A novel deterministic forecast model for COVID-19 epidemic based on a single ordinary integro-differential equation

Felix Köhler-Rieper
Goethe-Universität Frankfurt, Germany
(felixkoehler0049@gmx.de)
https://orcid.org/0000-0002-3099-1197

Claudius Röhl
Universität Leipzig, Germany
https://orcid.org/0000-0001-9598-5639

Enrico De Micheli
IBF - Consiglio Nazionale Delle Ricerche, Genova, Italy
https://orcid.org/0000-0002-8058-045X

April 29, 2020

Abstract

In this paper, we present a new approach to deterministic modelling of COVID-19 epidemic. Our model dynamics is expressed by a single prognostic variable in an integro-differential equation. All unknown parameters are described with a single, time-dependent variable $\kappa(t)$. We show that our model has similarities to classic compartmental models, such as SIR, and that the variable $\kappa(t)$ can be interpreted as the effective reproduction number R_{eff}. The advantages of our approach are the simplicity of having only one equation, the numerical stability due to an integral formulation and the reliability because the model is formulated with the most trustable statistical data variable: the number of cumulative diagnosed positive cases

of COVID-19. Once this dynamic variable is calculated, other non-dynamic variables, such as the number of heavy cases (hospital beds), the number of intensive-care cases (ICUs) and the fatalities, can be derived from it using a similarly stable, integral approach. The formulation with a single equation allows us to calculate from real data the values of the effective reproduction number, which can then be fitted. The extrapolated values of κ can be used in the model to make a reliable forecasts, though under the assumption that measures for reducing infections are maintained. We have applied our model to more than 15 countries and results are available via a web-based platform [6]. In this paper we focus on the data for two exemplary countries, Italy and Germany, and show that the model is capable of reproducing the course of the epidemic in the past and forecasting its course for a period of two to three weeks with a reasonable numerical stability.

1 Introduction

Our primary aim is to set up a deterministic model that can be easily tuned with available data in order to make numerically stable forecasts. We found that existing methods are not well-suited to reach this goal.

Empirical top-down modelling, *i.e.*, approaches that start from data and make prognoses, mostly ignore underlying dynamics. The easiest approach is curve fitting of available data. In [1] the number of cumulative diagnosed positive COVID-19 cases $P(t)$ was assumed to be an error function. This is true if the number of daily new cases $P'(t)$ can be described by a Gaussian distribution. As we will show, a symmetric distribution function $P'(t)$ corresponds to an effective reproduction number that converges rapidly to zero. This might be true for China data. In Italy and Germany we observed a final value for R_{eff} between 0.6 to 0.8, leading to an asymmetric function $P'(t)$ with a long tail. Although the peak date has been predicted well in [1], the predicted total cases and fatalities differ by more than 30%.

Current deterministic models were developed with the aim of simulating possible scenarios and showing the effect of containment and mitigation measurements. They are "bottom-up" in the sense that they are based on the knowledge of epidemiological parameters, such as the basic reproduction number R_0 or the time between contacts T_c, to name a few. Three reasons make it difficult to set up these complex models for forecasting:

- the epidemiological parameters are unknown and change in time;

- for most of the compartmental model variables, such as susceptible, exposed, infected or removed individuals, the availability of surveillance data is limited;
- model tuning requires fitting many variables simultaneously - making it difficult to find an optimum.

In [10] the classical SIR model has been applied to Italy, dividing the country into three parts: north, centre and south. The problem they face, in our opinion, is that the official number of infected individuals I contains people, who are officially not cured. But in Italy people enter the statistics as cured, when they have been tested twice or even three times in a week's distance as negative. Thus, they remain "infected" for too long - seen from a dynamic point of view. The model cannot capture this feature appropriately and to keep track of the statistical data, has to be re-tuned within days.

The German Robert-Koch Institut (RKI) uses an extended SEIR model to show various scenarios for the course of the COVID-19 epidemic in Germany by applying different seasonality of the epidemic and immunity of the population [12].

Another Italian team has set up a model with eight prognostic variables, SIDARTHE [11], taking also into account asymptomatic cases and detection issues. Again, these efforts allow precise simulation of scenarios but are difficult to be set up with real data to make forecasts. The comparison with real data looks good but is restricted to the initial period of the epidemic when the case numbers grew simply exponential.

In [15] statistical parameters are obtained to feed parametric models, though not explicitly specified. Ensemble calculations using various data sources and different models, allow for evaluating the statistical spread of the obtained forecasts - a procedure, which is widely used in meteorological forecasting. The overall approach seems successful but remains complex.

Therefore, if data-based forecasts are the scope, it seemed reasonable to us to develop a hybrid approach: a simple dynamic model that can be easily tuned with available data. This goal is obtained with our approach, based on a single prognostic variable, which satisfies an ordinary integro-differential equation.

To our knowledge, there are only a few approaches that are equally simple and effective. In [7] a delay model is presented with a single prognostic equation that even has an analytic solution. Arguments and results are comparable to ours, though our integral formulation is more general and more robust, when extracting parameters from available data to feed the prognostic model. But delay models [8, 14] can also be described with several

variables and many parameters, which again makes them difficult to set up as forecasting model.

The paper is organized as follows. In sect. 2 we derive the model and show that it can be interpreted as a generalisation of classical compartmental models, such as SIR. Sect. 3 is devoted to the analysis of real data from the COVID-19 epidemic in Italy and Germany. A summary of how the model is capable of handling data from other countries is given in sect. 4. We conclude our model presentation with some remarks on stability and numerical robustness in sect. 5. Finally, some conclusion are drawn in sect. 6.

2 The model

2.1 Derivation of model equations

Many compartmental models, such as SIR, use deterministic equations for susceptible S, removed R and currently infected individuals I - all these variables being difficult to obtain from real data for various reasons. In our opinion, the most reliable statistic variable is *the number of cumulative diagnosed positive cases*. We chose this entity as our model variable and denote it by P. We are aware of the fact that the diagnosed cases are only a part of all cases but we assume that

- they are a statistically relevant part of the population;
- the fraction of diagnosed to all cases does not change in time and therefore
- the dynamics applied to the *visible* part of the epidemic is representative for the entire epidemic.

We derive our model in a discrete version, using discrete daily values as they are given by various data sources. We refer to P_n as the number of cumulative diagnosed positive cases on day n and ΔP_n as the number of new infected COVID-19 cases on day n.

