Seeing the unseen—the iceberg phenomenon in the first months of the COVID19 pandemic

Dragoș-Victor Anghel¹ and Ioan Tudor Alexandru Anghel²
¹Horia Hulubei National Institute for R&D in Physics and Nuclear Engineering, Romania
²University of Bucharest, Faculty of Mathematics and Informatics, Romania
E-mail: dragos@theory.nipne.ro

Abstract. We analyze the evolution of the COVID19 infections in the first months of the pandemics and show that the basic compartmental SIR model cannot explain the data, some characteristic time series being by more than an order of magnitude different from the fit function over significant parts of the documented time interval. To correct this large discrepancy, we amend the SIR model by assuming that there is a relatively large population that is infected but was not tested and confirmed. This assumption qualitatively changes the fitting possibilities of the model and, despite its simplicity, in most cases the time series can be well reproduced. The observed dynamic is only due to the transitions between two infected compartments, which are the unconfirmed infected and the confirmed infected, and the rate of closing the cases (by recovery or death) in the confirmed infected compartment. We also discuss some relevant extensions of this model, to improve the interpretation and the fitting of the data. These findings qualitatively and quantitatively evidences the “iceberg phenomenon” in epistemology.

1. Introduction
In the absence of a vaccine and efficient antiviral drugs, the containment of the COVID19 pandemics relies mainly on non-pharmaceutical interventions. Such interventions depend upon the correct determination of the parameters that characterize the spread of the virus (like the transmission, recovery, and death rates), which are fed into mathematical models to predict the time evolution of the pandemics. But obtaining the correct numbers is difficult while the pandemics unfolds, because of incomplete data and uncertainties in their interpretation. Nevertheless, a careful analysis of the data may guide one to adapt the statistical models to consistently describe the observations and to recover the missing information. The data is only the tip of the iceberg [1]. Our challenge is to calculate the real size of the block of ice hiding beyond our observation possibilities—that is, to really find how fast the virus spreads and how many people are affected.

The mathematical modeling of the spreading of infectious diseases is centuries old [2] and eventually started with the work of Daniel Bernoulli on the variola spreading [3]. These models may be grouped in two categories: stochastic and deterministic. Whereas the stochastic models, based mainly on simulations, may be more suitable for smaller communities, the deterministic models are usually employed for large communities, like countries or cities, where the fluctuations around averages are small and the time evolution is accurately determined by a small number of parameters. Because the analyzed population is divided into compartments, the deterministic models are also called compartmental models [4]. The most simple such model is the SIR model,
composed of three compartments: 
Susceptible, Infected, and Recovered, which was introduced in 1927 by Kermack, McKendrick, and Walker [4, 5]. The Susceptible compartment includes the healthy persons that are not immune and may get infected by contact with the sick people from the Infected compartment. The persons in the Infected compartment are transferred into the Recovered compartment as they recover or die (we shall call this also the Closed cases compartment). In this model it is also assumed that the recovered people are immune and cannot get sick again (therefore they do not return to the Susceptible compartment upon recovery).

If we denote by $S$, $I$, and $R$ the number of people in the compartments Susceptible, Infected, and Recovered, respectively, then, in the simplest case one may assume that the total number of people, $N = S + I + R$ is constant. It is more simple to work with the normalized values, so we introduce the notations $s = S/N$, $y = I/N$, and $r = R/N$. Then, the time evolution of $s$, $y$, and $r$ is given by the set of equations (see Appendix A and Refs. [5, 6, 2])

$$\dot{s} \equiv \frac{ds}{dt} = -ps(t)y(t),$$  

$$(1a)$$

$$\dot{y} \equiv \frac{dy}{dt} = [ps(t) - \lambda]y(t),$$  

$$(1b)$$

$$\dot{r} \equiv \frac{dr}{dt} = \lambda y(t),$$  

$$(1c)$$

where $p$ and $\lambda$ will be called the infection rate and the recovery rate, respectively. In general, $p$ and $\lambda$ may also be functions of $t$ (for example, $p$ is influenced by regulations imposed to the population, such as social distancing, whereas $\lambda$ is influenced by the response of the public health system and the time passed since the infection), but since we discussed only the early stage of the pandemic, when the conditions did not improve significantly, we shall assume that they are constants. From Eqs. (1), only the first two are independent, the third one being a consequence of the normalization condition $s + y + r = 1$, which yields $\dot{r} = -\dot{s} - \dot{y}$. Equations (1) may be solved with the initial boundary conditions $s(0) = s_0$ and $y(0) = y_0$ (see Appendix A for more details).

