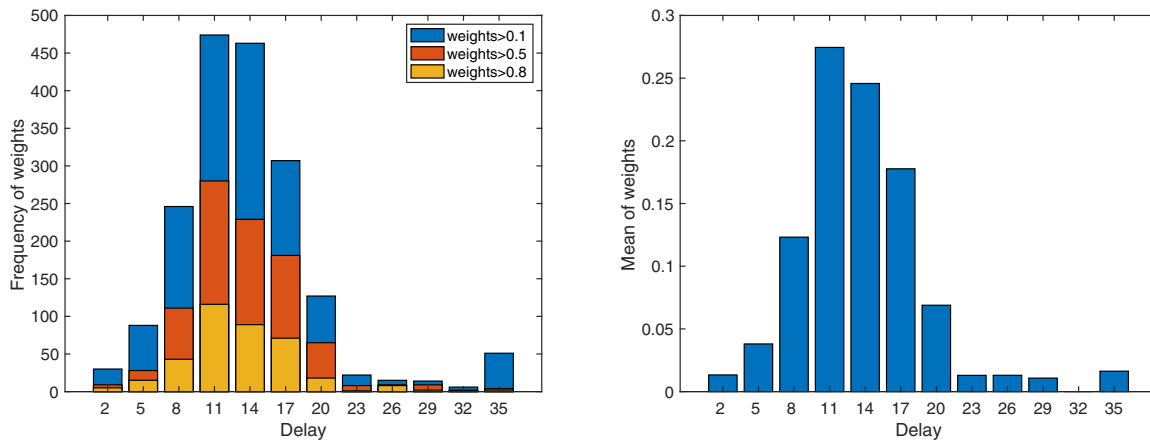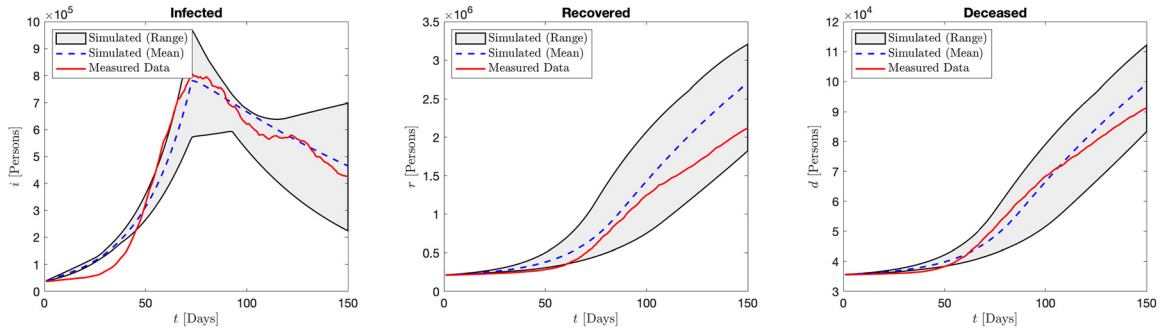
Figure 2: Italy: Aggregated results of the infected, recovered, and deceased compartments over the range of simulations when optimizing over the *i* and *d* compartments. We see good agreement across all compartments with the measured data, and particularly strong agreement (less than 10 % in L^2 norm) in the deceased compartment.

Table 2: Table of relative errors in L^2 norm against measured data for the different compartments over 1,000 simulations. The compartment towards which the optimization procedure is ran is shown in the columns.

		Optim. wrt *i, d*	Optim. wrt *i*	Optim. wrt *r*	Optim. wrt *d*
$\|s - \tilde{s}\|_{L^2} / \|\tilde{s}\|_{L^2}$	Mean	0.0051	0.0051	0.0025	0.0035
	Min	0.0022	0.0022	0.0016	0.0009
	Max	0.0104	0.0108	0.0041	0.0079
$\|i - \tilde{i}\|_{L^2} / \|\tilde{i}\|_{L^2}$	Mean	0.1101	0.1116	0.2347	0.1464
	Min	0.0798	0.0807	0.1580	0.0826
	Max	0.2757	0.2681	0.4672	0.4992
$\|r - \tilde{r}\|_{L^2} / \|\tilde{r}\|_{L^2}$	Mean	0.2602	0.2561	0.0586	0.1903
	Min	0.071	0.0650	0.0394	0.0440
	Max	0.5913	0.6054	0.1091	0.4769
$\|d - \tilde{d}\|_{L^2} / \|\tilde{d}\|_{L^2}$	Mean	0.0695	0.0684	0.0926	0.0430
	Min	0.0249	0.0266	0.0362	0.0170
	Max	0.2063	0.1659	0.1672	0.0864

Figure 3: Italy: Left: the different values of the w_j by frequency of occurrence when optimizing over *i* and *d*. We see a strong concentration of weights in the range of $\sigma = 8$ to $\sigma = 17$ days. Right: the mean values of the weights over the simulation ensemble. These results mirror those of the frequency plot, showing a concentration between 8 and 17 days. For both frequency and mean, $\sigma = 11$ appears as the most represented delay.
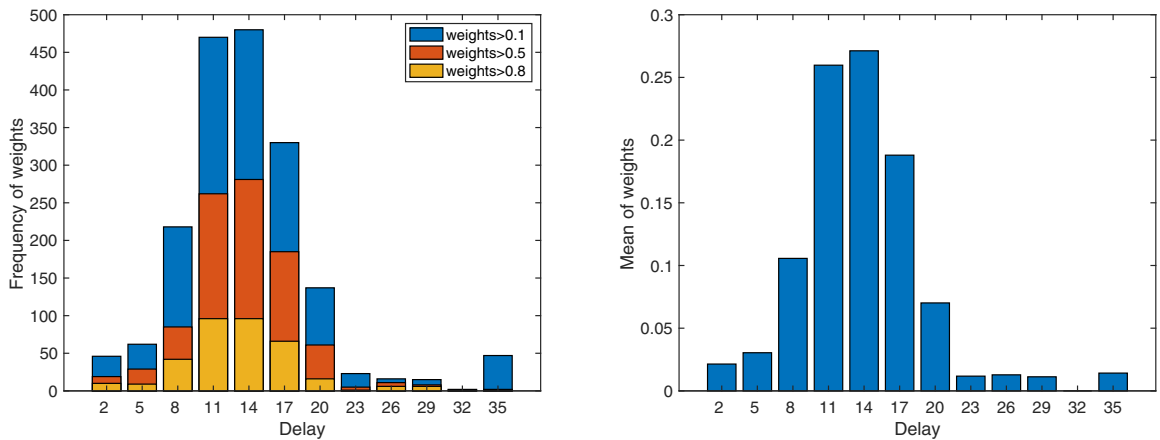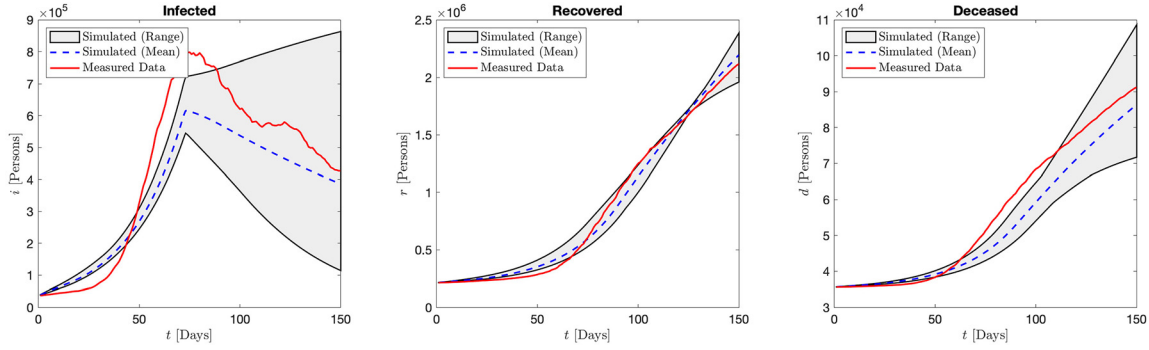
Figure 4: Italy: Aggregated results of the infected, recovered, and deceased compartments over the range of simulations when optimizing over the i compartment. We see good agreement across all compartments with the measured data, and particularly strong agreement (less than 10 % in L^2 norm) in the deceased compartment.

