

RESEARCH

Open Access



# Early stage COVID-19 disease dynamics in Germany: models and parameter identification

Thomas Götz<sup>1\*</sup>  and Peter Heidrich<sup>1,2</sup>

\*Correspondence:

[goetz@uni-koblenz.de](mailto:goetz@uni-koblenz.de)

<sup>1</sup>Mathematical Institute, University  
Koblenz-Landau, D-56070, Koblenz,  
Germany

Full list of author information is  
available at the end of the article

## Abstract

Since the end of 2019 an outbreak of a new strain of coronavirus, called SARS-CoV-2, is reported from China and later other parts of the world. Since January 21, World Health Organization (WHO) reports daily data on confirmed cases and deaths from both China and other countries ([www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports](http://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports)). The Johns Hopkins University ([github.com/CSSEGISandData/COVID-19/blob/master/csse\\_COVID\\_19\\_data/csse\\_COVID\\_19\\_time\\_series/time\\_series\\_COVID19\\_confirmed\\_global.csv](https://github.com/CSSEGISandData/COVID-19/blob/master/csse_COVID_19_data/csse_COVID_19_time_series/time_series_COVID19_confirmed_global.csv)) collects those data from various sources worldwide on a daily basis. For Germany, the Robert-Koch-Institute (RKI) also issues daily reports on the current number of infections and infection related fatal cases ([www.rki.de/DE/Content/InfAZ/N/Neuartiges\\_Coronavirus/Situationsberichte/Gesamt.html](http://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Situationsberichte/Gesamt.html)). However, due to delays in the data collection, the data from RKI always lags behind those reported by Johns Hopkins. In this work we present an extended SEIRD-model to describe the disease dynamics in Germany. The parameter values are identified by matching the model output to the officially reported cases. An additional parameter to capture the influence of unidentified cases is also included in the model.

**Keywords:** COVID-19; Epidemiology; Disease dynamics; SEIRD-model

There's an evil virus that's threatening  
mankind [...]

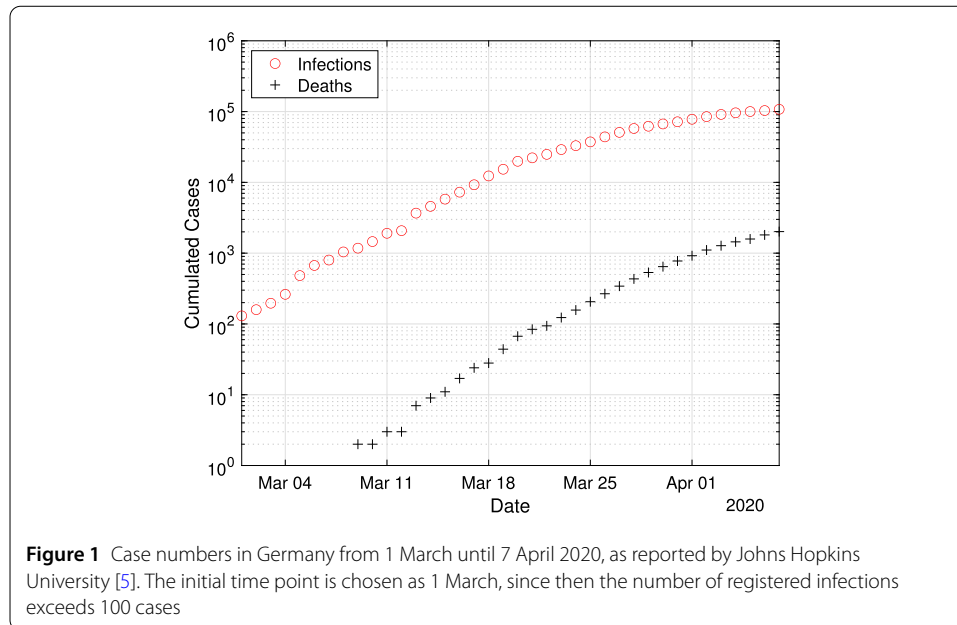
A menace to society

Iron Maiden, *Virus*, 1996.

## 1 Introduction

In December 2019, first cases of a novel *pneumonia of unknown cause* were reported from Wuhan, the seventh-largest city in China. In the meantime, these cases have been identified as infections with a novel strain of coronavirus, called SARS-CoV-2 and the disease it causes is called coronavirus disease 2019 (COVID-19). At the beginning of January 2020, the virus spread over mainland China and reached other provinces. Increased travel activities due to the Chinese new year festivities supported the expansion of the infection.

© The Author(s) 2020. This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.



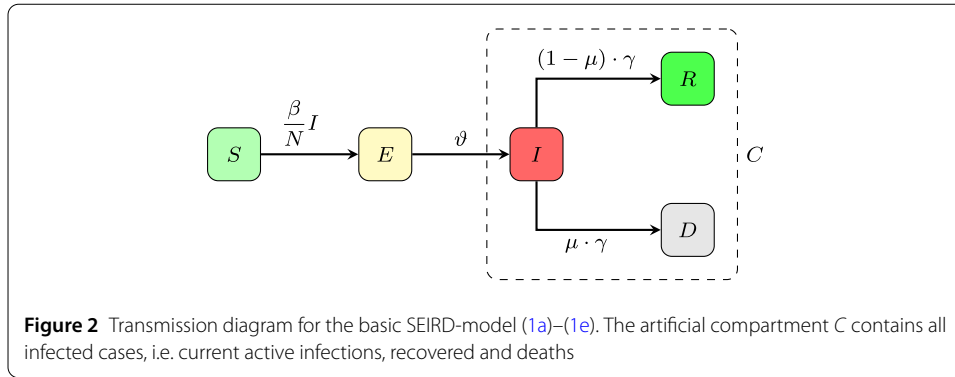
Since 21 January, WHO's daily situation reports contain the latest figures on confirmed cases and deaths, see [1].