2. Results

2.1. Method 1 (M1): fitting the time series of the active cases

With the formalism presented above, one may try to describe the evolution of the COVID19 pandemics. The official numbers of infected, dead, and recovered persons may be found on several databases. We use the one from GitHub (https://github.com/CSSEGISandData/COVID-19). In Fig. 1 we plot $y(t)$ (Active cases) for Romania, Hubei (China), Germany, and South Korea, together with the fit of the SIR model (minimum square deviation), with constant $p$ and $\lambda$ (or $a$). This simplified model does not fit the data very well and more complex models should be employed. But although the details of fitting the time series of the Active cases are important, here we shall focus on a more conspicuous problem. In Fig. 2 we show the time series of the Closed cases of the same countries and on the same periods as in Fig. 1. We observe that the SIR results for the same parameters as in Fig. 1 differ from the recorded data by orders of magnitude. Such discrepancies cannot be mended by adjustments of the parameters $p$ and $\lambda$ or by taking into consideration reasonable time variations of them.

We did similar fits also for many countries and we could not find any country in which the M1 method would provide a good approximation for the early stages of the pandemic evolution.

2.2. Method 2 (M2): fitting the time series of the total infections

One may try another approach. The time variation of $s$, and therefore of $y$, is determined by the number of infectious persons, which may not be accurately described by the number of active cases, as it is assumed in the SIR model. This may have happened because there was
Figure 1. The Active cases $y(t)$ of COVID19 for four examples of countries (blue circles) and the SIR model fit (red curve). The parameters are indicated in each plot. In each case, we use the data from the day in which the number of confirmed cases exceeded 100, until the date indicated.

Figure 1. The Active cases $y(t)$ of COVID19 for four examples of countries (blue circles) and the SIR model fit (red curve). The parameters are indicated in each plot. In each case, we use the data from the day in which the number of confirmed cases exceeded 100, until the date indicated.

no medical procedure that could eliminate a patient from the list of active cases immediately when it becomes noninfectious. Rather, one was eliminated from this list after a certain interval of time, when it was tested negative for the COVID19 disease. This may have allowed for an overestimation of the number of active cases at any moment in time. For this reason, we may assume that the (total) number of infected people, which is $y(t) + r(t) = 1 - s(t)$, is more appropriate for the fitting procedure. In this way we obtain the plots of Fig. 3. Obviously, this fits the time series of the infectious population much better than Method 1, but it fails to fit (by orders of magnitude) the time series of the active cases, as one can see in Fig. 4. Furthermore, for longer time intervals, the fitting functions for Romania and Germany of Fig. 3 are likely to differ significantly from the data. We investigated many countries but obtained similar results. Therefore, we observed that the SIR model cannot simultaneously fit the time series of the active cases and the total number of infections (that is, the active cases plus the closed cases). We may expect that by changing the constant $p$ and $\lambda$ parameters with functions of time we may be able to better fit the active cases or the closed cases, but not both of them at the same time, because of the huge differences observed between the predictions and the time series. This prompts us to make some substantial changes in the model.
Figure 2. The Closed cases $r(t)$. The model is fitted to the time series of the Active cases, as shown in Fig. 1.

3. Methods
To explain the discrepancies observed in the previous section, we extend the SIR model by assuming that the number of infected people is significantly different from the one we know, because, on one hand, a small fraction of the population has been tested and, on the other hand, the number of infected people was still very small. For simplicity, we shall refer to this as the SIR-Iceberg model or SIRI. We assume that beside the known infected population $I_1$, there are $I_2$ infected persons that we do not know about, such that the total infected population is $I = I_1 + I_2$. Similarly, the know and unknown recovered populations are denoted by $R_1$ and $R_2$, respectively (see Fig. 5 for the schematic representation of this model). If the total population is again denoted by $N$, then the susceptible population is $S = N - I_1 - I_2 - R_1 - R_2$. If we introduce the densities $s \equiv S/N$, $y_1 \equiv I_1/N$, $y_2 \equiv I_2/N$, $r_1 \equiv R_1/N$, $r_2 \equiv R_2/N$, then the system of coupled differential equations becomes