period of COVID is believed to have an average of 3–11 days (Byrne et al. 2020; Lauer et al. 2020; McAloon et al. 2020). When combined with natural delays coming from the testing and data reporting, a total time lag in the range of 8–17 days would appear consistent with this information.

We then repeat this experiment three more times: once each for the i, r, and d compartments considered individually in the optimization process. The error tables for the i, r and d compartments can be found in Table 2. The corresponding simulation outputs for each case are plotted in Figures 4, 6, and 8. The behavior is consistent with what one may intuitively expect; in each instance, the error behavior is minimized for the considered compartment, with the simulation variance greatly reduced for the isolated compartment. We note that one can observe from the relevant tables and figures that there is a much higher degree similarity between the i and d error behavior when compared to the r compartment, which appears to act more independently.

This is confirmed when looking at the frequency and mean distribution of the weights w_j in Figures 5, 7, and 9 for the i, r, and d compartments respectively. In the case of the i compartment, we see concentration between $\sigma = 8$ and $\sigma = 17$ days, similar to when i and d are considered jointly; however, for i alone, $\sigma = 14$ appears as the most dominant lag term, rather than $\sigma = 11$. The behavior appears similar but even more pronounced when considering d alone. In this instance $\sigma = 14$ is again the dominant delay term; however, the grouping is tightly clustered between $\sigma = 11$ and $\sigma = 17$, with terms outside this range showing very little influence.

Figure 5: Italy: Left: the different values of the w_j by frequency of occurrence when optimizing over i. We see a strong concentration of weights in the range of $\sigma = 8$ to $\sigma = 17$ days. Right: the mean values of the weights over the simulation ensemble. These results mirror those of the frequency plot, showing a concentration between 8 and 17 days. For both frequency and mean, $\sigma = 14$ appears as the most represented delay.

Figure 6: Italy: Aggregated results of the infected, recovered, and deceased compartments over the range of simulations when optimizing over the *r* compartment. We see good agreement in the *r* compartment; the agreement with the *d* compartment remains acceptable, though somewhat lesser than other optimization choices. The error in the infected compartment is notably higher. This different error behavior is reflected in the optimization results; when optimizing over *r*, the optimal time delays are different than when optimizing over *i* and *d*.
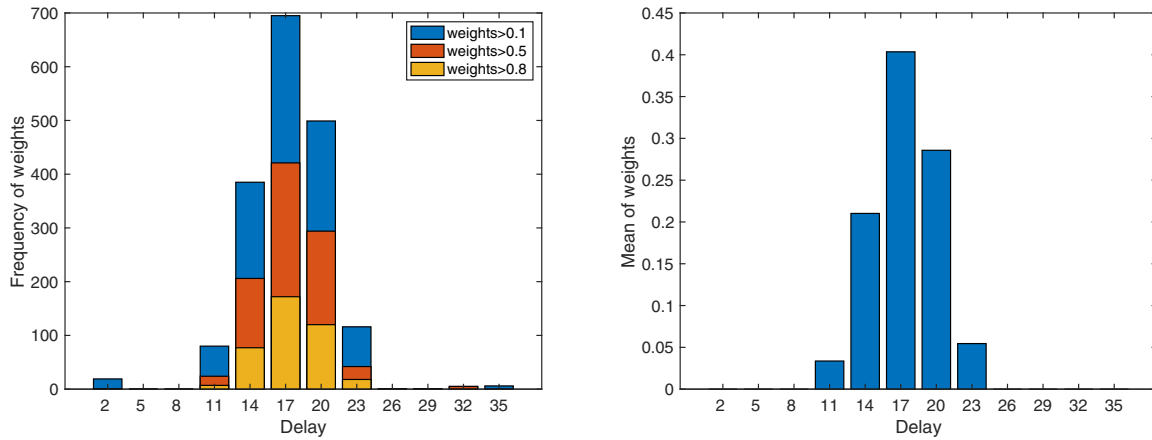
Figure 7: Italy: Left: the different values of the w_j by frequency of occurrence when optimizing over *r*. We see a strong concentration of weights in the range of $\sigma = 14$ to $\sigma = 20$ days. Right: the mean values of the weights over the simulation ensemble. These results mirror those of the frequency plot, showing a concentration between 14 and 20 days. For both frequency and mean, $\sigma = 17$ appears as the most represented delay. These results are less similar to those obtained when one optimizes over *i* and *d*, and suggests different time-delays present in the *r* compartment as compared to other compartments.

In contrast, when examining the case of isolated *r* in 7, we see a different distribution of delay terms. Indeed, in this instance, the dominant delay terms range from $\sigma = 14$ to $\sigma = 20$, and are very tightly concentrated within this range. This suggests a distinct and longer time scale when compared to the *i* and *d*, and may explain why the error behavior for the *r* compartment shows a distinctly different behavior in the plots and tables shown.
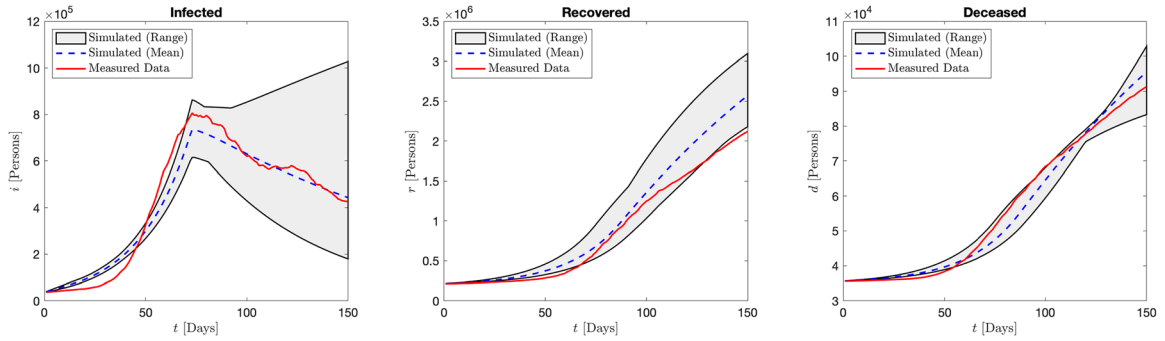
Austria