The first COVID-19 case in Germany was reported in late January 2020 in a company close to Munich, Bavaria. Later cases were imported by travelers from China, Iran or Italy as well as tourists returning from ski holidays in the Austria and Italy. By 1 March 2020 more than 100 cases were reported in Germany and since then the number of cases began to rise exponentially, see Fig. 1. The first deaths were reported on 9 March 2020 [2, 3].

By 16 March 2020 the federal government introduced first measures to reduce the spread of the disease: Schools, kindergartens and universities were closed. On 22 March these measures were tightened by implementing a national curfew and contact ban. People are advised to stay at home, leaving only for work related activities, necessary shopping, medical treatment or sports. All this should not be done in groups of more than two persons if they do not belong to the same household [4].

Our work is based on the data reported by Johns Hopkins University [5]. We refrain from using the official data from the Robert-Koch-Institute [2], since they suffer from a delay by several days due to the more complicate way of aggregating those data. For a detailed explanation of the difference between the data reported by Johns Hopkins and the Robert-Koch Institute we refer to the information given on the webpage of the Robert-Koch-Institute, see [6]. Johns Hopkins University continuously collects the data from internet queries at various sources (local health authorities, newspapers, etc.) whereas the Robert-Koch-Institute collects the data that are reported for the local health authorities to the district level, then state level and finally aggregates them to the federal statistics. Hence these data lag several days behind the ones collected by Johns Hopkins University.

The paper is organized as follows: In Sect. 2 we describe the model and the parameter identification problem. Our models consists of three variants of a five compartment SEIRD-system without demographic terms, where the transmission rate is either fixed (1a)–(1e) or time-dependent (3a)–(3e) and (4a)–(4e). The fatalities are either described by an ODE, see models (1a)–(1e) and (3a)–(3e), or via a delay term in model (4a)–



(4e). In the parameter estimation problem, we determine the transmission rate, detection rate and lethality together with the initial values for the exposed and infected compartment. In Sect. 3 we discuss the sensitivity of our model with respect to detection rate. Section 4 is devoted to the adjoint equations used for solving the optimization problem. The simulation results are presented in Sect. 5. Here we do compare the results obtained from the three models presented in Sect. 2.

## 2 Mathematical model

To model the dynamics of the spread of COVID-19 incidences, we propose a hierarchy of SEIRD models. For details regarding the original SIR- and SEIR-model we refer to classical works on mathematical epidemiology, e.g [7]. For our basic SEIRD-model, the total population of Germany with  $N \sim 83.000.000$  individuals is subdivided in to *susceptibles*  $S$ , *exposed*  $E$ , *infected*  $I$ , *recovered*  $R$  and *deaths*  $D$ . The susceptibles constitute the reservoir of persons that are not yet infected with SARS-CoV-2. After infection susceptible become exposed meaning that they already carry the virus but are not yet infectious. With a rate  $\vartheta$  exposed individuals become infectious and transmit the virus with rate  $\beta$  to susceptibles. An infected individual loses infectivity with  $\gamma$  and has a probability  $\mu$  of dying due to the disease [8]. Figure 2 shows the transmission structure. By  $C$  we denote all infected cases, independent of their current status. This artificial compartment is later on used to compare with the total number of registered cases reported by Johns Hopkins or RKI.

The resulting system of ordinary differential equations (ODE) for the above described SEIRD-model reads as

$$\frac{dS}{dt} = -\frac{\beta}{N}SI, \quad S(t_0) = S_0 := N - E_0 - I_0, \quad (1a)$$

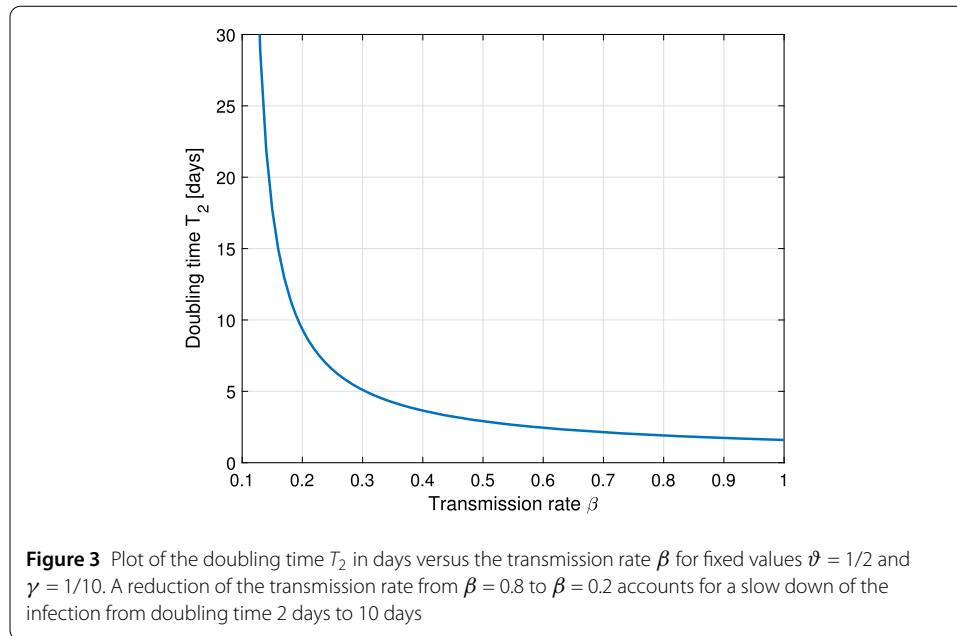
$$\frac{dE}{dt} = \frac{\beta}{N}SI - \vartheta E, \quad E(t_0) = E_0, \quad (1b)$$

$$\frac{dI}{dt} = \vartheta E - \gamma I, \quad I(t_0) = I_0, \quad (1c)$$

$$\frac{dR}{dt} = (1 - \mu) \cdot \gamma I, \quad R(t_0) = 0, \quad (1d)$$

$$\frac{dD}{dt} = \mu \cdot \gamma I, \quad D(t_0) = 0. \quad (1e)$$

The starting time  $t_0$  is chosen as 1 March and the initial conditions for the recovered and dead compartment are assumed to be zero, since in Germany the first COVID-19



related death was recorded on 9 March. Also we may assume that the number of recovered individuals by 1 March is negligible.