We define the ratio between today's new cases ΔP_n and the weighted sum of new cases of the last N_r days to be

$$\tilde{\kappa}_n = \frac{\Delta \bar{P}_n}{\sum_{i=0}^{N_r-1} \Delta P_{i+n-N_r} g_i}\,, \tag{1}$$

where g_i are a set of fixed weights with the property $\sum_{i=0}^{N_r-1} g_i = 1$, N_r is the average number of days until an infectious person is removed from the

infection process and $\Delta\bar{P}_n$ is the mean of ΔP_n. We will see in the next section that in the special case of constant weights $g_i = 1/N_r$ the variable $\kappa(t)$ can be seen as the effective reproduction number.

Let us first give the main idea supporting our hybrid approach. The numbers $\tilde{\kappa}_i$ can be calculated easily with existing epidemic data, then a regression curve $\kappa(t)$ can be fitted to give us a law, purely based on data, on how this variable of the epidemic evolves. The data we observed show that, prior to measurements, κ has a constant initial value corresponding to the basic reproduction number R_0 and, once contact behaviour changes (due to media information, measurements, quarantine, a.s.f.) from a certain time T_Q on, $\kappa(t)$ is no longer constant but decays towards a final asymptotic value κ_∞, *i.e.*,

$$\kappa(t) = \begin{cases} \kappa_0 & t < T_\mathrm{Q}\,, \\ (\kappa_0 - \kappa_\infty)e^{-\alpha(t-T_Q)} + \kappa_\infty & t \geq T_\mathrm{Q}\,. \end{cases} \tag{2}$$

Note the difference between the discrete values $\tilde{\kappa}_i$ that are calculated from data and show fluctuations, and κ_i that are the values of the regression curve evaluated at the sample day i, *i.e.*, $\kappa_i = \kappa(t_i)$. Also note that phases of different severity in mitigation measures lead to small intermediate plateaus for $\kappa(t)$. This behaviour of the data was also noticed in [9], especially for fatalities. Since the steps in $\tilde{\kappa}_i$ are strongly smeared out, we simply model a single step with an asymptotic decay. Obviously, for more precision, $\kappa(t)$ can be modelled with a more complex, piece-wise defined function. The time T_Q can be directly read from the course of the data as the time when $\kappa(t)$ starts decreasing.

Resolved for $\Delta\bar{P}_n$, eq. (1) turns into a prognostic model for future infection cases ΔP_n but delivers also a model-based, "smoothed" curve for describing available data:

$$\Delta P_n = \kappa_n \sum_{i=0}^{N_r-1} \Delta P_{i+n-N_r} g_i\,. \tag{3}$$

Introducing weights g_i is *sensible* because the probability to infect others is not equally distributed over time. The incubation time is known to be between 1 and 14 days, with an average of 5 days [12]. The infectiousness begins probably before symptoms manifest and is maximal at the beginning of the disease. All these characteristics can be captured with the summation weights g_i.

But introducing weights is also *practical* because they allow for tuning parameters. If, for example, we use a Gaussian distribution to model g_i, we have the variance σ^2 and the time shift t_shift as tuning parameters. Note that

throughout this paper we speak of Gaussian weights although, in practice, we use only a small sample of a full Gaussian distribution as can be seen in fig. 1.

We have introduced a prognostic model with six tuning parameter, *i.e.*, six degrees of freedom: N_r for the infection or removal time in days, σ and t_{shift} for summation weights and $\kappa_0, \beta, \kappa_\infty$ for fitting the regression curve $\kappa(t)$. All these parameters can either be seen as pure tuning numbers, and can also be interpreted in epidemiological terms. In fact, the model as a whole can be compared to standard compartmental models, such as SIR, as we show in the next section.

2.2 Comparison to other models

In order to compare our discrete model to continuous deterministic models, we write eq. (3) in continuous form:

$$\frac{\mathrm{d}P}{\mathrm{d}t}(t) = \kappa(t) \int_{t-T_r}^{t} \frac{\mathrm{d}P}{\mathrm{d}t}(\tau)\, g(\tau - t)\mathrm{d}\tau \,, \tag{4}$$

where $P(t)$ is the number of total COVID-19 cases, $\kappa(t)$ represents an effective reproduction number as we will show, T_r is the time during which infected individuals take part in the infection process and $g(t)$ is a weighting function:

$$g(t) : [-T_r, 0] \mapsto \mathbb{R}_+ \quad \text{with} \quad \int_{-T_r}^{0} g(\tau) d\tau = 1 \,. \tag{5}$$

The SIR model. SIR [2] is a classic compartmental epidemic model that works with three prognostic variables: the susceptible individuals S, infected individuals I and people removed from the infection process R. There are transitions from S to I to R, which lead to the following system of ODE's:

$$\frac{dS}{dt} = -\frac{\beta I S}{N} \,, \tag{6}$$

$$\frac{dI}{dt} = \frac{\beta I S}{N} - \gamma I \,, \tag{7}$$

$$\frac{dR}{dt} = \gamma I \,, \tag{8}$$

where $N = S + I + R$ is the total number of people, $\beta = 1/T_c$ is the contact frequency, T_c being the average time between contacts and $\gamma = 1/T_r$ is defined by the mean time T_r between infection and removal. We skip the discussion

of the SIR model and ask: How does our model compare to SIR? For this purpose let us rewrite the SIR model in terms of our prognostic variable $P = I + R$. If we add Eqs. (7) and (8) we obtain

$$\frac{d(I+R)}{dt} = \frac{\beta I S}{N}\,, \tag{9}$$

which, introducing P, becomes

$$\frac{dP}{dt} = \frac{\beta(P-R)S}{N}\,. \tag{10}$$

Eq. (10) can be written using the relative susceptible number $s = S/N$ in the form

$$\frac{dP}{dt} = \beta s P - \beta s R\,. \tag{11}$$

Now, let us rewrite our model eq. (4) by splitting the integral into two parts and replacing $g(t)$ with constant weights $g_0 = 1/T_r$:

$$\frac{\mathrm{d}P}{\mathrm{d}t}(t) = \kappa(t)\frac{1}{T_r}\int_0^t P'(\tau)\mathrm{d}\tau - \kappa(t)\frac{1}{T_r}\int_0^{t-T_r} P'(\tau)\mathrm{d}\tau\,. \tag{12}$$

After integration our simplified model reads

$$\frac{\mathrm{d}P}{\mathrm{d}t}(t) = \frac{1}{T_r}\kappa(t)P(t) - \frac{1}{T_r}\kappa(t)P(t-T_r)\,. \tag{13}$$

Comparing the SIR model in the form (11) with our model (13), we see that

$$\frac{1}{T_r}\kappa(t) = \beta s(t)\,, \tag{14}$$

$$P(t-T_r) = R(t)\,, \tag{15}$$

and, using standard definitions, $\gamma = 1/T_r$, $\beta = 1/T_c$ and the basic reproduction number $R_0 = T_r/T_c = \beta/\gamma$, can state the following epidemiological interpretation of the parameters:

1. $\kappa(t) = T_r\beta s(t) = \frac{\beta}{\gamma}s(t)$ corresponds to the effective reproduction number $R_{\text{eff}} = R_0 s(t)$.