$$
\begin{align}
\dot{s} &= -(p_1 y_1 + p_2 y_2)s(t), \\
\dot{y}_1 &= \xi y_2 - \lambda_1 y_1, \\
\dot{y}_2 &= (p_1 y_1 + p_2 y_2)s(t) - \xi y_2 - \lambda_2 y_2, \\
\dot{r}_1 &= \lambda_1 y_1(t) + \zeta r_2, \\
\dot{r}_2 &= \lambda_2 y_2(t) - \zeta r_2.
\end{align}
$$
Figure 3. The number of confirmed cases \(1 - s(t) = y(t) + r(t)\) and the fit with the SIR model in four populations: Romania, Hubei (China), Germany, and South Korea. The fit parameters and the time interval from which the data is taken are shown in the plots.

In the system (2) we assumed that the rates of contamination due to the infected populations \(I_1\) are \(I_2\), denoted by \(p_1 \) and \(p_2\), respectively, are different, because the population \(I_1\) is much better monitored and isolated from the healthy population (one may safely assume that \(p_1 \ll p_2\)). Furthermore, a contaminated person first enters the \(I_2\) population and it has the rate \(\xi\) (assumed constant) to enter the known infected population \(I_1\) (by random testing or because of severe symptoms). The rates by which the cases are closed (by recovery or death) in the populations \(I_1\) and \(I_2\), denoted by \(\lambda_1\) and \(\lambda_2\), respectively, may be different because of the following two (competing) reasons: (1) the population \(I_1\) gets better health care and (2) the symptoms of \(I_1\) may be (on average) more severe than the ones of the population \(I_2\), because they may have led to the observation of COVID19. Further, the rate \(\zeta\) of transferring population from \(R_2\) to \(R_1\) is mostly due to the postmortem detection of the SARS-Cov-2 virus, because in the early stages the pandemic antibody tests were only being developed. In practice, this may be very small, so we set \(\zeta = 0\) in the numerical calculations.

The initial conditions are \(s_0 \equiv s(0), y_{1,0} \equiv y_1(0), y_{2,0} \equiv y_2(0), r_{1,0} \equiv r_1(0),\) and \(r_{2,0} \equiv r_2(0)\). Since not all these parameters are independent \((s_0 + y_{1,0} + y_{2,0} + r_{1,0} + r_{2,0} = 1)\), we may choose, for example, \(r_{2,0}\) to be determined by the other ones. To fit the data, according to the arguments above we choose \(p_1 = p_2/10\) or \(p_2/20\) (the final differences are insignificant), \(\lambda_1 = \lambda_2 = ap_2\), and \(\zeta = 0\). In this way we are left with the fitting parameters \(p_2, a,\) and \(\xi\). The results of the fit are shown in Figs. 6 - 9.
Figure 4. The time series of the Active cases and the SIR predictions, for the parameters obtained by fitting the Confirmed cases with the SIR model as shown in Fig. 3.

Figure 5. The schematic representation of the SIRI model.

We observe that the SIRI model fits significantly better all the countries, despite its simplicity. Nevertheless, in some cases discrepancies between the model predictions and the data still exist. In Figs. 6 - 9 we show two contrasting situations. While the fit for European countries are quite good (see, for example, Romania and Germany in Figs. 6 and 7, respectively), the data for Hubei
are more uniformly distributed. Furthermore, even in the same geographical area, some people
and one should split countries (or regions) into more localized containers in which the infections
measures, time dependence of the recovery rates, etc.) and the testing of the population.

time evolution of the pandemic (people awareness, health system improvement, governmental
parameters that we use in the models should be replaced with functions of time, according to the
\( R \)
\( I \)
\( S \)
We estimate it in the model. If
\( s \)
\( I \)
\( I_2 \)
\( R_1 \)
\( y_2 \)
\( r_2 \)
\( (COVID19 \ diagnosis \ at \ some \ time \ after \ the \ infection) \) and from
\( (or, \ equivalently, \ I_2 \ and \ R_2) \).
We estimate it in the model. If \( s \) becomes too small, then the time variations of \( y_1 \) and \( r_1 \) are
not determined by the actual infections, but by the rate of people transfer from compartment
\( I_2 \) into compartment \( I_1 \) (COVID19 diagnosis at some time after the infection) and from \( I_1 \) to
\( R_1 \) (recovery or death of people confirmed with COVID19).