In the sequel, we will also consider two refined versions of the above basic model.

At the onset of the disease, the numbers of exposed, infected, recovered and dead are still small and the number of susceptibles is approximately equal to the entire population  $N$ . In this setting, the  $EI$ -part of the model reduces to

$$\begin{pmatrix} E \\ I \end{pmatrix}' = \begin{pmatrix} -\vartheta & \beta \\ \vartheta & -\gamma \end{pmatrix} \cdot \begin{pmatrix} E \\ I \end{pmatrix}.$$

The maximal eigenvalue  $\lambda$  of this linear system determines the initial growth rate and is given by

$$\lambda = \frac{1}{2}(-(\vartheta + \gamma) + \sqrt{(\vartheta - \gamma)^2 + 4\vartheta\beta})$$

and the *doubling time*  $T_2$  equals

$$T_2 = \frac{\ln 2}{\lambda}.$$

Figure 3 depicts the dependence of the doubling time on the transmission rate  $\beta$ . As of mid April, the doubling time in Germany is approximately 14 days compared to 2.5 days by mid March.

In the basic model (1a)–(1e), the transmission rate  $\beta$  is assumed to be fixed. The German state and federal governments introduced several measures to slow down the spread of the disease. Similar measures are nowadays taken in almost every country worldwide. As of 16 March schools, kindergartens and universities were closed and on 22 March a general contact ban was enforced in Germany. Both measures aim at reducing the transmission rate  $\beta$ .

To include this into the basic model (1a)–(1e), we also consider an alternative model for the transmission rate  $\beta$ : We assume  $\beta$  as a piecewise constant function on the time intervals prior to any measures, (until 15 March), after school closings (between 16 and 22 March) and after the contact ban (after March 22)

$$\beta(t) = \begin{cases} \beta_0 & t < 16 \text{ March}, \\ \beta_1 & 16 \text{ March} \leq t \leq 22 \text{ March}, \\ \beta_2 & t > 22 \text{ March}. \end{cases} \quad (2)$$

The resulting time-dependent SEIRD-model reads as

$$\frac{dS}{dt} = -\frac{\beta(t)}{N}SI, \quad S(t_0) = S_0 := N - E_0 - I_0, \quad (3a)$$

$$\frac{dE}{dt} = \frac{\beta(t)}{N}SI - \vartheta E, \quad E(t_0) = E_0, \quad (3b)$$

$$\frac{dI}{dt} = \vartheta E - \gamma I, \quad I(t_0) = I_0, \quad (3c)$$

$$\frac{dR}{dt} = (1 - \mu) \cdot \gamma I, \quad R(t_0) = 0, \quad (3d)$$

$$\frac{dD}{dt} = \mu \cdot \gamma I, \quad D(t_0) = 0. \quad (3e)$$

Setting  $\beta := \beta_0 = \beta_1 = \beta_2$ , the time-dependent model reduces to the basic one.

In order to validate our models and to identify the parameters involved therein, both the registered number of infections and the registered number of COVID-19 related deaths are important indications. The number of registered deaths is probably considerably more reliable, since the number of registered infections depends on the number of tests conducted and the dark figure of undetected, mostly asymptomatic cases, is assumed to be remarkably large [9]. We will discuss this point later in more detail. In the previous basic or time-dependent SEIRD-model, the actual increase of the disease related deaths  $\frac{dD}{dt}$  is assumed to be proportional to the current number of infected persons. The Robert-Koch-Institute specifies an average of 10 days between the onset of symptoms and admission to the intensive care unit [10]. Therefore, we assume  $\tau = 14$  for the time between the onset of infectiousness and death. In order to include this time lag into our model, we introduce a delay-term into the time-dependent model and obtain the final delayed time-dependent model:

$$\frac{dS}{dt} = -\frac{\beta(t)}{N}SI, \quad S(t_0) = S_0 := N - E_0 - I_0, \quad (4a)$$

$$\frac{dE}{dt} = \frac{\beta(t)}{N}SI - \vartheta E, \quad E(t_0) = E_0, \quad (4b)$$

$$\frac{dI}{dt} = \vartheta E - \gamma[(1 - \mu)I + \mu I(t - \tau)], \quad I(s) = I_0(s) \quad \text{for } s \leq t_0, \quad (4c)$$

$$\frac{dR}{dt} = (1 - \mu) \cdot \gamma I, \quad R(t_0) = 0, \quad (4d)$$

$$\frac{dD}{dt} = \mu \cdot \gamma I(t - \tau), \quad D(t_0) = 0. \quad (4e)$$

Note, that for solving this delay differential equation (DDE) we need an initial history of the infected compartment, i.e. values  $I_0(s)$  for  $t_0 - \tau \leq s \leq t_0$ .

In all the three models, the parameters  $\vartheta = 1/2$  [days<sup>-1</sup>],  $\gamma = 1/10$  [days<sup>-1</sup>] are assume to be fixed and resemble a latency period of 2 days and a recovery period of 10 days, see [2, Situation report 31 March 2020].

The parameters in the transmission rate, i.e.  $\beta$ , or  $\beta_0, \beta_1, \beta_2$  the lethality  $\mu$  and the initial values  $E_0, I_0$  resp. the initial history  $I_0(s)$  for the exposed and infected compartment are yet unknown to us. We will identify them together with the *detection rate*  $\delta$  by matching the model output to the given data. The detection rate  $\delta$  corresponds to the fraction of infected individuals which are positively tested for SARS-CoV-2 and hence appear in the official recordings. Various sources speculate that this detection rate is in the order of magnitude of 10–20% meaning that the true number of infected 5–10 times larger than the number published in the official statistics, see [9].