2. The positive cases $P(t-T_r)$ correspond to the removed individuals $R(t)$, thus T_r can be interpreted as the time until removal from the infection process.

Since T_r is usually not known precisely, we have set it to 11 days, which corresponds to the value for the time of infectiousness given in [12]. Two remarks are in order: First, a change of T_r slightly changes the values of $\tilde{\kappa}_n$, which makes it somewhat difficult to compare them to officially reported effective reproduction numbers R_{eff}. Nevertheless, we call it by this name as it has exactly this function in the model. Secondly, T_r is not to be confused with the recovery time but is the time until an infectious person is removed from the infection process.

Delay models. Delay models follow the same strategy, modelling the removal process not with a separate variable but with a time shift in the function describing the number of cumulative cases. In fact, eq. (13) is identical to the functional retarded differential eq. (11) in [7], though with our integration weights set to the constant value of $g_i = 1/T_r$.

The SEIR model. The SEIR model introduces a further group of people, *the exposed E*, *i.e.*, people who are infected but not yet infectious. This effect is accounted for in our model by excluding the first days in the integral in eq. (4) or, equivalently, by using zero-weights $g_i = 0$ for the according time. Though, for COVID-19 it seems that people are infectious right from the first days after infection [12].

What makes the difference?

1. First of all, our approach does not explicitly model $S(t)$ with a coupled prognostic equation. This is reasonable to us because the assumption that susceptible individuals are removed only by the infection process is wrong for the current COVID-19 epidemic. Severe quarantine measures, including lock downs, have been implemented in almost all countries.

2. Other compartmental models take such measures into account by introducing, e.g., direct transfers from S to R compartments. But, in our opinion, this makes models complicated and hides the fact that political measures and their effects are almost impossible to model. Our approach is a very practical one: we do not model $S(t)$. We focus on $\kappa(t)$, which we have seen being related to the product of the time-dependent basic reproduction number $R_0(t)$ and the relative susceptible number $s(t)$. We extract $\kappa(t)$ from real data using our model assumptions and apply a curve fitting procedure to allow for extrapo-

lation. Therefore, our approach can be called "hybrid": a mixture of curve fitting and modelling.

3. By using as prognostic variable the number of cumulative diagnosed positive cases $P(t)$, we automatically have the numbers of new infections as ΔP. In our opinion, this is the best variable to describe how the epidemic develops. However, this number is not automatically obtained in the SIR model because I represents the "currently infected people" and ΔI is a net difference mixing the "new positive cases", *i.e.*, the transfer form S to I, with the "removed cases", *i.e.*, the transfer from I to R.

4. By introducing weights g_i we can model an incubation time, *i.e.*, a time between "getting infected" and "being infectious" as well as a time before detection.

5. We define removed people $R(t)$, *i.e.*, individuals that no longer take part in the infection process, as all positive cases $P(t-T_r)$ at a certain time T_r ago. Again, we have no prognostic equation for $R(t)$. Therefore, we do not have to model and elaborate how these individuals are removed from the infection process. We simply assume they are removed after a certain time T_r because prior to curing or deceasing people are isolated in hospitals or, with weak symptoms, are put into quarantine.

6. Parameters or variables that are not known, such as $T_c, R_0, S(0)$ and N, are subsumed into a single function $\kappa(t)$, which is obtained from real data - without having to speculate on how it comes about. This makes the model simple - and setting the weights constant - even simpler so that it can be set up quickly to produce satisfying prognoses.

With these assumptions we have been able to describe the infection process with a single prognostic variable $P(t)$ in an integro-differential equation. Calculating deceased and cured people is a secondary process, which in our approach does not influence the dynamics of the epidemic. Nonetheless, they are important numbers to know and can be simply obtained from $P(t)$, as it will be shown in the next section.

2.3 Secondary variables

The number of fatalities. Let V_n be the total number of deceased individuals on day n from the beginning of the epidemic. We assume that the

casualties on the n-th day, *i.e.*, ΔV_n, is related to the weighted sum of new cases over the last N_V days, giving rise to the following ratio:

$$\tilde{\mu}_n := \frac{\Delta V_n}{\sum_{i=0}^{n_V-1} \Delta P_{i+n-N_V} h_i}\,, \tag{16}$$

where N_V is the maximum number of days after which the people decease and the weights h_i allow for taking a probability distribution into account. The numbers $\tilde{\mu}_n$ can be interpreted as case fatality ratio [13]. Their values can be obtained from existing data, fitted to $\mu(t)$ and used to extrapolate future values, $\mu_n = \mu(t_n)$, where t_n is a day in the future. The corresponding discrete equation is

$$\Delta V_n = \mu_n \sum_{i=0}^{N_V-1} \Delta P_{i+n-N_V} h_i\,. \tag{17}$$

Note that from the available data it can be seen that the peak of fatalities lacks about 7-8 days behind the peak of the daily new infections. This is accounted for with the weights h_i having a maximum at 7-8 days prior to the current day t_n.