We consider the following distributions for the parameters

$$\beta \sim N\big(.6/n_0, .05(.6/n_0)\big) \tag{32}$$

$$\phi_d \sim N\big(1/550, .05(1/550)\big) \tag{33}$$

$$\phi_r \sim N(1/14, .05(1/14)), \tag{34}$$

Figure 8: Italy: Aggregated results of the infected, recovered, and deceased compartments over the range of simulations when optimizing over the *d* compartment. We see good agreement in the *d* and *r* compartments, with somewhat less close agreement with the *i* compartment.
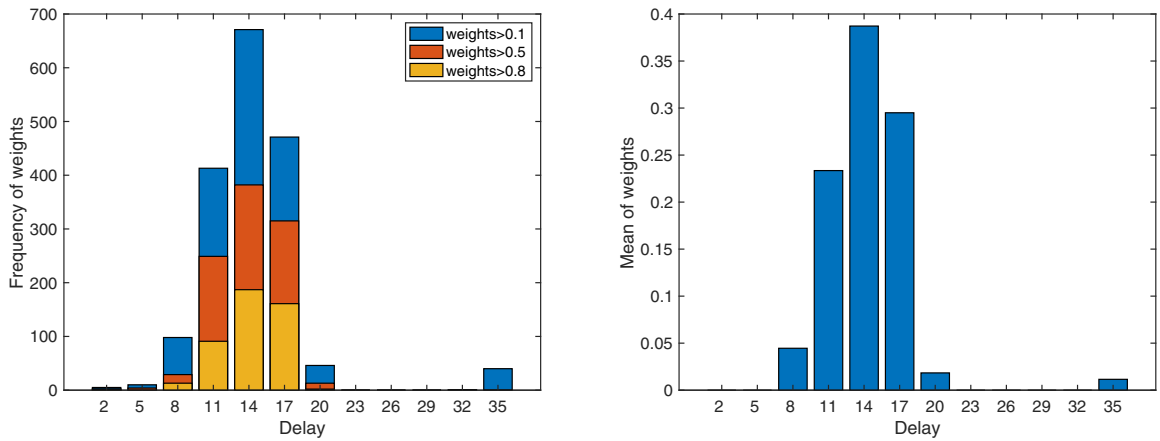
Figure 9: Italy: Left: the different values of the w_j by frequency of occurrence when optimizing over *d*. We see a tight concentration of σ in the range of 11–17 days. Right: the mean values of the weights over the simulation ensemble. These results mirror those of the frequency plot, showing a concentration between 11 and 16 days. For both frequency and mean, $\sigma = 14$ appears as the most represented delay.

where $N(\mu, \varsigma)$ denotes a normal distribution with mean μ and standard distribution ς and n_0 denotes the initial living population. Note that the parameters of the random variables are different from the ones related to Italy.

For $t > 66$, where $t_0 = 0$ we replace β by $\beta/4$ to model the introduced government restrictions, namely the second hard lockdown in Austria.

As in the case of Italy, the optimization is performed four times: once considering the *i* and *d* compartments together in the objective function, and for the *i*, *r* and *d* compartments, considering each compartment individually. For each ensemble, we show, as before, the aggregate results of the infected, recovered and deceased compartments, the mean and the frequency of the weights w_j of the delay.

Optimizing with respect to *i* and *d*, Figure 10, leads to a good agreement across all compartments with the measured data with a particular strong agreement in the deceased compartment, as in Figure 2.

Concerning the weights, shown in Figure 11, both the frequency and the mean of σ are concentrated between 8 and 20 days, similar to Figure 3.

For the optimization with respect to the infected (Figures 12 and 13) and recovered (Figures 14, 15) compartments considered individually, we see a good agreement with the results obtained in the case of Italy, Figure 4, 5 and Figure 6, 7 respectively.
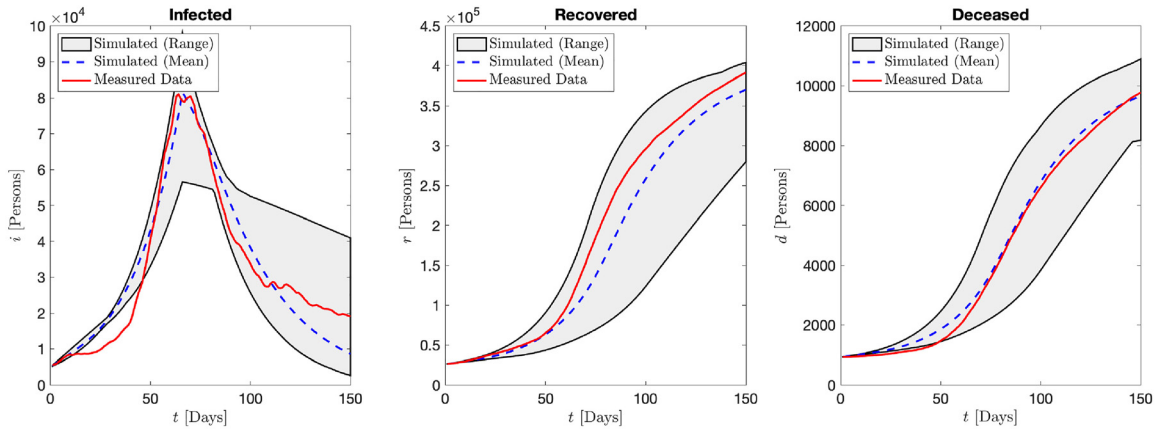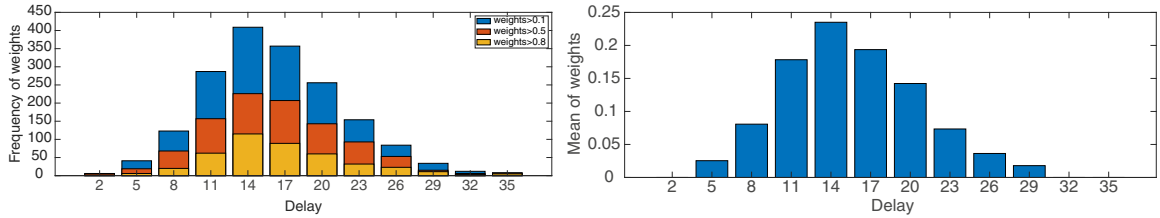
Figure 10: Austria: Aggregated results of the infected, recovered, and deceased compartments over the range of simulations when optimizing over the i and d compartment. We see good agreement across all compartments with the measured data, and particularly strong agreement in the deceased compartment.
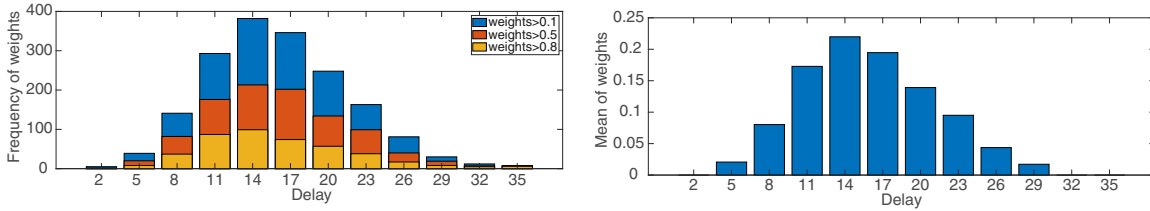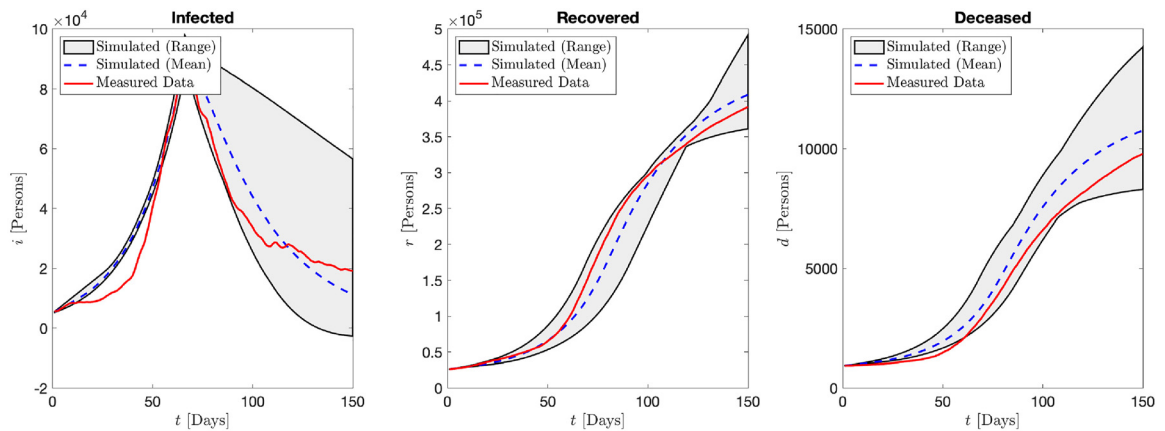

