Numerical Simulations of a Modified SIR Model Fitting Statistical Data for COVID-19

FLAVIUS GUIAŞ
Department of Mechanical Engineering
Dortmund University of Applied Sciences and Arts
Sonnenstr. 96, 44139 Dortmund
GERMANY

Abstract: We consider a system of ordinary differential equations obtained by modifying the classical SIR model in epidemiology in order to account for the particular features of COVID-19 and the structure of the available statistical data. Its main feature is that the infectious state is being split in two different stages. In the first one, which lasts a few days after being infected, the individuals are considered to be contagious and able to spread further the disease. After this, the individuals are considered to be isolated and this second stage lasts until either recovery or death is reported. The parameters of the model are fitted for several countries (Germany, Italy, Spain, Russia, USA, Romania) such that the solution matches the known number of new cases, active cases, recoveries and deaths. The values of these parameters give insight regarding the evolution of the pandemy and can reveal different policies and approaches in reporting the official data. For example one of them can indicate that in certain countries a substantial amount of cases were reported only post-mortem. The variation across several countries of another parameter, which models the average convalescence time (the duration of the second stage of the infectious state), points to the fact that the recoveries are reported at different rates, in some cases with significant delays. Since it can be assumed that this is only a matter of reporting, we also perform additional simulations for these countries by taking the average convalescence time the value of Germany, which is the smallest within the whole range. The conclusion is that under this assumption, the evolution of the active cases for example in Italy and Spain, is not significantly different to that in Germany, the comparison being based on the fact that these countries showed a similar number of cases within the considered period.

Key-Words: ordinary differential equations, SIR model, parameter fitting, numerical simulations, COVID-19

Received: June 1, 2020. Revised: November 30, 2020. Accepted: December 4, 2020. Published: December 15, 2020.

1 Introduction

The basic model which describes the time evolution of an epidemy is the SIR model introduced in 1927 by Kermack and McKendrick. Its principle lies at the core of the so called compartmental models in epidemiology, see [1], [2]. In the original setting the population is divided into several compartments according to the states: S - *susceptible*, I - *infected* and R - *recovered*. The variables S, I, R denote the number of individuals in the corresponding state, while the dynamics of the transitions between different states are described the following system of ordinary differential equations:

$$\begin{aligned} \frac{dS}{dt} &= -\beta \cdot I \cdot \frac{S}{N} \\ \frac{dI}{dt} &= \beta \cdot I \cdot \frac{S}{N} - \alpha \cdot I \\ \frac{dR}{dt} &= \alpha \cdot I \end{aligned} \tag{1}$$

Susceptible individuals get infected at a rate which is proportional to the numbers S and I, while the transition from I to R occurs at a rate proportional to the number I of infected individuals. The parameter β is called the *transmission rate* and α is the recovery rate. An individual which has recovered is considered to be immune and cannot contribute anymore to the spread of the disease. The number N denotes the size of the whole population and the presence of the factor N^{-1} is necessary for the correct scaling of the bilinear terms.

The SEIR model, see [1], [2], considers the additional state E of *exposed* individuals during the incubation time, situated between S and I, where they don't show yet the symptoms and cannot spread the disease.

In context of the outbreak of the COVID-19 pandemy at the beginning of 2020, several compartmental models of type SIR, SEIR and modifications of them were applied in order to describe the evolution of this disease. The papers [3] and [4] present

an overview of such models with applications to COVID-19. Applications to concrete countries are presented in [5], [6], [7], where the authors employ extensions of the SEIR model.

An important aspect related to the epidemic models based on differential equations is to fit the parameters in order that the components of the solution match given statistical data from certain countries. Here we mention the works [8], [9], [10], [11] which use different algorithms in order to compute the optimal parameters for models of SIR or SEIR type. Application of discrete models of SEIR type based on difference equations with parameter fitting are reported in [12], [13]. In the papers [11], [13] the transmission rate β is considered to be time-dependent, while in all other references mentioned above this parameter is taken as constant in time. However, this latter choice is realistic only for shorter time intervals or for the simulation of theoretical scenarios. In paper [10] the recovery and mortality rates are considered to be time-dependent.

In this paper we consider a new modification of the SIR model with time-dependent transmission rate in order to account for the features of the COVID-19 disease. The parameters of the ODE-system are fitted by an optimization algorithm based on the novel method introduced in [15] in order to match the statistical data available at https://www.worldometers.info/coronavirus/. The statistics record the numbers of new infections, the current infections, the recoveries and the deaths. Since data for E are not available, we will not consider a model of SEIR type. The argument for this choice is similar to that presented in [8]: since we want to identify the parameters of the model, but we do not know at what stage of the infection a person enters in the statistics, introducing a latent or exposed state beyond the active one would be an overcomplexification with respect to the actual data uncertainty. Instead of this approach, we introduce further states as described below.

Firstly, we add a component D, which counts the death cases and consider transitions into D from the state I and sometimes directly from S, in the case of infections reported only after the death of the individual. Secondly, unlike in the case of an usual epidemy, an individual infected by SARS-CoV-2 can be contagious also 1-2 days before the onset of the symptoms and in principle up to 10 days after that. However, in practice a symptomatic individual is in most cases isolated or quarantined, and therefore the time of effectively spreading the disease can be considered basically up to a few days after the onset of the symptoms. In [14] this time is considered to be of about only 4-5 days. This means that a person is highly infectious 1-2 days before the onset of the symptoms and 2-3 days after that.

For the infectious state we will consider therefore two stages. In the first one, I_1, which lasts relatively short, the individual can spread the disease, while death is unlikely to occure. The following stage I_2 lasts until the individual is reported as recovered or dead and for this category we assume that it does not contribute to the spread of the disease. The statistical data contain nevertheless only the aggregate value $I = I_1 + I_2$ and in order to initialize the simulations we need an additional parameter f defined as the ratio between I_1 and I at the beginning of the computation.