To match the model output and the reported data we use a least-squares approach. Let  $u = (\beta, \delta, \mu, E_0, I_0)$  resp.  $u = (\beta_0, \beta_1, \beta_2, \delta, \mu, E_0, I_0)$  denote the unknown model parameters to be determined. Furthermore, let  $Y(t)$  and  $Z(t)$  denote the data for the cumulated infected and dead cases at time  $t$  reported by Johns Hopkins University. The deviation between the model and the data is measured by the cost functional

$$\begin{aligned} J(u) &:= \frac{\|\delta(I + R) + D - Y\|_{L^2}^2}{\|Y\|_{L^2}^2} + c_1 \frac{\|D - Z\|_{L^2}^2}{\|Z\|_{L^2}^2} + c_2 \|u\|^2 \\ &= \frac{1}{\|Y\|_{L^2}^2} \left( \|\delta(I + R) + D - Y\|_{L^2}^2 + \omega_1 \|D - Z\|_{L^2}^2 + \omega_2 \|u\|^2 \right), \end{aligned} \quad (5)$$

where  $\|f\|_{L^2}^2 := \int_{t_0}^{T_{\text{Fit}}} f(t)^2 dt$  denotes the square of the  $L^2$ -norm of the function  $f$  on the interval  $[t_0, T_{\text{Fit}}]$  and  $\omega_1 = c_1 \frac{\|Y\|_{L^2}^2}{\|Z\|_{L^2}^2}$  as well as  $\omega_2 = c_2 \|Y\|_{L^2}^2$ . For the given data we have  $\|Y\|_{L^2}^2 \simeq 1.2 \cdot 10^{11}$  and  $\|Z\|_{L^2}^2 \simeq 6.5 \cdot 10^8$ , hence  $\omega_1 \simeq c_1 \cdot 185$ . The cumulated infected  $Y$ , i.e. total positive tests, are to be matched in the SEIRD-model to those individuals who had been infected until time  $t$ , i.e. the sum of the infected  $I$ , recovered  $R$  and deaths  $D$ . To account for the uncertainty in the *true* number of infected and recovered cases, we multiply both compartments by the detection rate  $\delta$ , which is itself part of the parameters to be identified. For the deaths we assume no undetected cases. By  $T_{\text{Fit}}$  we denote the time horizon used for the comparison between the model and the data. The regularization term  $\omega_2 \|u\|^2$  is included to ensure the convexity of the cost-functional. The weighting parameters  $c_1, c_2$  and hence  $\omega_1, \omega_2 > 0$  allow to balance the contributions from the least squares error in the fatalities and from the size of the parameter values themselves to the least squares error in the infected cases. The weight  $c_1$  for the fatal cases allows to compensate the different order of magnitude between the infected cases and the fatal cases, typically  $c_1 \simeq 2$ – $3$  leading to  $\omega_1 \simeq 500$ . The weight  $c_2$  is chosen small, such that the overall cost functional is still dominated by the least square fit between the model output and the given data.

The parameters  $u^*$  themselves are obtained from minimization problem

$$\min_u J(u) \text{ subject to one of the ODE-systems (1a)–(1e), (3a)–(3e) or (4a)–(4e).} \quad (6a)$$

$$u^* = \operatorname{argmin}_u J(u). \quad (6b)$$

### 3 A few analytical considerations

Due to the absence of demographic terms, our basic model (1a)–(1e) does not allow other equilibria besides the trivial disease free equilibrium  $X^0 = (N, 0, 0, 0, 0)$ . Since we focus only on the short-time behavior of the epidemics, demographic terms are excluded and equilibria do not play any important role.

An important issue is the question of whether we can identify the detection rate and lethality during the take-off period of the epidemics? The only data available for parameter identification are the total number of registered cases  $C = I + R + D$  and the deaths  $D$ . The total registered cases heavily depend on the number of tests conducted. If a person is infected, but not tested, this person will not appear in the official statistics. Hence, there is a presumably large dark figure in the officially recorded data. Our model parameter  $\delta$  takes this into account. The other, maybe more reliable, available data are the recorded deaths. Here we may assume that *all* COVID-19 related deaths are diagnosed and hence there is no dark figure in the  $D$ -compartment. A recent analysis by the Federal Statistical Office on the excess mortality in Germany for March and April 2020 confirms this assumption, see [11]. For other countries this assumption might be questionable, since they suffered from major COVID-19 outbreaks in care homes that did not enter the official statistics, e.g. in the UK, see [12].

However, one scenario could be possible. A large dark figure in the entire cases, i.e. a small detection rate  $\delta$  and a very small lethality could result in the same or at least similar observed data as a moderate or even small dark figure and hence large detection rate  $\delta$  combined with a higher lethality rate. In that setting a simultaneous identification of both, the detection rate  $\delta$  and the lethality  $\mu$  could be difficult due to their counteracting effects.

In order to investigate this scenario, we consider the simultaneous effect of the detection rate  $\delta$  scaling both the initial values of the  $E$  and  $I$  compartment to account for undetected cases together with a lethality  $\delta\mu$ . Removing the  $S$ -compartment by setting  $S = N - E - I - R - D$ , the basic SEIRD-system (1a)–(1e) reads as

$$\begin{aligned}\frac{dE}{dt} &= \frac{\beta}{N}(N - E - I - R - D)I - \vartheta E, & E(t_0) &= E_0/\delta, \\ \frac{dI}{dt} &= \vartheta E - \gamma I, & I(t_0) &= I_0/\delta, \\ \frac{dR}{dt} &= (1 - \delta\mu) \cdot \gamma I, & R(t_0) &= 0, \\ \frac{dD}{dt} &= \delta\mu \cdot \gamma I, & D(t_0) &= 0.\end{aligned}$$