Figure 11: Austria: Left: the different values of the w_j by frequency of occurrence when optimizing over i and d. We see a concentration of σ in the range of 8–20 days. Right: the mean values of the weights over the simulation ensemble. These results mirror those of the frequency plot, showing a concentration between 11 and 20 days. For both frequency and mean, $\sigma = 14$ appears as the most represented delay.
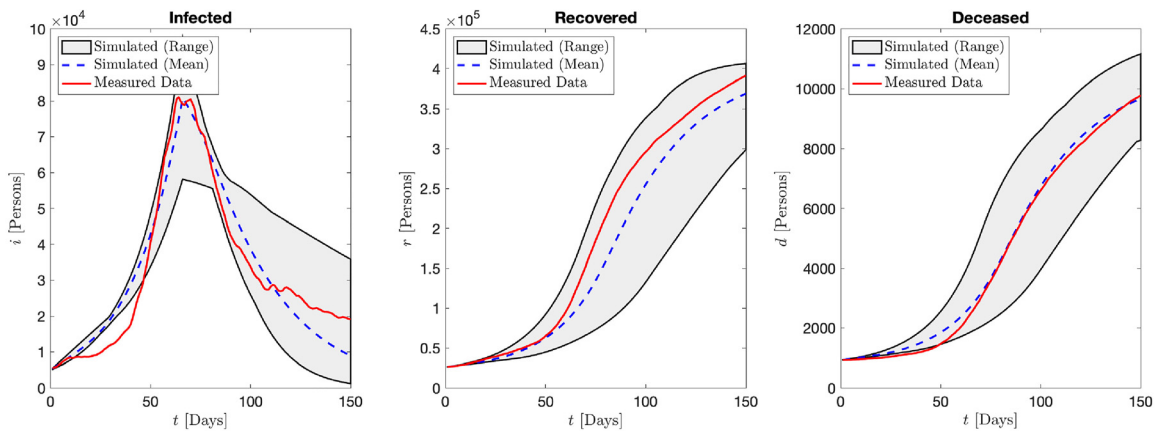

Figure 12: Austria: Aggregated results of the infected, recovered, and deceased compartments over the range of simulations when optimizing over the i compartment. We see good agreement across all compartments with the measured data, and particularly strong agreement in the deceased compartment.

Figure 13: Austria: Left: the different values of the w_j by frequency of occurrence when optimizing over i. We see more σ values in the range of 11–23 days. Right: the mean values of the weights over the simulation ensemble. These results mirror those of the frequency plot, showing a concentration between 11 and 20 days. For both frequency and mean, $\sigma = 14$ appears as the most represented delay.
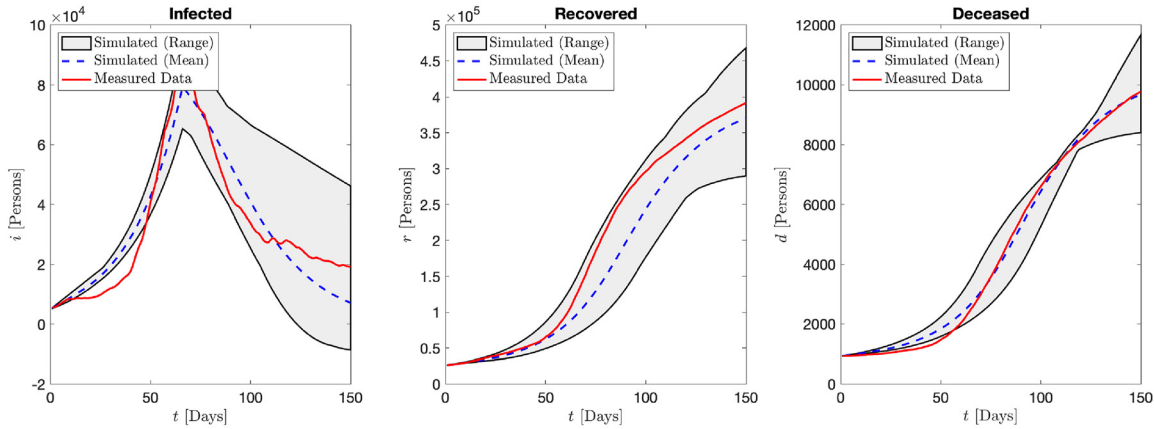