We consider therefore the following system of ordinary differential equations:

$$\begin{aligned}
\frac{dS}{dt} &= -\frac{R(t)}{T_{inf}} \cdot I_1 \cdot \frac{S}{N} - \tilde{\mu}_d \cdot S \\
\frac{dI_1}{dt} &= \frac{R(t)}{T_{inf}} \cdot I_1 \cdot \frac{S}{N} - T_{inf}^{-1} \cdot I_1 \\
\frac{dI_2}{dt} &= T_{inf}^{-1} \cdot I_1 - T_{conv}^{-1} \cdot I_2 - \mu_d \cdot I_2 \\
\frac{dR}{dt} &= T_{conv}^{-1} \cdot I_2 \\
\frac{dD}{dt} &= \mu_d \cdot I_2 + \tilde{\mu}_d \cdot S
\end{aligned} \tag{2}$$

In this model we consider the form $\beta(t) = R(t) \cdot T_{inf}^{-1}$. The term $R(t)$ denotes the time dependent *effective reproduction number*, i.e. the average number of further infections produced by the contacts with one infectious individual. At the beginning of the epidemy its value is equal to R_0, the so called *basic reproduction number*, but after that it may vary due to restrictions, social distancing, variable intensity of testing, etc. T_{inf} denotes the average time spent in the infectious state I_1 and consequently we have $\alpha = T_{inf}^{-1}$ for the transition rate from I_1 to I_2.

In the setting of the present paper we will consider that $R(t)$ has constant values over 4 consecutive days. Daily variations would increase significantly the number of parameters to be fitted and would also leed to sharp fluctuations of the solution curve, while considering larger time intervals would possibly smooth out too much from the details of the computed profile. We chose therefore the value which delivers a good compromise between a setting with a high computational complexity and possible computational artifacts and a setting with too much averaging, which misses essential details.

By observing the statistical data we can remark that while death cases are reported more or less timely, the average time elapsed until recovery is reported may vary from one country to another, being usually much larger than the effective time of being

physically ill. By T_{conv} we denote therefore the average convalescence time, i.e. the time spent in the state I_2 until being moved into the state R. Alternatively, individuals in state I_2 can die at the daily death rate μ_d. For countries where the reporting policy counts COVID-19 cases also if the positive testing occured only after the death of the person, we will assume a positive death rate $\tilde{\mu}_d$ which describes the transition directly from S to D. Otherwise this parameter is considered to be 0.

The goal of this paper is to perform numerical simulations of the model (2) with parameters chosen in order to fit the statistical data. Since $S + I_1 + I_2 + R + D = N$, for comparison we take as reference the available data for $I = I_1 + I_2$, R and D. The statistics record also the daily number of new cases, but this quantity influences the reproduction number $R(t)$ and thus indirectly the value of I. The daily changes in the statistical data set are used within the method of fitting the parameters introduced by the present author in [15] and applied also in the current paper. As numerical solver for the ODE system we use a variant of a stochastic Runge-Kutta method. For computing a predictor, instead of simulating a jump process as in the original stochastic method, or performing an Euler step as in the classical Runge-Kutta method, the scheme used here and in the mentioned reference takes into account the variations of the data set, after which the precision is improved by using correction steps of Runge-Kutta type. The solutions computed in this way are subjected to an optimization procedure which searches for the set of parameters which minimizes the mean square error between the computed and the given data.

In Sections 2-7 of the present paper we perform several numerical experiments for Germany, Italy, Spain, Russia, USA and Romania under different conditions and assumptions. The simulation for each country starts from the day where the number of cases was around 100. Unlike in other works in this area, where the considered time intervals are considerably shorter, in the present paper the parameter fitting is performed according to the method described above for 100 days but also for 64 days, in order to compare the values of $T_{inf}, T_{conv}, \mu_d, \tilde{\mu}_d$ which in the current model are assumed to be constant over time. The only exception is Spain, where these periods are of 80 and 56 days, since here from a certain date onwards the recoveries weren't reported anymore.

The parameter T_{inf} is intrinsic to the disease and this fact is confirmed also by the similar values between 3 and 4 for the considered countries, with the exception of the USA, where it is smaller by one unit. The values corresponding to the time intervals of 64 and 100 days do not exhibit significant changes.

The parameter T_{conv} describes basically the average time spent by an individual in the state I_2 until he is reported in the statistics as recovered. Its value shows major differences from country to country, reflecting significant differences in the reporting policies. In the cases of Germany (with $\tilde{\mu}_d = 0$) and Spain (with $\tilde{\mu}_d > 0$) the values of this parameter for the time intervals of 64 and 100 days show a remarkable stability, since the values don't change in a significant way. For the other countries we can notice that the value corresponding to the interval of 100 days is smaller than for 64 days, which reflects that the recoveries were not reported at a constant rate. At the beginning of the epidemy the average convalescence time was considered to be larger, while with the passing of time and presumably a better organization, the corresponding time interval could be reduced.

For Germany we have a value for T_{conv} of about 12 days, the smallest among all considered countries, which means that the reporting of the recoveries occurs relatively fast. This number plus the value of T_{inf} between 3 and 4 gives an average duration of the disease of about 16 days. As mentioned in [14], an infected individual can spread the virus also 1-2 days before the onset of the symptoms, so we have a time range of about 14 days after this moment, which corresponds basically to the recommended quarantine period for COVID-19.

Therefore, a value of T_{conv} smaller than the reference value of 12 can hardly be imagined and can not correspond to any reasonable reporting policy. However, for certain reasons which root in the different reporting policies, the value of T_{conv} can be larger, ranging between 18 in the case of Spain to up to 80 in the case of the USA.

Nevertheless, for this parameter we can take the reference value as being that of Germany and consider that the differences to other countries consist basically in delays in reporting the recoveries. The consequence of this fact is that the true profile of the curve of active cases I in these countries might be actually much lower than that reflected by the official statistics. We perform therefore simulations by taking for the other countries the reference value of T_{conv} from Germany. This approach can deliver a uniform basis in order to compare the evolution of the figures in the considered countries.

The parameter μ_d represents the transition rate from I_2 into D and the parameter $\tilde{\mu}_d$ is the rate of new cases reported only at the death. That is, in our model we have a direct transition from the state S into D. The basic assumption is $\tilde{\mu}_d = 0$, but for severals countries it turns out that assuming such transitions explains better the profile of the curve of the death cases D.