The discrepancies between the data and the SIRI estimates may be due to the fact that the parameters
that we use in the models should be replaced with functions of time, according to the
time evolution of the pandemic (people awareness, health system improvement, governmental
measures, time dependence of the recovery rates, etc.) and the testing of the population.
Another important aspect is that the spread of the disease is not geographically homogeneous
and one should split countries (or regions) into more localized containers in which the infections
are more uniformly distributed. Furthermore, even in the same geographical area, some people
are more exposed to the virus than others, which would lead to further splitting of the population

![Figure 6. The fitting of the SIRI model to the data in Romania, for the indicated period. The
obtained parameters are shown in the plots.](image)

![Figure 7. The fitting of the SIRI model to the data in Germany, for the indicated period. The
obtained parameters are shown in the plots.](image)
into containers. These aspects are not incorporated into our model, but our description, while it remains simple, is reliable.

Nevertheless, maybe the most important conclusion of this work is that there was a significant part of the population which was infected, but was not detected. Taking such aspects into account would significantly improve the estimation of parameters like the disease contagiousness, virulence, and healing rate, improving also the prediction of the new waves of pandemics.

Acknowledgments
We are grateful to Prof. Dan Pirjol for drawing our attention to this subject and for introducing us to the basic SIR model. The work was supported by the UEFISCDI project PN 19060101/2019 and Romania-JINR collaboration projects.

Authors’ contributions
DVA formulated the concept and the SIRI model, contributed to the numerical calculations, to the interpretation of the results, and wrote the first draft of the paper. ITAA discussed the models, contributed to the numerical calculations (he wrote and used the C++ codes), to the interpretation of the results, and wrote the first draft of the paper. ITAA discussed the figures.

Appendix A. The SIR model
As explained in the main text of the article, the SIR model consists of three compartments, Susceptible, Infected, and Recovered, each of them being populated by $S$, $I$, and $R$ persons. The

Figure 8. The fitting of the SIRI model to the data in Hubei, for the indicated period. The obtained parameters are shown in the plots.

Figure 9. The fitting of the SIRI model to the data in South Korea, for the indicated period. The obtained parameters are shown in the plots. The fluctuations observed in the last plot (susceptible population) are numerical error.
time evolution of these populations is given by the equations [5, 6, 2]

\[
\dot{S} \equiv \frac{dS}{dt} = -pI(t) \frac{S(t)}{N}, \quad (A.1a)
\]

\[
\dot{I} \equiv \frac{dI}{dt} = \left[p \frac{S(t)}{N} - \lambda\right] I(t), \quad (A.1b)
\]

\[
\dot{R} \equiv \frac{dR}{dt} = \lambda I(t), \quad (A.1c)
\]

where \( p \) and \( \lambda \) have been defined in Section 1. By dividing the system (A.1) by \( N \), we obtain the system (1). If we divide the Eqs. (1) by \( p \) and denote \( x \equiv pt \) and \( a \equiv \lambda/p \), Eqs. (1) get the simpler form [7]

\[
s''(x) \equiv \frac{ds}{dx} = -s(x)y(x), \quad (A.2a)
\]

\[
y'(x) \equiv \frac{dy}{dx} = [s(t) - a]y(x), \quad (A.2b)
\]

\[
r'(x) \equiv \frac{dr}{dx} = ar(x), \quad (A.2c)
\]

with \( a \) being the only parameter. From Eqs. (A.2) we see that \( a \) sets the shape of the curve whereas \( p \) sets the scale of the time interval or the evolution speed of the epidemic. In what follows, we shall omit to include explicitly the dependence of the functions on the \( t \) or \( x \) variables, unless necessary.

Other characteristics of the time evolution of \( s \) may be readily obtained from (A.2). For example, the first two equations may be combined into a single equation for \( s \),

\[
s'' = s'^2/s + (s - a)s', \quad (A.3)
\]

which may be further simplified if we introduce the notation \( w(x) = \ln[s(x)] \):

\[
w'' = (e^w - a)w' \quad (A.4)
\]

(by \( \cdot'' \) we denote the second derivative \( d^2/dx^2 \)). Starting from the initial conditions \( w(0) \equiv w_0 = \ln(s_0) = \ln(1 - y_0 - r_0) < 0 \) and \( w'(0) = -y_0 < 0 \), one may obtain an analytical expression for \( w(t) \). For this, we integrate both sides of Eq. (A.4) to obtain

\[
w'(x) = a \ln(1 - y_0 - r_0) - \left[1 - r_0 + aw(x) - e^{w(x)}\right] \equiv -D(w, a, y_0, r_0). \quad (A.5)
\]

In Eq. (A.5) we may separate the variables and obtain

\[
x = \int_w^{\ln(1 - y_0 - r_0)} \frac{dw'}{D(w', a, y_0, r_0)}. \quad (A.6)
\]

In the equations above, one can neglect \( r_0 \), since this is in general either zero or very small.