Figure 14: Austria: Aggregated results of the infected, recovered, and deceased compartments over the range of simulations when optimizing over the r compartment. We see good agreement in the r compartment; the agreement with the d compartment remains acceptable, though somewhat lower r than other optimization choices. The different error behavior is reflected in the optimization results; when optimizing over r, the optimal time delays are different than when optimizing over i and d.
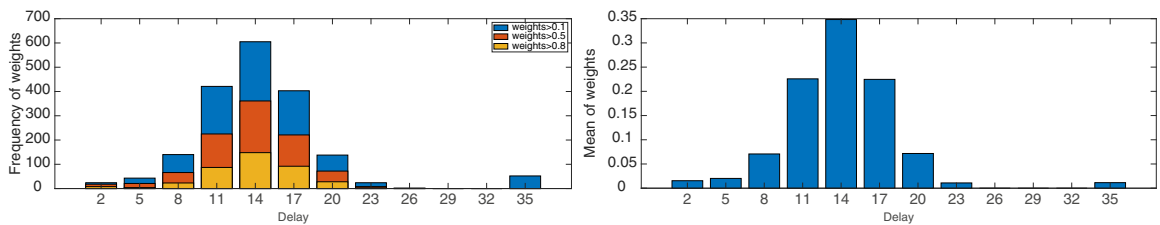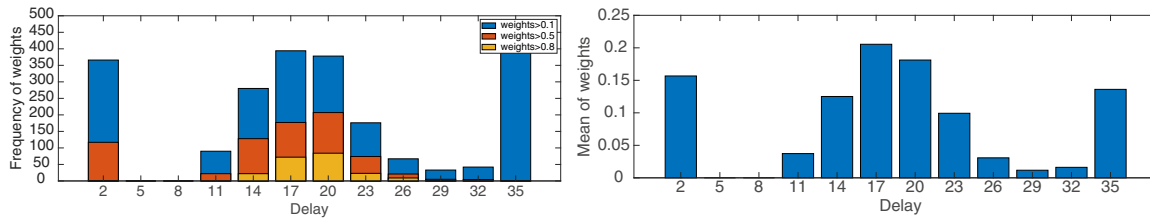

Figure 15: Austria: Left: the different values of the w_j by frequency of occurrence when optimizing over r. We see a tight concentration of σ in the range of 11–20 days. Right: the mean values of the weights over the simulation ensemble. These results mirror those of the frequency plot, showing a concentration between 11 and 16 days. For both frequency and mean, $\sigma = 14$ appears as the most represented delay These results are less similar to those obtained when one optimizes over i and d, and suggests different time-delays present in the r compartment as compared to other compartments.

However, this can not be said in the case of the deceased compartments (Figures 16 and 17). Here the two cases are quite different, see Figures 9 and 17. This is possibly explained by differences in how the various data are reported in the different countries causing delay distributions to differ; however, we are not sure as to the exact cause.

Figure 16: Austria: Aggregated results of the infected, recovered, and deceased compartments over the range of simulations when optimizing over the *d* compartment. We see good agreement in the *d* compartment, with somewhat less close agreement with the *r* and *i* compartments.

Figure 17: Austria: Left: the different values of the w_j by frequency of occurrence when optimizing over *d*. We see a concentration of higher weights for σ in the range of 14–23 days. Right: the mean values of the weights over the simulation ensemble. These results mirror those of the frequency plot, showing a concentration between 14 and 23 days, even though the presence of sigma values also in the first and the last time-bars is notable. According to the mean value, $\sigma = 20$ appears as the most represented delay.

## Sensitivity of delay distribution to discretization

In order to ensure the validity of our method, it is important to ensure that the results are insensitive to the discretization of $[\sigma_{min}, \sigma_{max})$. We perform a brief sensitivity analysis, considering the data from Austria described in the previous subsection. We consider two additional simulations, optimized over the *i* and *d* compartments (corresponding to Figures 10 and 11 above). The setup is identical to the previous simulations, except we vary the resolution of the delay discretization:

1. *Coarse-resolution case*: we discretize the interval [2, 35) with $\Delta = 5$ days.
2. *Fine-resolution case:* we discretize the interval [2, 35) with $\Delta = 1$ day.

We remind the reader that in the results previously shown, $\Delta = 3$ days. We show the results for the coarse-resolution case in Figures 18 and 19 and the fine-resolution case in Figures 20 and 21.

In both cases, we find that the solution of the system is similar to the one obtained in the preceding section. Similarly, the delay-distributions are also similar, both qualitatively and quantitatively, compared to the previous results. From our analysis, it appears that the resolution of the delay discretization does not result in significant changes to the qualitative or quantitative profile or the identified delay distribution.

We note that increasing the resolution of the delay discretization significantly increases the computational time necessary to complete the Monte Carlo analysis. On a MacBook Pro with Processor 2 GHz, Quad-Core, Intel Core i5, with $\Delta = 5$ (and hence 7 wt), each individual simulation required an average of 1.178 s; the average simulation instead required 3.435 s for $\Delta = 3$ (12 wt) and 48.994 s for $\Delta = 1$ (34 wt). Based on this evidence, it appears that one should increase the resolution of the delay term only if absolutely necessary to understand the underlying distribution, as the necessary computational cost appears to scale quadratically with the number of
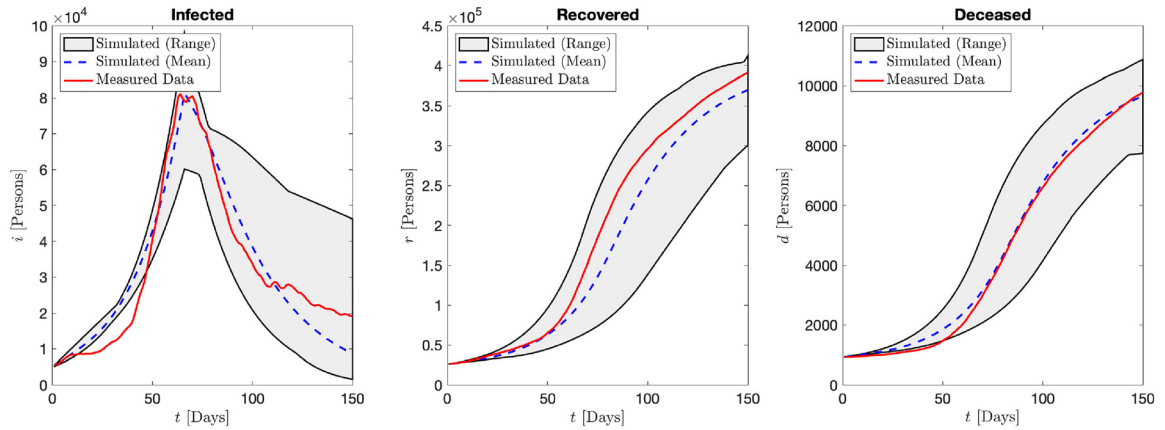
Figure 18: Austria: Aggregated results of the infected, recovered, and deceased compartments over the range of simulations when optimizing over the i and d compartment using a coarser discretization of the delay term. The results are similar, qualitatively and quantitatively, to those obtained with a finer discretization.
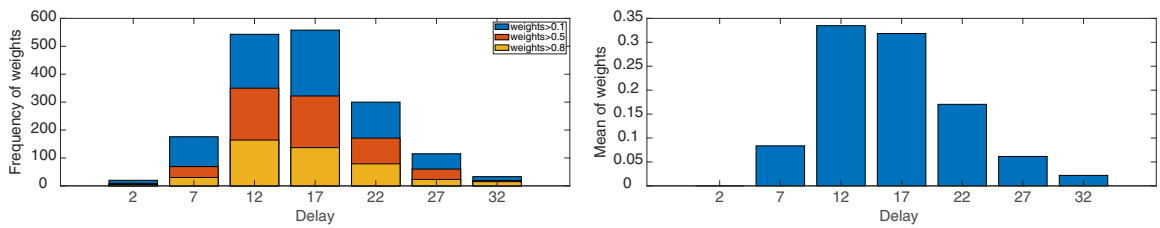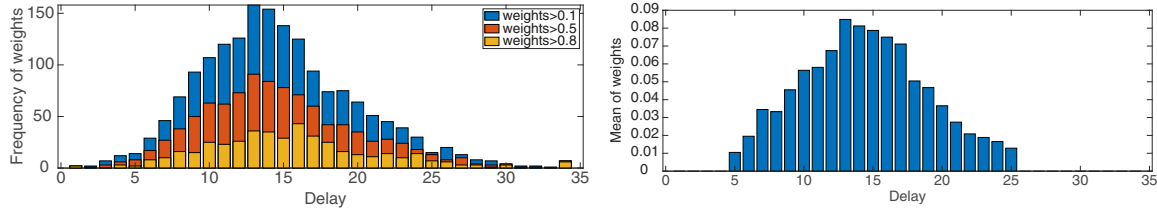