The Number of Cured. $C(t)$ is defined in an analogous way with the discrete curing ratio:

$$\tilde{\nu}_n := \frac{\Delta C_n}{\sum_{i=0}^{N_C-1} \Delta P_{i+n-N_C} j_i}\,. \tag{18}$$

The curing ratio can again be fitted with a continuous function $\nu(t)$ to obtain predictions of the model values $\nu_n = \nu(t_n)$ that are used to predict future number of cured people:

$$\Delta C_n = \nu_n \sum_{i=0}^{N_C-1} \Delta P_{i+n-N_C} j_i. \tag{19}$$

Active Cases. The number of currently infected individuals $I(t)$, also known as active cases, is the difference between all cases $P(t)$ and the deceased and cured cases

$$I(t) = P(t) - C(t) - V(t)\,. \tag{20}$$

Note that calculating the number of removed cases in the classical way is not valid for our model, that is:

$$R(t) \neq C(t) + V(t)\,, \tag{21}$$

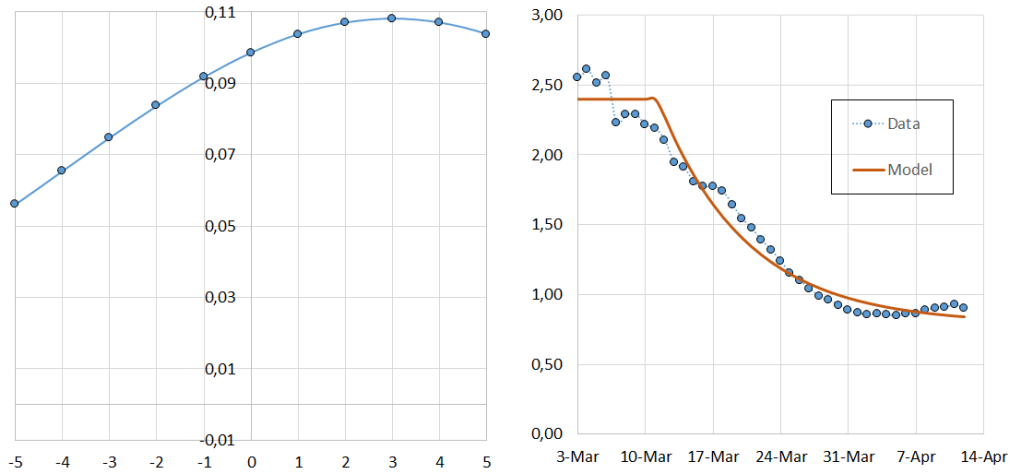

Figure 1: *Left: Gaussian weights used to calculate* $\tilde{\kappa}_n$ *with* $\sigma = 7$ *and* $t_{\text{shift}} = 3$. *Right: Effective reproduction number* $\kappa(t)$ *obtained from data (blue line) and the fitted curve (orange line) used for modelling the epidemic in Italy*

as we consider removal by cutting off the corresponding integral after a removal time T_r, taking into account other processes of removal than curing and deceasing, such as quarantine or isolation. But we have to assure with our tuning parameters that in the long term

$$R(t) = P(t) = C(t) + V(t) \quad \text{for} \quad t \to \infty \tag{22}$$

because, in this limit, $I(t) = 0$.

3 Discussion of model results

Data for the COVID-19 epidemic are made available by the John Hopkins University [5] and coincide, at least for Italy and Germany, with those from Worldometers [4]. The original time series data show significant weekly fluctuations, hence we only work with 7-day-means, which acts as a low-pass filter.

3.1 Italy

On the left of fig. 1 we see the Gaussian weights we used to obtain the effective reproduction number $\tilde{\kappa}_n$ shown on the right of the same figure. The integration time is $N_r = 11$ days, *i.e.* only individuals registered positive within this time period actually take part in the (model) infection process.

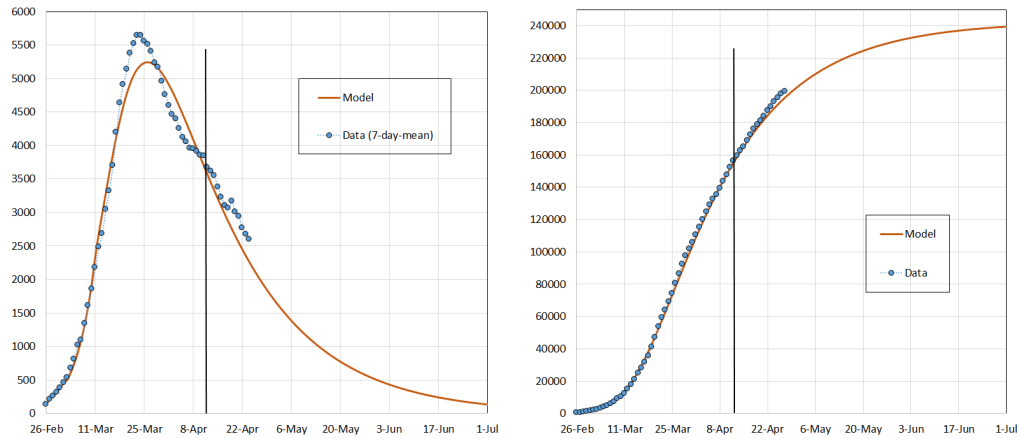

Figure 2: *Daily new infections (left) and total cases (right) of COVID-19 from February 26 to June 30, 2020 in Italy. Only data before the vertical black line, i.e., before April 13, have been used to tune the model.*

Note that the time-shift $t_{\text{shift}} = 3$ and $\sigma = 7$ turned out to produce a sensibly smooth result for $\tilde{\kappa}_n$. In terms of one person infecting others, this means that it infects majorly in the first five days and lesser in the following days. To us this makes sense, because with emerging symptoms people automatically retreat from social interaction.

The effective reproduction numbers are shown for the time period March 3 to April 14. The data-based $\tilde{\kappa}_n$ are shown with a blue line and the values used in the model with an orange line. The lock down in Italy was on March 8, 2020. Note that a couple of days before that date the curve has a plateau at $\kappa \approx 2.5$, which can be interpreted as the effective reproduction number R_{eff} before lock down. Also shown is the time before lock down, where κ has a short "swinging about" a value known as basic reproduction number $R_0 = 3.5$. After three weeks $\tilde{\kappa}_n$ seems to settle at a value of about 0.8. We decided to model this behaviour from day March 3, 2020, according to eq. (2) with the following parameter values:

$$\kappa_0 = 2.4\,, \quad \alpha = 0.11 \quad \text{and} \quad \kappa_\infty = 0.79\,.$$

The COVID-19 cases in Italy. The daily new diagnosed cases are shown on the left of fig. 2. The original time series data shows significant weekly fluctuations, therefore we show only the 7-day-mean. The model is capable of reproducing the exponential growth in the beginning of the epidemic, as well as the peak and the slow decay of the curve afterwards. The deviation of the curves originates in the deviation between data-based $\tilde{\kappa}_n$ and curve-fitted κ_n.