We see that \( D(w, a, y_0, r_0) \geq 0 \). Furthermore, the integral diverges if and only if the denominator \( D \) becomes zero at the lower integration limit. If we denote this limit by \( w_\infty \), then \( D(w_\infty, a, y_0, r_0) = 0 \) and \( \lim_{x \to \infty} w(x) = w_\infty \). Therefore, \( w_\infty \) may be determined from the equation

\[
1 - r_0 + aw_\infty - e^{w_\infty} = a \ln(1 - y_0 - r_0) \equiv aw_0. \quad (A.7)
\]

Since \( s' < 0 \) for any \( x \), then \( w' < 0 \), so \( w_\infty < w_0 < 0 \).
In the beginning of an epidemic, fitting the data with Eqs. (A.8) on a logarithmic scale, one can extract the model parameters \( a \) and \( y_0 \):

\[
\ln(y) = \ln(y_0) + (1 - y_0 - r_0 - a) pt \quad \text{and} \quad \ln(s) = \ln(1 - y_0) - y_0 pt. \tag{A.9}
\]

We assume that \( y_0 \ll 1 - a < 1 \), so we can use the approximations \( 1 - y_0 - r_0 - a \approx 1 - a \) and \( \ln(1 - y_0 - r_0) \approx 0 \).
If an epidemic had evolved for some time and is far from the beginning, as is the case of the COVID19 pandemic, one should fit the whole time series. The evolution of COVID19 was very complex and determined by many factors. We do not intend to describe here the whole evolution, but we analyze only time intervals shorter than 100 days. Day 1 is the day in which the number of confirmed cases exceeded 100 in the analyzed country, such that too large fluctuations that appear in the beginning due to the bad statistics may be partially avoided, while we still analyze the evolution from an early stage. During such a time interval, the evolution of the number of cases in each country may be modeled by parameters that do not vary in time (the parameters are different from country to country), since the measures taken by governments and governmental institutions were only being implemented and conditions did not vary much. The COVID19 started spreading in different countries or regions at different times, so we fit each country or region separately. As explained before, we take the data from https://github.com/CSSEGISandData/COVID-19 and form the vectors of data \( S_d, I_d, \) and \( R_d \), corresponding to the compartments \( S, I, \) and \( R \), respectively. The unit of time is 1 day and \( t = 0 \) at Day 1. If \( N \) is the total population of the country or region analyzed, we calculate the densities \( s_d = S_d/N, \ y_d = I_d/N, \) and \( r_d = R_d/N \). On the other hand, we calculate the time evolution \( s(t), \ y(t), \) and \( r(t) \), using Eqs. (1) and the initial conditions \( s_0 = s_d(1), \ y_0 = y_d(1), \) and \( r_0 = r_d(1) \). The functions \( s(t), \ y(t), \) and \( r(t) \) depend on the parameters \( p \) and \( \lambda \), which we determine by fitting the data using the least squares method.

The results of the fitting procedure were explained in Sections 2.1 and 2.2. Eventually the most important time series is \( y_d \), which represents the time evolution of the number of active cases. Therefore, in Section 2.1 we determined the parameters \( p \) and \( \lambda \) by fitting \( y(t) \) to \( y_d \) (with the initial conditions as specified above). The results are plotted in Fig. 1. We notice that while some of the details may not be fitted by this method, the calculated functions reproduce the general behavior of the time series. The details may be adjusted by assuming that \( p \) depends on time (because of the restrictions imposed and the reaction of the population), whereas \( \lambda \) depends both on the time of the infection and on the time elapsed from the infection (because of the time variation of the health care system efficiency and of the human body response). Using such adjustments, one may expect to be able to fit the time series \( y(t) \), but the real problem of this method is that the same parameters should be used to calculate the estimation of the other time series, namely of \( s(t) \) and \( r(t) \). It is shown in Fig. 2 that the difference between \( r(t) \) and the data \( y_d(t) \) is very big—in relevant time intervals, the estimation is more than one (or even two) orders of magnitude bigger than the reported values. Therefore, by adjusting the parameters to fit the \( y_d \) time series in more detail is not expected to make such a huge difference in the values of \( r(t) \) to compensate for the differences observed in Fig. 2.