Figure 19: Austria: Left: the different values of the w_j by frequency of occurrence when optimizing over i and d when using a coarser discretization of the delay term. Right: average value of each weight over the simulation ensemble. The results are similar, qualitatively and quantitatively, to those obtained with a finer discretization.
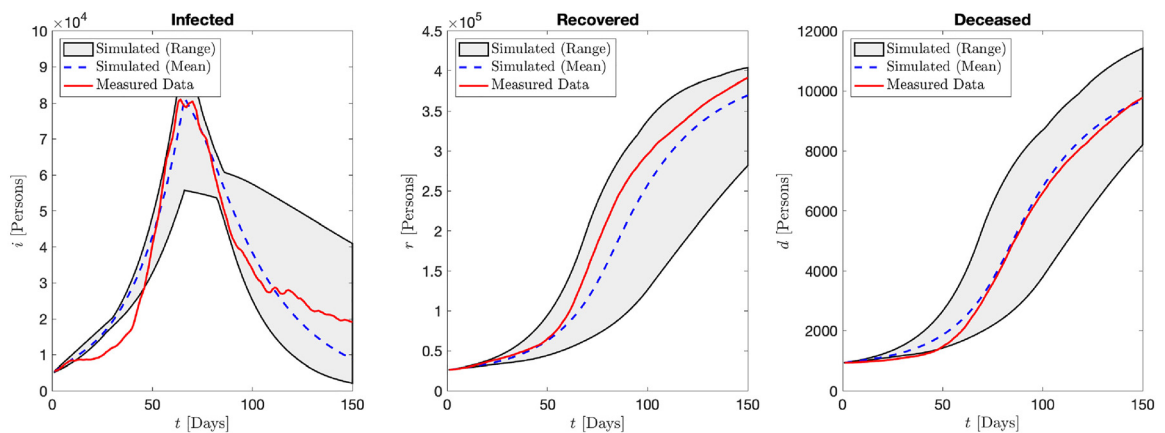
Figure 20: Austria: Aggregated results of the infected, recovered, and deceased compartments over the range of simulations when optimizing over the i and d compartment using a fine discretization of the delay term. The results are similar, qualitatively and quantitatively, to those obtained with a coarser discretization.

Figure 21: Austria: Left: the different values of the w_j by frequency of occurrence when optimizing over i and d when using a fine discretization of the delay term. Right: average value of each weight over the simulation ensemble. The results are similar, qualitatively and quantitatively, to those obtained with a coarser discretization.

weights. If a coarser resolution will suffice (as appears to be the case in the present example), the significantly shorter computational time make this a more attractive choice. The authors acknowledge, however, that the optimization algorithm employed in the present work are not necessarily the most efficient choice, and that a more carefully chosen optimization algorithm may provide better scaling behavior for increasing numbers of weights.

Key-points

In all, we can conclude three key takeaway points from these results:
1. The time delays present in COVID-19 data range from $\sigma = 8$ to $\sigma = 20$ days, with the most likely value for σ_j the i and d compartments in the range of $\sigma = 11$ to $\sigma = 14$ days, and appear to follow a normal or log-normal distribution.
2. The scale of the time delays across the compartments is not uniform. It is similar for new infections and for deceased persons, but occurs over a somewhat longer time scale for recovered individuals. This is not surprising due to the different recovery rate and the inherent difficulty of measuring this compartment, given delays in testing and in the reporting of results. This also suggests that the multiple delays present in the data are somewhat independent, and likely follow different distributions.
3. The time scales revealed by the optimization procedure shown here are longer than most estimates of the incubation period, suggesting that other delays are potentially present in the data, and a period of 14–17 days should be considered when analyzing the effects of interventions.
4. The results remain quantitatively and qualitatively similar when using both finer and coarser discretizations of the delay term, suggesting that the results obtained are not sensitive to this quantity.

Conclusions and future directions

In the present work, we have attempted to identify time delays present in COVID-19 data by means of newly-introduced method. This method consists of discretizing k possible time delays $\sigma_j, j = 1:k$ and expressing these terms as a weighted sum $\sum_{j=1}^{k} w_j i(t - \sigma_j)$. This formulation then yields an optimization problem which is linear in the weights w_j. After a preliminary parameter fitting, we then solve an ensemble of optimization problems, finding the weights w_j which minimize the difference between simulated and measured data over different samplings of parameter values. We recover a distribution of delay terms for the different variables, allowing us to recover the inherent time-delays within the measured data.

We tested this approach on COVID-19 data in both Italy and Austria, and found that the time-delays differ among different data points, but generally speaking range between 8 and 17 days for i and d, with the most likely value between 11 and 14 days, and over a somewhat longer time scale (14–20 days) for the recovered compartment. This is consistent with our expectations given the incubation period of the disease and the natural time delays one may expect from the testing and data collection process. Our results confirm that proper evaluation

of a given intervention for COVID-19 should allow for a sufficiently long time window to pass before conclusions are made. All of the identified distributions followed a bell-like distribution resembling a normal or lognormal distribution, and a sensitivity analysis suggests that this profile is independent of the discretization of the delay term.