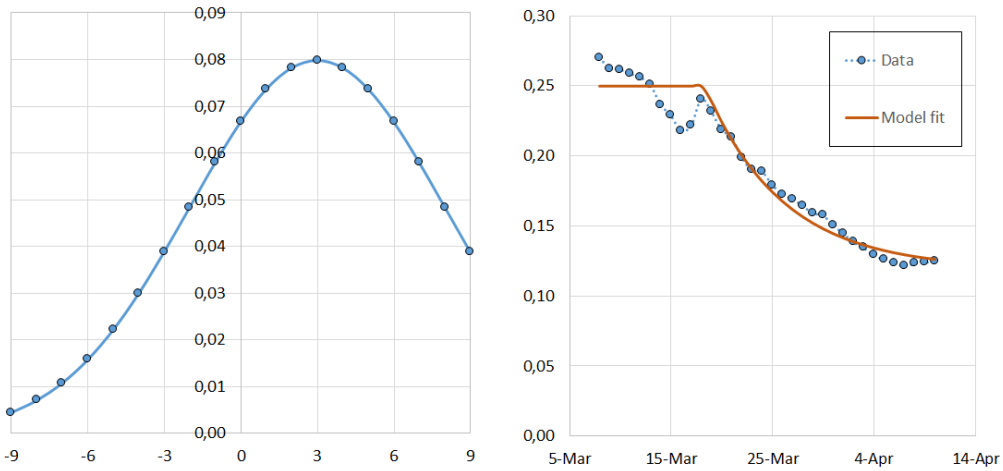

Figure 3: *Left: Gaussian weights used to calculate* $\tilde{\mu}_n$ *with* $\sigma = 5$ *and* $t_{\text{shift}} = 3$*. Right: case fatality rate* $\mu(t)$ *obtained from data (blue points) and the fitted curve (orange line) used for modelling the epidemic in Italy*

The cumulative number of diagnosed positive cases are shown on the right of fig. 2. The modelled curve follows accurately the data. Note that the last model-tuning was made on April 13, 2020. Two weeks later the relative deviation of the cumulative number of cases is about 2%.

Fatalities in Italy. On the left of fig. 3 we see the Gaussian weights we used to obtain the model case fatality rate $\tilde{\mu}_n$ shown on the right of the same figure. The integration time is $N_r = 18$ days, *i.e.*, only individuals diagnosed positive within the last 18 days are considered in the model calculation of fatalities. Note that setting $t_{\text{shift}} = 3$ for the Gaussian time-shift puts the maximum weight on patients that have been diagnosed positive 7 days ago. According to Germany's Robert-Koch Institut [12] the average time in hospital before deceasing is 14 days. We think, though, that 7 days is more realistic for Italy.

The number of daily and cumulative fatalities are shown in fig. 4. The course of fatalities is well represented by the model and 14 days after the last model-tuning the relative deviation for the cumulative number of fatalities is about 2%. Just one remark on the choice of the Gaussian weights: As observed from the data for Germany and Italy, the peak of daily fatalities lacks the peak of daily new infections by 7 to 8 days. This allows us to set the Gaussian weights h_i in eq. (16) appropriately, *i.e.*, with a maximum weight around 7 days prior to the calculation date of new fatalities.

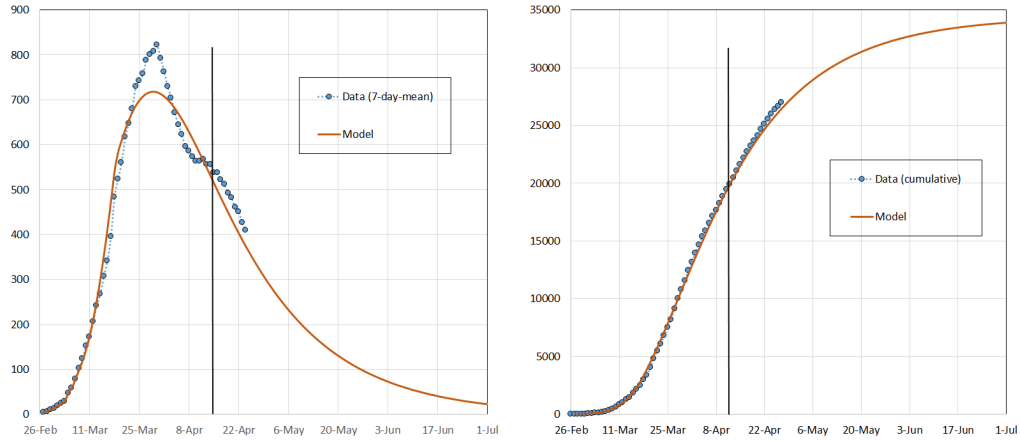

Figure 4: *Daily new fatalities (left) and the cumulative number of fatalities (right) of COVID-19 from February 26 to July 1, 2020 in Italy. Only data before the vertical black line, i.e., before April 13, have been used to tune the model*

3.2 Germany

On the left of fig. 5 we see the Gaussian weights we used to obtain the effective reproduction number $\tilde{\kappa}_n$ shown on the right for Germany. Again, the integration time is $N_r = 11$ days, $\mu = 3$ and $\sigma = 7$.

The effective reproduction number $\kappa(t)$ is shown for the time period March 3 to April 14, where the data-based $\tilde{\kappa}_n$ are shown with a blue line and the values used for the model in orange. The school closing in Germany was on March 14, 2020. Note that a couple of days before that date the curve has a short plateau at $R_{\text{eff}} \approx 3,0$, which can be interpreted as the basic reproduction number R_0 of COVID-19 in Germany. Also note the high values of κ from March 3 to 5. These values are not reliable in the sense that our integration period covers also days with 0 new infections, *i.e.*, time before the epidemic, where the model is not valid.
After three weeks, $\tilde{\kappa}_n$ seem to settle at a value of about 0.6. We decided to model this behaviour from day March 3, 2020 according to eq. (1) with the following parameter values

$$\kappa_0 = 3.0\,, \quad \alpha = 0.10 \quad \text{and} \quad \kappa_\infty = 0.58\,.$$

If we compare this to Italy we note that:

1. The initial effective reproduction rate was higher in Germany.

2. Its final value is higher in Italy.

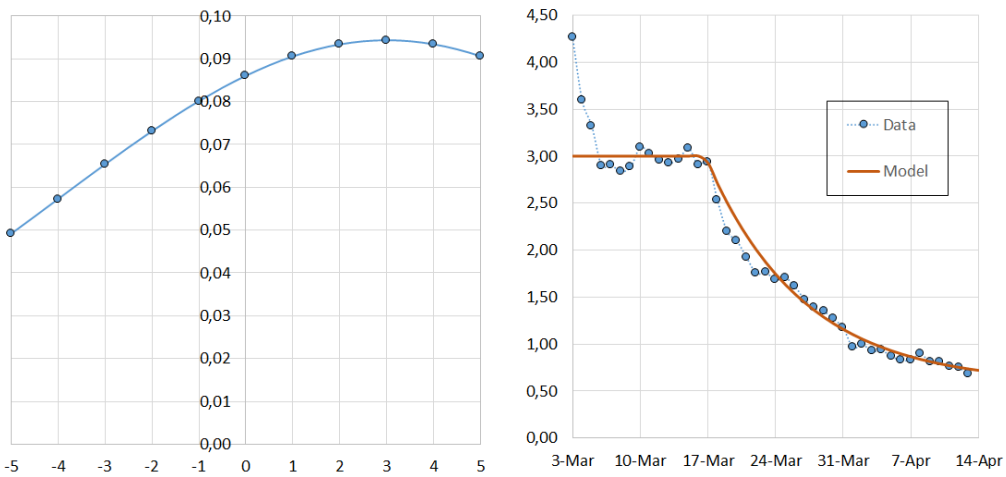

Figure 5: *Left: Gaussian weights used to calculate* $\tilde{\kappa}_n$ *with* $\sigma = 7$ *and* $\mu = 3$. *Right: Effective reproduction number* $\kappa(t)$ *as obtained from data (blue line) and the fitted curve (orange line) used for modelling the epidemic in Germany*

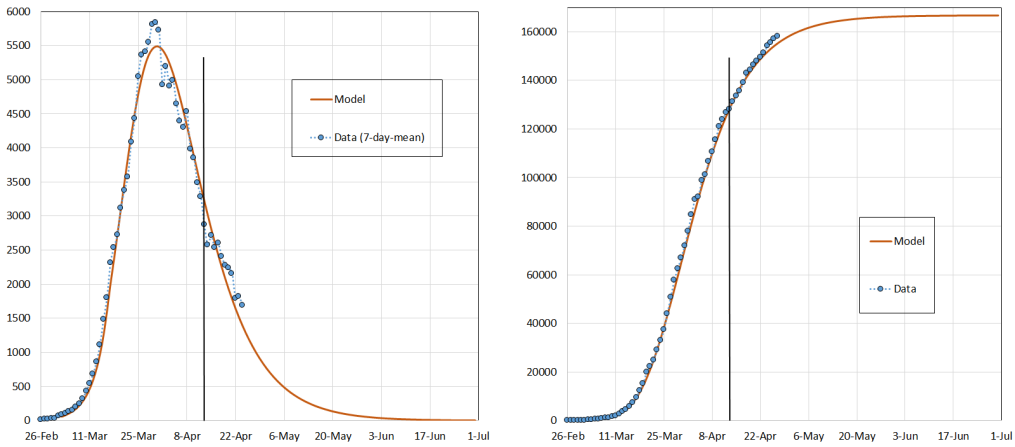

Figure 6: *Daily new infections (left) and total cases (right) of COVID-19 from February 26 to June 30, 2020 in Germany. Only data before the vertical black line, i.e., before April 13, have been used to tune the model*

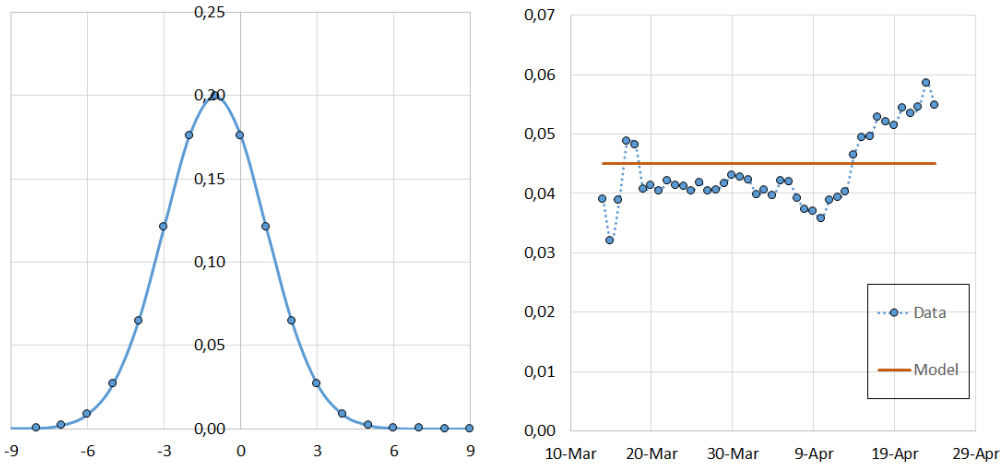

Figure 7: *Left: Gaussian weights used to calculate* $\tilde{\mu}_n$ *with* $\sigma = 2$ *and* $t_{\text{shift}} = -1$. *Right: case fatality rate* $\mu(t)$ *obtained from data (blue points) and the fitted curve (orange line) used for modelling the epidemic in Germany*

3. The transition velocity is in both countries approximately the same.

This is our tentative interpretation: The initial reproduction number in Germany was higher because in the beginning of the epidemic it was mainly spread among young people, who came from skiing in the Austrian Alps. If we assume more social contacts among young and sporty people this could be an explanation.

Although the measures taken by Italian politicians were more restrictive than in Germany, the final effective reproduction number κ_∞ is 30% higher in Italy, leading to a much slower decay of the numbers of infected.

The COVID-19 cases in Germany. They are shown in fig. 6. Original data are shown as a 7-day-mean (blue line). The model (orange line) is capable of reproducing the course of the epidemic. The Total Number of Infections are shown on the right of fig. 6. The modelled curve (orange line) follows again accurately the data (blue line).

Fatalities in Germany. On the left of fig. 7 we see the Gaussian weights we used to obtain the model case fatality rate $\tilde{\mu}_n$ shown on the right of the same figure. The integration time is $N_r = 18$ days again. Note a time shift of the Gaussian by $t_{\text{shift}} = -1$ putting a maximum weight on patients that have been diagnosed positive 10 days earlier. According to Germany's Robert-Koch Institut [12], the average time in hospital before deceasing is about 10 days.