The sensitivities  $\Sigma_E := \partial_\delta E$  and  $\Sigma_I, \Sigma_R, \Sigma_D$  of the solution with respect to the detection rate satisfy the system

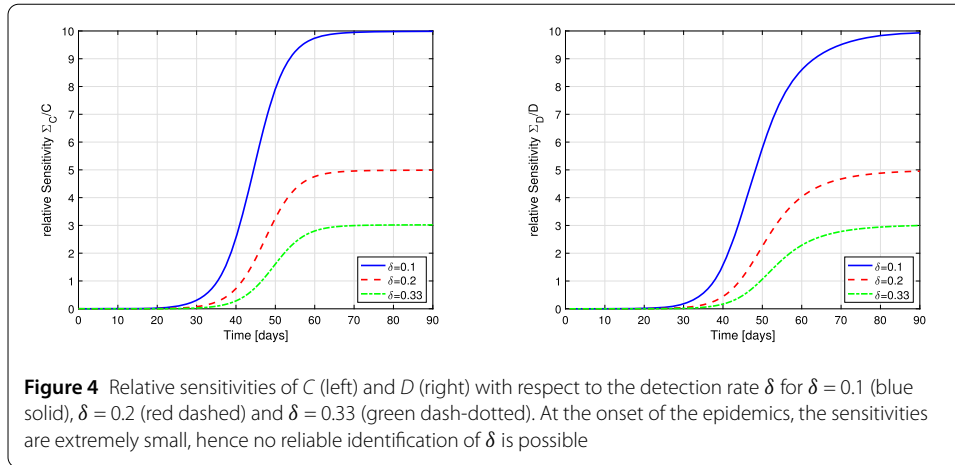
$$\Sigma'_E = \frac{\beta}{N}(N - E - 2I - R - D)\Sigma_I - \vartheta \Sigma_E - \frac{\beta}{N}I(\Sigma_E + \Sigma_R + \Sigma_D), \quad (7a)$$

$$\Sigma_E(t_0) = -E_0/\delta^2,$$

$$\Sigma'_I = \vartheta \Sigma_E - \gamma \Sigma_I, \quad \Sigma_I(t_0) = -I_0/\delta^2, \quad (7b)$$

$$\Sigma'_R = (1 - \delta\mu)\gamma \Sigma_I - \mu \gamma I, \quad \Sigma_R(t_0) = 0, \quad (7c)$$

$$\Sigma'_D = \delta\mu\gamma \Sigma_I + \mu \gamma I, \quad \Sigma_D(t_0) = 0. \quad (7d)$$



In Fig. 4 we show the relative sensitivities  $\Sigma_C/C$  and  $\Sigma_D/D$  for detection rates  $\delta = 0.1, 0.2$  and  $0.33$ . The chosen initial values are  $E_0 = 150$  and  $I_0 = 100$  (detected) cases at day 0. All other parameters resemble the assumed values for Germany. Note, that at the onset of the epidemics, i.e. in case of  $\delta = 0.1$  for  $t \lesssim 30$  and for  $\delta = 0.2, 0.33$  even for  $t \lesssim 40$ , the sensitivities are very small and hence the solution of the SEIR-model is almost *independent* of the particular value of the detection rate  $\delta$ . Hence  $\delta$  cannot be identified from the observed data in a reliable manner.

To illustrate these findings, we consider a linearization of a simplified SIR-model during the initial phase of the epidemics. We neglect the exposed compartment and assume that at the initial phase, the number of susceptibles is approximately equal to the entire population. Hence we get the linear system

$$\begin{aligned}\hat{I}' &= (\beta - \gamma)\hat{I}, & \hat{I}(0) &= \frac{1}{\delta}I_0, \\ \hat{D}' &= \delta\mu\gamma\hat{I}, & \hat{D}(0) &= 0\end{aligned}$$

with the solution

$$\hat{D}(t; \delta) = \frac{\mu\gamma}{\beta - \gamma} (e^{(\beta - \gamma)t} - 1) I_0.$$

In this linearized setting, the approximation  $\hat{D}$  for the dead compartment is independent of the detection rate  $\delta$ .

From the graphs in Fig. 4 one can conclude, the a significant dependence of the detected or dead compartment  $C$  resp.  $D$  is given only after the initial take-off period of the epidemic. In the setting of Germany, this implies, that during the month of March a reliable identification to the detection rate might not be possible.

#### 4 Adjoint equations and optimization

In order to solve the minimization problem (6a)–(6b), we use the adjoint equations, for details see [13, 14]. We introduce the Lagrangian

$$\mathcal{L}(t, x, u, z) = J(u) + \int_{t_0}^{T_{\text{Fit}}} z(t)^T \left( g(t, x, u) - \frac{dx}{dt} \right) dt.$$



Here  $z = (z_S, z_E, z_I, z_R, z_D)$  denotes the adjoint functions to the state variable  $x = (S, E, I, R, D)$  and  $g(t, x, u)$  denotes the right hand side of the ODE resp. DDE system. The gradient of  $\mathcal{L}$  with respect to the unknown parameters  $u$  is given by

$$\frac{\partial \mathcal{L}}{\partial \beta_i} = 2c_2 \beta_i + \frac{1}{N} \int_{t_0}^{T_{\text{Fit}}} \frac{\partial \beta(t)}{\partial \beta_i} S I (z_E - z_S) dt, \quad i = 0, 1, 2$$

$$\frac{\partial \mathcal{L}}{\partial \delta} = 2c_2 \delta + \frac{2}{\|Y\|_{L^2}^2} \int_{t_0}^{T_{\text{Fit}}} (I + R) [\delta(I + R) + D - Y] dt,$$

$$\frac{\partial \mathcal{L}}{\partial \mu} = 2c_2 \mu + \int_{t_0}^{T_{\text{Fit}}} \gamma I (z_D - z_I) dt,$$

$$\frac{\partial \mathcal{L}}{\partial E_0} = 2c_2 E_0 + z_E(t_0) - z_S(t_0),$$

$$\frac{\partial \mathcal{L}}{\partial I_0} = 2c_2 I_0 + z_I(t_0) - z_S(t_0).$$

Note, that in the case  $\beta = \beta_0 = \beta_1 = \beta_2$  we have  $\frac{\partial \beta(t)}{\partial \beta} = 1$ . By adding the time delay, we obtain

$$\frac{\partial \mathcal{L}}{\partial \mu} = 2c_2 \mu + \int_{t_0}^{T_{\text{Fit}}} \gamma I (z_I - z_R) + \gamma I(t - \tau) (z_D - z_I) dt.$$