In Section 2.2, we tried to fit the time series of the confirmed cases \( y_d + r_d \). We applied again the least squares method and we obtained the results presented in Fig. 3. Even though the calculations \( y(t) + r(t) \) are similar to the data \( y_d + r_d \), this time the active cases \( y(t) \) differ by orders of magnitude from the time series \( y_d \), as one can see in Fig. 4. As in the case of M1, the fine adjustments of \( p \) and \( \lambda \) to fit the time series of confirmed cases cannot compensate for the big differences obtained for the active cases.

We also tried to fit simultaneously \( y_d \) and \( r_d \), by applying the least square method to both sets, but by doing so, in the best case we could fit only one of the time series. Visually, in general this method led to worse results for both, \( y_d \) and \( r_d \), whereas quantitatively, the time series of bigger numbers was better fit than the time series consisting of the smaller numbers.

Appendix B. Fitting with the modified SIR model, SIRI
In order to fit all the observed time series, we had to make qualitatively important adjustments to the model, as explained in Section 3. We split the compartment \( I \) of the SIR model into \( I_1 \) and \( I_2 \), where \( I_1 \) represents the number of confirmed active cases, whereas \( I_2 \) is the number of
infected people which are not known and therefore are missed from the statistics. Similarly, $R$ is split into $R_1$ and $R_2$, $R_1$ representing the known closed cases (recovered persons plus deaths) and $R_2$ being the number of closed cases which are missing from the statistics. As in the SIR model, we define the densities $s \equiv S/N$, $y_1 = I_1/N$, $y_2 = I_2/N$, $r_1 = R_1/N$, and $r_2 = R_2/N$, whereas the rest of the parameters are defined in the main text. This model allowed us to fit quite well all the time series for most of the countries, as one can see in Section 3 and in Fig. B1, but there are also some countries and regions that cannot be fitted so well, like Hubei (China), South Korea (see main text), and the Netherlands (Fig. B1).

We also notice that the susceptible population decreases faster in some of the cases, which means that the dynamics of $y_1$ and $r_1$ in the time interval when $s(t) \ll 1$ is mainly due to the crossing from $I_2$ to $I_1$ (at rate $\xi$) and recovery from $I_1$ to $R_1$ (at rate $\lambda_1$). The oscillations and discontinuities of $s$, when $s \ll 1$, are numerical artifacts, since $s \geq 0$ and according to Eq. (2a) of the main text $\dot{s} < 0$ for any $t$. These artifacts are irrelevant for the time evolution of the other quantities because $s \ll 1$.

References

[1] J. M. Last. Commentary: The iceberg revisited. Int. J. Epidemiology, 42(6):1613, 01 2014.
[2] H. W. Hethcote. The mathematics of infectious diseases. SIAM Review, 42:599, 2000.
[3] D. Bernoulli. Essai d’une nouvelle analyse de la mortalité causée par la petite vérole et des avantages de l’inoculation pour la prévenir. Mémoires de Mathématiques et de Physique, Académie Royale des Sciences, Paris, 1760:1, 1760.
[4] Compartmental models in epidemiology. https://en.wikipedia.org/wiki/Compartmental_models_in_ epidemiology.
[5] W. O. Kermack, A. G. McKendrick, and G. T. Walker. A contribution to the mathematical theory of epidemics. Proc. R. Soc. Lond. A, 115:700, 1927.
[6] T. Harko, F. S. N. Lobo, and M. K. M. K. Mak. Exact analytical solutions of the Susceptible-Infected-Recovered (SIR) epidemic model and of the SIR model with equal death and birth rates. Applied Mathematics and Computation, 236:184 – 194, 2014.
[7] Dan Pirjol. Solutie analitica pentru modelul SIR. private communication.
Figure B1. Examples of time series of COVID19 infections fitted with the modified model SIRI. As the total populations, we used the values found on Wikipedia: Italy - $N = 60.36 \times 10^6$, Spain - $N = 46.94 \times 10^6$, the Netherlands - $N = 5.518 \times 10^6$, the United States - $N = 328 \times 10^6$, and Brazil - $N = 209.5 \times 10^6$. We see that in most cases, the fit is quite good (eventually with the exception of the Netherlands).