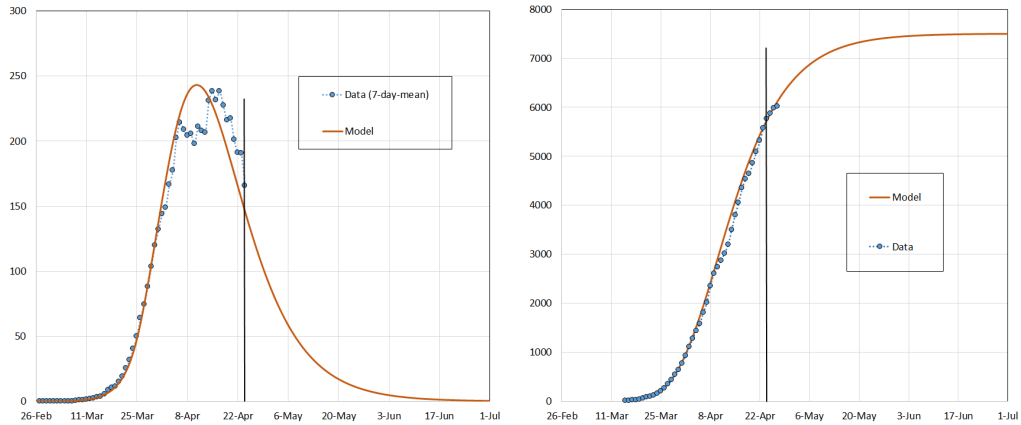

Figure 8: *Daily new fatalities (left) and the cumulative number of fatalities (right) of COVID-19 from February 26 to July 1, 2020 in Germany. Only data before the vertical black line, i.e., before April 23, have been used to tune the model*

The number of daily and cumulative fatalities for Germany are shown in fig. 8. The course of fatalities was not well represented due to a sudden rise in case fatality rate about April, 15. We therefore re-tuned the model on April 24, setting the mean case fatality rate to $\bar{\mu} = 0.45$.

Other Variables. As stated before, once the number of cumulative COVID-19 cases $P(t)$ are known, other variables can be derived from them, such as the number of

- cured individuals,
- hospital beds needed,
- intensive care units (ICU's) needed,
- active and closed cases

and so forth. Forecasting these variable is made with an analogous weighted-integral-approach as for the fatalities. Its forecast quality strongly depends on the quality of the available data and the knowledge of mean values, such as the time in hospital, in intensive care and so on. In fig. 9 we show an example for the number of cumulative cured cases and active cases for Germany. Note that the same diagram for Italy (not shown) was difficult to obtain because the data of cured cases did not seem to be reliable. Probably, cured individuals are registered with a very long time delay (as to date, cured cases make up only 65% of all closed cases).

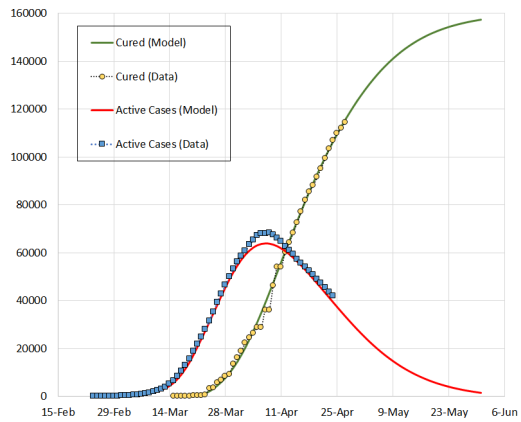

Figure 9: *Number of cumulative cured cases from data (yellow circles) and the model simulation (green line), along with the number of active COVID-19 cases from data (blue squares) and the model simulation (red line) from February 26 to June 5, 2020 in Germany*

4 Other Countries

We have applied our model to the COVID-19 data made available for other countries by European Centre for Disease Prevention and Control [3]. The corresponding graphs are made available on a web-based platform [6]. Here, we simply give some major results beginning with the fitting parameters to the temporal behaviour of the effective reproduction number $\kappa(t)$ given in 2. The date of the fitting was April 28, 2020. After the country name, the second column shows the initial value of the basic reproduction number, the third column indicates the slope for the exponential decay and the fourth column gives the final effective reproduction number.

Country	Initial R_0	Slope indicator α	Final R_{eff}
Austria	3.4	0.065	0.50
Belgium	3.2	0.045	0.64
Brazil	3.6	0.075	1.04
France	3.4	0.046	0.52
Germany	3.3	0.041	0.50
Iran	4.7	0.12	0.93
Italy	3.3	0.066	0.81
Russia	3.0	0.018	0.50
South Korea	2.3	0.34	0.76
Spain	2.9	0.074	0.63
Sweden	3.4	0.18	1.14
Turkey	3.9	0.10	0.84
UK	3.2	0.032	0.50
USA	2.8	0.12	0.91

Table 1: Fitting parameters for the effective reproduction number R_{eff}

The following facts can be observed:

- The lowest final reproduction number is 0.5.
- Some countries, such as Brazil and Sweden, to date still have $R_{\text{eff}} > 1$.
- The highest slope can be seen with South Korea, that has quickly introduced severe measures.

5 Stability

A global stability analysis for a generalized delayed SIR model was presented in [8]. It shows the behaviour of such systems for long terms. Speaking of numerical error propagation, we expect a high stability of our model due to its integral formulation. This is important when applying the model to noisy time series data because the parameter that governs the overall dynamic is not obtained from a single data point but from a weighted sum of data points, *i.e.*,

$$\sum_{i=0}^{N_r-1} \Delta P_{i+n-N_r} g_i \to \tilde{\kappa}_n \, .$$

This can also be interpreted as a statistical filter, that is based on empirical knowledge, *i.e.*, the fact, that infection is a process spread over time.

Here we do not make a stability analysis but limit ourselves to show empirically the sensitivity of the cumulative number of diagnosed cases $P(t)$ to changes in the curve fitting parameters, which are used to obtain the effective reproduction number $\kappa(t)$. To this purpose we disturb the parameters κ_0, α and κ_∞ of the fitting curve $\kappa(t)$ in 2 by 5%, one after the other, and observe the relative change δP in $P(t)$ after 2, 4, 6 and 8 weeks. The results are given in the following table:

	δP (2 weeks)	δP (4 weeks)	δP (6 weeks)	δP (8 weeks)
$\delta\alpha = +5\%$	1%	15%	21%	22%
$\delta\kappa_0 = +5\%$	20%	21%	21%	21%
$\delta\kappa_\infty = +5\%$	0%	4%	7%	9%

Table 2: Model reaction for a 5%-perturbation of the fitting parameters

Generally spoken, model deviations do not "explode" within mid-term forecasts up to two months. This is unimaginable, for instance, for weather forecasts, which have to cope with a chaotic system. The epidemic equation, though, is not chaotic and errors stay within reasonable limits even after long simulation times.