The adjoint system reads as

$$\frac{dz_S}{dt} = \frac{\beta(t)}{N} I (z_S - z_E),$$

$$\frac{dz_E}{dt} = \vartheta (z_E - z_I),$$

$$\frac{dz_I}{dt} = \frac{\beta(t)}{N} S (z_S - z_E) + \gamma [z_I - z_R + \mu (z_R - z_D)] - \frac{2\delta}{\|Y\|_{L^2}^2} [\delta(I + R) + D - Y],$$

$$\frac{dz_R}{dt} = -\frac{2\delta}{\|Y\|_{L^2}^2} [\delta(I + R) + D - Y],$$

$$\frac{dz_D}{dt} = -\frac{2}{\|Y\|_{L^2}^2} [\delta(I + R) + D - Y] - \frac{2c_1}{\|Z\|_{L^2}^2} (D - Z).$$

supplemented by the terminal condition  $(z_S, z_E, z_I, z_R, z_D)(T_{\text{Fit}}) = 0$ . In the case of the time delay we receive

$$\begin{aligned} \frac{dz_I}{dt} &= \frac{\beta(t)}{N} S (z_S - z_E) + (1 - \mu) \gamma (z_I - z_R) - \frac{2\delta}{\|Y\|_{L^2}^2} [\delta(I + R) + D - Y] \\ &\quad + \mu \gamma [z_I(t + \tau) - z_D(t + \tau)] \cdot \chi_{[t_0, T_{\text{Fit}} - \tau]}(t). \end{aligned}$$

Here  $\chi_{[a,b]}(t)$  denotes the characteristic function of the interval  $[a, b]$ , i.e. we define  $\chi_{[a,b]}(t) = 1$  for  $t \in [a, b]$  and  $= 0$  otherwise.

To solve the optimization problem (6a)–(6b) numerically, we apply the Forward-Backward Sweep method [13] combined with a Quasi-Newton method (BFGS) [15].

In each iteration step the ODEs and DDEs of the state variables and adjoint equations are solved with Runge–Kutta methods before the corresponding gradient and direction of descent can be determined. The algorithm stops as soon as the termination condition  $\|J(u_{k+1}) - J(u_k)\| < \text{TOL}$  is fulfilled.

As initial values we use  $\beta = \beta_0 = \beta_1 = \beta_2 = 0.3$  for the transmission rate. This is justified by the fact that an average Basic Reproduction Number of about  $\mathcal{R}_0 = 3$  is assumed and in our basic model we have

$$\mathcal{R}_0 = \frac{\beta}{\gamma}.$$

Epidemiologically,  $\mathcal{R}_0$  indicates the number of new infections an infected individual causes during the infectious period in an otherwise susceptible population. For the sake of simplicity, we assume the same starting value for  $I_0$  and  $E_0$ . This corresponds to the value at the first data point of our measurement, i.e. 130 registered infected persons on 1st March. As already mentioned, we assume that for the recovered and deaths at this time  $R_0 = D_0 = 0$  holds. The possible problems with the optimization of  $\delta$  and  $\mu$  were already mentioned in the previous section. To increase the probability of generating a global minimum, we use  $n = 1000$  normally distributed start values for both parameters fulfilling  $\delta \sim \mathcal{N}(0.25, 0.25^2)$  and  $\mu \sim \mathcal{N}(0.03, 0.03^2)$  with  $\delta, \mu > 0$ . The algorithm selects the best result of these  $n$  data fits. The reason for this is the assumption that the proportion of detected cases is between 1–50% and the lethality below 6%. For the case fatality rate  $\mu$  we have

$$\frac{Z}{Y} \leq \mu \leq \frac{Z}{R_{\text{reported}}/\delta + Z}, \quad (8)$$

where  $R_{\text{reported}}$  stands for the reported recovered at time  $T_{\text{Fit}}$ . The approach for this estimation can be found in [10]. The smaller the detection rate  $\delta$ , the lower the upper bound for the case fatality gets.

In case of the time delay we choose as initial history for  $s \in [t_0 - \tau, t_0]$

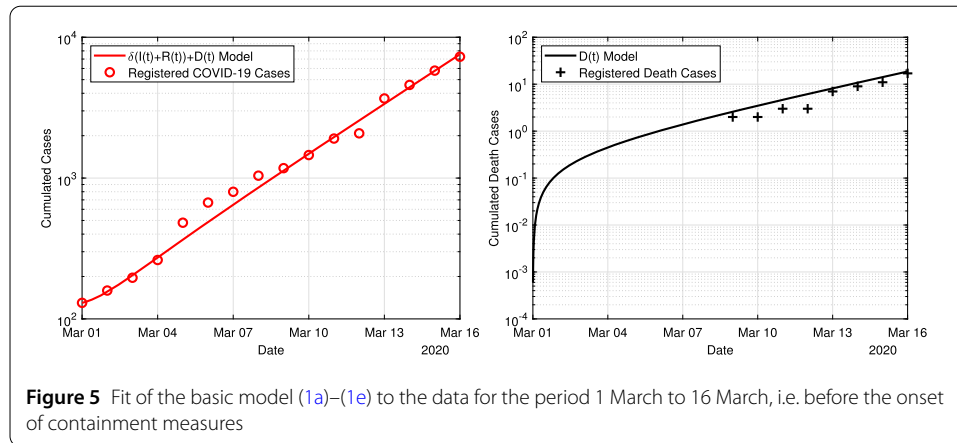
$$I(s) = I_0 \exp\left(-\frac{\ln(0.1)}{\tau}(s - t_0)\right).$$

This is justified by the fact that the number of registered cases has increased tenfold during this period and we assume an exponential growth in this time span.