In Section 8 we perform a comparison of the infection dynamics between the different countries, as

well as a sensitivity analysis regarding the reproduction number. In our model we assume that the reproduction number $R(t)$ is time dependent, having a constant value over 4 consecutive days. This ensures a high degree of flexibility in order to obtain a parameter set providing a good approximation of the profile of the curve of the active infections $I = I_1 + I_2$. Plotting the time evolution of $R(t)$ for several countries shows basically a similar behaviour. Although the differences between the profiles are not major ones, a sensitivity analysis for this parameter shows that the model is highly sensitive with respect to the reproduction number. Comparing the curve of active infections obtained for the parameter vectors $0.95R$, R and $1.05R$, where R is the computed optimized parameter vector for Germany, we can see that the peak of the curve of I has for $0.95R$ about half of the height of the curve for R, while for $1.05R$ the increasing factor is of about 2.

The results of this paper are summarized and discussed in Section 9, while the final section is dedicated to conclusions and outlook.

In this paper all values of the parameters are presented rounded to two decimal digits, but in the computations we used the full precision.

2 Simulations for Germany

The values of the time-independent parameters for which our model fits the data for Germany starting at 01.03.2020 are given in Table 1.

days	T_{inf}	T_{conv}	$\mu_d/\tilde{\mu}_d = 0$	$\mu_d/\tilde{\mu}_d$	imp.err.
64	3.97	12.38	0.0043	-	-
100	3.92	11.95	0.0044	0.0034 / 2.54e-7	0.078%

Table 1: Parameters for Germany

The following explanations of the structure of the table are valid also for the simulations performed for the other countries.

We compare first the fitted values of the time independent parameters over time ranges of 64 and 100 days in order to check the assumption of time homogeneity. The values in the case of Germany suggest strongly that this is indeed the case. Nevertheless, our main focus will be laid on the simulation over 100 days and the values for the other parameters are considered only for this situation. For this reason, the row in the table for each country corresponding to the simulations over 64 days contains only the data for the most relevant of the alternatives $\tilde{\mu}_d = 0$ or $\neq 0$.

The 4th column of the table represents the value of μ_d estimated under the assumption $\tilde{\mu}_d = 0$, while the 5th column contains the estimated values of μ_d and $\tilde{\mu}_d$ without any restriction. One one these two columns is marked with grey, which means that the

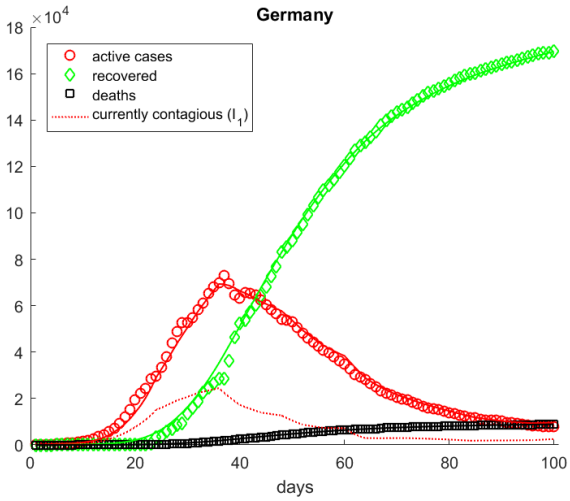

Figure 1: Simulated curves and data for Germany

corresponding assumption will not be considered in our final analysis. The choice for Germany is motivated by the fact that considering $\tilde{\mu}_d \neq 0$ leads to no significant improvement in the error relative to the data set, which in this case is of only 0.078%. This improvement factor is written in the last column of the table. Another indicator for the fact that we can consider $\tilde{\mu}_d = 0$ is the order of magnitude of $\tilde{\mu}_d$ which is of 10^{-7}. As we will see for other countries, the assumption $\tilde{\mu}_d \neq 0$ can lead to an improvement of the error up to 15-19% and to an order of magnitude for $\tilde{\mu}_d$ of 10^{-6}. The small values of this parameter in both cases is explained by the fact that it represents the transition rate from S to D, where S denotes the number of susceptibile individuals, which is a very large one, close to the total population of the country.

The other parameters for Germany are $f = 0.66$ and R =[2.62 1.20 2.14 2.17 1.81 1.66 1.24 1.15 1.09 0.66 0.79 0.90 0.64 0.83 0.98 0.12 0.99 1.01 0.90 0.87 0.77 1.04 1.02 1.05 1.20]. The vector R has 25 elements, since we perform the simulation for 100 days and assume that $R(t)$ is constant over 4 consecutive days.

The results of the simulation with the given values of the parameters are plotted in Figure 1.

The plots corresponding to each country contain the simulated curves and the data for $I = I_1 + I_2, R, D$, that is the active cases, the recoveries and the deaths, as well as the simulated values for I_1. For Germany we can notice a very good agreement between data and simulations and we can conclude that our model, altough minimalistic, is in principle a correct approach for describing the time evolution of the COVID-19 epidemy.

3 Simulations for Italy

The values of the time-independent parameters for which our model fits the data for Italy starting at 22.02.2020 are presented in Table 2.

days	T_{inf}	T_{conv}	$\mu_d/\tilde{\mu}_d = 0$	$\mu_d/\tilde{\mu}_d$	imp.err.
64	3.61	45.23	0.0070	-	-
100	3.58	37.83	0.0077	0.0028 / 4.03e-6	19.83%

Table 2: Parameters for Italy

The other parameters are $f = 0.78$ and R =[2.70 1.97 1.59 1.62 1.58 1.41 1.46 1.07 1.09 1.03 0.96 0.97 0.97 0.93 0.84 0.90 0.85 0.64 0.50 0.55 0.24 0.34 0.30 0.31 0.42].

We can notice that the assumption $\tilde{\mu}_d = 0$ does not capture correctly the profile of the curve of the deaths, as can be seen in Figure 2. However, if we allow transitions from S to D, according to the table above (last column) the approximation error is improved by 19.83% , which is a significant figure, and the value of $\tilde{\mu}_d$ is of order 10^{-6}, being larger by a factor 10 compared to the same value estimated for Germany. The simulation under this assumption shows that the curve corresponding to D is approximated much better, as can be seen in Figure 3. Here the dotted lines correspond to the deaths originating form the states I_2 and S.