If we double the perturbation from 5% to 10% we observe the following changes of the cumulative number of diagnosed positive cases $P(t)$:

	δP (2 weeks)	δP (4 weeks)	δP (6 weeks)	δP (8 weeks)
$\delta\alpha = +10\%$	2%	26%	35%	38%
$\delta\kappa_0 = +10\%$	43%	46%	47%	47%
$\delta\kappa_\infty = +10\%$	0%	8%	16%	21%

Table 3: Model reaction for a 10%-perturbation of the fitting parameters

It can be noticed that doubling the perturbation roughly leads to doubling the deviation of the function. Thus, the model can be seen as numerically robust. The same results are expected for secondary variables, such as fatalities, because they depend linearly on $P(t)$.

6 Conclusions

The advantage of our approach is the simplicity of its formulation, the precision with which the real course can be reproduced and the effectiveness to make mid-term forecasts.

We were able to show that our model has similarities to classic compartmental models, in particular:

- The variable κ_n can be interpreted as an effective reproduction number R_{eff}.
- Limiting the integral term in the deterministic equation models the removal process.
- The prognostic variable $P(t)$ is the sum of active cases I and removed ones R.
- The integration weights account, among others, for an incubation time, *i.e.*, a time between being infected and infectious. This avoids an additional equation for *exposed individuals* as used in the SEIR model.

Although our model consists of only one dynamic equation, it is simpler compared to most approaches and still able to capture probability distributions of the infection process by using integration weights. In addition, the integral formulation leads to more numerical robustness.

The simple formulation allows us to solve the single deterministic equation for the effective reproduction number, so that we can calculate its discrete values $\tilde{\kappa}_n = R_{\text{eff}}$ from real time series data. Doing this we found that epidemic can be modelled by basically three parameters: the initial κ_0, corresponding to the basic reproduction number R_0 without measures; the final κ_∞, corresponding to the effective reproduction number due to inevitable contacts during mitigation measures; and the transition parameter, describing the velocity with which the effective reproduction number approaches its final value.

We applied our model to many countries and gave a special focus on the data of Italy and Germany. After extracting the three parameters we have been able to model the course of the epidemic in both countries rather well. We found it astonishing how the parameters κ_0 and κ_∞ differ between the two countries - inviting us to interpret them in terms of effectiveness of measures, social organisation (in Italy elderly vulnerable people are more likely found to live with the younger part of the family) as well as organisation and quality of the health system.

We set up the hypothesis that three parameters suffice to model the epidemic from the outbreak, over the period of social distancing measures until the end - under the assumption that the measures remain effective with respect to infections till the end, *i.e.*, zero new infections. It remains to be shown that this hypothesis remains valid for longer periods of time, especially when mitigation measures are loosened.

We finally demonstrated with "numerical experiments" that our approach is numerically robust in terms of changes to the fitting of the parameter function $\kappa(t)$. A profound analysis of the stability of the model is still to be done.

We hope that our approach facilitates forecasting of ongoing epidemics for long term periods but also for now-casting because incorporating new daily data is easy and can also be automated. Our results will be available on website [6].

References

[1] I. Ciufolini and A. Paolozzi. "A Mathematical prediction of the time evolution of the Covid-19 pandemic in some countries of the European Union using Monte Carlo simulations". In: *Eur Phys J Plus.* (135(4):355 2020). DOI: 10.1140/epjp/s13360-020-00383-y.

[2] *Compartmental Models.* 2017. URL: https://en.wikipedia.org/wiki/Compartmental_models_in_epidemiology.

[3] *COVID-19 Coronavirus data.* European Centre for Disease Prevention and Control. 2020. DOI: 10.2906/101099100099/1. URL: https://data.europa.eu/euodp/de/data/dataset/covid-19-coronavirus-data.

[4] *COVID-19 Coronavirus Pandemic.* 2020. URL: https://www.worldometers.info/coronavirus/.

[5] *COVID-19 Dashboard.* 2020. URL: https://coronavirus.jhu.edu/map.html.

[6] *COVID-19 Prognostic Model.* 2020. URL: http://www.roehlnet.de/corona/countries-all.

[7] Luca Dell'Anna. *Solvable delay model for epidemic spreading: the case of Covid-19 in Italy.* 2020. arXiv: 2003.13571 [q-bio.PE].

[8] A. Elazzouzi et al. "Global stability analysis for a generalized delayed SIR model with vaccination and treatment". In: *Advances in Difference Equations.* 532nd ser. (2019), pp. 1687–1847. DOI: 10.1186/s13662-019-2447-z. URL: https://doi.org/10.1186/s13662-019-2447-z.

[9] Seth Flaxman et al. *Report 13 - Estimating the number of infections and the impact of non-pharmaceutical interventions on COVID-19 in 11 European countries.* report. Imperial College, 2020.

[10] *Forecaste of the COVID-19 epidemic spread in Italy.* 2020. URL: http://web.math.unifi.it/users/brugnano/covid19/.

[11] Giulia Giordano et al. *A SIDARTHE Model of COVID-19 Epidemic in Italy.* Mar. 2020.

[12] M. an der Heiden and U. Buchholz. *Modellierung von Beispielszenarien der SARS-CoV-2-Epidemie 2020 in Deutschland.* report. Robert-Koch-Institut, Germany, 2020. DOI: 10.25646/6571.2.

[13] O. S. Miettinen. *Epidemiological research: terms and concepts.* Dordrecht Heidelberg London New York: Springer, 2011. ISBN: 9400711700. URL: https://books.google.it/books?hl=de&lr=&id=GT-XHKhrrwgC&oi=fnd&pg=PR5&dq=miettinen+epidemiological+&ots=NkDvjhbcOA&sig=mV7TNuFOGIV_wayuGTOMO9xUC50&redir_esc=y#v=onepage&q=miettinen%20epidemiological&f=false.

[14] Nian Shao et al. "Dynamic models for Coronavirus Disease 2019 and data analysis". In: *Mathematical Methods in the Applied Sciences* 43.7 (2020), pp. 4943–4949. DOI: 10.1002/mma.6345. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/mma.6345. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/mma.6345.

[15] Steffen Uhlig et al. "Modeling projections for COVID-19 pandemic by combining epidemiological, statistical, and neural network approaches". In: *medRxiv* (2020). DOI: 10.1101/2020.04.17.20059535. eprint: https://www.medrxiv.org/content/early/2020/04/22/2020.04.17.20059535.full.pdf. URL: https://www.medrxiv.org/content/early/2020/04/22/2020.04.17.20059535.