## 5 Simulation results

To estimate the unknown parameters  $u$ , we match the data reported on a daily basis by Johns Hopkins [5] to our simulation results for a time period starting on 1 March.

The first results in Fig. 5 show a parameter estimation using the basic model (1a)–(1e) and the time period before the onset of any containment measures, i.e. before the closing of schools on 16 March. We fitted the parameters  $\beta$ ,  $\delta$  and  $\mu$  along with the initial values  $E_0$  and  $I_0$  over the time period 1 March to 16 March. The initial values  $E_0$  and  $I_0$  are also subject to fitting, since the official data does not provide information about the active infections at a given day. The weight  $\omega_2 = 1$  to keep the cost functional dominated by the two least square errors. The other weight is chosen as  $\omega_1 = 500$  to compensate the significantly smaller value of the least square error in the fatal cases.



**Table 1** Optimal parameter values for the three models (1a)–(1e), (3a)–(3e) and (4a)–(4e) obtained from the minimization problem (6a)–(6b)

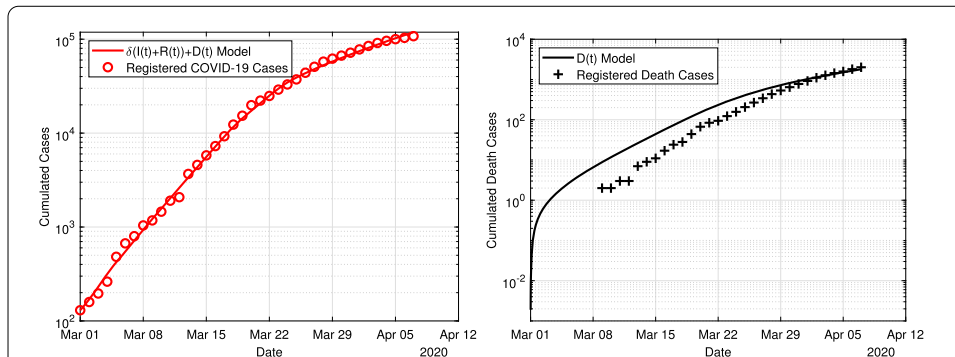
Parameter		$\beta_0$	$\delta$	$\mu$	$E_0 + I_0$	$\beta_1$	$\beta_2$	$L^2$ -diff
Fit until	Model							
16.03.20	basic	0.566	0.372	0.0034	418	–	–	0.3771
07.04.20	time-dep	0.523	0.308	0.0087	659	0.3561	0.1788	0.2721
07.04.20	delay	0.553	0.202	0.0389	930	0.3578	0.1415	0.2242

For the given time period of the fit, the model prediction and the observed data are in good accordance. The estimated parameter values are given in Table 1. The detection rate was estimated as  $\delta = 0.37$  implying that the *true* number of infections exceeds the registered cases by a factor 3. The transmission rate  $\beta = 0.57$  accounts for a doubling time of 2.6 days at the initial, uncontrolled phase of the epidemic in Germany.

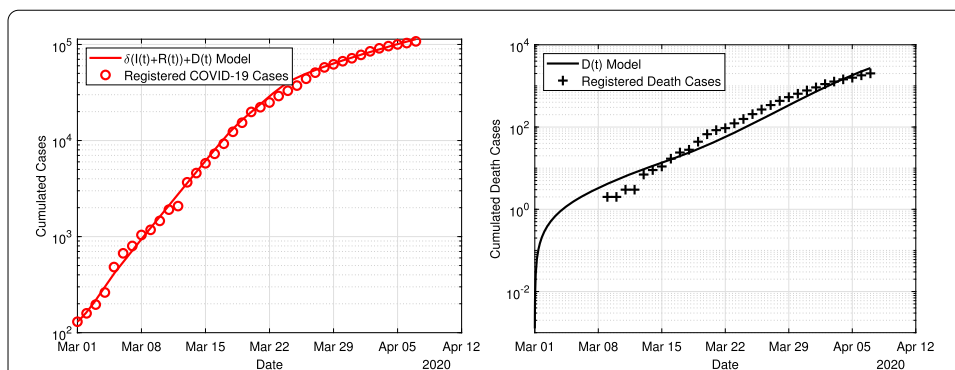
In Fig. 6 we show the results obtained with the time-dependent model (3a)–(3e). In this case, the fitting period equals to the entire simulation period starting from 1 March to 7 April. The weights  $\omega_1, \omega_2$  are identical to the previous simulation. The obtained transmission rate and according doubling times change from  $\beta_0 = 0.5232$  and  $T_2(\beta_0) = 2.8$  days at the initial uncontrolled phase to  $\beta_2 = 0.18$  and  $t_2(\beta_2) = 11.4$  days after the contact ban has been introduced. The effect of the contact ban effectively reduces the transmission rate by a factor of about 3 and significantly slows down the speed of the epidemics by increasing the doubling time by a factor 4.

In Fig. 7 we show the result obtained with the delay model (4a)–(4e). For the delay model, we assume a delay of 14 days between entering the class of infected and death. Again, we show the simulation results compared to the reported cases for the infections and deaths. Quite good agreement is found between the model and the simulation for both, infections and deaths. Compared to the time-dependent model, shown in Fig. 6, the delay model agrees better in particular for the fatal cases. In Table 1 we have listed the estimated parameter values in for the three models. We have also included the normalized  $L^2$ -difference between the simulation outcome and the given data, i.e. the first two summands from the cost functional (5). A  $t$ -test revealed that the deviations of the simulation to the reported data is not normal distributed at a significance level of 5%.