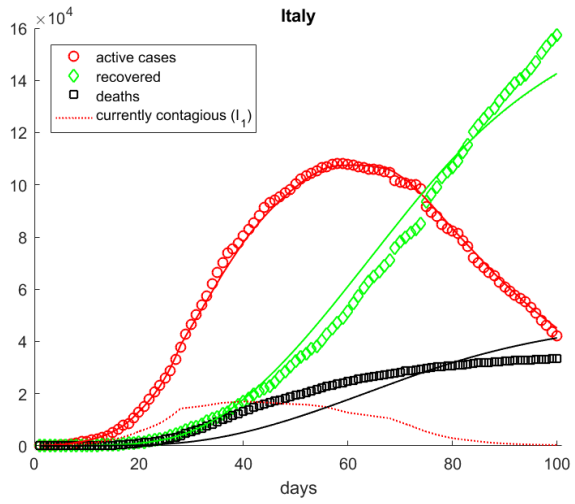

Figure 2: Simulated curves and data for Italy, $\tilde{\mu}_d = 0$

We conclude therefore that our simulations reflect the known fact that in Italy the tests for COVID-19 were performed also after the death of the individuals, which in this case enter into the statistics of the total cases only post mortem, directly from S into the category D. According to the results of the present model, which have to be interpreted essentially in a qualitative manner, more than half of the deaths in

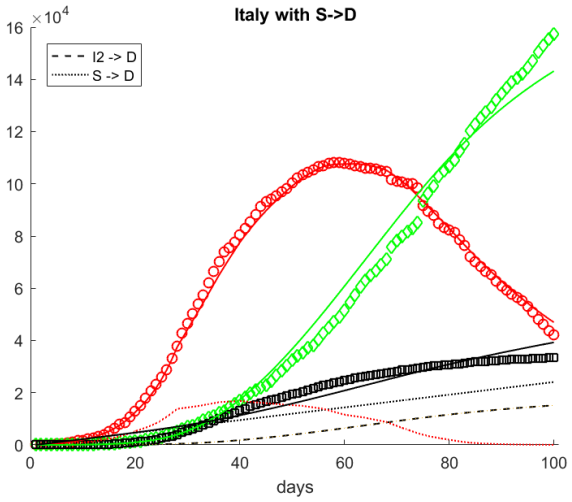

Figure 3: Simulated curves and data for Italy, $\tilde{\mu}_d \neq 0$

Italy which were associated to COVID-19 were not of patients which were positively tested during their lifetime, but only after their death.

Moreover, comparing the values of T_{conv} estimated for 64 and 100 days, we can remark that the difference is also significant: approximately 45 and 38 days respectively. This is an indicator that in Italy the reporting of the recoveries didn't take place at a constant rate as it was the case in Germany.

Secondly, the value of T_{conv} for Italy is much larger than 12 (the value for Germany), which indicates that the official statistics don't keep up the pace with the effective situation of the recoveries.

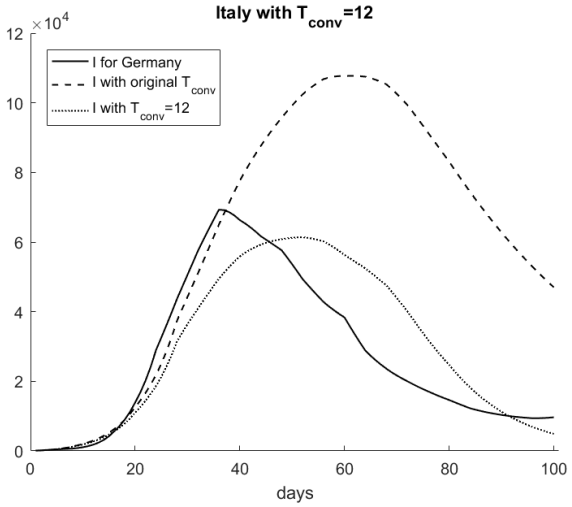

Figure 4: Active cases for Italy with $T_{conv} = 12$.

Figure 4 shows therefore a comparison of the curves of the active cases for Germany and Italy with the estimated parameters in order to fit the data, but

additionally also with the curve computed for Italy by taking $T_{conv} = 12$. This is basically the same value as for Germany and on average it corresponds to the real-life duration of the disease in its second part, after the initial stage of $T_{inf} \approx 4$ days. We can remark that in this case the curve of the active cases shrinks significantly and its behaviour is not substantially different from the curve corresponding to Germany.

4 Simulations for Spain

The values of the time-independent parameters for which our model fits the data for Spain starting at 01.03.2020 are given in Table 3. Here we perform the simulations only over maximally 80 days, since for Spain no further data regarding the recovered cases are available.

days	T_{inf}	T_{conv}	$\mu_d/\tilde{\mu}_d = 0$	$\mu_d/\tilde{\mu}_d$	imp.err.
56	3.06	19.59	-	0.0084/ 3.14e-6	-
80	3.19	18.08	0.0077	0.0036 / 5.20e-6	16.84%

Table 3: Parameters for Spain

The other parameters are $f = 0.25$ and R =[2.50 2.39 2.19 1.61 1.42 1.56 1.16 1.04 0.96 0.91 0.93 1.22 0.98 0.63 0.75 0.98 0.95 0.99 0.91 1.07].

In the case of Spain we can remark that the values of the parameters fit into the assumption of time-homogeneity, since their values for 56 and 80 days don't exhibit significant differences. By allowing transitions from S to D we obtain an improvement in the error of almost 17% and the corresponding death rate is also of order 10^{-6}, so in this case we will stick to the same assumption as for Italy, namely that $\tilde{\mu}_d \neq 0$. This time we plot only the results of the simulations under these conditions, which can be seen in Figure 5. In contrast to the case of Italy, where the last part of the curve of the recoveries is not well approximated due to the lack of time-homogeneity of the parameter T_{conv}, here the same curve is approximated much better, with the exception of the middle part, where we can notice for a period a lower frequency in reporting the recoveries, followed by a sudden jump. But before and after this part, the estimated rate of reporting the recoveries fits the available data in a very good manner.