There are several important directions for future work. Given the increased importance of waning vaccine efficacy over time, a similar analysis given vaccine data and data on breakthrough infections could be helpful in identifying similar such time-delay trends for waning immunity. While we have elected to avoid vaccines in the current work, preferring to establish a proof of concept on a more simplified problem, the approach offered is extremely general and may be applied to these important problems. Outside of COVID-19, time delays are an important part of other types of epidemiological modeling; similar methods could be used, for example, to improve back-calculation methods for estimating HIV incidence, in which incidence is estimated by reconstructing the distribution of the time-delays between infection and diagnosis (Hall et al. 2017; Song et al. 2017; Viguerie et al. 2022b; Xia et al. 2020). Further, given the generality of the proposed optimization procedure, it may have wide applications outside of epidemiology, and may be applied to any model where time delays are present, including applications in traffic modeling, pedestrian flow, remote sensing, and other fields. From the theoretical and methodological standpoint, an extension of the framework shown here for problems in which the delays are state-dependent is also an important direction of future research. Finally, we note that the general idea shown, in which general sojourn time-distributions are approximated by a weighted sum of delay-differential equations, may be applied to approximate systems featuring more complex delay-distributions.

Author contribution: All authors have accepted responsibility for the entire content of this manuscript and approved its submission.
Competing interests: Authors state no conflict of interest.
Informed consent: Informed consent was obtained from all individuals included in this study.
Ethical approval: The local Institutional Review Board deemed the study exempt from review.

References

Albi, G., G. Bertaglia, W. Boscheri, G. Dimarco, L. Pareschi, G. Toscani, and M. Zanella. 2022. "Kinetic Modelling of Epidemic Dynamics: Social Contacts, Control with Uncertain Data, and Multiscale Spatial Dynamics." In *Predicting Pandemics in a Globally Connected World, Volume 1: Toward a Multiscale, Multidisciplinary Framework Through Modeling and Simulation*, 43−108. Birkhäuser: Springer.

Barros, G. F., M. Grave, A. Viguerie, A. Reali, and A. L. G. A. Coutinho. 2022. "Dynamic Mode Decomposition in Adaptive Mesh Refinement and Coarsening Simulations." *Engineering with Computers* 38 (5): 4241−68.

Bastos, L. S., T. Economou, M. F. C. Gomes, D. A. M. Villela, F. C. Coelho, O. G. Cruz, O. Stoner, T. Bailey, and C. T. Codeço. 2019. "A Modelling Approach for Correcting Reporting Delays in Disease Surveillance Data." *Statistics in Medicine* 38 (22): 4363−77.

Bertaglia, G., and L. Pareschi. 2021. "Hyperbolic Compartmental Models for Epidemic Spread on Networks with Uncertain Data: Application to the Emergence of COVID-19 in Italy." *Mathematical Models and Methods in Applied Sciences* 31 (12): 2495−531.

Bertaglia, G., and L. Pareschi. 2021. "Hyperbolic Models for the Spread of Epidemics on Networks: Kinetic Description and Numerical Methods." *ESAIM: Mathematical Modelling and Numerical Analysis* 55 (2): 381−407.

Bertrand, F., and E. Pirch. 2021. "Least-Squares Finite Element Method for a Meso-Scale Model of the Spread of COVID-19." *Computation* 9 (2): 18.

Bhouri, M. A., F. S. Costabal, H. Wang, K. Linka, M. Peirlinck, E. Kuhl, and P. Perdikaris. 2021. "Covid-19 Dynamics across the US: A Deep Learning Study of Human Mobility and Social Behavior." *Computer Methods in Applied Mechanics and Engineering* 382: 113891.

Brauer, F., and C. Castillo-Chavez. 2012. *Mathematical Models in Population Biology and Epidemiology*, 2. New York, NY: Springer.

Breda, D., O. Diekmann, W. F. De Graaf, A. Pugliese, and R. Vermiglio. 2012. "On the Formulation of Epidemic Models (An Appraisal of Kermack and McKendrick)." *Journal of Biological Dynamics* 6: 103−17.

Buonomo, B., A. d'Onofrio, and D. Lacitignola. 2008. "Global Stability of an SIR Epidemic Model with Information Dependent Vaccination." *Mathematical Biosciences* 216 (1): 9−16.

Byrne, A. W., D. McEvoy, A. B. Collins, K. Hunt, M. Casey, A. Barber, F. Butler, J. Griffin, E. A. Lane, C. McAloon, K. O'Brien, P. Wall, K. A. Walsh, and S. J. More. 2020. "Inferred Duration of Infectious Period of SARS-CoV-2: Rapid Scoping Review and Analysis of Available Evidence for Asymptomatic and Symptomatic COVID-19 Cases." *BMJ Open* 10 (8): e039856.

Calafiore, G. C., C. Novara, and C. Possieri. 2020. "A Modified SIR Model for the COVID-19 Contagion in Italy." In *2020 59th IEEE Conference on Decision and Control (CDC)*, 3889−94. IEEE.

Choi, S., and M. Ki. 2020. "Estimating the Reproductive Number and the Outbreak Size of COVID-19 in Korea." *Epidemiology and Health* 42: 1−10.

2021, *Coronavirus in Italy: Updated Map and Case Count*. https://lab24.ilsole24ore.com/coronavirus/en/ (accessed November 20, 2021).

Dell'Anna, L. 2020. "Solvable Delay Model for Epidemic Spreading: The Case of Covid-19 in Italy." *Scientific Reports* 10 (1): 1−10.

Devipriya, R., S. Dhamodharavadhani, and S. Selvi. 2021. "SEIR Model for COVID-19 Epidemic Using Delay Differential Equation." *Journal of Physics: Conference Series* 1767: 012005.

Ferguson, N., D. Laydon, G. G. Nedjati, N. Imai, K. Ainslie, M. Baguelin, S. Bhatia, A. Boonyasiri, P. Z. Cucunuba, G. Cuomo-Dannenburg, A. Dighe, I. Dorigatti, H. Fu, K. Gaythorpe, W. Green, A. Hamlet, W. Hinsley, L. C. Okell, S. van Elsland, H. Thompson, R. Verity, E. Volz, H. Wang, Y. Wang, P. G. T. Walker, C. Walters, P. Winskill, C. Whittaker, C. A. Donnelly, S. Riley, and A. C. Ghani. 2020. *Impact of Non-pharmaceutical Interventions (NPIs) to Reduce COVID19 Mortality and Healthcare Demand*. *Technical Report*. London: Imperial College London.

Forde, J. E. 2005. *Delay Differential Equation Models in Mathematical Biology*. Ann Arbor: University of Michigan.

Gatto, M., E. Bertuzzo, L. Mari, S. Miccoli, L. Carraro, R. Casagrandi, and A. Rinaldo. 2020. "Spread and Dynamics of the COVID-19 Epidemic in Italy: Effects of Emergency Containment Measures." *Proceedings of the National Academy of Sciences* 117 (19): 10484−91.

Grave, M., and A. L. G. A. Coutinho. 2021. "Adaptive Mesh Refinement and Coarsening for Diffusion−Reaction Epidemiological Models." *Computational Mechanics* 67 (4): 1177−99.

Grave, M., A. Viguerie, G. F. Barros, A. Reali, and A. L. G. A. Coutinho. 2021. "Assessing the Spatio-Temporal Spread of COVID-19 via Compartmental Models with Diffusion in Italy, USA, and Brazil." *Archives of Computational Methods in Engineering* 28 (6): 4205−23.