In the simulations the detection rate is found to be 20–40%, indicating that the *true* number of SARS-CoV-2 infections might be 3–5 times higher than the officially recorded



**Figure 6** Fit of the time-dependent model (3a)–(3e) to the data for the period 1 March to 7 April



**Figure 7** Fit of the delay model (4a)–(4e) to the data for the period 1 March to 7 April

data suggest. The lethality rate is found to be rather small, taking into account the large number of *true cases*.

Comparing the obtained values for the lethality, the value for the delay-model seems to be most realistic, since in this model we compare the fatal cases today to the infections that occurred two weeks ago. The two other models related the fatal cases of today to the infected cases *today*, hence to a significantly larger number. Therefore in these two models, the lethality rate seems to be smaller.

## 6 Conclusions and outlook

We present three SIR-based models for describing the outbreak of the SARS-CoV-2 outbreak in Germany. Besides a standard SEIR-model, we consider an extension taking into account the effect of social distancing by a time-dependent reduction of the transmission rate. The third model introduces a delay-term to accurately describe the deaths depending on infected cases that occurred several days in the past. Comparing the simulation results to the data published by Johns Hopkins University allows an estimation of the unknown model parameters. Best results are obtained using the delay equation model. In this setting, we find a detection rate of about 20% and a lethality of about 4%. The social distancing measures were leading to an effective reduction of the transmission rate by a factor 4. That is, after the introduction of the measures roughly just 25% of the social contact compared to the initial period were leading to infections.

**Acknowledgements**

The authors would like to thank the MOCOS (MOdeling CORona Spread) International Group consisting of research groups at the universities in Kaiserslautern, Koblenz, Trier and Wrocław for valuable discussions on this urgent and internationally highly relevant issue. We also express our gratitude to the unknown referees for helping with their comments to improve the manuscript.

**Funding**

Not applicable.

**Availability of data and materials**

The used data on registered COVID-19 cases from Johns Hopkins university is available, see [5].

**Competing interests**

The authors declare that they have no competing interests.

**Authors' contributions**

Both authors contributed equally. All authors read and approved the final manuscript.

**Authors' information**

Not applicable.

**Author details**

<sup>1</sup>Mathematical Institute, University Koblenz-Landau, D-56070, Koblenz, Germany. <sup>2</sup>Magister Laukhart IGS Herrstein-Rhaunen, D-55756, Herrstein, Germany.

**Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 30 April 2020 Accepted: 6 July 2020 Published online: 10 July 2020

**References**

1. World Health Organization (WHO). Novel Coronavirus (2019-nCoV) situation reports. [www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports](http://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports).
2. Robert-Koch-Institute. Daily situation reports. [www.rki.de/DE/Content/InfAZ/N/Neuartiges\\_Coronavirus/Situationsberichte/Gesamt.html](http://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Situationsberichte/Gesamt.html).
3. Wikipedia. 2020 coronavirus pandemic in Germany. [https://en.wikipedia.org/wiki/2020\\_coronavirus\\_pandemic\\_in\\_Germany#Statistics](https://en.wikipedia.org/wiki/2020_coronavirus_pandemic_in_Germany#Statistics)
4. Federal Government of Germany. Guidelines for reducing social contacts. [www.bundesregierung.de/breg-de/themen/coronavirus/besprechung-der-bundeskanzlerin-mit-den-regierungschefinnen-und-regierungschefs-der-laender-1733248](http://www.bundesregierung.de/breg-de/themen/coronavirus/besprechung-der-bundeskanzlerin-mit-den-regierungschefinnen-und-regierungschefs-der-laender-1733248) (in German).
5. Johns Hopkins University. Time series of confirmed COVID-19 cases globally. [github.com/CSSEGISandData/COVID-19/blob/master/csse\\_COVID\\_19\\_data/csse\\_COVID\\_19\\_time\\_series/time\\_series\\_COVID19\\_confirmed\\_global.csv](https://github.com/CSSEGISandData/COVID-19/blob/master/csse_COVID_19_data/csse_COVID_19_time_series/time_series_COVID19_confirmed_global.csv).
6. Robert-Koch-Institute. Why differ the information reported by RKI and Johns Hopkins University? (in German). <https://www.rki.de/SharedDocs/FAQ/NCOV2019/gesamt.html>.
7. Martcheva M. An introduction to mathematical epidemiology. Berlin: Springer; 2015.
8. WHO. Q&A on coronaviruses. [www.who.int/news-room/q-a-detail/q-a-coronaviruses](http://www.who.int/news-room/q-a-detail/q-a-coronaviruses)
9. Magal P, Webb G. Predicting the number of reported and unreported cases for the COVID-19 epidemic in South Korea, Italy, France and Germany. medRxiv <https://doi.org/10.1101/2020.03.21.20040154>, 2020.
10. Robert-Koch-Institute. Corona fact sheet. [https://www.rki.de/DE/Content/InfAZ/N/Neuartiges\\_Coronavirus/Steckbrief.html](https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Steckbrief.html).
11. Federal Statistical Office: Press release No. 179 of 22 May 2020. [https://www.destatis.de/EN/Press/2020/05/PE20\\_179\\_12621.html](https://www.destatis.de/EN/Press/2020/05/PE20_179_12621.html).
12. The Guardian: UK care home Covid-19 deaths may be five times government estimate'. <https://www.theguardian.com/world/2020/apr/18/uk-care-home-covid-19-deaths-may-be-five-times-government-estimate> (18 April 2020).
13. Lenhart S, Workman JT. Optimal control applied to biological models. Boca Raton: CRC Press; 2007.
14. Vinter R. Optimal control. Berlin: Springer; 2010.
15. Nocedal J, Wright S. Numerical optimization. Berlin: Springer; 2006.