Comparing the values of T_{inf}, we note that they are similar to Germany and Italy, while $T_{conv} \approx 18$ is much closer to the corresponding value for Germany. As in the case of Italy, for Spain we perform also a simulation with $T_{conv} = 12$ and the result is plotted in Figure 6. Of course, when comparing such figures, one has also to take into account that the population size of Germany is by a factor of about 1.4 larger than in the case of Italy and by a factor of 1.8 larger than the population size of Spain.

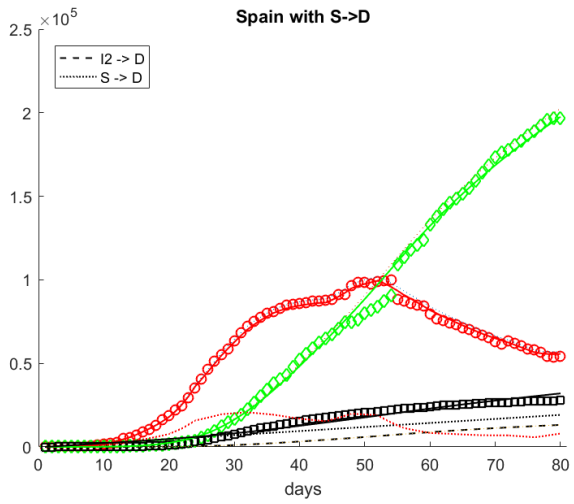

Figure 5: Simulated curves and data for Spain

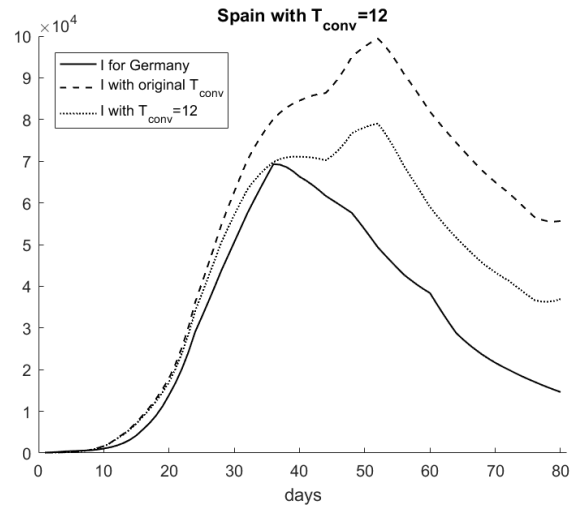

Figure 6: Active cases for Spain with $T_{conv} = 12$.

5 Simulations for Russia

The values of the time-independent parameters for which our model fits the data for Russia starting at 17.03.2020 are given in Table 4.

days	T_{inf}	T_{conv}	$\mu_d/\tilde{\mu}_d = 0$	$\mu_d/\tilde{\mu}_d$	imp.err.
64	3.80	43.51	0.0011	-	-
100	3.53	29.42	0.0009	0.0006 / 1.31e-7	0.21%

Table 4: Parameters for Russia

We further have $f = 0.17$, R =[1.74 1.42 2.65 1.43 2.05 1.38 1.46 1.56 1.36 1.29 1.15 1.27 1.25 1.00 1.00 0.96 0.85 0.83 1.05 1.00 1.02 0.97 1.06 0.79 1.06];

The fitted parameters show the fact that for Russia T_{conv} is not time homogeneous and is substantially

larger than the standard value in the case of Germany. But, similar to this country, we may assume also that $\tilde{\mu}_d = 0$, that is, we don't have transitions $S \to D$.

Figure 7 plots the simulation results for Russia. The virtual curve of active cases by considering the reference value $T_{conv} = 12$ is represented here in the same plot. The same will be the case for the following countries.

The low profile of the death curve shows that the case fatality rate for Russia is significantly smaller than in the other countries, which might indicate different criteria for recording the deaths in conjunction with COVID-19. Probably deaths are reported as such only if the disease had a decisive role in causing them and not only in the context of a positive test result.

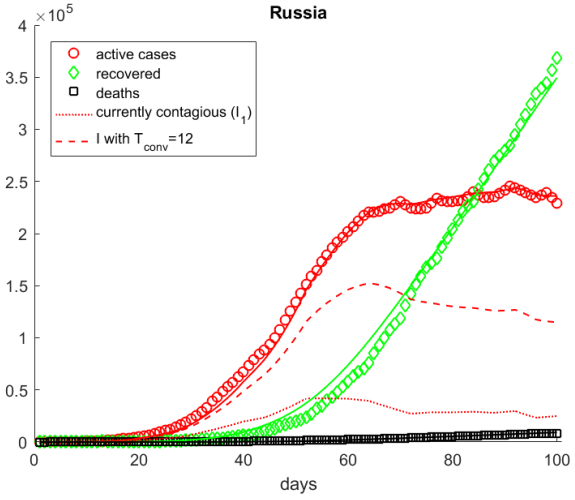

Figure 7: Simulated curves and data for Russia

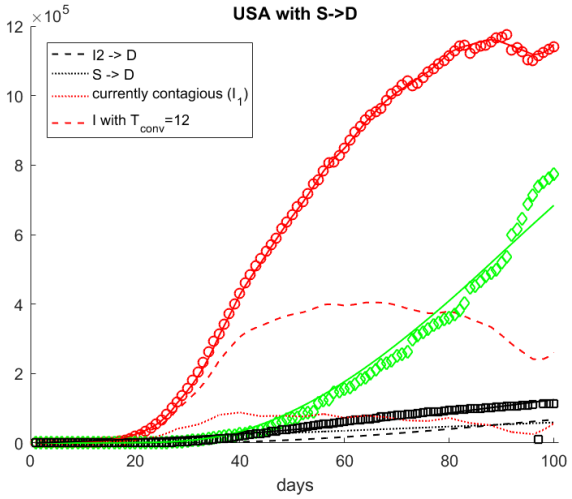

Figure 8: Simulated curves and data for the USA

6 Simulations for the USA

The values of the time-independent parameters for which our model fits the data for the USA starting at 01.03.2020 are given in Table 5.

days	T_{inf}	T_{conv}	$\mu_d/\tilde{\mu}_d = 0$	$\mu_d/\tilde{\mu}_d$	imp.err.
64	2.59	95.35	0.0046	-	-
100	2.59	79.85	0.0025	0.0012 / 1.81e-6	6.87%

Table 5: Parameters for the USA

We further have $f = 0.56$, R =[2.64 1.93 1.31 1.57 1.58 1.42 1.52 1.17 1.27 1.06 0.92 1.01 0.98 1.06 0.91 1.05 0.97 0.92 0.99 1.09 0.85 0.93 0.69 0.86 1.52].