Guglielmi, N., E. Iacomini, and A. Viguerie. 2022. "Delay Differential Equations for the Spatially Resolved Simulation of Epidemics with Specific Application to COVID-19." *Mathematical Methods in the Applied Sciences* 45 (8): 4752−71.

Hall, H. I., R. Song, T. Tang, Q. An, J. Prejean, P. Dietz, A. L. Hernandez, T. Green, N. Harris, E. McCray, and J. Mermin. 2017. "HIV Trends in the United States: Diagnoses and Estimated Incidence." *JMIR Public Health and Surveillance* 3 (1): e7051.

Iannelli, M., and A. Pugliese. 2015. *An Introduction to Mathematical Population Dynamics: Along the Trail of Volterra and Lotka*, 79. Edinburgh: Springer.

Ivorra, B., M. R. Ferrández, M. Vela-Pérez, and A. M. Ramos. 2020. "Mathematical Modeling of the Spread of the Coronavirus Disease 2019 (COVID-19) Taking into Account the Undetected Infections. The Case of China." *Communications in Nonlinear Science and Numerical Simulation* 88: 105303.

Jha, P. K., L. Cao, and J. T. Oden. 2020. "Bayesian-based Predictions of Covid-19 Evolution in Texas Using Multispecies Mixture-Theoretic Continuum Models." *Computational Mechanics* 66 (5): 1055−68.

Kumar, P., and V. S. Erturk. 2020. "The Analysis of a Time Delay Fractional Covid-19 Model via Caputo Type Fractional Derivative." *Mathematical Methods in the Applied Sciences* 46: 7613−8429.

Lauer, S. A., K. H. Grantz, Q. Bi, F. K. Jones, Q. Zheng, H. R. Meredith, A. S. Azman, N. G. Reich, and J. Lessler. 2020. "The Incubation Period of Coronavirus Disease 2019 (COVID-19) from Publicly Reported Confirmed Cases: Estimation and Application." *Annals of Internal Medicine* 172 (9): 577−82.

Linka, K., P. Rahman, A. Goriely, and E. Kuhl. 2020. "Is it Safe to Lift Covid-19 Travel Bans? the Newfoundland Story." *Computational Mechanics* 66 (5): 1081−92.

McAloon, C., Á. Collins, K. Hunt, A. Barber, A. W. Byrne, F. Butler, M. Casey, J. Griffin, E. Lane, D. McEvoy, P. Wall, M. Green, L. O'Grady, and S. J. More. 2020. "Incubation Period of COVID-19: A Rapid Systematic Review and Meta-Analysis of Observational Research." *BMJ Open* 10 (8): e039652.

Murray, J. D. 2007. *Mathematical Biology I: An Introduction*, 3rd ed.. New York: Springer.

Ogilvy Kermack, W., and A. G. McKendrick. 1927. "A Contribution to the Mathematical Theory of Epidemics." *Proceedings of the Royal Society of London - Series A: Containing Papers of a Mathematical and Physical Character* 115 (772): 700−21.

Pareschi, L., and G. Toscani. 2013. *Interacting Multiagent Systems: Kinetic Equations and Monte Carlo Methods*. New York: OUP Oxford.

Parolini, N., L. Dede, P. F. Antonietti, G. Ardenghi, A. Manzoni, E. Miglio, A. Pugliese, M. Verani, and A. Quarteroni. 2021 "SUIHTER: A New Mathematical Model for COVID-19. Application to the Analysis of the Second Epidemic Outbreak in Italy." *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences A.* (477).

Piccolomini, E. L., and F. Zama. 2020. "Monitoring Italian COVID-19 Spread by a Forced SEIRD Model." *PLoS One* 15 (8): e0237417.

Remuzzi, A., and G. Remuzzi. 2020. "COVID-19 and Italy: What Next?" *The Lancet* 395 (10231): 1225−8.

Sarnaglia, A. J. Q., B. Zamprogno, F. A. F. Molinares, L. G. de Godoi, and N. A. J. Monroy. 2021. "Correcting Notification Delay and Forecasting of COVID-19 Data." *Journal of Mathematical Analysis and Applications* 514 (2): 125202.

Song, R., H. I. Hall, T. A. Green, C. L. Szwarcwald, and N. Pantazis. 2017. "Using CD4 Data to Estimate HIV Incidence, Prevalence, and Percent of Undiagnosed Infections in the United States." *JAIDS Journal of Acquired Immune Deficiency Syndromes* 74 (1): 3−9.

Takeuchi, Y., W. Ma, and E. Beretta. 2000. "Global Asymptotic Properties of a Delay SIR Epidemic Model with Finite Incubation Times." *Nonlinear Analysis: Theory, Methods & Applications* 42 (6): 931−47.

Viguerie, A., A. Veneziani, G. Lorenzo, D. Baroli, N. Aretz-Nellesen, A. Patton, T. E. Yankeelov, A. Reali, T. J. R. Hughes, and F. Auricchio. 2020. "Diffusion−Reaction Compartmental Models Formulated in a Continuum Mechanics Framework: Application to Covid-19, Mathematical Analysis, and Numerical Study." *Computational Mechanics* 66 (5): 1131−52.

Viguerie, A., G. Lorenzo, F. Auricchio, D. Baroli, T. J. R. Hughes, A. Patton, A. Reali, T. E. Yankeelov, and A. Veneziani. 2021. "Simulating the Spread of COVID-19 via a Spatially-Resolved Susceptible−Exposed−Infected−Recovered−Deceased (SEIRD) Model with Heterogeneous Diffusion." *Applied Mathematics Letters* 111: 106617.

Viguerie, A., G. F. Barros, M. Grave, A. Reali, and A. L. G. A. Coutinho. 2022a. "Coupled and Uncoupled Dynamic Mode Decomposition in Multi-Compartmental Systems with Applications to Epidemiological and Additive Manufacturing Problems." *Computer Methods in Applied Mechanics and Engineering* 391: 114600.

Viguerie, A., R. Song, A. S. Johnson, C. M. Lyles, A. Hernandez, and P. G. Farnham. 2022b. "Isolating the Effect of Covid-19 Related Disruptions on HIV Diagnoses in the United States in 2020." *JAIDS Journal of Acquired Immune Deficiency Syndromes* 92 (4): 10−1097.

Wang, Z., X. Zhang, G. H. Teichert, M. Carrasco-Teja, and K. Garikipati. 2020. "System Inference for the Spatio-Temporal Evolution of Infectious Diseases: Michigan in the Time of COVID-19." *Computational Mechanics* 66 (5): 1153−76.

Xia, Q., S. Lim, B. Wu, L. A. Forgione, A. Crossa, A. B. Balaji, S. L. Braunstein, D. C. Daskalakis, B. W. Tsoi, G. Harriman, L. V. Torian, and R. Song. 2020. "Estimating the Probability of Diagnosis within 1 Year of HIV Acquisition." *AIDS* 34 (7): 1075−80.

Zohdi, T. I. 2020. "An Agent-Based Computational Framework for Simulation of Global Pandemic and Social Response on Planet X." *Computational Mechanics* 66 (5): 1195−209.