We first notice that, in contrast to the other countries, the value $T_{inf} = 2.59$ is with one unit smaller. Since the values of the reproduction number $R(t)$ are more or less similar, we conclude that the spreading of the epidemy in the USA occured at a higher speed. A possible explanation for this is the fact that here the disease has spread mostly in agglomerated districts of major cities.

The different values of T_{conv} show that there is no time homogeneity and that the reporting of the recoveries occur at an extremely slow rate compared to all other countries. However, in the last part of the considered period, the rate of reporting the recoveries shows a significant increase which is reflected in higher figures in the data set than in the computed curve. We can also notice a slight effect of transitions of the type $S \to D$, but not as accentuated as in the case of Italy or Spain. However, since the value of $\tilde{\mu}_d$ is of magnitude 10^{-6} and not 10^{-7}, we plot the results of the simulations under this assumption.

7 Simulations for Romania

The values of the time-independent parameters for which our model fits the data for Romania starting at 14.03.2020 are given in Table 6.

days	T_{inf}	T_{conv}	$\mu_d/\tilde{\mu}_d = 0$	$\mu_d/\tilde{\mu}_d$	imp.err.
64	4.18	26.70	0.0055	-	-
100	3.69	22.85	0.0044	0.0015 / 5.42e-7	4.84%

Table 6: Parameters for Romania

For the other parameters we have $f = 0.43$, R =[1.47 1.04 2.17 1.63 1.24 1.53 1.08 0.98 1.09 1.00 1.15 0.77 0.94 1.00 0.72 0.12 1.68 0.90 0.70 0.98 1.13 1.09 1.17 1.39 0.96].

We can notice slight effects of time inhomogeneity in the parameters T_{inf} and T_{conv}. The value of the latter is however not so large, ranking on the third place in an increasing order, after Germany and Spain,

and closer to these two countries than the values of 38-80 for Italy and USA. For Romania we have also a small amount of transitions of the type $S \to D$, but the effect is not significant (error improvement of only 4.84%) and therefore for the plot we will consider that $\tilde{\mu}_d = 0$. Moreover, the estimated value for $\tilde{\mu}_d$ is of order 10^{-7}, similar to Germany and Russia, in contrast to the order of 10^{-6} for the other three countries.

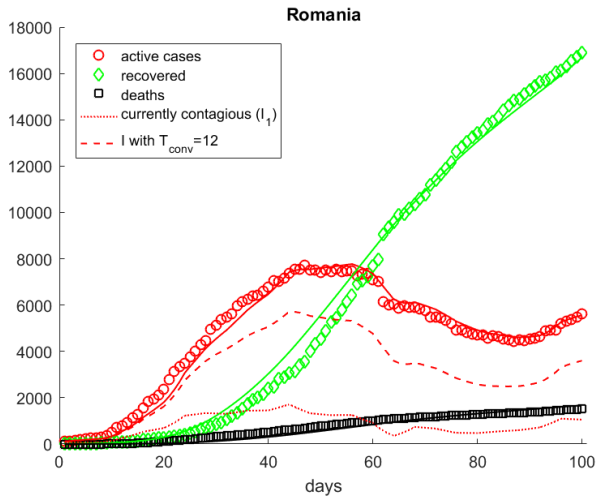

Figure 9: Simulated curves and data for Romania

8 Comparison of the infection dynamics

We will compare next the infection dynamics between the considered countries. The apparition of new infection cases is driven by the transmission rate $\beta(t) = R(t)/T_{inf}$, where the reproduction number $R(t)$ describes the average number of further infections produced by an infected individual and T_{inf} is the average time spent in the contagious state.

If we consider a single country, the value of $R(t)$ is sufficiently relevant in order to evaluate the future evolution of the epidemy. If its value is larger than 1, then the epidemy is expanding and the number of new infections is increasing, while for values less than 1 the number of cases is decreasing.

Figure 10 shows the time evolution of the reproduction number $R(t)$. We note that the profiles for the considered countries are basically similar, the differences between the curves being small. However, the solutions of the equation system (2) turn out to be highly sensitive regarding this parameter. Figure 11 illustrates a sensitivity analysis with respect to $R(t)$. The results show a high sensitivity of the model regarding the reproduction number. If we increase or decrease the time-depedenent vector $R(t)$

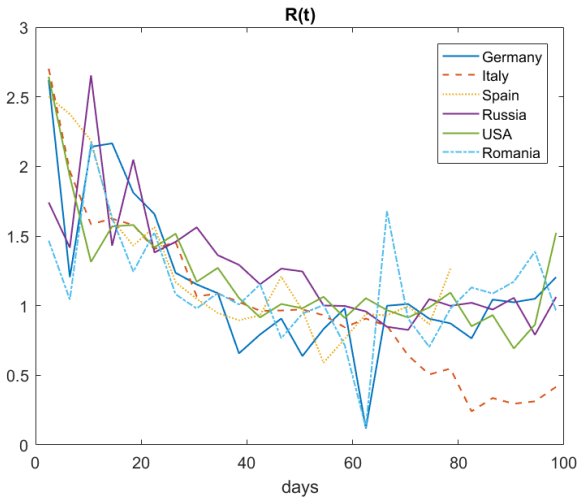

Figure 10: The evolution of the reproduction number $R(t)$ in different countries

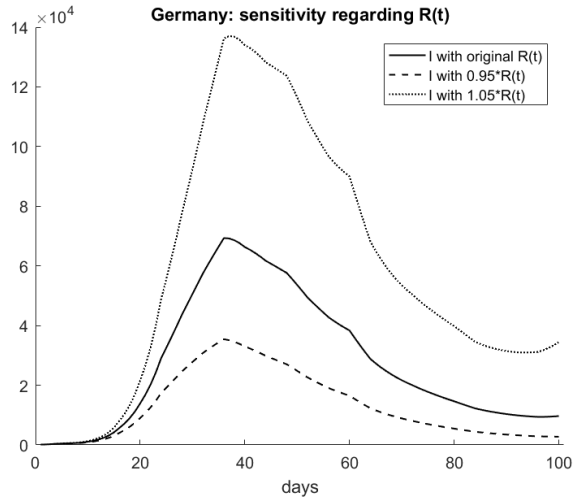

Figure 11: Comparing the dynamics of $I = I_1 + I_2$ for $R, 0.95 \cdot R$ and $1.05 \cdot R$ (Germany)

by only 5%, the maximum of the curve of the active cases increases or decreases roughly by the factor 2.

If we compare different countries, this high sensitivity shows that even small differences in the values of the reproduction number can lead to large differences in the numbers of new infections. Moreover, the comparison of the reproduction number $R(t)$ is relevant only if for the considered countries we have the same value for T_{inf}. This parameter is intrinsic to the disease and in our examples it has similar values, between 3 and 4 (less that 3 only for the USA).

But even these small differences in T_{inf} and therefore in $\beta(t) = R(t)/T_{inf}$ can reflect in considerable changes in the apparition of the new cases, so a more realistic comparison of the dynamics in the different

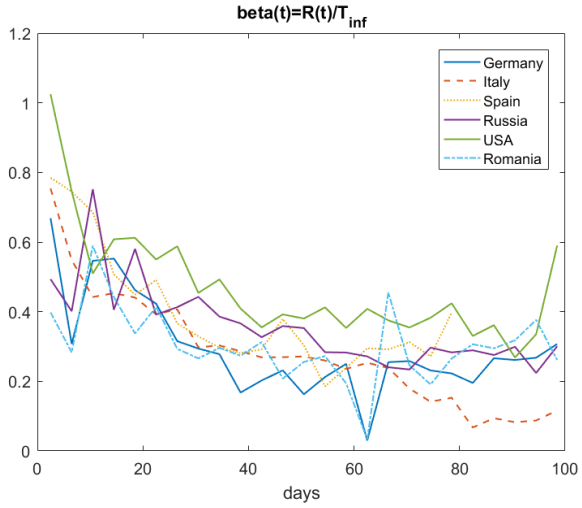

Figure 12: The evolution of the transmission rate $\beta(t) = R(t)/T_{inf}$ in different countries

countries can be read from the plot of the transmission rate $\beta(t)$. This is illustrated in Figure 12. We can see that on average the largest values of the factor β are associated to the USA and secondly to Russia. This provides an explanation for the fact that here the epidemy has spread at a higher speed than in the other countries considered in this paper.

9 Discussion

In this section we will summarize the results of the present paper.

The comparison of the statistical data for Germany and Spain with the results of the simulations shows a very good agreement, which indicates that the considered model (with its two variants regarding the death rates $\mu_d, \tilde{\mu}_d$) is in principle suitable for describing the spreading of the COVID-19 disease among the population.

While the assumption of time homogeneity for the infectious period T_{inf} can be verified for each country, in case of the convalescence period T_{conv} (average time until an infected individual is reported as recovered) this assumption holds basically for Germany, Spain and Romania, while for Italy, Russia and USA a comparison of parameters fitted for 64 and 100 days shows significantly different values. Moreover, if for Germany the value of this parameter is of about 12 days, for the other countries it ranges from 18 in the case of Spain to about 80 in the case of the USA, showing therefore more or less significant delays in reporting the recoveries, if we take as reference the rate for Germany.

If for Italy and Spain, countries where the number of cases was in the considered period similar to Germany, we run a scenario with the value $T_{conv} = 12$ corresponding to Germany, we can conclude that the real curves of active cases of these three countries would be in fact not so far away from each other. Of course, this interpretation has also to take into account that the population of Germany is by a factor of 1.4 larger than the population of Italy and 1.8 times larger than the population of Spain.

We also considered the possibility that new cases of COVID-19 may be reported only after the death of the individuals, that is that we have transitions directly from the state S (susceptible) into D (dead). The data for Italy and Spain indicate a strong effect of this type. In the case of the USA we can observe also this effect, but not as accentuated, while for Romania only a slight one, which we finally chose to neglect. In contrast to these situations, the data for Germany and Russia suggest that here we have no such effect at all.

A comparative plot of the effective reproduction number $R(t)$ for the considered countries exhibited basically similar profiles for all of them. However, a sensitivity analysis regarding this parameter showed that the curve of active cases is highly sensitive with respect to the reproduction number. By raising or lowering the values of $R(t)$ with only 5%, the maximum of the curve of active cases increases, respectively decreases, with the factor 2.

Due to this fact and to the slight differences of the parameter T_{inf} between the different countries, a better comparison of the infection dynamics can be read out from the plot of the transmission rates $\beta(t) = R(t)/T_{inf}$. This picture shows that on average this parameter is largest for the USA and then for Russia, the two countries where the epidemy has spread at a higher speed than in the others. This can be seen either from the original data reporting the daily new cases (which are not plotted in this paper) or by the number of active cases, even in the scenario with $T_{conv} = 12$ as in Germany, which for the USA and Russia is higher than in the other countries analyzed in the present paper.

10 Conclusion

In this paper we use ordinary differential equations in order to describe the evolution of the COVID-19 pandemy. The approach belongs to the so-called *compartmental models*, where the individuals can change from one state to another according to certain rates. We consider a modified SIR model which, to our knowledge, is novel in the literature dedicated to this topic. We distinguish the state S for susceptible individuals and instead of considering only one infectious state I as in the classical approach, we split it into the substates I_1 of contagious individuals and I_2 where the individuals are isolated and can not spread the disease anymore. The reason for this is the fact

that the recoveries, corresponding to the state R, are reported in different countries, due to their peculiarities, at different rates. This means that a unique infectious state I would last until the recovery or death is reported (the latter corresponding to the state D), which may correspond possibly to a long time spent in the infectious state, much beyond the period where one is effectively contagious. Additionally we introduce another novel feature, namely the possibility of transitions from S into D, which means that a certain number of cases is discovered or reported only after the death of the individual.

Another feature of the present model, which is ignored by many similar approaches in the literature, is the fact that the transmission rate of the infection is considered to be time-dependent. Especially in conjunction with a relatively long time interval of 100 days, this fact poses a challenging problem for the algorithm of parameter fitting. The goal of this step is to find an optimal set of parameters such that the solutions delivered by the model matches the given statistical data. The optimization procedure used here relies on a novel algorithm based on stochastic Runge-Kutta methods, where the predictor is computed neither by Euler steps as in the classical schemes, nor by a Markov process, as in the original stochastic method, but uses the variations of the data set, after which the approximation is enhanced by steps of Runge-Kutta type.

In this paper the model was kept as simple as possible, in order to be able to describe the evolution according to the publicly available data sets. Nevertheless, it can be improved concerning several aspects, which might open further research topics for the future.

The first approach in this direction would be to consider that also other parameters than the transmission rate are time-dependent, like the convalescence time or death rates. The simulations in the present paper showed that while for some countries one can consider these latter parameters as constant, for others they show a change in time, so this assumption would be a natural enhancement of the current model.

However, in this way one is faced with the challenging task of improving also the optimization algorithm used for parameter fitting, since in this case one has more than one time-dependent parameter. Further research topics may open therefore in the direction of developing such algorithms for parameter fitting for ordinary differential equations in general situations, not only in epidemic models.

Another direction of improvement of the present model may consist in introducing more states, corresponding to different degrees of severity of the illness (asymptomatic, symptomatic, hospitalized) as in [11], or a further state of patients on intensive care units.

The susceptible population can also be split in more compartments, as confined or not, see [10], or one can handle different states after the infection, such as symptomatic/asymptomatic in combination with detected/undetected, as in [6]. However, in the case of a more complex model, statistical data for the corresponding compartments should be available and one has to decide if the parameters of the model are considered as time-dependent or not.

Therefore, considering both directions simultaneously, i.e. combining more states with time-dependent parameters which have to be fitted to given data, would be the most challenging task in extending the present model.

11 Acknowledgement

I would like to thank the reviewers for their helpful comments and constructive suggestions, which allowed a considerable improvement of the structure and of the clarity of the exposition of this paper.

References:

[1] F. Brauer and C. Castillo-Chávez, *Mathematical Models in Population Biology and Epidemiology*. Springer, New-York, 2001.

[2] F. Brauer, Compartmental Models in Epidemiology, in F. Brauer, P. van den Driessche and J. Wu (Eds.) *Mathematical Epidemiology*, Springer, Berlin, Heidelberg 2008, Chapter 2, pp. 19–79

[3] A. Abou-Ismail, Compartmental Models of the COVID-19 Pandemic for Physicians and Physician-Scientists, *SN Comprehensive Clinical Medicine*, Vol. 2, 2020, pp.852–858. https://doi.org/10.1007/s42399-020-00330-z

[4] A. Anirudh, Mathematical modeling and the transmission dynamics in predicting the Covid-19 - What next in combating the pandemic, *Infectious Disease Modelling*, Vol. 5, 2020, pp. 366–374. https://doi.org/10.1016/j.idm.2020.06.002

[5] J.M. Carcione, J.E. Santos, C. Bagaini, J. Ba, A Simulation of a COVID-19 Epidemic Based on a Deterministic SEIR Model. *Frontiers in Public Health*, Vol. 8, 2020, Article 230. https://doi.org/10.3389/fpubh.2020.00230

[6] G. Giordano, F. Blanchini, R. Bruno et al., Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. *Nature Medicine* Vol. 26, 2020, pp. 855–860. https://doi.org/10.1038/s41591-020-0883-7

[7] A. Menon, N.K. Rajendran, A. Chandrachud, G. Setlur, Modelling and simulation of COVID-19 propagation in a large population with specific reference to India, *medRxiv, May 5, 2020*. https://doi.org/10.1101/2020.04.30.20086306

[8] E. Postnikov, Estimation of COVID-19 dynamics "on a back-of-envelope": Does the simplest SIR model provide quantitative parameters and predictions? *Chaos, Solitons & Fractals.* Vol. 135, 2020, 109841. https://doi.org/10.1016/j.chaos.2020.109841

[9] A.L. Kapetanović, D. Poljak, Modeling the Epidemic Outbreak and Dynamics of COVID-19 in Croatia, *arXiv, May 4, 2020*. https://arxiv.org/pdf/2005.01434.pdf

[10] L. López, X. Rodó, A Modified SEIR Model to Predict the COVID-19 Outbreak in Spain and Italy: Simulating Control Scenarios and Multi-Scale Epidemics, *medRxiv, April 16, 2020*. https://doi.org/10.1101/2020.03.27.20045005

[11] P. Teles, A time-dependent SEIR model to analyse the evolution of the SARS-covid-2 epidemic outbreak in Portugal, *Bulletin of the World Health Organization*, E-pub: April 7, 2020. http://dx.doi.org/10.2471/BLT.20.258392

[12] C. Anastassopoulou, L. Russo, A. Tsakris, C. Siettos, Data-based analysis, modelling and forecasting of the COVID-19 outbreak, *PLOS ONE* Vol. 15(3), 2020, e0230405. https://doi.org/10.1371/journal.pone.0230405

[13] D. Efimov, R. Ushirobira, On an interval prediction of COVID-19 development based on a SEIR epidemic model, *Research Report, Inria. hal-02517866v6, June 3, 2020*. https://hal.inria.fr/hal-02517866/file/SEIR.pdf

[14] M. an der Heiden, O. Hamouda, Schätzung der aktuellen Entwicklung der SARS-CoV-2-Epidemie in Deutschland – Nowcasting. *Epidemiologisches Bulletin* 17, 2020, pp. 10–16. https://doi.org/10.25646/6692.4

[15] F. Guiaş, Estimating parameters for systems of ordinary differential equations using the principle of stochastic Runge-Kutta solvers, 6^{th} *SMTDA Conference Proceedings, 2-5 June 2020, Barcelona, Spain*